Robert Fletcher
Marie-Josée Fortin

Spatial Ecology and Conservation Modeling

Applications with R



Springer

Spatial Ecology and Conservation Modeling

Libros de Estadística en PDF | Statistics Books in PDF

Robert Fletcher • Marie-Josée Fortin

Spatial Ecology and Conservation Modeling

Applications with R

Springer

Libros de Estadística en PDF | Statistics Books in PDF

Robert Fletcher
Department of Wildlife Ecology and Conservation
University of Florida
Gainesville, FL, USA

Marie-Josée Fortin
Department of Ecology and Evolutionary Biology
University of Toronto
Toronto, ON, Canada

ISBN 978-3-030-01988-4 ISBN 978-3-030-01989-1 (eBook)
https://doi.org/10.1007/978-3-030-01989-1

Library of Congress Control Number: 2018965604

© Springer Nature Switzerland AG 2018
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Libros de Estadística en PDF | Statistics Books in PDF

To Christine, Hayden, and Ava:
You inspire me everyday
—Robert Fletcher

To Ian
—Marie-Josée Fortin

Libros de Estadística en PDF | Statistics Books in PDF

Preface

All questions in ecology and problems in conservation play out in space. This simple fact has generated a long-standing interest from scientists and practitioners, and it is now well known that explicitly accounting for space in ecology, or spatial ecology, is necessary for accurately answering ecological questions and providing useful conservation solutions.

Only in the past few decades, however, has spatial ecology really matured. With advances in data collection across both broad and fine spatial scales, data are now widely available that are spatially explicit. At the same time, the growth in ways to analyze, model, and interpret spatial data has also increased substantially in the past two decades. The synergism of data availability and new models and methods for interpreting spatial data has led to rapid advancements in our understanding of spatial ecology and the way in which spatial concepts are applied to conservation.

Why Do We Need a New Book on Spatial Ecology and Conservation?

Several books have been published over the years that are relevant to spatial ecology and conservation (see Chap. 1). Some of these books have focused on spatial ecology explicitly while others have been on related topics, such as landscape ecology and metapopulation ecology. In addition, there have been several books published on spatial analysis, and some focus specifically on ecological data. There have also been books that emphasize spatial theory in ecology. These books have been very helpful to the ecological community.

Despite these efforts, a book on spatial ecology that provides a general, applied overview and allows readers to go "under-the-hood" to interpret concepts and models with real data has remained absent. Our book attempts to fill this gap by providing an introduction to relevant concepts and modeling techniques used by

Libros de Estadística en PDF | Statistics Books in PDF

applied ecologists and conservation biologists to address ecological and conservation problems. We provide this introduction by using empirical examples to illustrate not just *what* spatial ecology is, but *how* it is implemented. Our book covers several topics, and in doing so, it illustrates to readers many of the commonalities among topics to address the problem of space.

Who Should Read this Book?

This book is intended to provide a gentle introduction to the concepts and modeling involved in the field of spatial ecology and conservation. The target audience of the book includes students and professionals. Scientists and conservation practitioners who seek to understand ecology and conservation should benefit from the overviews provided. This is not a book written for experts in spatial ecology and modeling. There are several advanced books written for experts, and our intent is not to re-create those books. Rather, students and professionals who would like to find out what spatial ecology is about and how it is applied to conservation problems are the target audience.

Readers should have a general familiarity with statistics and models. Such knowledge is essential for further extending quantitative techniques to the problem of space. We do not expect all readers to be especially quantitative. Throughout this book, we provide examples of spatial concepts, models, and analyses using the program R. Readers do not need to be proficient in R, however, to use this book. We provide an Appendix that covers some foundational topics for the use of R that should provide readers enough information on R to interpret the examples in this book.

The material from this book has been developed from graduate courses that the authors teach. We have used this material both for lecture development and lab development, where students work through the examples provided in the book. We expect that this book could provide a useful complement to courses on spatial ecology, landscape ecology, species distributions, spatial analysis for non-statisticians, and applied conservation.

Organization of the Book

The goal of this book is to provide an overview of several (but not all) issues in spatial ecology and conservation. Our book was written as an introduction to these topics with an emphasis on learning-by-doing, where we illustrate these topics with real data and the application of spatial modeling to these topics. For each application, key R code is provided as an illustration of the aspects needed to understand the essence of each topic and its potential relevance to conservation applications.

Libros de Estadística en PDF | Statistics Books in PDF

The first part of the book (*Quantifying Spatial Pattern in Ecological Data*) focuses primarily on spatial pattern analysis and modeling of relevance to ecology and conservation. Issues such as the problem of scale, spatial dependence, land-cover pattern and change quantification, as well as spatial prediction are introduced and illustrated using case studies.

The second part of the book (*Ecological Responses to Spatial Pattern and Conservation*) focuses on spatial ecological disciplines and topics, ranging from species distributions and resource selection, movement and connectivity, to metapopulation and metacommunity dynamics. Again, an overview of these ecological processes and resulting spatial patterns are presented and case studies are used to depict how to analyze them spatially.

For each chapter, we first provide a concise overview of the relevant terms and concepts. We focus on practical issues and concepts, only briefly providing the relevant spatial ecological theory and concepts needed to understand the spatial analyses presented. Key terms that are required for understanding the major concepts are highlighted in italics and defined in tables and text. The latter portion of each chapter depicts these concepts and attempts to demystify them via worked examples using real data and R code. We end each chapter with a "Next Steps and Advanced Issues" section that directs readers to further approaches and issues of relevance to each topic.

Throughout the book, R code is provided as well as the interpretation of model outputs. We use R for several reasons: (1) R is an open-source environment that works on Windows, Mac OS X, and Linux; (2) the R environment is flexible enough to offer statistical, visualization, and simple programming language features; and (3) a series of R packages have been developed by scientists, including ecologists, offering more advanced analytical methods to analyze spatial ecological data than other software. The major drawback of R is that there is a steep learning curve to master data manipulation within the R environment. This is why our book aims to knock down this learning roadblock by providing not only the relevant R code but also its interpretation. By providing key foundations in R, we hope that students and researchers will be able to adapt the R code to analyze their own data. In general, we take a practical approach to programming in R: in many cases, the code provided could be made more efficient and elegant, but we use code to provide what we hope are the most intuitive ways to go from an idea or model to the actual coding of it. We provide an Appendix that provides some necessary background on using R for those readers that are not familiar with this software. All data and R code can be downloaded at the first author's website (http://www.fletcherlab.com under "Products") and at the University of Florida's Institutional Repository (http://ufdc.ufl.edu/ufirg).

This book could be used in a variety of ways. For readers interested primarily in concepts, the first portion of each chapter will provide a concise overview on terms and concepts needed for understanding spatial ecology and conservation. For those readers who want to use spatial ecology in their own work, the examples will provide insight to the applications of terms, concepts, and modeling approaches. For those

Libros de Estadística en PDF | Statistics Books in PDF

unfamiliar with R, readers should first go through materials in the Appendix prior to working through the main text.

We hope that this book will be a useful guide for learning spatial ecology and solving applied problems facing biodiversity throughout the world.

Gainesville, FL, USA — Robert Fletcher
Toronto, ON, Canada — Marie-Josée Fortin
August 2018

Libros de Estadística en PDF | Statistics Books in PDF

Acknowledgments

Our research has been supported by several agencies over the years, and much of this research has been directly or indirectly incorporated into the development of this book. Fletcher thanks the National Science Foundation (DEB-1343144, DEB-1655555), the US Department of Agriculture (USDA-NIFA Initiative Grant No. 2012-67009-20090), the US Army Corps of Engineers, the Florida Fish and Wildlife Conservation Commission, and the University of Florida for support. Most of the data used in the book come from this research. We also thank Richard Hutto for the use of the landbird data (Chaps. 6, 7, and 11), Dave Ornato and the Florida Fish and Wildlife Conservation Commission for collecting data on panthers (Chap. 8), and Raymond Tremblay and Elvia Melendez-Ackerman for data on orchids (Chap. 10).

Fletcher is grateful for all of the support and feedback he received from his lab and students in his graduate courses over the years. Many students and colleagues encouraged the production of this book, and it would not have materialized without their encouragement. Fletcher also thanks many collaborators and mentors, including Jim Austin, Matthew Betts, Bill Clark, Brent Danielson, Nick Haddad, Richard Hutto, Rolf Koford, Tom Martin, Bob McCleery, John Orrock, and Katie Sieving. Fletcher also thanks his family for encouraging him to pursue the development of this book and his father for his support over the years.

Fortin thanks NSERC Discovery Grant Program and a Canada Research Chair in Spatial Ecology for financial support and the Department of Ecology and Evolutionary Biology of the University of Toronto for their support. Fortin also thanks her son for his support.

We thank several people for reviewing chapters of this book. Special thanks to Dan Thornton, Kirk Moloney, Kevin McGarigal, Matthew Betts, Chris Rota, Brian Reichert, Ben Baiser, Chevonne Reynolds, Rob Ahrens, Divya Vasudev, Ellen Robertson, Jessica Hightower, Brad Udell, Julian Resasco, and Noah Burrell. Their attention to detail and useful suggestions improved each chapter considerably.

Libros de Estadística en PDF | Statistics Books in PDF

Contents

1 Introduction to Spatial Ecology and Its Relevance for Conservation 1
1.1 What Is Spatial Ecology? 1
1.2 The Importance of Space in Ecology 3
1.3 The Importance of Space in Conservation 5
1.4 The Growth of Frameworks for Spatial Modeling 6
1.5 The Path Ahead 8
References 9

Part I Quantifying Spatial Pattern in Ecological Data

2 Scale 17
2.1 Introduction 17
2.2 Key Concepts and Approaches 18
2.2.1 Scale Defined and Clarified 18
2.2.2 Why Is Spatial Scale Important? 21
2.2.3 Multiscale and Multilevel Quantitative Problems 23
2.2.4 Spatial Scale and Study Design 25
2.3 Examples in R 28
2.3.1 Packages in R 28
2.3.2 The Data 28
2.3.3 A Simple Simulated Example 28
2.3.4 Multiscale Species Response to Land Cover 31
2.4 Next Steps and Advanced Issues 47
2.4.1 Identifying Characteristic Scales Beyond Species–Environment Relationships 47
2.4.2 Sampling and Scale 48
2.5 Conclusions 48
References 49

Libros de Estadística en PDF | Statistics Books in PDF

3 Land-Cover Pattern and Change 55
3.1 Introduction 55
3.2 Key Concepts 56
3.2.1 Land Use Versus Land Cover 56
3.2.2 Conceptual Models for Land Cover and Habitat Change 56
3.2.3 Habitat Loss and Fragmentation 60
3.2.4 Quantifying Land-Cover Pattern 62
3.3 Examples in R 65
3.3.1 Packages in R 65
3.3.2 The Data 65
3.3.3 Quantifying Land-Cover Variation at Different Scales 66
3.3.4 Simulating Land Cover: Neutral Landscapes 86
3.4 Next Steps and Advanced Issues 90
3.4.1 Testing for Pattern Differences Between Landscapes 90
3.4.2 Land-Cover Quantification via Image Processing 91
3.4.3 Categorical Versus Continuous Metrics 91
3.5 Conclusions 92
References 92

4 Spatial Dispersion and Point Data 101
4.1 Introduction 101
4.2 Key Concepts and Approaches 103
4.2.1 Characteristics of Point Patterns 103
4.2.2 Summary Statistics for Point Patterns 105
4.2.3 Common Statistical Models for Point Patterns 107
4.3 Examples in R 113
4.3.1 Packages in R 113
4.3.2 The Data 113
4.3.3 Creating Point Pattern Data and Visualizing It 114
4.3.4 Univariate Point Patterns 116
4.3.5 Marked Point Patterns 120
4.3.6 Inhomogeneous Point Processes and Point Process Models 123
4.3.7 Alternative Null Models 125
4.3.8 Simulating Point Processes 127
4.4 Next Steps and Advanced Issues 128
4.4.1 Space-Time Analysis 128
4.4.2 Replicated Point Patterns 128
4.5 Conclusions 129
References 129

5 Spatial Dependence and Autocorrelation 133
5.1 Introduction 133
5.2 Key Concepts and Approaches 134

Libros de Estadística en PDF | Statistics Books in PDF

5.2.1 The Causes of Spatial Dependence . . . 134
5.2.2 Why Spatial Dependence Matters . . . 135
5.2.3 Quantifying Spatial Dependence . . . 137
5.3 Examples in R . . . 142
5.3.1 Packages in R . . . 142
5.3.2 The Data . . . 143
5.3.3 Correlograms . . . 144
5.3.4 Variograms . . . 150
5.3.5 Kriging . . . 155
5.3.6 Simulating Spatially Autocorrelated Data . . . 157
5.3.7 Multiscale Analysis . . . 159
5.4 Next Steps and Advanced Issues . . . 165
5.4.1 Local Spatial Dependence . . . 165
5.4.2 Multivariate Spatial Dependence . . . 165
5.5 Conclusions . . . 166
References . . . 166

6 Accounting for Spatial Dependence in Ecological Data . . . 169
6.1 Introduction . . . 169
6.2 Key Concepts and Approaches . . . 170
6.2.1 The Problem of Spatial Dependence in Ecology and Conservation . . . 170
6.2.2 The Generalized Linear Model and Its Extensions . . . 171
6.2.3 General Types of Spatial Models . . . 175
6.2.4 Common Models that Account for Spatial Dependence . . . 176
6.2.5 Inference Versus Prediction . . . 183
6.3 Examples in R . . . 183
6.3.1 Packages in R . . . 183
6.3.2 The Data . . . 184
6.3.3 Models that Ignore Spatial Dependence . . . 185
6.3.4 Models that Account for Spatial Dependence . . . 194
6.4 Next Steps and Advanced Issues . . . 204
6.4.1 General Bayesian Models for Spatial Dependence . . . 204
6.4.2 Detection Errors and Spatial Dependence . . . 204
6.5 Conclusions . . . 205
References . . . 206

Part II Ecological Responses to Spatial Pattern and Conservation

7 Species Distributions . . . 213
7.1 Introduction . . . 213
7.2 Key Concepts and Approaches . . . 214
7.2.1 The Niche Concept . . . 214
7.2.2 Predicting Distributions or Niches? . . . 219

Libros de Estadística en PDF | Statistics Books in PDF

7.2.3 Mechanistic Versus Correlative Distribution Models 219
7.2.4 Data for Correlative Distribution Models 220
7.2.5 Common Types of Distribution Modeling Techniques 222
7.2.6 Combining Models: Ensembles 229
7.2.7 Model Evaluation 230
7.3 Examples in R 235
7.3.1 Packages in R 235
7.3.2 The Data 235
7.3.3 Prepping the Data for Modeling 236
7.3.4 Contrasting Models 240
7.3.5 Interpreting Environmental Relationships 251
7.3.6 Model Evaluation 253
7.3.7 Combining Models: Ensembles 257
7.4 Next Steps and Advanced Issues 258
7.4.1 Incorporating Dispersal 258
7.4.2 Integrating Multiple Data Sources 258
7.4.3 Dynamic Models 259
7.4.4 Multi-species Models 259
7.4.5 Sampling Error and Distribution Models 259
7.5 Conclusions 260
References 261

8 Space Use and Resource Selection 271
8.1 Introduction 271
8.2 Key Concepts and Approaches 271
8.2.1 Distinguishing Among the Diversity of Habitat-Related Concepts and Terms 271
8.2.2 Habitat Selection Theory 273
8.2.3 General Types of Habitat Use and Selection Data 278
8.2.4 Home Range and Space Use Approaches 278
8.2.5 Resource Selection Functions at Different Scales 281
8.3 Examples in R 286
8.3.1 Packages in R 286
8.3.2 The Data 286
8.3.3 Prepping the Data for Modeling 287
8.3.4 Home Range Analysis 289
8.3.5 Resource Selection Functions 298
8.4 Next Steps and Advanced Issues 310
8.4.1 Mechanistic Models and the Identification of Hidden States 310
8.4.2 Biotic Interactions 311
8.4.3 Sampling Error and Resource Selection Models 311
8.5 Conclusions 311
References 312

Libros de Estadística en PDF | Statistics Books in PDF

9 Connectivity . . . 321
9.1 Introduction . . . 321
9.2 Key Concepts and Approaches . . . 321
9.2.1 The Multiple Meanings of Connectivity . . . 322
9.2.2 The Connectivity Concept . . . 324
9.2.3 Factors Limiting Connectivity . . . 326
9.2.4 Three Common Perspectives on Quantifying Connectivity . . . 328
9.3 Examples in R . . . 335
9.3.1 Packages in R . . . 335
9.3.2 The Data . . . 336
9.3.3 Functional Connectivity Among Protected Areas for Florida Panthers . . . 336
9.3.4 Patch-Based Networks and Graph Theory . . . 345
9.3.5 Combining Connectivity Mapping with Graph Theory . . . 356
9.4 Next Steps and Advanced Issues . . . 358
9.4.1 Connectivity in Space and Time . . . 358
9.4.2 Individual-Based Models . . . 359
9.4.3 Diffusion Models . . . 359
9.4.4 Spatial Capture–Recapture . . . 359
9.5 Conclusions . . . 360
References . . . 360

10 Population Dynamics in Space . . . 369
10.1 Introduction . . . 369
10.2 Key Concepts and Approaches . . . 371
10.2.1 Foundational Population Concepts . . . 371
10.2.2 Spatial Population Concepts . . . 372
10.2.3 Population Viability Analysis . . . 382
10.2.4 Common Types of Spatial Population Models . . . 384
10.3 Examples in R . . . 387
10.3.1 Packages in R . . . 387
10.3.2 The Data . . . 388
10.3.3 Spatial Correlation and Synchrony . . . 391
10.3.4 Metapopulation Metrics . . . 394
10.3.5 Estimating Colonization–Extinction Dynamics . . . 395
10.3.6 Projecting Dynamics . . . 403
10.3.7 Metapopulation Viability and Environmental Change . . . 406
10.4 Next Steps and Advanced Issues . . . 408
10.4.1 Spatial Population Matrix Models . . . 408
10.4.2 Diffusion and Spatial Dynamics . . . 408
10.4.3 Agent-Based Models . . . 408
10.4.4 Integrated Population Models . . . 409
10.5 Conclusions . . . 409
References . . . 410

Libros de Estadística en PDF | Statistics Books in PDF

11 Spatially Structured Communities 419
11.1 Introduction 419
11.2 Key Concepts and Approaches 420
11.2.1 Spatial Community Concepts 420
11.2.2 Common Approaches to Understanding Community–Environment Relationships 428
11.2.3 Spatial Models for Communities 431
11.3 Examples in R 437
11.3.1 Packages in R 437
11.3.2 The Data 438
11.3.3 Modeling Communities and Extrapolating in Space 438
11.3.4 Spatial Dependence in Communities 460
11.3.5 Community Models with Explicit Accounting for Space 461
11.4 Next Steps and Advanced Issues 463
11.4.1 Decomposition of Space–Environment Effects 463
11.4.2 Accounting for Dependence Among Species 463
11.4.3 Spatial Networks 463
11.5 Conclusions 464
References 464

12 What Have We Learned? Looking Back and Pressing Forward 475
12.1 The Impact of Spatial Ecology and Conservation 475
12.2 Looking Forward: Frontiers for Spatial Ecology and Conservation 477
12.3 Where to Go from Here for Advanced Spatial Modeling? 478
12.4 Beyond R 479
12.5 Conclusions 480
References 481

Appendix A: An Introduction to R 489

Index 513

Libros de Estadística en PDF | Statistics Books in PDF

Chapter 1
Introduction to Spatial Ecology and Its Relevance for Conservation



1.1 What Is Spatial Ecology?

"Space: The final frontier" Kareiva (1994)

All aspects of ecology play out in space. From Darwin's entangled bank to Hutchinson's ecological theater (Hutchinson 1965; Darwin 1859), space is inherent to all processes and research in ecology. The importance of space has captured the imagination of biologists interested in a wide variety of topics, such as migration, species coexistence, deforestation, and the spread of invasive species. Therefore, how space directly and indirectly affects biodiversity and ecosystem functioning is implicitly and/or explicitly the focus of several subdisciplines in the life sciences (Fig. 1.1).

All of these subdisciplines share concepts and analytical methods that stem from the field of spatial ecology: a field coined by Tilman and Karieva in 1997. Since then, the term "spatial ecology" has been used in a wide range of ways depending on each ecological subdiscipline and field. Biogeography focuses on species geographic distributions (Lomolino 2017). Landscape ecology relates spatial heterogeneity to ecological processes and species distribution (Turner and Gardner 2015). Movement ecology focuses on organismal dispersal and migration (Nathan et al. 2008). Macroecology investigates the relation of processes and species at large spatial scales (Gaston and Blackburn 2000). Metaecology considers dispersal and spatial interactions at different spatial scales to model ecological processes that affect species distribution and dynamics (i.e., metapopulations, metacommunities, metaecosystems; Massol et al. 2011). Spatial and landscape genetics relate how landscape features affect gene glow and local adaptation (Manel et al. 2003; Guillot et al. 2009). Finally, conservation biology develops and applies spatial solutions to a variety of problems, including mitigating the effects of roads, protected area networks, and spatial prioritization in conservation planning (Primack 2014) (Fig. 1.1).

Throughout this book, we use the term spatial ecology in a broad sense referring to the study and modeling of the role(s) of space on ecological processes (e.g.,

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_1

Libros de Estadística en PDF | Statistics Books in PDF

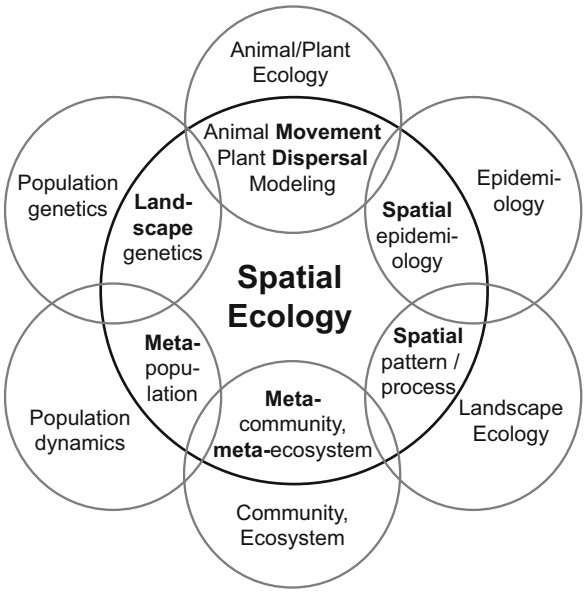


Fig. 1.1 Spatial subdisciplines derived from ecological disciplines using a spatial ecology framework to tackle current conservation issues

population dynamics, species interactions, dispersal) that in turn affects ecological patterns, such as species distributions. This definition shares similarities with some early definitions of landscape ecology (Pickett and Cadenasso 1995; Turner 1989). Yet over the years, landscape ecology evolved to include socio-economic aspects of landscapes as well (Wu 2017).

Research in spatial ecology aims to understand the processes that affect species distributions and dynamics, and how these processes play out across space. Endogenous processes are related to the dynamics of each ecological entity (e.g., movement, dispersal, and migration) and the interactions among entities within and across species (population demographics, genetic variation, behavior, competition, facilitation, trophic interactions, etc.). Exogenous processes are related to the response of organisms to environmental factors that are themselves spatially structured (climate, local habitat features, microhabitat heterogeneity, patch disturbance-succession, environmental filtering, historical contingencies, etc.). Overall, it is the combined action and feedback effects of these endogenous and exogenous processes that result in the spatial patterns observed at different levels of organization though space (e.g., metapopulations, metacommunities, and metaecosystems) (Fig. 1.2).

Spatial ecology is increasingly applied to conservation and management to help deliver more effective ways to conserve biodiversity. The rapid rate at which landscapes are altered is creating spatially heterogeneous environmental conditions that affect species ability to disperse and ultimately persist. Yet, even in homogeneous environments, endogenous processes alone can shape species spatial distributions (Okubo 1974). This is why many of the core ecological theories and analytical models used in spatial ecology are process-based ones. Therefore, one of the most important cornerstones of spatial ecology as a discipline is the way in

Libros de Estadística en PDF | Statistics Books in PDF

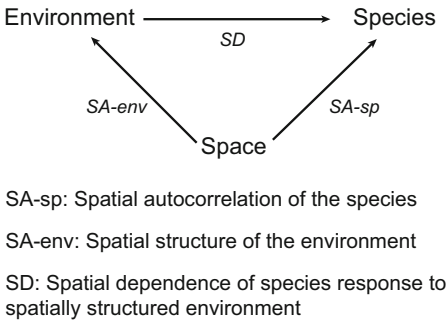


Fig. 1.2 How space affects both the spatial structure of the environmental conditions and species distribution. Species distribution is also affected by the spatial structure of the environmental data (adapted from Wagner and Fortin 2005)

which the challenges of understanding the processes underlying the spatial distribution of ecological entities are tackled. Spatial ecology offers concepts and tools to understand, predict, and map how biodiversity responds to environmental change.

1.2 The Importance of Space in Ecology

Species dynamics occur over space and time. Space affects species in multiple ways from how they use resources and occupy space within their home range and throughout their geographical range, how they move, disperse, and migrate through heterogeneous landscapes, as well as how they interact with other species (Table 1.1).

To determine the relative importance of space on ecological patterns and processes, both mathematical and statistical models are frequently used (Dale and Fortin 2014; Cantrell et al. 2009; Fortin et al. 2012; Ovaskainen et al. 2016). These two modeling approaches encompass stark differences from data needs, model assumptions, and epistemologies (Fig. 1.2). Both process-based (e.g., mathematical, stochastic simulations and computational models) and phenomenological approaches (e.g., statistical regression models) have a long history of contributing to our understanding of the spatial distribution of ecological entities from fine to broad scales (Levin 1976; MacArthur and Wilson 1967). Such spatial models aim to improve our understanding of the underlying processes acting on species distributions (e.g., to estimate the relative importance of environmental drivers versus dispersal to species distributions) and to perform ecological forecasting (e.g., to predict species distributions based on such processes; Pagel and Schurr 2012; Dietze 2017).

The foundation for spatial ecology can be traced largely to the seminal paper of Watt (1947) on the relationship between spatial pattern and ecological processes. Watt (1947) emphasized that plants occurred in bounded communities—patches—that form a dynamic mosaic across the landscape, what has become known as the

Libros de Estadística en PDF | Statistics Books in PDF

Table 1.1 Examples of how space can be incorporated into spatial analyses and their effects on ecological processes (adapted from Fortin et al. (2012))

Spatial aspects	Effects on ecological processes and data
x–*y* coordinates	Location of data according to positions of other locations (Euclidean or relative distance)
Spatial autocorrelation	The magnitude, spatial scale, and directionality of data values as a function of distances between data point locations
Spatial relationship	Locations of abiotic predictors affect the responses of biotic/ecological variables
Spatial legacy	Influence of past spatial pattern on current ecological processes and species current spatial pattern
Spatial contingency	Influence of nearby locations (local neighbors) on ecological processes and species spatial pattern
Spatial perception	How the intervening landscape features affect daily animal movement and species dispersal ability
Multiple spatial scales	Additive spatial scales influence current spatial pattern

"shifting-mosaic steady state" concept (Bormann and Likens 1979). Then, in the 1950s and 1960s, there were three key areas of research that emphasized the importance of space for ecological processes and its relevance for conservation. First, some influential experimental studies highlighted the importance of space for ecology. In a seminal experiment, Huffaker (1958) showed how predator–prey dynamics could be stable when including the potential for spatial refugia of prey, while stability was not possible in small, homogenous habitats. This result was important because prior to that time, spatial concepts had not been formally considered in theory and concepts regarding species coexistence. This experiment emphasized the role of movement in altering species interactions and community structure, a theme that has persisted and grown over time.

A second area of conceptual development came from theoretical ecology (Hastings and Gross 2012), where ecologists investigated how diffusion of organisms through space can alter population and community dynamics (Skellam 1951; Okubo and Levin 2001; Hilborn 1979). Skellam (1951) pioneered these ideas by applying reaction–diffusion models originally derived for molecular processes to the problem of dispersal and population dynamics. In this model, Skellam (1951) assumed diffusion (or random movement) of organisms. While it is clear that organisms do not move in a simple random manner, the utility of this approach is that this simple formulation can go a long way in explaining observed patterns in ecology (Kareiva 1982, 1983), and it can be extended to capture non-random issues (e.g., advection; Reeve et al. 2008). In addition, Skellam's work set the stage for modeling invasive spread, a topic of great importance to conservation biology.

The third area is the application of biogeographic concepts to our understanding of species–area relationships by Preston (1948, 1962) and later MacArthur and

Libros de Estadística en PDF | Statistics Books in PDF

Wilson (1963, 1967). This area was particularly crucial in developing the application of spatial ecology to practical issues of conservation (Higgs 1981). Indeed, many ecological theories and conservation concepts, including practical solutions, stem from island biogeography theory, where the size of islands/patches and their spatial configuration (spacing/isolation) are critical for species persistence through variation in colonization and extinction events (MacArthur and Wilson 1967; Laurance 2008).

The current era of spatial ecology has grown from island biogeography, where dispersal of individuals is key and can act as a rescue effect or spatial insurance (Loreau et al. 2003a) that protects a population from local extinction. Here, species are often considered to act as metapopulations (Hanski 1999; Levins 1969). The concept of spatial insurance has been extended to dispersal of several species to maintain species assemblages and communities as metacommunities (Leibold et al. 2017, 2004) and to maintain ecosystem functions as metaecosystems (Loreau et al. 2003b; Guichard 2017).

1.3 The Importance of Space in Conservation

Conservation biologists have increasingly embraced the importance of space in the conservation of biodiversity and ecosystem services (Schagner et al. 2013; Moilanen et al. 2009). Space is relevant for conservation in four major ways: (1) it is essential for spatial mapping of biodiversity and ecosystem services; (2) it provides guidance for mitigating effects of environmental change; (3) it facilitates effective prioritization of areas for conservation; and (4) it provides key components of tools and models used in conservation.

Several biogeography and macroecology theories provide spatial foundations for understanding and mapping biodiversity across the planet. The emphasis on spatial components first emerged in the field of biogeography, where there was interest in identifying and understanding species distributions and geographic gradients in biodiversity throughout the world. For instance, early on scientists emphasized the latitudinal gradient of diversity, where diversity was greater in the tropics than in the temperate zone (Currie and Paquin 1987). Understanding this and other biogeographic (and macroecological) patterns have been, and continue to be, of interest in conservation as it helps identify hotspots of biodiversity and endemism of conservation relevance (Myers et al. 2000; Dawson et al. 2017; Orme et al. 2005).

Many approaches to mitigating the effects of environmental change embrace spatial concepts. For example, the use of corridors in conservation explicitly emphasizes how the spatial configuration of the environment can promote biodiversity (Crooks and Sanjayan 2006). Translocations and re-introduction programs require understanding how potential release locations may inhibit or foster the success of such programs (Seddon et al. 2014). Adaptation strategies to mitigate the effects of climate change often emphasize spatial ecological concepts (Heller and Zavaleta 2009).

Libros de Estadística en PDF | Statistics Books in PDF

Conservation prioritization and planning, one of the major foci for conservation biology, also emphasizes the importance of spatial ecology. Early rules for conservation planning embraced the need to limit isolation of protected areas and maximize their area (Diamond 1975). Later work has embraced explicit mapping of conservation prioritization strategies and how issues such as complementarity of biodiversity among protected areas is essential for efficient conservation planning (Margules and Pressey 2000). More recently, conservation planning for climate change emphasizes how key areas are currently connected and how connectivity may change as climate and land use continue to change (Pressey et al. 2007; Schmitz et al. 2015; Carroll et al. 2017). Throughout, spatial concepts are essential for guiding effective strategies for both biodiversity and ecosystem service conservation (Chan et al. 2006; Moilanen and Wintle 2007).

Ecological concepts and analytical tools developed in the fields of landscape ecology, geography, and spatial statistics are now commonly used in conservation so that informed decisions about planning strategies and management can be made (e.g., Moilanen et al. 2009). Indeed, most conservation planning and management requires knowledge and the explicit spatial modeling of space and its major consequences on species spatial variation and responses to global change. The inclusion of space is therefore crucial when modeling species ecology and responses to a changing world such as (1) species dispersal, (2) species interactions, (3) disturbance dynamics, and (4) environmental change. Furthermore, as the field of conservation aims to provide better management recommendations to mitigate threats to biodiversity, implicit and explicit aspects of space need to be incorporated into applied solutions such as restoration, species reintroductions, and maintaining connectivity among habitat patches. In all these conservation applications the spatial scale of implementation is key (Wiens 1989; Levin 1992, 2000; McGarigal et al. 2016; Doak et al. 1992; Fletcher et al. 2013; Gering et al. 2003).

1.4 The Growth of Frameworks for Spatial Modeling

Before modeling species dispersal, response to environmental conditions, and species interactions, quantification of their spatial distribution is needed. This is why in ecology and conservation the first steps toward a better understanding and management of biodiversity often consist of (1) mapping species distributions, and (2) quantifying spatial patterns of both species distributions and environmental conditions (Ferrier 2002; Gaston and Blackburn 2000; Guisan and Thuiller 2005). Once such quantitative information is obtained, the next modeling steps frequently aim at relating and modeling the responses of species to environmental conditions across space and/or the species (intraspecific and interspecific) spatial interactions (Synes et al. 2017).

Modeling the processes that affect species distribution can be done using different degrees of complexity in the analytical tools used. The level of complexity depends on the processes modeled and ecological theories considered. Then, knowledge gaps about

Libros de Estadística en PDF | Statistics Books in PDF

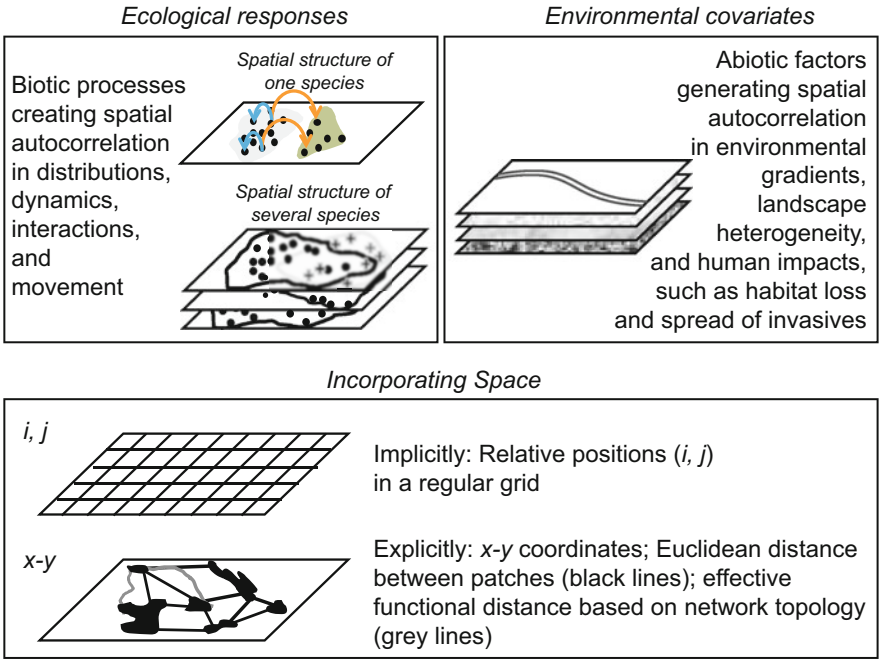


Fig. 1.3 How spatial processes affect species (response variables) and covariates (predictors), and how space can be incorporated into models

species distribution can be gained by combining data on species behavior from empirical studies and theoretical models of dispersal and related flows across space. Early dispersal models set the stage for the development of ecological theories that embrace space (Fig. 1.3), such as island biogeography (MacArthur and Wilson 1967), patch dynamics (Pickett and White 1984), hierarchical theory (Wu and Loucks 1995; Allen and Starr 1982), species coexistence (Chesson 2000), metapopulation (Hanski 1999), metacommunity (Leibold and Chase 2017), and metaecosystem theory (Guichard 2017). Although these disciplines can be seen as separate fields, spatial ecology brings them together through theory, models, and data analysis (Massol et al. 2011).

The emergence of modeling frameworks for spatial ecology was also fostered by several technological advances ranging from the availability of aerial photographs, remote sensing captors, and computing power. This allowed for conceptual and modeling developments in spatial ecology to advance more realistic ways to represent and incorporate space into statistical and modeling approaches (Fig. 1.3). Indeed, the ability to explicitly include the effects of space in ecological models was also pivotal in the explosion of novel ecological questions and analytical ways to address them over the last few decades.

Libros de Estadística en PDF | Statistics Books in PDF

The quantum leap in spatial ecology modeling frameworks involved considering and incorporating space into modeling: implicitly (kernels, moving windows, relative topological position, etc.), explicitly (*x*–*y*, diffusion, spread, individual/agent-based models, etc.), and realistically (explicit network structure, spatial weights, multiple spatial scales, etc.) (Fig. 1.3). It started by considering space as discrete units. Such discretization of space opened a multitude of novel ways to model ecological systems either in a spatially implicit fashion, where species occupancy and abundance are modeled considering the effects of relative neighbors based on grid topology (e.g., cellular automata models), or in a spatially explicit way, where the actual Euclidean distances among cells (quadrats, pixels, sampling locations) are used to model the spread of disturbance, disease, or species using dispersal kernels. Then space was represented by the exact *x*–*y* coordinates of each individual in a given area such that the spatially explicit movement of individuals could be modeled using individual/agent-based modeling approaches (Grimm et al. 2005; Matthews et al. 2007). For example, this approach enabled modeling the dynamics and succession of tree species at the tree-level using SORTIE (Pacala et al. 1996). Using *x*–*y* coordinates of individuals or sampling locations also allowed the spatially explicit modeling of movement and connectivity while accounting for species dispersal ability through spatially heterogeneous landscapes (Urban and Keitt 2001). Lastly, the spatially explicit representation of space permits us to model processes acting over several spatial scales using meta-models (Urban 2005; Talluto et al. 2016). The ability to model species and their responses to global change explicitly in space opens the door to investigate the effects of the spatial legacy (Wallin et al. 1994; James et al. 2007; Peterson 2002) of heterogeneity on ecological processes and species persistence.

1.5 The Path Ahead

Spatial ecology and conservation has rapidly advanced over the past 20 years. With an increasing emphasis on the use of spatial data and modeling to address both fundamental and applied problems, the topic has matured. Spatial ecology embraces spatial modeling and analysis, which is often applied to conservation issues.

In the remainder of this book, our path will be to provide an introduction to several issues in spatial ecology and conservation, with an emphasis on spatial modeling of applied ecological problems. We emphasize learning-by-doing, where we illustrate these topics with real data and the application of spatial modeling to these topics. We first cover topics regarding the quantification of spatial pattern in ecological data and we then focus more specifically on topics regarding how species respond to spatial pattern and its relevance for conservation (Table 1.2). We hope that this coverage will deliver a strong foundation for students and professionals alike to begin tackling ongoing issues of ecological and conservation importance.

Libros de Estadística en PDF | Statistics Books in PDF

Table 1.2 Examples of spatial analytical methods (and book chapter(s) where they are presented) used in spatial ecology and the quantitative components that they are estimating

Spatial analytical methods	Spatial components addressed
Multiscale analysis (Chap. 2)	Determine key spatial scales affecting the response variables
Categorical pattern analysis (Chap. 3)	Quantify land-use and land-cover patterns
Spatial point processes (Chap. 4)	Identifying the spatial pattern of points (events) and understand the potential processes generating those patterns
Spatial-, geo-statistics (Chap. 5)	Magnitude, range, and directionality of spatial variance
Spatial regressions (Chap. 6)	Accounting for spatial structure of the response (spatial nuisance) and independent variables (spatial contingency) in estimating relationships
Species distribution models (Chap. 7)	Interpolation, projections, and forecasting
Animal movement models (Chaps. 8 and 9)	Accounting for spatial heterogeneity and quantifying trajectories
Spatial network analysis (Chap. 9)	Topological network, Euclidean, and functional distances
Spatial population dynamics (Chap. 10)	Population dynamics accounting for spatial heterogeneity
Beta diversity (Chap. 11)	Spatial species turnover
Spatial community analysis (Chap. 11)	Spatial components of species interactions and environmental filtering

References

Allen TFH, Starr TB (1982) Hierarchy: perspectives for ecological complexity. University of Chicago Press, Chicago

Bormann FH, Likens GE (1979) Catastrophic disturbance and the steady-state in northern hardwood forests. Am Sci 67(6):660–669

Cantrell S, Cosner C, Ruan S (2009) Spatial ecology. Chapman & Hall/CRC, Boca Raton, FL

Carroll C, Roberts DR, Michalak JL, Lawler JJ, Nielsen SE, Stralberg D, Hamann A, McRae BH, Wang TL (2017) Scale-dependent complementarity of climatic velocity and environmental diversity for identifying priority areas for conservation under climate change. Glob Chang Biol 23(11):4508–4520. https://doi.org/10.1111/gcb.13679

Chan KMA, Shaw MR, Cameron DR, Underwood EC, Daily GC (2006) Conservation planning for ecosystem services. PLoS Biol 4(11):2138–2152. https://doi.org/10.1371/journal.pbio.0040379

Chesson PL (2000) General theory of competitive coexistence in spatially varying environments. Theor Popul Biol 58:211–237

Crooks KR, Sanjayan M (eds) (2006) Connectivity conservation. Cambridge University Press, New York

Currie DJ, Paquin V (1987) Large-scale biogeographical patterns of species richness of trees. Nature 329(6137):326–327. https://doi.org/10.1038/329326a0

Dale MRT, Fortin MJ (2014) Spatial analysis: a guide for ecologists, 2nd edn. Cambridge University Press, Cambridge

Darwin C (1859) On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. John Murray, London

Libros de Estadística en PDF | Statistics Books in PDF

Dawson W, Moser D, van Kleunen M, Kreft H, Pergl J, Pysek P, Weigelt P, Winter M, Lenzner B, Blackburn TM, Dyer EE, Cassey P, Scrivens SL, Economo EP, Guenard B, Capinha C, Seebens H, Garcia-Diaz P, Nentwig W, Garcia-Berthou E, Casal C, Mandrak NE, Fuller P, Meyer C, Essl F (2017) Global hotspots and correlates of alien species richness across taxonomic groups. Nat Ecol Evol 1(7):0186. https://doi.org/10.1038/s41559-017-0186

Diamond JM (1975) The island dilemma: lessons of modern biogeographic studies for the design of natural reserves. Biol Conserv 7(2):129–146. https://doi.org/10.1016/0006-3207(75)90052-x

Dietze M (2017) Ecological forcasting. Princeton University Press, Princeton

Doak DF, Marino PC, Kareiva PM (1992) Spatial scale mediates the influence of habitat fragmentation on dispersal success: implications for conservation. Theor Popul Biol 41(3):315–336. https://doi.org/10.1016/0040-5809(92)90032-o

Ferrier S (2002) Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? Syst Biol 51(2):331–363. https://doi.org/10.1080/10635150252899806

Fletcher RJ Jr, Revell A, Reichert BE, Kitchens WM, Dixon JD, Austin JD (2013) Network modularity reveals critical scales for connectivity in ecology and evolution. Nat Commun 4:2572. https://doi.org/10.1038/ncomms3572

Fortin MJ, James PMA, MacKenzie A, Melles SJ, Rayfield B (2012) Spatial statistics, spatial regression, and graph theory in ecology. Spatial Stat 1:100–109. https://doi.org/10.1016/j.spasta.2012.02.004

Gaston KJ, Blackburn TM (2000) Pattern and process in macroecology. Blackwell Science, Oxford, UK

Gering JC, Crist TO, Veech JA (2003) Additive partitioning of species diversity across multiple spatial scales: implications for regional conservation of biodiversity. Conserv Biol 17 (2):488–499. https://doi.org/10.1046/j.1523-1739.2003.01465.x

Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke HH, Weiner J, Wiegand T, DeAngelis DL (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310(5750):987–991. https://doi.org/10.1126/science.1116681

Guichard F (2017) Recent advances in metacommunities and meta-ecosystem theories. F1000 Research 6:610

Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. Mol Ecol 18(23):4734–4756. https://doi.org/10.1111/j.1365-294X.2009.04410.x

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8(9):993–1009

Hanski I (1999) Metapopulation ecology. Oxford University Press, Oxford

Hastings A, Gross L (eds) (2012) Encyclopedia of theoretical ecology. UC Press, Berkeley, CA

Heller NE, Zavaleta ES (2009) Biodiversity management in the face of climate change: a review of 22 years of recommendations. Biol Conserv 142(1):14–32. https://doi.org/10.1016/j.biocon.2008.10.006

Higgs AJ (1981) Island biogeography theory and nature reserve design. J Biogeogr 8(2):117–124. https://doi.org/10.2307/2844554

Hilborn R (1979) Some long-term dynamics of predator-prey models with diffusion. Ecol Model 6 (1):23–30. https://doi.org/10.1016/0304-3800(79)90055-3

Huffaker CB (1958) Experimental studies on predation: dispersion factors and predator-prey oscillations. Hilgardia 27:343–383

Hutchinson GE (1965) The ecological theater and the evolutionary play. Yale University Press, New Haven

James PMA, Fortin MJ, Fall A, Kneeshaw D, Messier C (2007) The effects of spatial legacies following shifting management practices and fire on boreal forest age structure. Ecosystems 10 (8):1261–1277. https://doi.org/10.1007/s10021-007-9095-y

Kareiva P (1982) Experimental and mathematical analyses of herbivore movement: quantifying the influence of plant spacing and quality on foraging discrimination. Ecol Monogr 52(3):261–282. https://doi.org/10.2307/2937331

Libros de Estadística en PDF | Statistics Books in PDF

Kareiva PM (1983) Local movement in herbivorous insects - applying a passive diffusion-model to mark-recapture field experiments. Oecologia 57(3):322–327. https://doi.org/10.1007/bf00377175
Kareiva P (1994) Space: the final frontier for ecological theory. Ecology 75(1):1–1. https://doi.org/10.2307/1939376
Laurance WF (2008) Theory meets reality: how habitat fragmentation research has transcended island biogeographic theory. Biol Conserv 141(7):1731–1744. https://doi.org/10.1016/j.biocon.2008.05.011
Leibold MA, Chase JM (2017) Metacommunity ecology. Princeton University Press, Princeton, NJ
Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman D, Loreau M, Gonzalez A (2004) The metacommunity concept: a framework for multi-scale community ecology. Ecol Lett 7(7):601–613. https://doi.org/10.1111/j.1461-0248.2004.00608.x
Leibold MA, Chase JM, Ernest SKM (2017) Community assembly and the functioning of ecosystems: how metacommunity processes alter ecosystems attributes. Ecology 98(4):909–919
Levin SA (1976) Population dynamic models in heterogeneous environments. Annu Rev Ecol Syst 7:287–310. https://doi.org/10.1146/annurev.es.07.110176.001443
Levin SA (1992) The problem of pattern and scale in ecology. Ecology 73(6):1943–1967. https://doi.org/10.2307/1941447
Levin SA (2000) Multiple scales and the maintenance of biodiversity. Ecosystems 3(6):498–506. https://doi.org/10.1007/s100210000044
Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. Bull Entomol Soc Am 15:237–240
Lomolino MV (2017) Biogeography: biological diversity across space and time, 5th edn. Sinauer, Sunderland, MA
Loreau M, Mouquet N, Gonzalez A (2003a) Biodiversity as spatial insurance in heterogeneous landscapes. Proc Natl Acad Sci U S A 100(22):12765–12770. https://doi.org/10.1073/pnas.2235465100
Loreau M, Mouquet N, Holt RD (2003b) Meta-ecosystems: a theoretical framework for a spatial ecosystem ecology. Ecol Lett 6(8):673–679. https://doi.org/10.1046/j.1461-0248.2003.00483.x
MacArthur RH, Wilson EO (1963) Equilibrium theory of insular zoogeography. Evolution 17(4):373. https://doi.org/10.2307/2407089
MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press, Princeton, NJ
Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18(4):189–197. https://doi.org/10.1016/s0169-5347(03)00008-9
Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405(6783):243–253. https://doi.org/10.1038/35012251
Massol F, Gravel D, Mouquet N, Cadotte MW, Fukami T, Leibold MA (2011) Linking community and ecosystem dynamics through spatial ecology. Ecol Lett 14(3):313–323. https://doi.org/10.1111/j.1461-0248.2011.01588.x
Matthews RB, Gilbert NG, Roach A, Polhill JG, Gotts NM (2007) Agent-based land-use models: a review of applications. Landsc Ecol 22(10):1447–1459. https://doi.org/10.1007/s10980-007-9135-1
McGarigal K, Wan HY, Zeller KA, Timm BC, Cushman SA (2016) Multi-scale habitat selection modeling: a review and outlook. Landsc Ecol. https://doi.org/10.1007/s10980-016-0374-x
Moilanen A, Wintle BA (2007) The boundary-quality penalty: a quantitative method for approximating species responses to fragmentation in reserve selection. Conserv Biol 21(2):355–364. https://doi.org/10.1111/j.1523-1739.2006.00625.x
Moilanen A, Wilson KA, Possingham H (eds) (2009) Spatial conservation prioritization: quantitative methods and computational tools. Oxford University Press, Oxford

Libros de Estadística en PDF | Statistics Books in PDF

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403(6772):853–858. https://doi.org/10.1038/35002501

Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE (2008) A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci U S A 105 (49):19052–19059. https://doi.org/10.1073/pnas.0800375105

Okubo A (1974) Diffusion-induced instability in model ecosystems, another possible explanation of patchiness. Technical Report 86. Chesapeake Bay Institute, MD

Okubo A, Levin SA (2001) Diffusion and ecological problems: modern perspectives. Springer, New York

Orme CDL, Davies RG, Burgess M, Eigenbrod F, Pickup N, Olson VA, Webster AJ, Ding TS, Rasmussen PC, Ridgely RS, Stattersfield AJ, Bennett PM, Blackburn TM, Gaston KJ, Owens IPF (2005) Global hotspots of species richness are not congruent with endemism or threat. Nature 436(7053):1016–1019. https://doi.org/10.1038/nature03850

Ovaskainen O, De Knegt HJ, del Mar Delgado M (2016) Quantitative ecology and evolutionary biology: integrating models with data. Oxford University Press, Oxford

Pacala SW, Canham CD, Saponara J, Silander JA, Kobe RK, Ribbens E (1996) Forest models defined by field measurements: estimation, error analysis and dynamics. Ecol Monogr 66 (1):1–43. https://doi.org/10.2307/2963479

Pagel J, Schurr FM (2012) Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. Glob Ecol Biogeogr 21(2):293–304. https://doi.org/10.1111/j.1466-8238.2011.00663.x

Peterson GD (2002) Contagious disturbance, ecological memory, and the emergence of landscape pattern. Ecosystems 5(4):329–338. https://doi.org/10.1007/s10021-001-0077-1

Pickett STA, Cadenasso ML (1995) Landscape ecology: spatial heterogeneity in ecological systems. Science 269(5222):331–334

Pickett STA, White PS (1984) The ecology of natural disturbance and patch dynamics. Academic Press, New York

Pressey RL, Cabeza M, Watts ME, Cowling RM, Wilson KA (2007) Conservation planning in a changing world. Trends Ecol Evol 22(11):583–592. https://doi.org/10.1016/j.tree.2007.10.001

Preston FW (1948) The commonness, and rarity, of species. Ecology 29(3):254–283. https://doi.org/10.2307/1930989

Preston FW (1962) The canonical distribution of commonness and rarity: Part I. Ecology 43 (2):185–215,431–432

Primack RB (2014) Essentials of conservation biology, 6th edn. Sinauer Associates, Sunderland, MA

Reeve JD, Cronin JT, Haynes KJ (2008) Diffusion models for animals in complex landscapes: incorporating heterogeneity among substrates, individuals and edge behaviours. J Anim Ecol 77 (5):898–904. https://doi.org/10.1111/j.1365-2656.2008.01411.x

Schagner JP, Brander L, Maes J, Hartje V (2013) Mapping ecosystem services' values: current practice and future prospects. Ecosyst Serv 4:33–46. https://doi.org/10.1016/j.ecoser.2013.02.003

Schmitz OJ, Lawler JJ, Beier P, Groves C, Knight G, Boyce DA, Bulluck J, Johnston KM, Klein ML, Muller K, Pierce DJ, Singleton WR, Strittholt JR, Theobald DM, Trombulak SC, Trainor A (2015) Conserving biodiversity: practical guidance about climate change adaptation approaches in support of land-use planning. Nat Areas J 35(1):190–203

Seddon PJ, Griffiths CJ, Soorae PS, Armstrong DP (2014) Reversing defaunation: restoring species in a changing world. Science 345(6195):406–412. https://doi.org/10.1126/science.1251818

Skellam JG (1951) Random dispersal in theoretical populations. Biometrika 28:196–218

Synes NW, Brown C, Watts K, White SM, Gilbert MA, Travis JM (2017) Emerging opportuniteis for landscape ecological modelling. Curr Landsc Ecol Rep 1:146–167

Talluto MV, Boulangeat I, Ameztegui A, Aubin I, Berteaux D, Butler A, Doyon F, Drever CR, Fortin MJ, Franceschini T, Lienard J, McKenney D, Solarik KA, Strigul N, Thuiller W, Gravel D (2016) Cross-scale integration of knowledge for predicting species ranges: a metamodelling framework. Glob Ecol Biogeogr 25(2):238–249. https://doi.org/10.1111/geb.12395

Tilman D, Kareiva P (1997) Spatial ecology: the role of space in population dynamics and interspecific interactions. Princeton University Press, Princeton, NJ

Libros de Estadística en PDF | Statistics Books in PDF

Turner MG (1989) Landscape ecology: the effect of pattern on process. Annu Rev Ecol Syst 20:171–197. https://doi.org/10.1146/annurev.ecolsys.20.1.171
Turner MG, Gardner RH (2015) Landscape ecology in theory and practice, 2nd edn. Springer, New York
Urban DL (2005) Modeling ecological processes across scales. Ecology 86(8):1996–2006
Urban D, Keitt T (2001) Landscape connectivity: a graph-theoretic perspective. Ecology 82 (5):1205–1218
Wagner HH, Fortin MJ (2005) Spatial analysis of landscapes: concepts and statistics. Ecology 86 (8):1975–1987. https://doi.org/10.1890/04-0914
Wallin DO, Swanson FJ, Marks B (1994) Landscape pattern response to changes in pattern generation rules: land-use legacies in forestry. Ecol Appl 4(3):569–580. https://doi.org/10.2307/1941958
Watt AS (1947) Pattern and process in the plant community. J Ecol 35(1–2):1–22. https://doi.org/10.2307/2256497
Wiens JA (1989) Spatial scaling in ecology. Funct Ecol 3(4):385–397
Wu JG (2017) Thirty years of landscape ecology (1987–2017): retrospects and prospects. Landsc Ecol 32(12):2225–2239. https://doi.org/10.1007/s10980-017-0594-8
Wu JG, Loucks OL (1995) From balance of nature to hierarchical patch dynamics: a paradigm shift in ecology. Q Rev Biol 70(4):439–466. https://doi.org/10.1086/419172

Libros de Estadística en PDF | Statistics Books in PDF

Part I
Quantifying Spatial Pattern in Ecological Data

Libros de Estadística en PDF | Statistics Books in PDF

Chapter 2
Scale



2.1 Introduction

All questions in ecology and problems in conservation have a spatial and temporal dimension, and scale is a concept that attempts to capture those dimensions and make sense of them. *Scale* describes the spatiotemporal dimension of a pattern or process. By understanding and quantifying scale, it can profoundly influence our understanding of ecological patterns and processes, altering conclusions regarding behavior, population viability, species interactions, evolutionary dynamics, and conservation decisions. In addition, many patterns and processes occur at fundamentally different scales in space and time (Fig. 2.1). Of great interest are critical (or characteristic) scales—scales that govern the dynamics of key ecological and evolutionary processes (Urban et al. 1987).

The problem of scale in ecology and conservation has roots dating back for many decades (Greig-Smith 1952; Preston 1962), yet it was in the late 1980s and early 1990s that the concept of scale emerged as a central theme. Prior to this time, there was frequently an assumption of *scale invariance*, that is, that spatial patterns and processes did not change with scale, and therefore the problem of scale was not essential to address. Yet with seminal syntheses by John Wiens (1989), Simon Levin (1992), and others, several arguments were put forward regarding how pervasive the problem of scale is to ecology and conservation. At that time, it was offered that explicit consideration of scale can help solve some debates in ecology, such as the role of competition on community assembly, and could also help conservation and management problems through explicit accounting for processes operating at different scales. Since that seminal work, there have been major advances in our understanding of scale (Chave 2013; Jackson and Fahrig 2015), as well as new quantitative methods to interpret the role of scale (Keitt and Urban 2005; Dray et al. 2012; Fortin et al. 2012; Chandler and Hepinstall-Cymerman 2016).

In this chapter, we provide an overview of the issues related to scale, why it is important, and how it is often considered in spatial ecology and conservation. Our

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_2

Libros de Estadística en PDF | Statistics Books in PDF

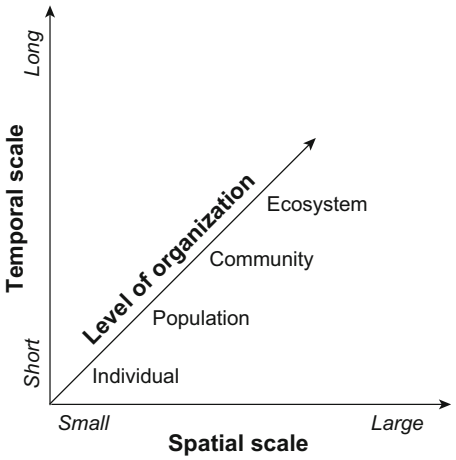


Fig. 2.1 The problem of scale and levels of organization, viewed through a modified space-time diagram. Scale can be interpreted in the context of the temporal dimension, the spatial dimension, and the level of organization of a pattern or process. Temporal dimensions can be viewed as short (e.g., daily variation in temperature) or long (e.g., glacial cycles). Spatial dimensions can be viewed as small (or fine; e.g., spatial variation in humidity) or large (or coarse; e.g., spatial variation in rainfall). Level of organization refers to the place in the biotic hierarchy for which a pattern or process occurs, such as at the individual-level or at the population-level

goals are to highlight terminology used in understanding spatial scale, provide examples of why scale is important, and illustrate some simple approaches to multiscale and multilevel modeling in R with an example of reptile distribution in the Southeastern USA. This chapter is intended to introduce this topic, but we dive into various issues and methods related to scale and its quantification throughout the book.

2.2 Key Concepts and Approaches

2.2.1 Scale Defined and Clarified

The term *scale* refers to the spatiotemporal dimension or domain of a process or a pattern (Table 2.1). In landscape ecology, scale is frequently described by its components: grain and extent (Fortin et al. 2012; Turner and Gardner 2015). *Grain* is the finest spatial unit of measurement for a pattern or process, while *extent* describes the length or area under investigation. *Scope*, or the ratio of grain to extent, is also frequently mentioned in the description of scale because this ratio can play a large role on observed patterns and processes (see below; Schneider 2001).

Grain and extent tend to covary—investigations with larger extent tended to have larger grain as well (Wiens 1989). Part of this covariance is practical: it is difficult to

Libros de Estadística en PDF | Statistics Books in PDF

Table 2.1 Terms and definitions of scale concepts

		Applicability to		
Term	Definition	Phenomenon	Sampling	Analysis
Cartographic ratio	Ratio between the distance on a map and the real-world distance. Sometimes referred to as "cartographic scale".		X	X
Characteristic scale	The scale at which the dominant pattern emerges.	X		X
Cross-scale interaction	Processes at one scale interact with processes at another scale.	X		
Ecological fallacy	A logical fallacy in which inferences about individual units are deduced from higher-level group characteristics or averages.	X		X
Ecological neighborhood	The region within which that an organism is active or has some influence during an appropriate period of time.	X		
Extent	The total length or area under consideration.	X	X	X
Grain	The finest level of spatial resolution of a data set, pattern, or process.	X	X	X
Hierarchy	A system of interconnections wherein the higher levels constrain the lower levels to various degrees, depending on the time constraints of the behavior.	X		
Lag	The interval or spacing (distance) between units.		X	X
Level of organization	The place within a biotic hierarchy.	X		
Modifiable aerial unit problem	A bias that occurs when point-based (or pixel-based) variables are aggregated, such that the summaries are influenced by the shape and scale of the aggregation unit.		X	X
Resolution	The finest level of measurement; for spatial measurements; resolution is equivalent to grain size.		X	X
Scale	The spatial or temporal domain of a pattern or process.	X	X	X
Scale of effect	The scale (typically extent considered around a sampling location) at which most variability is explained. Sometimes referred to as characteristic scale.	X		
Scale invariance	When spatial patterns and processes do not change with scale.	X	X	X
Scope	The ratio of grain to extent.	X	X	X
Support	The size, shape, and orientation of a sample of data. Can be applied to both spatial and temporal support.		X	X

For each term, we highlight whether it is relevant to scales of phenomena, sampling, and/or analysis (*sensu* Dungan et al. 2002)

Libros de Estadística en PDF | Statistics Books in PDF

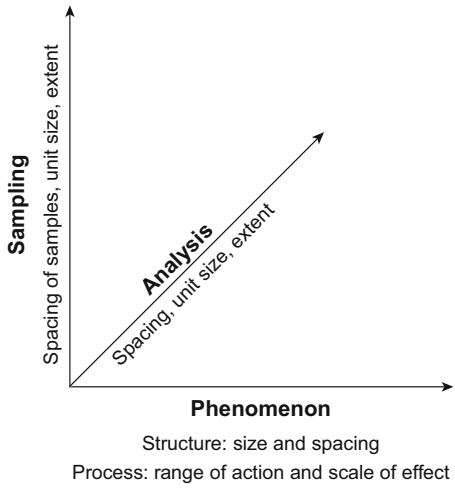


Fig. 2.2 The dimensions of spatial scale concepts. Spatial scale can be interpreted based on the phenomenon of interest, the sampling that occurs, or the analysis. These different aspects of spatial scale have led to some confusion in the spatial scale literature. Modified from Dungan et al. (2002) and Dale and Fortin (2014)

work at large extents with data collected at fine grain sizes. However, part of this covariance is also conceptual: often at large extents, we may expect processes operating at very fine grains to provide mostly "noise" in the systems. As computational challenges decline and availability of high-resolution data has increased, the covariance between grain and extent in investigations has decreased.

Dungan et al. (2002) emphasized that scale concepts have been used differently across scientific disciplines (e.g., ecology, geography, spatial statistics). They categorized this variability based on three dimensions: scale of the phenomenon of interest, scale in the sampling of these phenomena, and scale of the analysis based on sampling (Fig. 2.2). *Scale of the phenomenon* of interest emphasizes the pattern or process of interest, such as the scale of plankton dispersal or scales in vertebrate movement and habitat selection (Nams 2005; Mayor et al. 2009; Shurin et al. 2009). *Scale of sampling* emphasizes the characteristics of the sampling unit and sampling design used to measure and interpret the phenomenon, such as a quadrat size or study area extent. *Scale of analysis* is related to sampling scale but focuses on how samples are used for statistical inferences of the statistical population. For example, one possibility in analysis of spatial scale is that we might pool sampling units to different resolutions (i.e., increase grain size) for inferences on the role of scale on spatial pattern or process (Thompson and McGarigal 2002). Throughout this book, we emphasize scale concepts from an ecological perspective, with most focus on the scale of phenomenon and scale of analysis.

Scale should not be confused with the concept of "level of organization" (Allen and Hoekstra 1990, Levin 1992). *Level of organization* in ecology refers to the place within a biotic hierarchy, such as the categorization of individuals, populations, and communities (Turner et al. 1989a). Consequently, we may apply the problem of scale to each of these levels of organization. For example, our interpretation of the spatial scale of competition may vary in regard to competitive interactions between

Libros de Estadística en PDF | Statistics Books in PDF

individual heterospecifics versus competitive interactions at the meta-community level (see Chap. 11; Holyoak et al. 2005).

Hierarchy theory is frequently used to interpret issues that cross levels of organization (Urban et al. 1987; O'Neill et al. 1989). In this context, a *hierarchy* is a system that can be viewed as sets of interconnections operating at different levels. Typically, higher levels in the hierarchy constrain lower levels while lower levels provide details and mechanisms of relevance that may explain higher levels. In a seminal article, Johnson (1980) emphasized four hierarchical orders in habitat selection for animals: the geographic range, the home range, the territory, and the foraging patch (see Chap. 8). In general, higher levels are thought to have larger spatiotemporal scales than lower levels. For example, Cohen et al. (2016) found that climate factors changed slowly over space, impacting disease prevalence (West Nile virus, Lyme disease, and chytrid fungus) at broad spatial scales, while biotic interactions changed more rapidly only explaining disease prevalence at local scales. The hierarchy concept provides a foundation for issues of hierarchical patch dynamics (Wu and Loucks 1995) and multilevel problems (Cushman and McGarigal 2002, see below; McGarigal et al. 2016) in ecology. Often times hierarchy theory is used more as a heuristic framework for organizing hypotheses for species and ecosystems rather than a quantitative framework for modeling (but see Wu and David 2002).

2.2.2 *Why Is Spatial Scale Important?*

Spatial ecology and conservation emphasize the importance of scale to understand ecological processes, biodiversity patterns, and to better inform conservation decisions. The most common reasons for the importance of scale include that it may alter: (1) the role of biotic and abiotic interactions; (2) the degree to which systems are "open" versus "closed"; (3) the quantitative relationships of ecological patterns and processes; and (4) conservation and management decisions.

There are several examples where the scale at which questions are analyzed alter the conclusions regarding the role of biotic interactions (Wiens 1989; Levin 1992). For example, at a small scale (fine grain), field experiments have shown that American redstarts (*Setophaga ruticilla*) and least flycatchers (*Empidonax minimus*), two species of migratory birds that coexist in deciduous forest habitats in North America, are interspecific competitors, where presence of the dominant flycatcher may inhibit the redstart from establishing territories (Sherry and Holmes 1988; Fletcher 2007). This behavior results in a negative correlation in occurrence and abundance at a fine grain. In contrast, at a broader, regional scale there is a positive correlation in the occurrence and abundance of these two species, which is likely driven by their convergence on similar foraging strategies and prey (Sherry and Holmes 1988). To understand processes and species distribution at one scale, it is important to account for the processes acting at both a finer and broader scale (Allen and Hoekstra 1992).

Libros de Estadística en PDF | Statistics Books in PDF

Second, the interpretation of "open" versus "closed" dynamics varies tremendously with scale. Open dynamics emphasize that flow into and out of an area of interest (e.g., patch, landscape, and ecosystem) occurs. This can include the flow of energy (Cadenasso et al. 2003), resources (e.g., spatial subsidies; Polis et al. 1997), individuals (immigration/emigration; Pulliam 1988), or alleles (gene flow; Slatkin 1985). The roles of localized movement on population dynamics have received considerable attention to delineate local populations (Waples and Gaggiotti 2006), where the scales of movement locally and regionally influence metapopulation persistence (see Chap. 10). For example, giant kelp (*Macrocystis pyrifera*) forests in southern California illustrate that within-patch dynamics can occur, where local sites within patches are colonized more frequently than sites across patches, implying that within-patch dynamics are more "open" than between patch dynamics (Cavanaugh et al. 2014). Similarly, at the community and ecosystem levels, there has been considerable interest in understanding the roles of dispersal limitations and flows of nutrients across space to alter metacommunity and metaecosystem processes (Loreau and Holt 2004; Jacobson and Peres-Neto 2010).

Third, quantitative aspects of data change fundamentally with scale, in terms of both grain and extent (Turner et al. 1989b). For example, as grain gets larger in ecological investigations (while holding extent constant), the spatial variance tends to get smaller, metrics of land-cover diversity decreases, and rare land-cover types tend to disappear (Turner et al. 1989b; Horne and Schneider 1995). The form of decrease (e.g., linear or exponential decrease) will depend on several issues, such as the spatial patterning of heterogeneity and sampling design. Changing the grain of data also sometimes causes quantitative bias in patterns through what has been termed the *modifiable areal unit problem* (Openshaw 1984; Jelinski and Wu 1996; Dark and Bram 2007). This problem highlights that aggregated data can have different properties than the sample data from which they were derived, and biases can occur particularly in situations where data are aggregated into irregular sampling units, such as county-level polygon data. In geostatistics, aggregating (or resampling) data is referred to as change of support (Cressie 1996), where *support* refers to the size, shape, and orientation of a sample of data. The modifiable areal unit problem is related to the *ecological fallacy*, where inappropriate inferences occur for individual sample units from aggregated data for which no individual-level data occur (Piantadosi et al. 1988). For extent, several quantitative changes patterns can also occur. As extent gets larger (while holding grain constant), spatial diversity tends to increase as more spatial heterogeneity is captured in regards to habitat types and species (Wiens 1989). These issues can have profound impacts on the quantitative patterns in data. For example, species richness often exhibits a hump-shaped pattern with altitude, where high species richness occurs at moderate altitudes; however, if the extent of the investigation only captures a subset of an altitudinal gradient, the pattern in species richness can frequently appear linear (Rahbek 2005).

Finally, conservation strategies and the effectiveness of conservation decisions vary with scale. Prioritizing patches for connectivity conservation can vary based on the grain and extent of areas considered, as well as assumptions regarding the scale of dispersal for population connectivity (Pascual-Hortal and Saura 2007; Fletcher

Libros de Estadística en PDF | Statistics Books in PDF

et al. 2013; Maciejewski and Cumming 2016). Similarly, conservation planning for biodiversity and ecosystem services in Britain suggests that national priorities can vary dramatically, depending on the extent (i.e., region) considered within the country (Anderson et al. 2009). More broadly, the scale at which conservation planning is viewed may alter the roles of complementarity (i.e., the degree to which a species assemblage in an area complements assemblages in other protected areas) and irreplaceability (i.e., the uniqueness of an area to overall biodiversity in a region of interest) as key factors in the conservation planning process (Margules and Pressey 2000; Larsen and Rahbek 2003). Based on a review of 4239 species of vertebrates, Boyd et al. (2008) argued that the effectiveness of conservation planning among species varies with spatial scale.

2.2.3 Multiscale and Multilevel Quantitative Problems

Because of the large role of scale in ecology and conservation, there has been considerable focus on the development and application of quantitative methods that address the problem of scale. A primary emphasis has been in determining the critical scales, or *characteristic scales*, for ecological patterns and processes (Keitt and Urban 2005). Critical scales have been identified in organism responses to habitat (Holland et al. 2004; Jackson and Fahrig 2015), movement and dispersal (Reichert et al. 2016), population dynamics (Liebhold et al. 2004), community interactions (Andersen 1992), and conservation planning (Minor and Urban 2008).

In this context, multiscale and multilevel modeling have emerged (Fig. 2.3; McGarigal et al. 2016). Multiscale modeling quantifies environment conditions at multiple scales by altering either the grain or extent of the analysis, and then evaluates which of the considered scales best explains a pattern or process (Holland et al. 2004). This can be accomplished in several ways, the most common being based on considering multiple extents around locations of interest (e.g., buffers around habitat patches) to test the relative contribution of the different extents to explain species occurrence in a region (Jackson and Fahrig 2015). For example, Weaver et al. (2012) modeled the invasive mute swan (*Cygnus olor*) distribution based on biologically relevant spatial scales—average territory size radius, the median dispersal distance of cygnets, and average activity distance of adult males—finding that different environmental variables were correlated with swan distribution at different scales.

In multilevel modeling, the focus is on interpreting effects at different levels in an organizational hierarchy (Mayor et al. 2009; Wheatley and Johnson 2009; McGarigal et al. 2016). Such approaches often encapsulate three different perspectives. First, models may include contrasting levels of organization (e.g., effects of forest cover at different scales on population abundance versus genetic diversity; Jackson and Fahrig 2014). Second, a model may focus in situations where the environment is hierarchically structured and questions are asked regarding how this hierarchy might differentially influence organisms. For example, individuals may use different cues to select

Libros de Estadística en PDF | Statistics Books in PDF

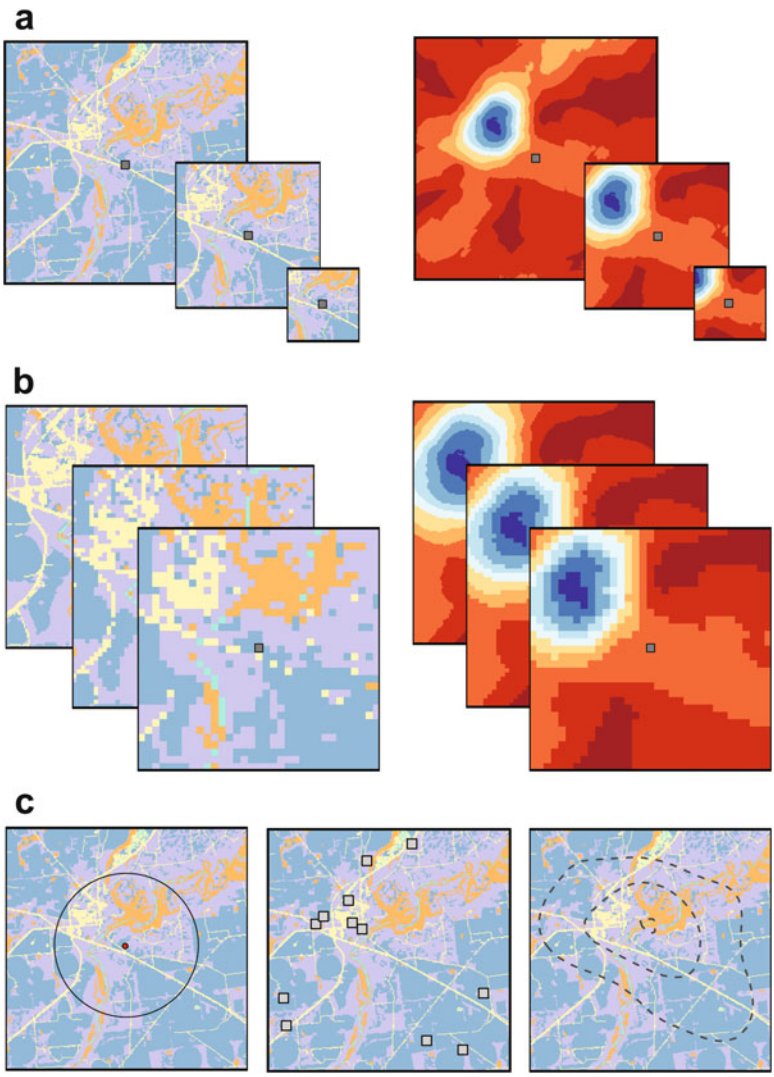


Fig. 2.3 Multiscale and multilevel modeling. Spatial modeling is increasing multiscale, where the (**a**) extent and/or (**b**) grain of the environment is varied, with the aim to understand what grain and extent best explains a pattern or process. Shown are two different environmental factors, one categorical (e.g., land-use variation) and another that reflects a continuous gradient. (**c**) Multilevel modeling occurs when responses at different levels of organization or hierarchies are considered. Shown here is an example for the same landscape where an individual location is considered, several populations are considered, and the entire spatial range of the metapopulation is considered

nest sites within patches versus habitat at the patch-scale or those available in the surrounding area (Chalfoun and Martin 2007). Third, a model may aim in quantifying the variation in different responses by focal organisms that are hierarchical in nature. Multilevel modeling can occur in space and/or time. For example, temporal multilevel

Libros de Estadística en PDF | Statistics Books in PDF

modeling might include asking how environmental relationships explain animal use between day and night, within seasons versus between seasons, and across an annual time scale (Rettie and Messier 2000; Schooley 1994; McLoughlin et al. 2002; Guyot et al. 2017). McGarigal et al. (2016) also classified multilevel models as those where different levels were modeled separately, largely because each level may focus on different types of responses. However, hierarchical and/or multilevel statistical modeling can provide a means where different levels can be modeled simultaneously (Gelman and Hill 2007).

Finally, an active area of interest includes how to translate or predict across scales (and levels of organization). These problems often focus on identifying potential scaling coefficients that can be used to translate patterns across scales (Miller et al. 2004). For example, scale transition theory (Chesson 2012) aims to identify equations that can link population and community processes across scales based on nonlinearities and variation at finer scales (grains) to predict emergent properties. Currently, this work is largely theoretical (but see Melbourne and Chesson 2005, 2006) but it has the potential to be applied to situations where information is limited. Other approaches include identifying *cross-scale interactions* and/or correlations, where patterns and processes at one scale covary or interact with patterns and processes at other scales (Falk et al. 2007; Peters et al. 2004, 2007; Schooley and Branch 2007; Soranno et al. 2014).

2.2.4 Spatial Scale and Study Design

Given the role that scale can play in our understanding of ecological patterns and processes, as well as its importance for conservation problems, how should spatial scale be considered in study design? Clearly, the answer to this question will vary depending on the phenomena of interest, yet ecologists and statisticians have provided some important guidance (e.g., Dungan et al. 2002; Dale and Fortin 2014). Key issues include the size of the sampling unit (grain), the type of the sample unit, and sample unit locations, including the spatial lag between samples (distance between samples) and the size of the study area.

The sample unit size, or the grain of the data, will set the lower limit on the *resolution* of inference for an investigation. It is frequently recommended that the grain be 2×–5× smaller than the spatial extent of the phenomenon of interest (O'Neill et al. 1996). While smaller grains might intuitively be preferred, too small of sample units relative to the phenomenon of interest can add noise to the data, causing potential challenges for inference. Nonetheless, if sample units can easily be aggregated, then smaller unit size is generally preferable because they may be pooled or aggregated as needed for phenomena under investigation (Dale and Fortin 2014). In contrast, disaggregating data (i.e., resampling data from a coarser to finer resolution) can sometimes be limited in terms of providing reliable information at a fine grain (see below).

Libros de Estadística en PDF | Statistics Books in PDF

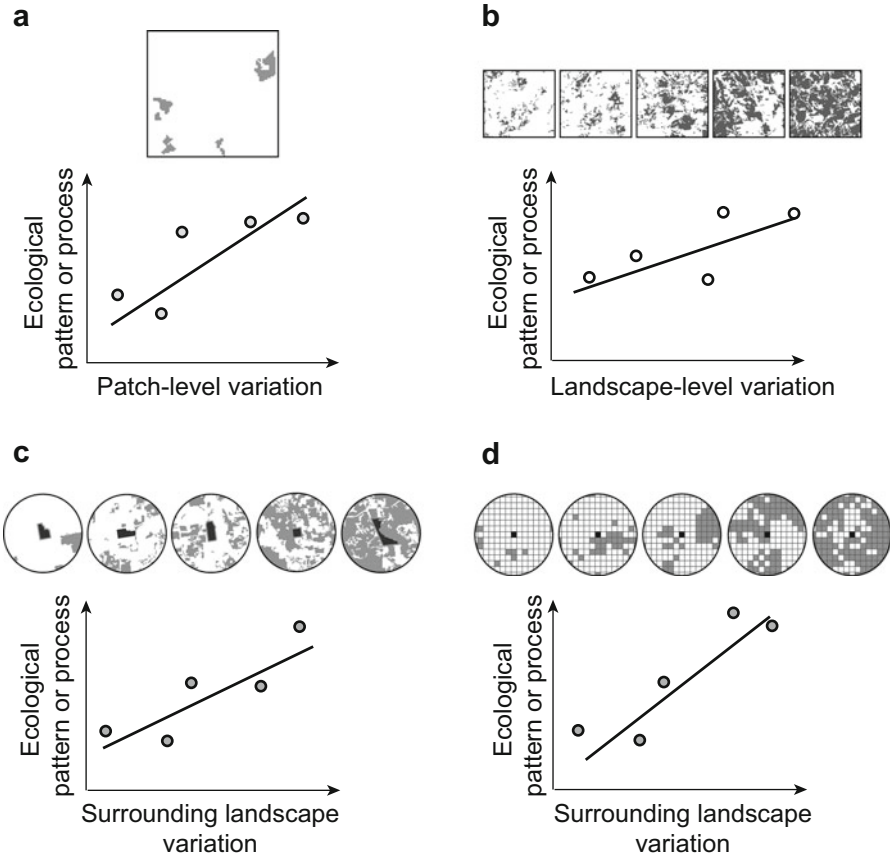


Fig. 2.4 Some alternative study designs for investigating processes that may act at different levels. (**a**) In patch-scale studies, ecological patterns and processes are summarized among patches within the study area. (**b**) In landscape-scale studies, ecological patterns are processes are summarized among landscapes or regions, and not among patches. (**c**) In patch-landscape (or "focal-patch") studies, the patch is the sample unit, but the surrounding environment is summarized, typically within a buffer of a radius thought to be relevant to the ecological pattern or process, taken from either the centroid of the patch (shown here) or from the patch boundary. (**d**) In "local-landscape" studies, the pixel (or related area) is the sample unit, rather than a patch, which standardizes the area sampled, and the surrounding landscape is considered. In all panels, maps show examples of variation in habitat and/or the land cover. Extended from Fahrig (2003)

A related issue regarding multiscale effects of land cover (Thornton et al. 2011) and the sample unit size is whether the unit should be a pixel, a patch (Fleishman et al. 2002), or an entire landscape (Villard et al. 1999). Fahrig (2003) referred to the latter two designs as "patch-scale" versus "landscape-scale" designs (Fig. 2.4), and also described "patch-landscape" designs (also known as "focal patch" designs), where the patch is the focal unit, while covariates may be measured in the surrounding landscape. This latter design can be readily extended to a "local-landscape"

Libros de Estadística en PDF | Statistics Books in PDF

design, where the pixel or local sampling unit (within a patch or irrespective of patch boundaries) is the focus and the surrounding landscape is considered. When interest is in predicting species distributions across landscapes or regions (Guisan and Thuiller 2005), it is natural to view the sampling unit as the grain (e.g., remote sensing pixel, GIS raster cell) because this allows for straightforward mapping of the data across a landscape or region without the need to alter the summary statistics or predictions. In contrast, some questions operate at patch scales (Diamond 1975), while others operate at larger scales (Fahrig 2013), such that study designs might benefit from using sampling units that reflect underlying questions or phenomena of interest.

The extent of an investigation is frequently recommended to be at least 2×–5× greater than the spatial extent of the phenomenon of interest (O'Neill et al. 1996), with some studies suggesting up to 10× (Jackson and Fahrig 2015; Miguet et al. 2016). If a study area is too small, there will likely not be enough variability in the samples to identify meaningful patterns and processes. Yet, if it is too large, there may be several processes involved, generating patterns at multiple scales (Dale and Fortin 2014).

The location of sample units can be described by both the spatial *lag*, or distance between samples (Table 2.1), and the sampling strategy, such as spatially random sampling, sampling across a regular grid, or spatially stratified sampling (e.g., stratifying random samples based on land-cover types to ensure environmental variation is captured). While random or stratified random sampling is often done in ecology, spatial ecologists frequently use regular/systematic or nested grids for sampling. The benefit of regular or nested grids is that it may ensure that the gradient-wide spectrum of spacing between sampling locations is captured, which can potentially allow for a better detection of the spatial scales of the patterns. Limitations of systematic sampling designs can occur if there is periodicity in the environment that is poorly captured based on the position and lag distance used in sampling. The spatial lag between samples will be driven based in part on the total number of samples and the extent of the investigation. Sample lag should be smaller than the average distance between units of the phenomenon under investigation (Dungan et al. 2002). In general, for most scaling analyses, inferences are typically made at distances less than ⅓ to ½ of the extent length under investigation (see Chap. 5). Consequently, ensuring adequate lag distances of samples within this domain can be helpful. For instance, nested grid sampling can help increase replication for lag distances used in inference with smaller total effort (Fortin et al. 1989). The spatial lag between samples has also been debated in the context of determining scales of effects in multiscale investigations (Holland et al. 2004; Zuckerberg et al. 2012). Some argue that when testing for scale effects of land cover on species responses, the description of that cover (e.g., the proportion of forest within a distance of the sampling location) should not overlap between sampling locations due to non-independence of samples (Holland et al. 2004; Eigenbrod et al. 2017). However, Zuckerberg et al. (2012) argued that this concern is misleading: the lack of independence in explanatory variables (e.g., forest cover) is not relevant but rather the lack of independence in the response variables is of critical importance.

Libros de Estadística en PDF | Statistics Books in PDF

In Chaps. 5 and 6, we provide an overview of the issue of spatial dependence and its influence on spatial analyses, such as the use of spatial regression (Dormann et al. 2007; Beale et al. 2010).

2.3 Examples in R

2.3.1 Packages in R

In R, we will address some general aspects related to spatial scale using the `raster` package (Hijmans 2017). `raster` is a package dedicated to the use of raster (or grid) GIS layers and allows for many types of summaries, analyses, and visualizations. We will use this package throughout this book, but here we will introduce some of the foundational techniques when using the `raster` package. In later chapters, we will use other important packages for interpreting some problems of scale, but here we introduce the topic primarily with the `raster` package.

2.3.2 The Data

We will first use simulated data to illustrate some ways in which we can alter the scale of raster data and interpret summaries based on these changes. Simulating data can be very helpful for a variety of problems in spatial ecology and conservation because it provides a means to simplify the task or problem, use an example where we "know the truth," formulate hypotheses, and test or interpret methods that we may later apply to real data.

We will then illustrate a simple, multiscale analysis of reptile response to the amount of forest cover in the Southeast USA. These data come from drift fence arrays at 78 forested sites in Alabama, Georgia, and Florida, USA.

2.3.3 A Simple Simulated Example

First, we create a toy landscape. To do so, we set up an empty raster layer and then populate the empty cells with randomly generated values taken from a Poisson distribution. The Poisson distribution is a discrete probability distribution (i.e., a probability mass distribution) that is relevant for count-based (integer) data, where the data, y, can take on the values of 0, 1, 2, etc. It assumes that the mean equals the variance. By using a Poisson distribution, we will set values of the cells to non-negative integers, which is a common format for storing land-cover information.

First, we load the `raster` package and use `set.seed` to set a random number seed in R. This allows users to be able to replicate analyses where random number

Libros de Estadística en PDF | Statistics Books in PDF

generators are used. We then create a 6×6 raster, specifying the numbers of rows (`nrow`) and columns (`ncol`), as well as the minimum and maximum coordinates. We then populate the raster by taking random draws (i.e., random deviates) from the Poisson distribution with the `rpois` function.

```
> library(raster)
> set.seed(16)#sets random number seed for repeatability
> toy <- raster(ncol=6, nrow=6, xmn=1, xmx=6, ymn=1, ymx=6)
> values(toy) <- rpois(ncell(toy), lambda=3)
> ncell(toy)
> plot(toy)
> text(toy, digits=2)
```

In the above code, we generate 36 values (`ncell(toy)` $= 36$) from the Poisson distribution, where the mean value $= 3$. We could also use a multinomial distribution for simulating land-cover types. A multinomial distribution models the frequency of K land-cover types based on their probabilities of occurrence (where $\sum_K = 1$). This can be accomplished with the `rmultinom` function, but we will focus on the Poisson distribution for now because it is more straightforward to implement. We discuss simulating raster layers with more realistic spatial patterns in Chap. 3.

Note that when raster populates the raster layer with the Poisson data (`rpois`), it will start from the top left of the layer and populate right and then down. We can check to make sure this is the case by creating a second raster layer of the same dimensions, but rather than loading randomly generated values from a distribution, we will populate the raster layer with a vector string:

```
> ncell(toy)
> toy2 <- toy
> values(toy2) <- 1:ncell(toy)
> plot(toy2)
> text(toy2, digits=2)
```

Altering the grain of a raster layer is straightforward. We can increase the grain size using the `aggregate` function (Fig. 2.5). Two common approaches are to: (1) take the mean value of the cells being aggregated; or (2) use a "majority rule," where we take the most frequent value in the cells being aggregated. We can illustrate each of these approaches as:

```
> toy_mean <- aggregate(toy, fact=2, fun=mean) #mean value
> toy_maj <- aggregate(toy, fact=3, fun=modal) #majority rule
```

Note that these rules are helpful for different situations. For categorical data (e.g., vegetation types), a majority rule might be helpful because it will aggregate cells based on the most frequent category. For continuous data (e.g., canopy cover), taking the mean value might be more helpful than a majority rule (which would take the mode of the values). In addition, the `modal` function can be limited when

Libros de Estadística en PDF | Statistics Books in PDF

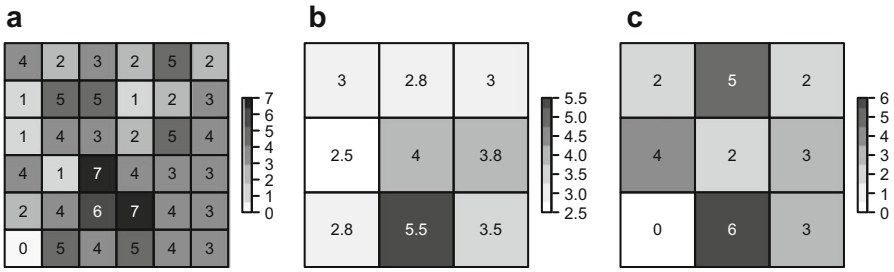


Fig. 2.5 Increasing grain in a toy landscape. (**a**) A landscape where values are generated from a Poisson distribution. Aggregating cells to increase the grain based on (**b**) the mean and (**c**) using a majority rule

ties are common (no majority value), which will occur more frequently when aggregating fewer cells. In this situation, the function defaults to randomly picking one of the values, but it can be altered to return the lowest, highest, or first value.

For these toy landscapes, we can formally ask whether the means and variances change as we increase grain size, as described above. It is straightforward to do mathematical operations on values of raster layers. Here, we contrast means and variances of the original raster to that of the raster where we used the mean values to increase grain size using the `cellStats` function (results not shown):

```
> cellStats(toy, mean)
> cellStats(toy, var)

> cellStats(toy_mean, mean)
> cellStats(toy_mean, var)
```

In this situation, the mean value remains identical (3.412), whereas the variance decreases as we increase the grain size (from 2.82 to 0.86).

We can reduce the grain size by resampling the data (Fig. 2.6). This can be accomplished with the `disaggregate` function. When using this function, several approaches can be used to resample the data into smaller grains. Two common approaches are to use a simple disaggregation, which simply replicates values, or using bilinear interpolation, which is based on a distance-weighted average of values in both the *x* and *y* directions (hence "bi'" linear).

```
> toy_dis2 <- disaggregate(toy, fact=2)
> toy_dis2_bilinear <- disaggregate(toy, fact=2,
  method='bilinear')
```

Bilinear interpolation can be useful when working with continuous data but would not be helpful if data were based on land-cover categories.

Altering the extent is also straightforward. We can reduce the extent of the map by use of the `crop` function. To do so, we need to create a new extent for cropping.

Libros de Estadística en PDF | Statistics Books in PDF

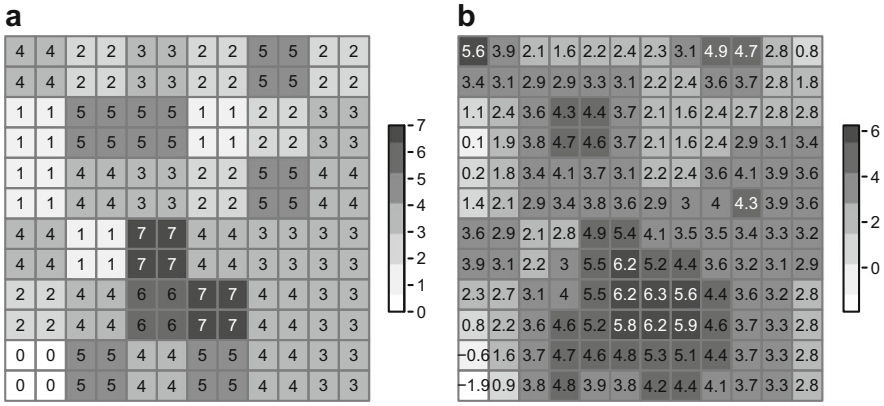


Fig. 2.6 Reducing grain size in a toy landscape. (**a**) Disaggregating cells to increase the grain. (**b**) Reducing grain size using a bilinear interpolation

This can be a simply rectangle of coordinates or we could use a polygon file (e.g., a shapefile). In contrast, the extent can be increased using the `extent` function. For this toy example, we use new coordinates for changing the extent:

```
#decrease the extent
> e <- extent(2, 4, 2, 4) #first create new, smaller extent
> toy_crop <- crop(toy, e)
> plot(toy_crop)
> text(toy_crop)

#increase the extent
> e <- extent(0, 7, 0, 7) #first create new, larger extent
> toy_big <- extend(toy, e)
> plot(toy_big)
> text(toy_big)
```

In this case, increasing the extent is not helpful unless we also populate the data in the new extent.

This simple example illustrates how to change the grain and extent of a raster map, as well as some of the consequences of changing the grain and extent. Simple examples like this one can be generally helpful when starting a new analysis problem in spatial ecology because they can provide a tractable means of understanding what different functions and models do.

2.3.4 Multiscale Species Response to Land Cover

Now we interpret the scale at which species may respond to habitat, such as forest cover. To do so, we quantify the amount of forest that occurs at various distances

Libros de Estadística en PDF | Statistics Books in PDF

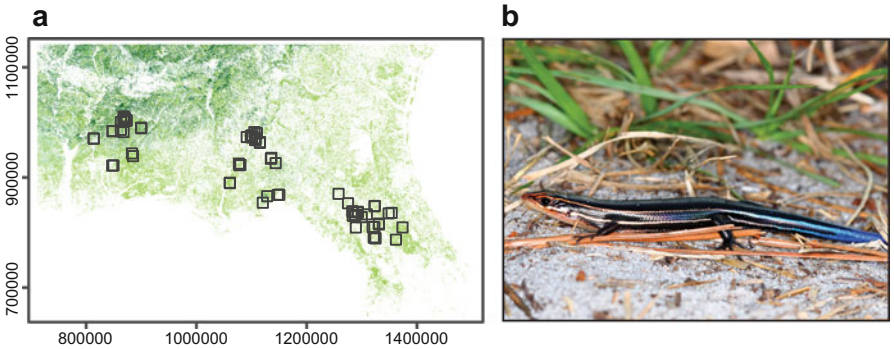


Fig. 2.7 The case study for interpreting the scale of effects of land-cover on species distribution. (**a**) The southeastern USA, with study sites shown. (**b**) The five-lined skink, a common reptile, which was sampled with drift-fence arrays

from sampling points and then determine the scale at which forest cover best predicts species occurrence.

To illustrate this general approach, we use the 2011 National Land Cover Database (NLCD) (Homer et al. 2015) and link these land-cover data to reptile sampling in managed forests for the Southeast USA. NLCD is a standardized land-cover dataset created from Landsat data (grain = 30 × 30 m) for the entire USA. NLCD classifies land cover into 20 categories. For the purpose here, we are primarily interested in forest land-cover, which is classified as deciduous (ID: 41), evergreen (ID: 42), and mixed (ID: 43) forest. This land-cover database is particularly useful because it provides a consistent land-cover classification scheme across states. For this exercise, we have clipped the 2011 NLCD layer to the southeastern USA (`nlcd2011SE`).

Reptiles were sampled with drift-fences at 85 sites (Fig. 2.7). Sampling occurred in mature, naturally regenerated longleaf pine (*Pinus palustris*) savannas, slash (*P. elliottii*) and loblolly (*P. taeda*) pine plantations, and corn (*Zea mays*) fields distributed among three geographic regions within the Southeastern Plains and Southern Coastal Plains ecoregions in the states of Florida, Georgia, and Alabama (Fig. 2.7a; Gottlieb et al. 2017). At each site, two drift fences were set up along two transects, with one transect located along the edge and a second located in the interior of the sites. Drift fences were opened for 3 days each month, April to July, 2013–2015. For this exercise, we will not consider data collected in corn fields and we will pool data within sites (using the centroids of the sampling locations for the two transects). Here, we will focus on data collected for the southeastern five-lined skink (*Plestiodon inepectatus*; Fig. 2.7b).

Libros de Estadística en PDF | Statistics Books in PDF

2.3.4.1 Multiscale Analysis of Forest Cover

We start by loading the data, defining its projection and inspecting it. We can read the land-cover data with the `raster` function.

```
> nlcd <- raster("nlcd2011SE")
> proj4string(nlcd)

##
[1] "+proj=aea +lat_1=29.5 +lat_2=45.5 +lat_0=23 +lon_0=-96 +x_0=0
+y_0=0 +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +units=m +no_defs"
```

This projection contains lots of information. Most importantly, `aea` refers to Albers Equal Area projection. We define the projection so that we can make sure the transect data are considered to be in the same projection as the land-cover data (see Appendix for more on projections).

```
> nlcd_proj <- projection(nlcd)
```

We can inspect other aspects of the raster layer as well, including the resolution (grain size), extent, and number of cells, with the `res`, `extent`, and `ncell` functions. For example,

```
> res(nlcd)

##
[1] 30 30
```

As expected, we find that the grain of the land-cover data is 30×30 m. Note that R does not consider the land-cover data as categorical, which can be shown with `is.factor(nlcd)`. So, we convert the layer to be considered categorical with `as.factor` and inspect the number of categories of land cover with `levels` (note output is suppressed here).

```
> nlcd <- as.factor(nlcd) #convert to factor
> levels(nlcd)
```

There are 16 categories of land cover, with labels reflecting the NLCD IDs. For example, deciduous forest is ID = 41.

We read in the transect data (a `SpatialPointsDataFrame` object, which can be determined with the `class` function) with the `readOGR` function from the `rgdal` package.

```
> library(rgdal)
#reptile data
```

Libros de Estadística en PDF | Statistics Books in PDF

```
> sites <- readOGR("reptiledata")
> class(sites)

##
[1] "SpatialPointsDataFrame"
attr(,"package")
[1] "sp

> summary(sites)

##
Object of class SpatialPointsDataFrame
Coordinates:
 min max
coords.x1 812598.9 1373597
coords.x2 786930.5 1014229
Is projected: NA
proj4string : [NA]
Number of points: 85
Data attributes:
 site management coords_x1 coords_x2
 AL1 : 1 Reference :10 Min. : 812599 Min. : 786931
 AL10 : 1 Thinned :10 1st Qu.: 872612 1st Qu.: 838447
 AL11 : 1 Young : 8 Median :1106233 Median : 933843
 AL12 : 1 Corn : 7 Mean :1094814 Mean : 918191
 AL13 : 1 Unthinned : 6 3rd Qu.:1288328 3rd Qu.: 982365
 AL14 : 1 Clear cut, debris LEFT: 4 Max. :1373597 Max. :1014229
 (Other):79 (Other) :40
```

The `summary` function provides a lot of relevant information. First, it provides the extent of the layer. Notice that this data set includes data collected in eight different land-uses, seven of which are different types of conifer forest, whereas one is corn. It also shows that there are 85 sites (points) and that it does not know what the projection is for the layer. We set the projection to be consistent with the other data (note that this does not change the projection; the layer was already in a consistent projection, but R did not recognize it).

```
> proj4string(sites) <- nlcd_proj #set projection
```

We can call the `SpatialPointsDataFrame` using the generic functions. For example,

```
> head(sites, 2)

##
 site management coords_x1 coords_x2
1 AL1 Reference 846279.4 921444.9
2 AL10 Clear cut, residues removed 899063.5 989168.9
```

Libros de Estadística en PDF | Statistics Books in PDF

We use the `subset` function that is often used on data frames to remove the corn land use:

```
> sites <- subset(sites, management!="Corn")
```

With this subset, we work with 78 sites (`nrow(sites)`). Next, we crop the nlcd layer to make the extent only be 10 km beyond the sampling transects to increase computing speed.

```
#define reduced extent
> x.min <- min(sites$coords_x1)-10000
> x.max <- max(sites$coords_x1)+10000
> y.min <- min(sites$coords_x2)-10000
> y.max <- max(sites$coords_x2)+10000
> extent.new <- extent(x.min, x.max, y.min, y.max)
> nlcd <- crop(nlcd, extent.new)
```

To simplify our consideration of scaling issues, we reclassify the nlcd layer into a binary forest/non-forest layer. This task can be accomplished in at least two ways. First, we could reclassify land cover categories (pooling different forest land-cover types) using some generic R commands to create a new layer that captures the forest cover across the study region. To do so, we create a map of the same grain and extent, and then we can reset values of the map. In this case, we want to pool land-cover categories 41, 42, and 43 (Deciduous, Evergreen, and Mixed Forest, respectively).

```
#create a new forest layer
> forest <- nlcd
> values(forest) <- 0
> forest[nlcd==41 | nlcd==42 | nlcd==43] <- 1 #forest categories
```

Note how with the `raster` package we can easily reclassify land-cover using simple operations in R similar to other operations for vectors and matrices (Appendix). In this situation, we are populating our new forest raster, which is initially set to all 0 values, as 1 if the nlcd layer at a location is either 41, 42, or 43 (using the OR statement, `|`). The result is a new raster where forest values are 1 and all other values are 0. Alternatively, we can use the `reclassify` function in the `raster` package to do the same operation, which is much quicker computationally. This function requires creating a matrix where the first column is the original land-cover categories and the second provides information on the reclassification categories. In this case we need to make sure that in the second column, there are all zeros except for the rows representing nlcd categories 41, 42, and 43 (conifer, deciduous and mixed forests). In `levels(nlcd)[[1]]`, we find that IDs 41–43 are the 8th–10th values in the vector. Consequently, we can create a reclassification vector as:

Libros de Estadística en PDF | Statistics Books in PDF

```
#reclassification vector
> reclass <- c(rep(0,7), rep(1,3), rep(0,6))
```

We then create a reclassification matrix and reclassify the nlcd layer with the `reclassify` function.

```
#create reclassification matrix
> reclass.mat <- cbind(levels(nlcd)[[1]], reclass)
> head(reclass.mat, 3)

##
 ID reclass
1 0 0
2 11 0
3 21 0

> forest <- reclassify(nlcd, reclass.mat)
```

We then take point coordinates of sample locations and calculate the amount of forest that surrounds each sampling location at different extents. To do so, we set the local extents to 1000 and 5000 m (Fig. 2.8). We can then use the `buffer` function to create circular buffers of different extents surrounding the sites. We consider the first site (`sites[1, ]`) to illustrate and then use the same logic to apply to all sites.

```
#buffer sites
> buf1km <- 1000
> buf5km <- 5000

#buffer only first site
> buffer.site1.1km <- buffer(sites[1,], width=buf1km)
> buffer.site1.5km <- buffer(sites[1,], width=buf5km)
```

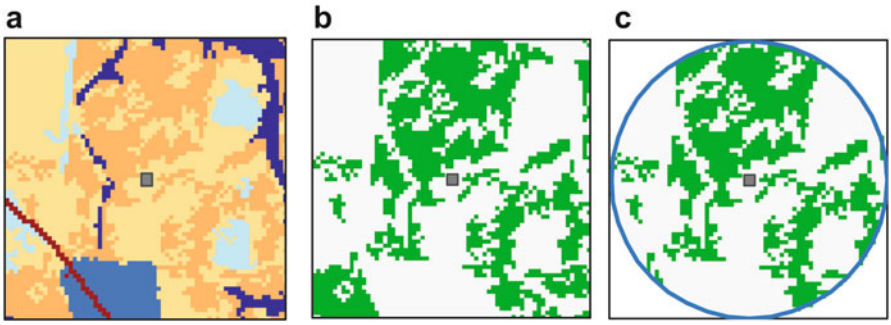


Fig. 2.8 Illustration of determining the amount of habitat surrounding a point. For a given study extent (**a**), the habitat of interest is isolated (**b**). A buffer is placed surrounding a point (**c**) or polygon and the number of cells (pixels) that contain the habitat is summed and multiplied by the area of each cell

Libros de Estadística en PDF | Statistics Books in PDF

The raster package has a useful function for viewing portions of raster layers. Here we use the zoom function to zoom into the buffer we just created.

```
#zoom into area for viewing
> zoom(nlcd,buffer.site1.5km)
> plot(buffer.site1.5km, border="red", lwd = 5, add=T)
> plot(buffer.site1.1km, border="red", lwd = 3, add=T)
> points(sites[1,], pch=19, cex=2, add=T)
```

Note that the rgeos package (Bivand and Rundall 2017) has a function gBuffer that is similar to buffer function in the raster package (in some situations it is slightly faster). With gBuffer you can also adjust how the circular buffer is approximated in R (see below), which can be helpful.

How can we extract appropriate information at different scales? Let us focus on this first site. Once we can capture the information we need for one point, we then repeat for all sites. There are several ways to accomplish this task. The simplest way is to take the buffered layer we just created and use the crop and mask functions:

```
> buffer.forest1.1km <- crop(forest, buffer.site1.1km)
> buffer.forest1.1km <- mask(forest, buffer.site1.1km)
```

Extracting the forest area is straightforward from here. Given that the map is a binary map of forest cover, we can use the cellStats function in the raster package to sum the amount of forest cover (sum the 1's for each cell to obtain the total number of forest cells). We can then multiple this number by the grain to get the forest area and divide by the buffer size to get the proportion of forest:

```
#area of each cell, in ha
> grainarea <- res(forest)[[1]]^2/10000

#area of 1km buffer
> bufferarea <- (3.14159*buf1km^2)/10000

#calculation of forest cover and % cover
> forestcover1km <- cellStats(buffer.forest1.1km, 'sum') *
  grainarea
> percentforest1km <- forestcover1km / bufferarea * 100
```

That's it! Now to repeat this for all of the points, we use a for loop, where we iteratively go through all points, calculating buffers, extracting forest area for each buffer, and then append the proportion of forest area for each point (and/or buffer size) for each sample location. To do this more efficiently for so many points, we will use the rasterize function to convert the buffer into a raster layer, which can be computationally quicker than not rasterizing the buffer. We make a generic function that can then be used to automate all of the steps for a given point. In this function, we first crop the layer to the buffer so that we can work on a smaller extent, then we create an empty raster that we use for rasterizing the buffer. With that new layer, we

Libros de Estadística en PDF | Statistics Books in PDF

can use the `mask` function to create a new raster that only includes forest cover within the buffer.

```
> BufferCover <- function(coords, size, landcover, grain){
  bufferarea.i <- pi*size^2/10000
  coords.i <- SpatialPoints(cbind(coords[i, 1], coords[i, 2]))
  buffer.i <- gBuffer(coords.i, width=size)
  crop.i <- crop(landcover, buffer.i)
  crop.NA <- setValues(crop.i, NA) #for the rasterization
  buffer.r <- rasterize(buffer.i, crop.NA) # rasterize buffer
  land.buffer <- mask(x=crop.i, mask=buffer.r)
  coveramount <- cellStats(land.buffer, 'sum')*grain
  percentcover <- 100*(coveramount/bufferarea.i)
  return(percentcover)
}
```

So this function requires *x*–*y* locations of a point, the buffer distance of interest (size), a binary land cover raster layer, and the grain area of the map (note the latter could build into the function, but it would recalculate grain area each iteration, which is not necessary for this example). We use this function, nesting it in a `for` loop, to calculate forest cover for all the points:

```
#create empty vector for storing output first
> f1km <- vector(NA, length = nrow(sites))
> f2km <- vector(NA, length = nrow(sites))

> for(i in 1:nrow(sites)) {
  f1km[i] <- BufferCover(coords = sites, size = 1000,
  landcover = forest, grain = grainarea)
  f1km[i] <- BufferCover(coords = sites, size = 2000,
  landcover = forest, grain = grainarea)
  print(i) #print iteration in for loop
}
#make data frame with associated site data
> forest.scale <- data.frame(site = sites$site, x =
  sites$coords_x1, y = sites$coords_x2, f1km = f1km, f2km = f2km)
```

We then use the above function to calculate the proportion of forest cover at different buffer sizes for all of the points. The above code shows calculations for 1000 and 2000 m, but we also ran 500, 3000, 4000, and 5000 m. In doing so, we find that the percent of forest cover at different scales tends to be highly correlated (Fig. 2.9). This is not surprising, given that calculations at a larger buffer size include area considered at smaller buffer sizes. However, this correlation has implications for the interpretation of scale effects (see below). Note that in R (and in other GIS), as the buffer increases, the computation time also increases. See a recent package in R, `spatialEco` (Evans 2017), for similar functionality regarding calculating landscape metrics surrounding points.

Libros de Estadística en PDF | Statistics Books in PDF

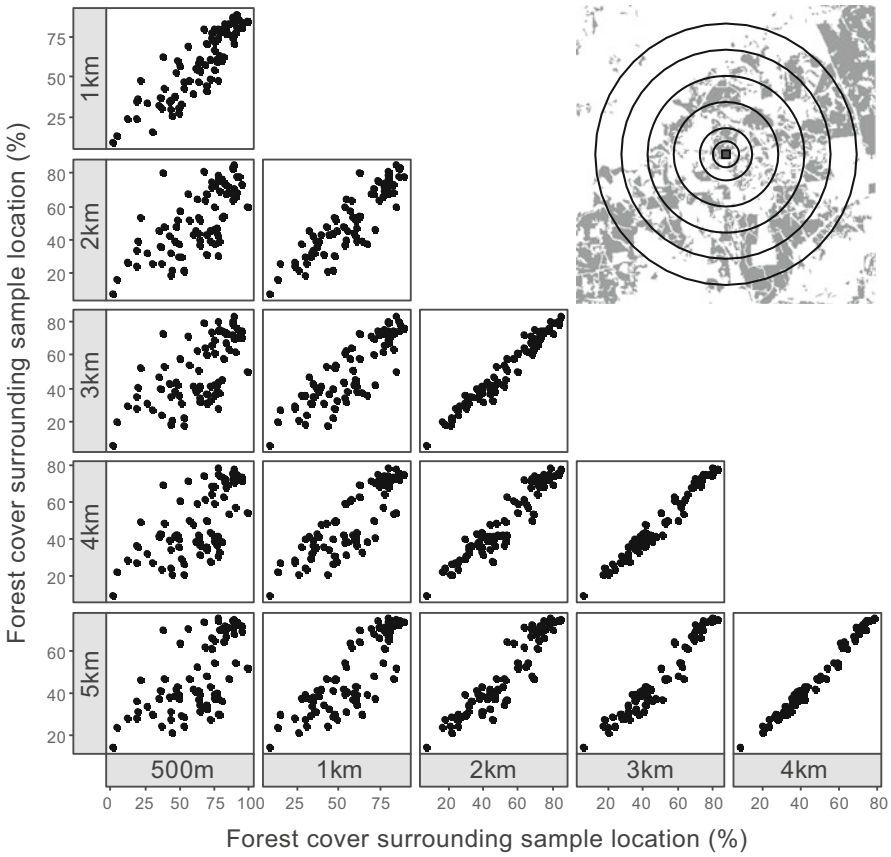


Fig. 2.9 Forest cover surrounding sampling locations, calculated at different scales. Shown are scatter plots of percent forest cover for each pairwise combination of scales (buffers; inset shows an example). Note the high degree of correlation between scales

We can also repeat this process for different grains by using the `aggregate` function to coarsen the map. Why might we want to do this? A primary reason is to translate the map to a resolution of data being collected in the field that we are using for making inferences. In this case, we are considering data collected along two, 200 × 100 m transects within forest patches, or 4 ha. If we wish to make predictions of species–environment relationships, we may want our map grain to reflect the sampling grain. Consequently, we would want the map to have an approximate 200 × 200 m grain. We can do this as:

```
#changing the grain
> forest200 <- aggregate(forest, fact=7, fun=modal)
```

Libros de Estadística en PDF | Statistics Books in PDF

2.3.4.2 Multiscale Analysis of Species Responses

We now consider how we might relate these differences in grain and extent to species occurrence to help identify the characteristic scale (or *scale of effect*) of forest cover on species occurrence. This is an increasingly considered problem in applied ecology (Holland et al. 2004; Jackson and Fahrig 2015; McGarigal et al. 2016; Miguet et al. 2016).

Several approaches have been advanced to quantify the scale(s) of effect. Pearson (1993) was one of the first to address this problem by measuring the landscape surrounding survey locations based on buffers of different sizes. This approach became very popular and has been used extensively (Holland et al. 2004; Jackson and Fahrig 2015). More recently, the use of spatial kernels has been suggested (Heaton and Gelfand 2011). Kernels can be applied to weight land-cover data as a function of distance from the survey location (Aue et al. 2012; Miguet et al. 2017). Not only might kernels better capture neighborhood effects by weighting nearby locations more than distant ones (Fig. 2.10), Chandler and Hepinstall-Cymerman (2016) showed how this formulation can be implemented to optimally select the scale of effect without resorting to a priori binning of different neighborhood sizes (e.g., the 1 and 5 km buffers described above). Here, we first illustrate the use of buffers of different sizes, and then proceed to illustrate the use of kernel-based approaches.

Buffer Analysis. To illustrate quantifying the scale of effect, we first use a buffer-based analysis, which is a commonly used technique (Holland et al. 2004; Jackson and Fahrig 2015). We fit a logistic regression model for binary (0, 1) data to interpret the relationship of forest cover and species occurrence. Logistic regression is a type of generalized linear model for binary (or binomial) data that is analogous to linear regression. We will discuss generalized linear models in more detail in Chap. 5, but for the purpose of illustrating scale effects here, we can describe a logistic regression for the effect of forest cover on species occurrence as:

$$\text{logit}(p_i) = \alpha + \beta \text{forest}_i, \tag{2.1}$$

where logit is defined as $\log(p/(1-p))$, p_i is the probability of presence of a species at location i, α is the intercept and β is the coefficient for the relationship of forest cover surrounding locations. In a logistic regression, we assume that the errors (residuals) come from a binomial distribution. We note that this model is a very simple model for occurrence data and it does not account for observation errors (MacKenzie et al. 2002) and other model complexities (Dormann et al. 2007). But it will be useful for illustrating scale effects here.

We fit this model to the data and contrast different models based on measurements of forest cover at different grains and local extents (buffer sizes). To contrast models, we can use measures of model fit (e.g., the likelihood of the model, given the data; Fletcher et al. 2016; Stuber et al. 2017), variation explained (Holland et al. 2004), or measures of predictive success (e.g., how well the model may predict to

Libros de Estadística en PDF | Statistics Books in PDF

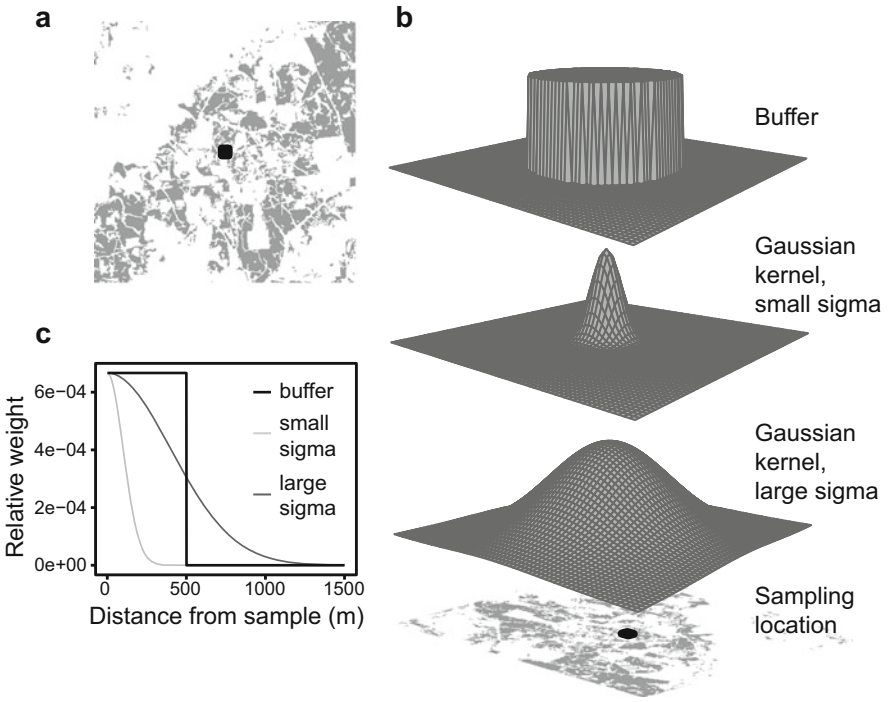


Fig. 2.10 An illustration of using buffer and kernel approaches to estimate the scale of effect. (**a**) A sampling location and the surrounding forest cover shown in light gray. (**b**) The spatial weighting scheme for the use of circular buffers and kernels relative to the sampling location. (**c**) The relative weights as a function of distance from the sampling location for circular buffers and kernels, where a kernel with a small and a large sigma is shown

new locations; Fielding and Bell 1997). We use the log-likelihood, which has a strong philosophical foundation in statistics. Log-likelihoods are based on the concept of maximum likelihood and quantify the plausibility (or likelihood) of a model's parameters, given the data. Note that the use of model selection criteria, such as Akaike's Information Criterion, that penalize the log-likelihood based on the number of parameters (Burnham and Anderson 1998), would provide identical results to use of the log-likelihood here because we contrast models that have the same number of parameters to identify scales of effects.

We first load the reptile data on southeastern five-lined skinks (FLSK) and merge it to our summaries of forest cover calculated at different scales.

```
> flsk <- read.csv("reptiles_flsk.csv", header=T)
> flsk <- merge(flsk, forest.scale, by = "site", all = F)
```

Libros de Estadística en PDF | Statistics Books in PDF

Generalized linear models, like logistic regression, can be then be implemented in R with the `glm` function. For instance, a logistic regression model with forest cover calculated at the 1 km scale can be fit as:

```
> pres.1km <- glm(pres ~ f1km, family = "binomial", data = flsk)
> logLik(pres.1km)

##
'log Lik.' −33.83839 (df=2)
```

We plot the log-likelihoods based on fitting different models like that above as a function of forest cover calculated at different scales, ranging from 500 m to 5 km (Fig. 2.11a). Frequently, the scale with the best fit to the data (e.g., highest log-likelihood) is deemed the "scale of the effect" (Jackson and Fahrig 2015). In this case, we find that, based on the log-likelihoods of the models, forest cover within 2 km is most supported by the data (Fig. 2.11a). However, when plotting the β terms at different scales (Fig. 2.11b), we find that relationships are identical when forest cover is measured at 2 km or greater buffer sizes. In this situation, only at 500 m do we see a weak relationship with forest cover. In this case, there is great uncertainty regarding the scale of the effect, which is likely a consequence of forest cover being highly correlated across scales (Fig. 2.9). Nonetheless, there is consistent evidence that occurrence of five-lined skinks increases with forest cover on the landscape (based on cover measured at 1 km and greater scales).

Kernel Analysis. In contrast to using a buffer-based analysis, we can use a kernel-based approach for estimating the scale of effect (Fig. 2.10). To do so, we extend the approach outlined in Chandler and Hepinstall-Cymerman (2016) for logistic regression and applications on large landscapes. This approach requires customizing the log-likelihood function for estimating parameters of the logistic regression model. Here we briefly describe this process, but for more on the use of likelihood functions, see Bolker (2008).

First, we illustrate how a logistic regression model can be fit from scratch. The log-likelihood of a logistic regression model like that described in Eq. (2.1) can be formalized as:

$$L(\theta|x) = \sum_i y_i \log(p_i) + (1 - y_i)\log(1 - p_i), \tag{2.2}$$

where p_i is the probability of occurrence for survey *i,* taken from Eq. (2.1), and y_i is the observed presence or absence of the species at location *i*. In R, we code this as a negative log-likelihood function:

```
> nll <- function(par, cov, y) {
  alpha <- par[1]
  beta <- par[2]
  lp <- alpha + beta*cov #linear predictor
   p <- plogis(lp) #back-transform
```

Libros de Estadística en PDF | Statistics Books in PDF

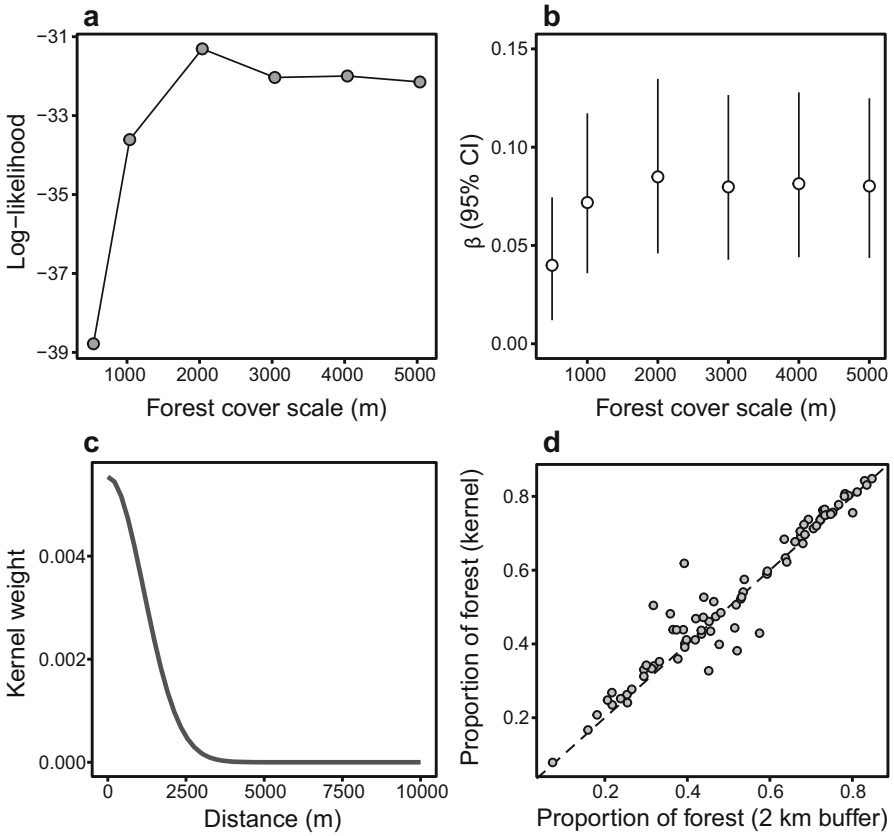


Fig. 2.11 Scale of effect of forest cover on the occurrence of the five-lined skink based on a buffer analysis and a kernel analysis. (**a**) The log-likelihood for logistic regression models, where the probability of skink occurrence is modeled as a function of the percent forest, calculated at different scales using a circular buffer around points (0.5–5 km). A higher log-likelihood suggests a better fit of the model to the data. (**b**) The parameter estimates for the effect of forest cover on the probability of skink occurrence using different sized buffers. Note the similarity in estimates among scales. (**c**) The kernel weight estimated from the data, where a maximum distance of 10 km was considered. This optimal weight emphasizes that most weight should occur for sample locations <2 km from the survey points. (**d**) The proportion of forest cover based on the optimal kernel weight and the best-fitting buffer considered ($r = 0.97$)

```
loglike <- -sum(y*log(p) + (1-y)*log(1-p)) #negative ||
return(loglike)
}
```

In this function, we first formalize the linear predictor based on the two parameters, α and β, as shown in Eq. (2.1). This linear predictor, `lp`, is on the logit scale; we back-transform `lp` to a probability scale with the `plogis` function. We then quantify the negative log-likelihood. We use the negative log-likelihood because

Libros de Estadística en PDF | Statistics Books in PDF

when we fit this model to the data, we use the `optim` function, which minimizes the log-likelihood rather than maximizing it. We can then fit the model to the data, where we consider the proportion of forest within 2 km of survey locations to find the parameter estimates using the `optim` function as:

```
#fit logistic model
> lr.buffer <- optim(par = c(0, 0), fn = nll, cov = flsk$f2km,
  y = flsk$pres, hessian = T)

> lr.buffer$par

##
[1] -6.16271714 0.08561185

> lr.buffer.vc <- solve(lr.buffer$hessian) #var-cov matrix
> lr.buffer.se <- sqrt(diag(lr.buffer.vc)) #SE
> lr.buffer.se

##
[1] 1.46565540 0.02225678
```

Note that the `optim` function requires that we provide starting values for the parameters and that requesting `hessian = T` allows for calculating the standard errors of the parameter estimates based on the inverse of the Hessian matrix (Bolker 2008). We can contrast these estimates to the use of the `glm` function described above as:

```
> summary(pres.2km)$coefficients

##
Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.1626986 1.46453460 -4.207957 2.576899e-05
f2km 0.0856115 0.02223369 3.850530 1.178626e-04
```

The estimates for α and β parameters and associated SEs are identical (up to four decimal places) to that given by the `glm` function. The log-likelihoods are also identical.

We can extend the log-likelihood function to estimate the scale of effect using a kernel approach. Different types of kernels could be considered; here we use a Gaussian kernel as illustrated in Chandler and Hepinstall-Cymerman (2016). In this approach, we summarize the landscape variable based on a weighted average using the weight, w, is based on a Gaussian kernel:

Libros de Estadística en PDF | Statistics Books in PDF

$$w(d_{ij}|\sigma) = \frac{\exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right)}{\sum_{i \neq j} \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right)}, \tag{2.3}$$

where σ is the scale parameter that determines the shape of the kernel (Fig. 2.10) and d_{ij} is the distance between sample location i and location j surrounding i. Smaller values of σ reflect greater weight to nearby locations, whereas larger values place an increasingly large weight on distant values of the covariate of interest. We can add this weighting scheme to our logistic log-likelihood function as:

```
> nll.kernel <- function(par, D, cov, y) {
  sig <- exp(par[1]) #ensures sig > 0
  alpha <- par[2]
  beta <- par[3]
  cov.w <- apply(D, 1, function(x) {
  w0 <- exp(-x^2 / (2 * sig^2)) #Gaussian kernel
  w0[w0==1] <- 0 #for truncated data
  w <- w0/sum(w0) #kernel weights; sums to 1
  sum(cov * w) #weighted average of raster
})
  lp <- alpha + beta * cov.w #linear predictor
  p <- plogis(lp) #back-transform
  loglike <- -sum(y*log(p) + (1-y)*log(1-p)) #nll
  return(loglike)
}
```

This function is similar to the above `nll` function, with the primary difference being the calculation of `cov.w`, a Gaussian kernel-weighted covariate (in our case, the proportion of forest). To calculate `cov.w`, the `apply` function is used on a distance matrix, `D`, where `D` is a site × NC matrix and NC is the number of cells in the raster being considered. The `apply` function takes each row of `D`, in which the column values represent the distance from the survey location to each raster cell. This `D` matrix can be very large: in the case of `forest200`, it is a 78 × 3215254 matrix. In our situation, such a large matrix would bog down computations, making the analysis impractical. But we can trick R to make this computationally efficient by restructuring this matrix. We can truncate our distance calculations by setting a maximum distance that we will consider, which will allow the matrix to become a "sparse matrix." Sparse matrices are matrices with many zeros, which can be stored and manipulated in an efficient way (Golub and Van Loan 1996). In our case, we will set values in `D` above some maximum distance, say 10 km, as zero and then ignore these values in the `cov.w` calculation (by setting `w0` to zero at these locations; see `nll.kernel` above). We will calculate `D` with the `rdist` function in the `fields` package (Nychka et al. 2017). To do so, we first convert our raster covariate, `forest200`, to a data frame with the `rasterToPoints` function.

Libros de Estadística en PDF | Statistics Books in PDF

```
> for200.df <- data.frame(rasterToPoints(forest200))
> library(fields)
> D <- rdist(as.matrix(flsk[,c("x","y")]),
  as.matrix(for200.df[,c("x","y")]))
```

We then convert the matrix to a sparse matrix format with the `Matrix` package (Bates and Maechler 2017).

```
> library(Matrix)
> D <- D/1000 #in km
> D[D > 10] <- 0 #truncate to only consider max dist
> D <- Matrix(D, sparse = TRUE)
```

This matrix is now "sparse," which will speed the fitting of the model. Because of the broad extent, there are many cells that are >10 km from all survey locations; for these locations, all column values are 0. We can further speed model fitting by removing these columns in `D` that are >10 km from all survey locations. We first identify the columns in which all values are not zero for subsetting our covariate values and then subset `D`:

```
> cov.subset <- which(colSums(D)!=0, arr.ind = T)
> D <- D[, cov.subset]
```

With this truncated, sparse matrix, we can now fit the kernel-based logistic regression as:

```
> lr.kernel <- optim(fn = nll.kernel, hessian = T, par = c(0,-
  6,8), D = D, cov = for200.df$layer[cov.subset], y = flsk$pres)
> lr.kernel$par

##
[1] 0.118360 -6.271160 8.563042
```

Note in our call to `optim`, we subset our forest data frame to only consider the raster data that are within 10 km of at least one survey location and we use starting values for the alpha and beta parameters that are based loosely on estimates from the buffer analysis. From this analysis, the $\log(\sigma) = 0.118$ and the effect of this covariate on occurrence is $\beta = 8.56$ (0.85 SE; note that `cov.w` is a weighted proportion of forest cover while above when using buffer analysis, we used the percent forest cover as a covariate). In general, the estimated `cov.w` is highly correlated with the proportion of forest in the 2 km buffer ($r = 0.97$) and the log-likelihoods are nearly identical. However, if we contrast AIC calculated from the `optim` function, where AIC $= -2\text{LL} + 2K$ (LL = log-likelihood, K = number of parameters), there is more support for using a simple buffer approach at the 2 km scale because it has one fewer parameter:

Libros de Estadística en PDF | Statistics Books in PDF

```
> AICkernel <- 2 * lr.kernel$value + 2 * length(lr.kernel$par)
> AICbuffer <- 2 * lr.buffer$value + 2 * length(lr.buffer$par)
> c(AICkernel, AICbuffer)

##
[1] 68.90598 66.73604
```

Despite the support for the 2 km buffer based on AIC due to the additional parameter being estimated, the kernel approach is useful for three primary reasons. First, it intuitively provides greater weight to nearby locations than distant ones, which is a more plausible assumption regarding how the landscape may influence organisms in many situations. Second, it delivers a more objective means to identify the scale of effect that does not require a priori decisions regarding potential scales to consider (as in the buffer approach). Third, it provides estimates of uncertainty in the scale of effect—an issue that is not directly estimated with buffer approaches. Despite these benefits, in practice there may be high uncertainty in estimates of σ, and estimates may be sensitive to starting values. More work on optimal estimation of the scale of effect is needed.

This example is intended to illustrate the general approach of identifying scales of effects in environmental relationships using R, an issue of increasing interest (Jackson and Fahrig 2015; Miguet et al. 2017). However, there are several improvements to the approach that could be considered, such as accounting for detection errors (MacKenzie et al. 2002) and considering the problem of spatial dependence from the use of overlapping landscapes (Zuckerberg et al. 2012). In addition, other covariates could be considered, and the scale of the effects may differ for each covariate (McGarigal et al. 2016).

Common multilevel models may be approached in a similar way as described here, but different models and data may be used for different levels (McGarigal et al. 2016). Alternatively, formal hierarchical statistical models could be considered (Gelman and Hill 2007). We will discuss aspects of such models in Chap. 6.

2.4 Next Steps and Advanced Issues

2.4.1 Identifying Characteristic Scales Beyond Species–Environment Relationships

Beyond regression techniques that correlate variables measured at different scales to response variables, there are several other techniques that are aimed at identifying characteristic scales of variability. The approaches vary depending on the specific questions being addressed and the type of data used. We will address some of these in later chapters. Some common examples are the use of spatial point pattern analysis (Chap. 4), semivariograms (Chap. 5), spatial eigenvector mapping (Chap. 6), wavelet analysis (Chap. 5), and certain types of network metrics (Chap. 9).

Libros de Estadística en PDF | Statistics Books in PDF

2.4.2 *Sampling and Scale*

Above we discussed issues of sampling and the problem of scale. In general, different sampling designs could be considered to understand scaling issues. The `raster` package has several functions for sampling raster maps, including the `SampleRandom`, `SampleRegular`, and `SampleStratified` functions. The `SampleRandom` function generates a number of random samples within the extent of a raster of interest. The `SampleRegular` function applies a grid of regularly spaced points within the extent of a raster, setting the lag distance based on the total number of points considered. The `SampleStratified` function generates random samples within strata defined by the raster layer, such as selecting 20 random points within each land-cover type on a map. Each of these functions can be useful for study design and for interpreting how variation in study design (e.g., lag distance) may influence the variation being captured across a region of interest. We note that one could also develop a sampling design outside of the raster package by generating *x*–*y* coordinates of interest (e.g., points every 100 m) and then create a `SpatialPoints` data frame with this information that could then be used with other geographic data.

2.5 Conclusions

Spatial scale is fundamental to many problems in ecology and conservation, and it is a concept that permeates this entire book. Problems of scale require interpreting the spatiotemporal domain of the phenomenon of interest, the sampling of the phenomenon, and the analysis (Dungan et al. 2002). Spatial ecology and conservation often explicitly deals with spatial scale in addressing questions and problems of conservation concern. Indeed, there has been a great deal of advancements in our understanding and quantification of scale and its effects.

Despite these advances, understanding how organisms respond to issues operating at different spatial scales is challenging. Here, we illustrate that summaries of forest structure at different spatial scales are highly correlated, a common problem in identifying relevant spatial scales for species responses to the environment. Kernel-based approaches (Heaton and Gelfand 2011; Chandler and Hepinstall-Cymerman 2016) provide an objective means to identifying the scale of effect when using data on species–environment relationships, yet there can still be a great deal of uncertainty in such estimates. We will continue to address the problem of scale throughout most of this book.

Libros de Estadística en PDF | Statistics Books in PDF

References

Allen TFH, Hoekstra TW (1990) The confusion between scale-defined levels and conventional levels of organization in ecology. J Veg Sci 1(1):5–12. https://doi.org/10.2307/3236048
Allen TFH, Hoekstra T (1992) Toward a unified ecology. Columbia University Press, New York
Andersen M (1992) Spatial analysis of two species interactions. Oecologia 91(1):134–140
Anderson BJ, Armsworth PR, Eigenbrod F, Thomas CD, Gillings S, Heinemeyer A, Roy DB, Gaston KJ (2009) Spatial covariance between biodiversity and other ecosystem service priorities. J Appl Ecol 46(4):888–896. https://doi.org/10.1111/j.1365-2664.2009.01666.x
Aue B, Ekschmitt K, Hotes S, Wolters V (2012) Distance weighting avoids erroneous scale effects in species-habitat models. Methods Ecol Evol 3(1):102–111. https://doi.org/10.1111/j.2041-210X.2011.00130.x
Bates D, Maechler M (2017) Matrix: sparse and dense matrix classes and methods. R package version 1, pp 2–12
Bivand R, Rundall C (2017) Rgeos: interface to geometry engine - open source (GEOS). R package version 0.3–26
Beale CM, Lennon JJ, Yearsley JM, Brewer MJ, Elston DA (2010) Regression analysis of spatial data. Ecol Lett 13(2):246–264. https://doi.org/10.1111/j.1461-0248.2009.01422.x
Bolker B (2008) Ecological models and data in R. Princeton University Press, Princeton, NJ
Boyd C, Brooks TM, Butchart SHM, Edgar GJ, da Fonseca GAB, Hawkins F, Hoffmann M, Sechrest W, Stuart SN, van Dijk PP (2008) Spatial scale and the conservation of threatened species. Conserv Lett 1(1):37–43. https://doi.org/10.1111/j.1755-263X.2008.00002.x
Burnham KP, Anderson DR (1998) Model selection and inference: a practical information-theoretic approach. Springer, New York
Cadenasso ML, Pickett STA, Weathers KC, Jones CG (2003) A framework for a theory of ecological boundaries. Bioscience 53(8):750–758
Cavanaugh KC, Siegel DA, Raimondi PT, Alberto F (2014) Patch definition in metapopulation analysis: a graph theory approach to solve the mega-patch problem. Ecology 95(2):316–328. https://doi.org/10.1890/13-0221.1
Chalfoun AD, Martin TE (2007) Assessments of habitat preferences and quality depend on spatial scale and metrics of fitness. J Appl Ecol 44(5):983–992. https://doi.org/10.1111/j.1365-2664.2007.01352.x
Chandler R, Hepinstall-Cymerman J (2016) Estimating the spatial scales of landscape effects on abundance. Landsc Ecol 31(6):1383–1394. https://doi.org/10.1007/s10980-016-0380-z
Chave J (2013) The problem of pattern and scale in ecology: what have we learned in 20 years? Ecol Lett 16:4–16. https://doi.org/10.1111/ele.12048
Chesson P (2012) Scale transition theory: its aims, motivations and predictions. Ecol Complex 10:52–68. https://doi.org/10.1016/j.ecocom.2011.11.002
Cohen JM, Civitello DJ, Brace AJ, Feichtinger EM, Ortega CN, Richardson JC, Sauer EL, Liu X, Rohr JR (2016) Spatial scale modulates the strength of ecological processes driving disease distributions. Proc Natl Acad Sci U S A 113(24):E3359–E3364. https://doi.org/10.1073/pnas.1521657113
Cressie N (1996) Change of support and the modifiable areal unit problem. J Geogr Syst 3 (2–3):159–180
Cushman SA, McGarigal K (2002) Hierarchical, multi-scale decomposition of species-environment relationships. Landsc Ecol 17(7):637–646. https://doi.org/10.1023/a:1021571603605
Dale MRT, Fortin MJ (2014) Spatial analysis: a guide for ecologists, 2nd edn. Cambridge University Press, Cambridge
Dark SJ, Bram D (2007) The modifiable areal unit problem (MAUP) in physical geography. Prog Phys Geogr 31(5):471–479. https://doi.org/10.1177/0309133307083294
Diamond JM (1975) The island dilemma: lessons of modern biogeographic studies for the design of natural reserves. Biol Conserv 7(2):129–146. https://doi.org/10.1016/0006-3207(75)90052-x

Libros de Estadística en PDF | Statistics Books in PDF

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kuehn I, Ohlemueller R, Peres-Neto PR, Reineking B, Schroeder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30(5):609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

Dray S, Pelissier R, Couteron P, Fortin MJ, Legendre P, Peres-Neto PR, Bellier E, Bivand R, Blanchet FG, De Caceres M, Dufour AB, Heegaard E, Jombart T, Munoz F, Oksanen J, Thioulouse J, Wagner HH (2012) Community ecology in the age of multivariate multiscale spatial analysis. Ecol Monogr 82(3):257–275. https://doi.org/10.1890/11-1183.1

Dungan JL, Perry JN, Dale MRT, Legendre P, Citron-Pousty S, Fortin MJ, Jakomulska A, Miriti M, Rosenberg MS (2002) A balanced view of scale in spatial statistical analysis. Ecography 25(5):626–640. https://doi.org/10.1034/j.1600-0587.2002.250510.x

Eigenbrod F, Hecnar SJ, Fahrig L (2017) Sub-optimal study design has major impacts on landscape-scale inference. Biol Conserv 144(1):298–305. https://doi.org/10.1016/j.biocon.2010.09.007

Evans JS (2017) spatialEco. R package version 0.0.1-7

Fahrig L (2003) Effects of habitat fragmentation on biodiversity. Annu Rev Ecol Evol Syst 34:487–515. https://doi.org/10.1146/annurev.ecolsys.34.011802.132419

Fahrig L (2013) Rethinking patch size and isolation effects: the habitat amount hypothesis. J Biogeogr 40(9):1649–1663. https://doi.org/10.1111/jbi.12130

Falk DA, Miller C, McKenzie D, Black AE (2007) Cross-scale analysis of fire regimes. Ecosystems 10(5):809–823. https://doi.org/10.1007/s10021-007-9070-7

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24(1):38–49

Fleishman E, Ray C, Sjogren-Gulve P, Boggs CL, Murphy DD (2002) Assessing the roles of patch quality, area, and isolation in predicting metapopulation dynamics. Conserv Biol 16(3):706–716

Fletcher RJ Jr (2007) Species interactions and population density mediate the use of social cues for habitat selection. J Anim Ecol 76(3):598–606

Fletcher RJ Jr, Revell A, Reichert BE, Kitchens WM, Dixon JD, Austin JD (2013) Network modularity reveals critical scales for connectivity in ecology and evolution. Nat Commun 4:2572. https://doi.org/10.1038/ncomms3572

Fletcher RJ, McCleery RA, Greene DU, Tye CA (2016) Integrated models that unite local and regional data reveal larger-scale environmental relationships and improve predictions of species distributions. Landsc Ecol 31(6):1369–1382. https://doi.org/10.1007/s10980-015-0327-9

Fortin MJ, Drapeau P, Legendre P (1989) Spatial autocorrelation and sampling design in plant ecology. Vegetation 83(1–2):209–222. https://doi.org/10.1007/bf00031693

Fortin MJ, James PMA, MacKenzie A, Melles SJ, Rayfield B (2012) Spatial statistics, spatial regression, and graph theory in ecology. Spat Stat 1:100–109. https://doi.org/10.1016/j.spasta.2012.02.004

Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York

Golub GH, Van Loan CF (1996) Matrix computations, 3rd edn. Johns Hopkins University Press, Baltimore

Gottlieb IGW, Fletcher Jr RJ, Nunez-Regueiro MM, Ober H, Smith L, and Brosi BJ (2017) Alternative biomass strategies for bioenergy: implications for bird communities across the southeastern United States. Glob Change Biol Bioenergy 9:1606–1617

Greig-Smith P (1952) The use of random and contiguous quadrats in the study of the structure of plant communities. Ann Bot 16(62):293–316

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8(9):993–1009

Guyot C, Arlettaz R, Korner P, Jacot A (2017) Temporal and spatial scales matter: circannual habitat selection by bird communities in vineyards. PLoS One 12(2):e0170176. https://doi.org/10.1371/journal.pone.0170176

Libros de Estadística en PDF | Statistics Books in PDF

Heaton MJ, Gelfand AE (2011) Spatial regression using kernel averaged predictors. J Agric Biol Environ Stat 16(2):233–252. https://doi.org/10.1007/s13253-010-0050-6
Hijmans RJ (2017) raster: geographic data analysis and modeling. R package version 2.6-7
Holland JD, Bert DG, Fahrig L (2004) Determining the spatial scale of species' response to habitat. Bioscience 54(3):227–233. https://doi.org/10.1641/0006-3568(2004)054[0227:dtssos]2.0.co;2
Holyoak M, Leibold MA, Holt RD (2005) Metacommunities: spatial dynamics and ecological communities. University of Chicago Press, Chicago
Homer C, Dewitz J, Yang LM, Jin S, Danielson P, Xian G, Coulston J, Herold N, Wickham J, Megown K (2015) Completion of the 2011 National Land Cover Database for the Conterminous United States - representing a decade of land cover change information. Photogramm Eng Remote Sensing 81(5):345–354. https://doi.org/10.14358/pers.81.5.345
Horne JK, Schneider DC (1995) Spatial variance in ecology. Oikos 74(1):18–26
Jackson ND, Fahrig L (2014) Landscape context affects genetic diversity at a much larger spatial extent than population abundance. Ecology 95(4):871–881. https://doi.org/10.1890/13-0388.1
Jackson HB, Fahrig L (2015) Are ecologists conducting research at the optimal scale? Glob Ecol Biogeogr 24(1):52–63. https://doi.org/10.1111/geb.12233
Jacobson B, Peres-Neto PR (2010) Quantifying and disentangling dispersal in metacommunities: how close have we come? How far is there to go? Landsc Ecol 25(4):495–507. https://doi.org/10.1007/s10980-009-9442-9
Jelinski DE, Wu JG (1996) The modifiable areal unit problem and implications for landscape ecology. Landsc Ecol 11(3):129–140. https://doi.org/10.1007/bf02447512
Johnson DH (1980) The comparison of usage and availability measurements for evaluating resource preference. Ecology 61(1):65–71
Keitt TH, Urban DL (2005) Scale-specific inference using wavelets. Ecology 86(9):2497–2504. https://doi.org/10.1890/04-1016
Larsen FW, Rahbek C (2003) Influence of scale on conservation priority setting - a test on African mammals. Biodivers Conserv 12(3):599–614. https://doi.org/10.1023/a:1022448928753
Levin SA (1992) The problem of pattern and scale in ecology. Ecology 73(6):1943–1967. https://doi.org/10.2307/1941447
Liebhold A, Koenig WD, Bjørnstad ON (2004) Spatial synchrony in population dynamics. Annu Rev Ecol Evol Syst 35:467–490. https://doi.org/10.1146/annurev.ecolsys.34.011802.132516
Loreau M, Holt RD (2004) Spatial flows and the regulation of ecosystems. Am Nat 163 (4):606–615. https://doi.org/10.1086/382600
Maciejewski K, Cumming GS (2016) Multi-scale network analysis shows scale-dependency of significance of individual protected areas for connectivity. Landsc Ecol 31(4):761–774. https://doi.org/10.1007/s10980-015-0285-2
MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. Ecology 83(8):2248–2255
Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405(6783):243–253. https://doi.org/10.1038/35012251
Mayor SJ, Schneider DC, Schaefer JA, Mahoney SP (2009) Habitat selection at multiple scales. Ecoscience 16(2):238–247. https://doi.org/10.2980/16-2-3238
McGarigal K, Wan HY, Zeller KA, Timm BC, Cushman SA (2016) Multi-scale habitat selection modeling: a review and outlook. Landsc Ecol. https://doi.org/10.1007/s10980-016-0374-x
McLoughlin PD, Case RL, Gau RJ, Cluff HD, Mulders R, Messier F (2002) Hierarchical habitat selection by barren-ground grizzly bears in the central Canadian Arctic. Oecologia 132 (1):102–108. https://doi.org/10.1007/s00442-002-0941-5
Melbourne BA, Chesson P (2005) Scaling up population dynamics: integrating theory and data. Oecologia 145(2):179–187. https://doi.org/10.1007/s00442-005-0058-8
Melbourne BA, Chesson P (2006) The scale transition: scaling up population dynamics with field data. Ecology 87(6):1478–1488. https://doi.org/10.1890/0012-9658(2006)87[1478:tstsup]2.0.co;2

Libros de Estadística en PDF | Statistics Books in PDF

Miguet P, Jackson HB, Jackson ND, Martin AE, Fahrig L (2016) What determines the spatial extent of landscape effects on species? Landsc Ecol. https://doi.org/10.1007/s10980-015-0314-1
Miguet P, Fahrig L, Lavigne C (2017) How to quantify a distance-dependent landscape effect on a biological response. Methods Ecol Evol 8(12):1717–1724. https://doi.org/10.1111/2041-210x.12830
Miller JR, Turner MG, Smithwick EAH, Dent CL, Stanley EH (2004) Spatial extrapolation: the science of predicting ecological patterns and processes. Bioscience 54(4):310–320. https://doi.org/10.1641/0006-3568(2004)054[0310:setsop]2.0.co;2
Minor ES, Urban DL (2008) A graph-theory framework for evaluating landscape connectivity and conservation planning. Conserv Biol 22(2):297–307. https://doi.org/10.1111/j.1523-1739.2007.00871.x
Nams VO (2005) Using animal movement paths to measure response to spatial scale. Oecologia 143(2):179–188. https://doi.org/10.1007/s00442-004-1804-z
Nychka D, Furrer R, Paige J, Sain S (2017) fields: tools for spatial data. R package version 9.6
O'Neill RV, Johnson AR, King AW (1989) A hierarchical framework for the analysis of scale. Landsc Ecol 3(3–4):193–205. https://doi.org/10.1007/bf00131538
O'Neill RV, Hunsaker CT, Timmins SP, Jackson BL, Jones KB, Riitters KH, Wickham JD (1996) Scale problems in reporting landscape pattern at the regional scale. Landsc Ecol 11(3):169–180. https://doi.org/10.1007/bf02447515
Openshaw S (1984) The modifiable areal unit problem. Geo Books, Norwich
Pascual-Hortal L, Saura S (2007) Impact of spatial scale on the identification of critical habitat patches for the maintenance of landscape connectivity. Landsc Urban Plan 83(2–3):176–186. https://doi.org/10.1016/j.landurbplan.2007.04.003
Pearson SM (1993) The spatial extent and relative influence of landscape-level factors on wintering bird populations. Landsc Ecol 8(1):3–18. https://doi.org/10.1007/bf00129863
Peters DPC, Pielke RA, Bestelmeyer BT, Allen CD, Munson-McGee S, Havstad KM (2004) Cross-scale interactions, nonlinearities, and forecasting catastrophic events. Proc Natl Acad Sci U S A 101(42):15130–15135. https://doi.org/10.1073/pnas.0403822101
Peters DPC, Bestelmeyer BT, Turner MG (2007) Cross-scale interactions and changing pattern-process relationships: consequences for system dynamics. Ecosystems 10(5):790–796. https://doi.org/10.1007/s10021-007-9055-6
Piantadosi S, Byar DP, Green SB (1988) The ecological fallacy. Am J Epidemiol 127(5):893–904. https://doi.org/10.1093/oxfordjournals.aje.a114892
Polis GA, Anderson WB, Holt RD (1997) Toward an integration of landscape and food web ecology: the dynamics of spatially subsidized food webs. Annu Rev Ecol Syst 28:289–316. https://doi.org/10.1146/annurev.ecolsys.28.1.289
Preston FW (1962) The canonical distribution of commonness and rarity: Part I. Ecology 43(2):185–215, 431–432
Pulliam HR (1988) Sources, sinks, and population regulation. Am Nat 132(5):652–661
Rahbek C (2005) The role of spatial scale and the perception of large-scale species-richness patterns. Ecol Lett 8(2):224–239. https://doi.org/10.1111/j.1461-0248.2004.00701.x
Reichert BE, Fletcher RJ Jr, Cattau CE, Kitchens WM (2016) Consistent scaling of population structure despite intraspecific variation in movement and connectivity. J Anim Ecol 85:1563–1573
Rettie WJ, Messier F (2000) Hierarchical habitat selection by woodland caribou: its relationship to limiting factors. Ecography 23(4):466–478. https://doi.org/10.1034/j.1600-0587.2000.230409.x
Schneider DC (2001) The rise of the concept of scale in ecology. Bioscience 51(7):545–553. https://doi.org/10.1641/0006-3568(2001)051[0545:trotco]2.0.co;2
Schooley RL (1994) Annual variation in habitat selection: patterns concealed by pooled data. J Wildl Manage 58(2):367–374. https://doi.org/10.2307/3809404
Schooley RL, Branch LC (2007) Spatial heterogeneity in habitat quality and cross-scale interactions in metapopulations. Ecosystems 10(5):846–853. https://doi.org/10.1007/s10021-007-9062-7

Libros de Estadística en PDF | Statistics Books in PDF

Sherry TW, Holmes RT (1988) Habitat selection by breeding American redstarts in response to a dominant competitor, the least flycatcher. Auk 105:350–364

Shurin JB, Cottenie K, Hillebrand H (2009) Spatial autocorrelation and dispersal limitation in freshwater organisms. Oecologia 159(1):151–159. https://doi.org/10.1007/s00442-008-1174-z

Slatkin M (1985) Gene flow in natural populations. Annu Rev Ecol Syst 16:393–430. https://doi.org/10.1146/annurev.ecolsys.16.1.393

Soranno PA, Cheruvelil KS, Bissell EG, Bremigan MT, Downing JA, Fergus CE, Filstrup CT, Henry EN, Lottig NR, Stanley EH, Stow CA, Tan PN, Wagner T, Webster KE (2014) Cross-scale interactions: quantifying multiscaled cause-effect relationships in macrosystems. Front Ecol Environ 12(1):65–73. https://doi.org/10.1890/120366

Stuber EF, Gruber LF, Fontaine JJ (2017) A Bayesian method for assessing multi-scale species-habitat relationships. Landsc Ecol 32(12):2365–2381. https://doi.org/10.1007/s10980-017-0575-y

Thompson CM, McGarigal K (2002) The influence of research scale on bald eagle habitat selection along the lower Hudson River, New York (USA). Landsc Ecol 17(6):569–586. https://doi.org/10.1023/a:1021501231182

Thornton DH, Branch LC, Sunquist ME (2011) The influence of landscape, patch, and within-patch factors on species presence and abundance: a review of focal patch studies. Landsc Ecol 26 (1):7–18. https://doi.org/10.1007/s10980-010-9549-z

Turner MG, Gardner RH (2015) Landscape ecology in theory and practice, 2nd edn. Springer, New York

Turner MG, Dale VH, Gardner RH (1989a) Predicting across scales: theory development and testing. Landsc Ecol 3(3–4):245–252. https://doi.org/10.1007/bf00131542

Turner MG, O'Neill RV, Gardner RH, Milne BT (1989b) Effects of changing spatial scale on the analysis of landscape pattern. Landsc Ecol 3(3–4):153–162. https://doi.org/10.1007/bf00131534

Urban DL, O'Neill RV, Shugart HH (1987) Landscape ecology. Bioscience 37(2):119–127

Villard MA, Trzcinski KM, Merriam G (1999) Fragmentation effects on forest birds: relative influence of woodland cover and configuration on landscape occupancy. Conserv Biol 13 (4):774–783

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol 15 (6):1419–1439. https://doi.org/10.1111/j.1365-294X.2006.02890.x

Weaver JE, Conway TM, Fortin MJ (2012) An invasive species' relationship with environmental variables changes across multiple spatial scales. Landsc Ecol 27(9):1351–1362. https://doi.org/10.1007/s10980-012-9786-4

Wheatley M, Johnson C (2009) Factors limiting our understanding of ecological scale. Ecol Complex 6(2):150–159. https://doi.org/10.1016/j.ecocom.2008.10.011

Wiens JA (1989) Spatial scaling in ecology. Funct Ecol 3(4):385–397

Wu JG, David JL (2002) A spatially explicit hierarchical approach to modeling complex ecological systems: theory and applications. Ecol Model 153(1–2):7–26

Wu JG, Loucks OL (1995) From balance of nature to hierarchical patch dynamics: a paradigm shift in ecology. Q Rev Biol 70(4):439–466. https://doi.org/10.1086/419172

Zuckerberg B, Desrochers A, Hochachka WM, Fink D, Koenig WD, Dickinson JL (2012) Overlapping landscapes: a persistent, but misdirected concern when collecting and analyzing ecological data. J Wildl Manage 76(5):1072–1080. https://doi.org/10.1002/jwmg.326

Libros de Estadística en PDF | Statistics Books in PDF

Chapter 3
Land-Cover Pattern and Change



3.1 Introduction

Understanding spatial and temporal variation in land use and land cover is a topic that bridges disciplines such as ecology, geography, sociology, and economics (Lambin et al. 2001; Rindfuss et al. 2004; Turner et al. 2007). It is also a topic that is well integrated into spatial ecology and conservation (Vitousek 1994; Blair 1996). The topics of habitat loss and fragmentation, agricultural intensification, agroforestry, and urbanization all involve land-use and land-cover change (Brockerhoff et al. 2008; Grimm et al. 2008; Ewers et al. 2009). Land-use change has had major effects on biodiversity (Newbold et al. 2015) and it is predicted to have unprecedented effects on biodiversity in the coming decades (Tilman et al. 2017).

Interpreting the effects of land use and land cover (hereafter LULC) requires quantifying its spatial patterns. This quantification typically focuses on how to interpret patterns from categorical maps (McGarigal et al. 2002). In some cases, interest lies in quantifying continuous data (McGarigal et al. 2009), such as quantifying the Normalized Difference Vegetation Index (NDVI) and its influence on animals (Pettorelli et al. 2005). In general, quantifying spatial patterns of LULC is complex, and there are hundreds of metrics and a variety of frameworks used to guide the quantification of LULC (Vogt et al. 2007; Cushman et al. 2008; Walz et al. 2016).

Here, we focus on the quantification of land use and land cover, with an eye toward then relating these patterns to ecological processes. Our goals for this chapter are to introduce key concepts regarding land-use change, provide insight to the quantification of such change, and illustrate how spatial patterns can be quantified at different scales. We first discuss some foundational concepts and terms that capture key aspects of LULC. We then provide an overview on common ways in which spatial patterns of LULC are quantified, focusing primarily on the quantification of land-cover variation and heterogeneity (Li and Reynolds 1995), which is typically

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_3

Libros de Estadística en PDF | Statistics Books in PDF

based on categorical maps (McGarigal et al. 2002).We provide examples of these approaches with land-cover data from the Southeast USA. We also provide an overview on the use of neutral landscapes for interpreting spatial patterns (Gardner et al. 1987; Gardner and Urban 2007; Etherington et al. 2015).

3.2 Key Concepts

3.2.1 Land Use Versus Land Cover

"Land use" and "land cover" are often used interchangeably, but each concept captures different issues (Table 3.1) (Lambin et al. 2001). *Land use* refers specifically to how humans utilize landscapes of the Earth, and it often incorporates socioeconomic issues. *Land cover* refers specifically to the physical material of Earth at a location, such as water, vegetation, or concrete. Some common examples of land use include various types of agriculture, urbanized landscapes, and community-owned forestry practices. Land-use change can occur through land-use extensification or through land-use intensification (Pinto-Correia and Mascarenhas 1999; Tilman et al. 2011; Macedo et al. 2012). Extensification refers to the spread in land-use practices across a region, such that a greater total area of a land use occurs. Land-use intensification refers to changes in an existing land use, where a greater amount of inputs or changes occur. For example, land-use change from corn (*Zea mays*) agriculture can occur via extensification where land uses, such as pastures, are converted to corn agriculture (Wright et al. 2017), or through intensification, where corn is grown for greater yield through increased use of fertilizers or greater irrigation (Grassini and Cassman 2012).

3.2.2 Conceptual Models for Land Cover and Habitat Change

To interpret land-cover change in ecology, several conceptual models have been used (Lindenmayer and Fischer 2007). These models have different fundamental assumptions regarding variation in LULC patterns. These models vary in the complexity for which they interpret the environment and whether the environment is assumed to be discrete or is considered as continuous environmental gradients (Fig. 3.1).

Perhaps the earliest conceptual model applied to land cover was the *island model*. This simple model only considered discrete islands or patches of a focal *cover type*, ignoring all other variation in land use and land cover. Spatial pattern is typically quantified based on island size (or patch size, see below) and *isolation*. This general conceptual model emerged from island biogeography and metapopulation theories (MacArthur and Wilson 1967; Hanski 1999; Diamond 1975) (see Chaps. 10 and 11).

Libros de Estadística en PDF | Statistics Books in PDF

Table 3.1 Common terms and concepts used in land-use and land-cover problems in ecology

Term/concept	Description
Aggregation	Tendency of patch or land-cover types to be spatially adjacent or in close proximity. Can capture several related concepts: dispersion, interspersion, subdivision, and isolation.
Boundary	A zone composed of the edges of adjacent ecosystems (see edge below).
Composition	The amount and number of elements, cover types, or habitats.
Configuration	Specific arrangement of spatial elements: often used synonymously with physiognomy or spatial structure.
Contrast	The magnitude of difference in a variable between adjacent patch types. Also known as "edge contrast."
Corridor	A relatively narrow strip of a particular type that differs from the areas adjacent on both sides.
Cover type	Category within a classification scheme defined by the user that distinguishes among the different habitats, ecosystems, or vegetation types on a landscape.
Dispersion	The spatial distribution of a land-cover type without explicit reference to other land-cover types.
Ecotone	A region of transition between two biological communities. Typically more gradual than an edge or boundary.
Edge	Proportion of an ecosystem or cover type near its perimeter and within which environmental conditions may differ from interior locations in the ecosystem.
Fragmentation	The breaking up of habitat or cover types into smaller, disconnected parcels. Most appropriately used as a measure for a given amount of habitat loss.
Functional metrics	Landscape metrics that use information on the species or ecological process being considered to alter the quantification of spatial pattern, such that the same landscape may have different patterns for different species or processes. Examples include using information on dispersal distances or distance of edge effects to alter connectivity and core area metrics.
Landscape heterogeneity	Variation in in landscape structure, typically measured with landscape metrics.
Interspersion	The spatial mixing of different land-cover types without explicit reference to the dispersion of any single land-cover type.
Isolation	Degree to which patches are separated from each other. Focuses on distance between patches.
Land cover	A description or classification of the physical material of Earth at a location.
Land use	A description or classification of how humans utilize areas of the Earth.
Landscape metrics	Group of indices (e.g., mean patch size) used to characterize composition and spatial configuration of landscape such as diversity, homogeneity, and fragmentation.
Matrix	Background cover type(s) in a landscape, characterized by extensive cover and high connectivity; not all landscapes have a definable matrix.
Neutral landscape	A neutral or null model of landscape pattern, in which expected patterns are generated in the absence of specific biological or ecological processes.
Patch	Surface area that differs from its surroundings in nature or appearance.

(continued)

Libros de Estadística en PDF | Statistics Books in PDF

Table 3.1 (continued)

Term/concept	Description
Structural metrics	Landscape metrics that quantify the physical structure of the environment without respect to the species or process under considerations, such that only one value occurs for a given landscape (or patch). Contrast with functional metrics.
Subdivision	The breaking apart of land-cover types into distinct patches. The distance between patches is not considered.
Thematic resolution	The resolution of land-cover and land-use classifications. As thematic resolution increases, there is an increasing number of categories.

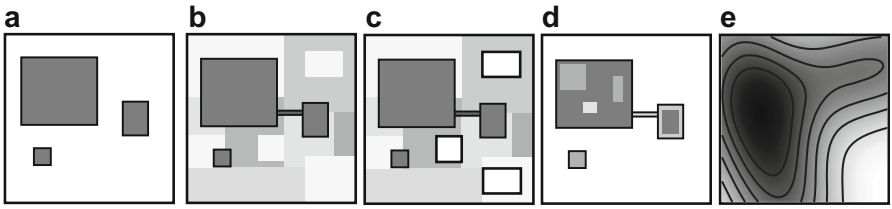


Fig. 3.1 Some common conceptual models for land-cover used for quantifying patterns and interpreting its effects on ecological processes. Shown is an example where the real world is conceptualized based on (**a**) the island model, (**b**) the patch-corridor-matrix model, (**c**) the landscape mosaic model, (**d**) the habitat variegation model, and (**e**) the continuum model

It has been frequently used to address problems of habitat loss and fragmentation (Haila 2002; Fahrig 2003; Fisher and Lindermayer 2007).

The island model has been extended and largely replaced by the *patch-matrix-corridor model* (Forman 1995b), also known as the *patch mosaic* model (Turner 1989). In general, a *patch* represents a relatively discrete area containing homogeneous conditions, where, within patches, conditions are similar enough to effectively ignore the internal variability (Wiens 1976; Forman and Godron 1981). Patches are delineated based on discontinuities or changes in land-cover or land-use conditions. This delineation may be contingent on the scale being considered (see Chap. 2). Patch delineation can be done visually or through a set of rules based on adjacencies of a land use or land cover relative to the focal location (see below). The *matrix* is considered the non-focal land-cover types or elements in the landscape (Kupfer et al. 2006). It sometimes refers to the most dominant land-cover type that is not the focal land-cover type. For example, if we were interested in dipterocarp forests of Southeast Asia, the matrix is often oil palm forestry (Sodhi et al. 2010). In recent years, several lines of evidence suggest that the matrix can be more important than patch area and other local factors in explaining biodiversity (Haynes et al. 2007; Prugh et al. 2008). A *corridor* typically refers to linear landscape element that connects patches and can be defined based on the structure of the landscape (i.e., structural corridors) or through their function (i.e., functional corridors), where function is inferred if corridors promote movement or flow (see Chap. 9). The *landscape mosaic model* (Wiens 1995) is related to the patch-matrix-corridor model, but it emphasizes

Libros de Estadística en PDF | Statistics Books in PDF

that there are different types of patches from various cover types, such that it de-emphasizes a single focal habitat or cover type. For instance, different types of natural land cover (e.g., forest and wetland), agricultural land uses, and urban areas may all be simultaneously considered (Fahrig et al. 2011; Gottlieb et al. 2017). This conceptual model is now often used in a variety of conservation settings, particularly in situations where multiple objectives for land use and conservation are considered (Polasky et al. 2008; Phalan et al. 2011). These models simplify the landscape as a series of patches and their surrounding context. Because of this simplification, it is a tractable conceptual model for land cover and landscape variation, but it may not be appropriate in some situations, particularly when there are important environmental gradients that are more continuous in form or when patch delineation is difficult.

Conceptual models have therefore been advanced that attempt to better capture environmental gradients. These include the habitat variegation model and the continuum model. The *habitat variegation model* extends some of the ideas of the patch-corridor-matrix model to consider the fact that disturbances can modify environments in a continuous manner, rather than simply the wholesale destruction of habitat (McIntyre and Barrett 1992; McIntyre and Hobbs 1999). Thus, rather than simple classifications of habitat/non-habitat, habitat may be considered unmodified or modified from disturbances. This model emphasizes that landscapes may be generally intact, variegated (where disturbances modify, but do not necessarily destroy habitat), fragmented (where habitat loss creates patch fragments), or relictual, where landscapes have nearly all habitat lost. This framework has been used to understand species distributions in a variety of human-dominated landscapes (Fischer and Lindenmayer 2002; Thornton et al. 2013; Vergara et al. 2017).

The *continuum model* and related *gradient model* emphasize that landscapes can be considered a combination of several environmental gradients, which are often continuous in their form (Fisher et al. 2004; Fischer and Lindenmayer 2006; Cushman et al. 2010). These models de-emphasize the concept of patches because environmental conditions may not be so similar within patches or disrupted between them to generate meaningful patch delineations. In doing so, these models also abandon the habitat/non-habitat dichotomy, such that focal habitats are also de-emphasized. Rather, these models assume that habitat is a species-specific concept (Hall et al. 1997) and that several aspects of the environment drive variation in habitat. These models also assume that species may respond to the environment at different spatial grains (Kotliar and Wiens 1990). This general framework has been useful to explain communities in landscapes lacking strong contrast among land uses (Brudvig et al. 2017). These models fit more naturally into niche concepts in ecology (see Chap. 7) and suggest that landscape patterns should be quantified in species-specific ways using *functional metrics* (rather than *structural metrics*; Table 3.1).

Libros de Estadística en PDF | Statistics Books in PDF

3.2.3 *Habitat Loss and Fragmentation*

The issue of habitat loss and fragmentation is one common type of land-use and land-cover change. Understanding the role of *habitat loss* and *habitat fragmentation* on ecological patterns and processes has been of long-standing interest in spatial ecology and conservation (Diamond 1975; Fahrig 2003; Tscharntke et al. 2012; Haddad et al. 2015). Early views on habitat fragmentation tended to focus on how land conversion, such as the clearing of forests, could result in smaller, more isolated patches, with a greater proportion of habitat *edge* (Fig. 3.2) (Wilcove 1985; Ries et al. 2004). More recently, the reduction in habitat (habitat loss) is often distinguished from the breaking apart of habitat (habitat fragmentation), such that fragmentation is quantified for a given amount of habitat loss (Fahrig 2003, 2017;

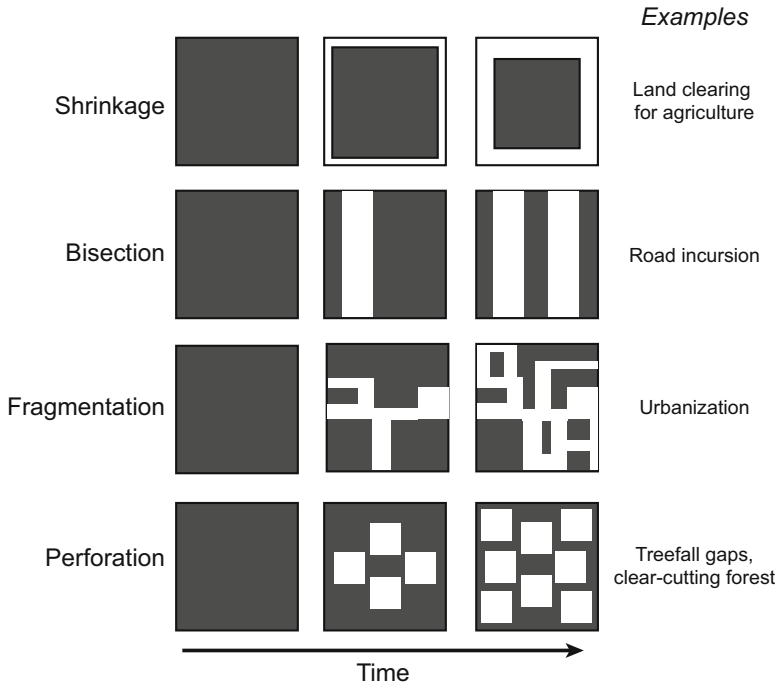


Fig. 3.2 Ways in which habitat can be lost and fragmented over time. Habitat loss and fragmentation are often conceptualized as processes that occur over time, where landscapes that are initially contiguous undergo different patterns of loss. This land conversion can arise from the shrinkage of habitat, bisection of habitat from land changes such as road infrastructure and development, fragmentation (in its narrow sense) where contiguous habitat is broken into pieces, and perforation where habitat has internal loss where holes are cut into contiguous habitat. In practice, many of these changes occur simultaneously, such as road networks bisecting habitats, leading to fragmentation of remaining habitat. Shown are changes across three time periods, with the same amount of loss for each time period occurring for each pattern. Modified from Collinge and Forman (1998)

Libros de Estadística en PDF | Statistics Books in PDF

Hadley and Betts 2016). In practice, these two issues are often confounded because habitat fragmentation requires that habitat loss occurs over time (Ewers and Didham 2006; Didham et al. 2012; Villard and Metzger 2014; Fletcher et al. 2018). Technically, we might want to consider habitat fragmentation effects conditional on the amount of loss.

There are several ways in which habitat can be lost and fragmented over time, leading to variation in the spatial patterns of land cover, which can have major consequences for biodiversity. Forman (1995a) and Collinge and Forman (1998) emphasized four general patterns in a conceptual framework for land conversion: shrinkage, bisection, fragmentation (in its narrow sense), and perforation (Fig. 3.2). Shrinkage occurs when a contiguous land cover reduces in size, without land cover being broken into pieces; such changes can arise from land clearing for agriculture. Bisection occurs when a contiguous land cover becomes split into two or more patches from road incursion or related types of land change. Fragmentation in its narrow sense occurs when contiguous habitat is broken into pieces. Finally, perforation occurs when internal holes are punched into formerly contiguous habitat, such as treefall in forest creating gaps or clear-cutting of contiguous forest. In this framework, Collinge and Forman (1998) envisioned these patterns of habitat loss would initially occur and then the shrinkage of remaining habitat would continue over time. Alternatively, bisection, fragmentation, and perforation could continue over time without shrinkage of remaining habitat (Fig. 3.2). Overall, Collinge and Forman (1998) emphasized that each of these patterns of habitat loss resulted in different amounts of edge and connectivity over time. For instance, perforation and fragmentation can result in the greatest proportion of edge, whereas bisection and fragmentation can lead to the greatest decreases in connectivity of land cover. In general, each of these types of loss likely occurs together over time with ongoing habitat loss.

There are many reasons why habitat loss and fragmentation has received considerable focus and several concepts throughout this book relate directly or indirectly to the problem of habitat loss and fragmentation. Habitat loss is considered to be one of the most important—if not the most important—threat to biodiversity both locally and world-wide (Wilcove et al. 1998; Brooks et al. 2002; Jetz et al. 2007). The reason is simple: as habitats get smaller, there are fewer resources for organisms, which ultimately reduces population sizes and alters community structure (Fahrig 2003). Understanding effects of fragmentation can provide conservation strategies to mitigate the effects of loss, such as the use of conservation corridors to connect remaining habitats.

Conceptually, there has also been a great deal of elegant theoretical development to interpret the effects of habitat loss and fragmentation (Hill and Caswell 1999; Flather and Bevers 2002). This theoretical development has emerged from two different fields. First, the development of the equilibrium theory of island biogeography (MacArthur and Wilson 1967) was incredibly influential. In this theory, island area and isolation (distance from mainland) were emphasized as being critical for extinction and immigration dynamics on islands. Shortly after its description by MacArthur and Wilson, it was applied to terrestrial systems where habitat loss and

Libros de Estadística en PDF | Statistics Books in PDF

fragmentation were occurring (Diamond 1975). See Chap. 11 for more on the influence of island biogeography for community ecology and conservation planning. Second, theoretical developments in spatial ecology began focusing on the role of patchy resources on populations and species interactions (Huffaker 1958; Roff 1974). This theoretical development was then applied to interpret the effects of habitat loss and fragmentation.

3.2.4 Quantifying Land-Cover Pattern

A variety of approaches have been developed to quantify LULC patterns. These approaches typically start with a map that has been classified into different land-use or land-cover categories. The number and detail of the categories is referred to as the *thematic resolution*. Ultimately, patterns will vary depending on the thematic resolution of the data. In addition, the scale of the map, both grain and extent, can affect the quantification of pattern. Here, we focus on general issues regarding the quantification of land-cover pattern via *landscape metrics*. In Sect. 3.3.3 we discuss specific metrics frequently used.

3.2.4.1 Composition Versus Configuration

When quantifying variation in land use and land cover across the landscape, pattern can arise from variation in composition and/or in configuration (Gustafson 1998). *Composition* emphasizes the amount and variety of different land-use or land-cover types, without explicit consideration of land-use or land-cover arrangement. In contrast, *configuration* focuses on the arrangement and/or position of land uses or land covers across landscapes.

Distinguishing the effects of landscape composition versus configuration is highly relevant to many problems in ecology and conservation. Based on this distinction, habitat loss focuses on changes in the composition of the landscapes, whereas habitat fragmentation focuses on changes in the configuration of the landscape (Fahrig 2003). Another issue where this distinction is relevant is related to the habitat heterogeneity hypothesis in agricultural landscapes. This hypothesis posits that landscapes with greater heterogeneity, either from compositional and/or configurational heterogeneity, may harbor greater biodiversity than those with less heterogeneity (Benton et al. 2003; Oliver et al. 2010; Fahrig et al. 2011; Fahrig 2017; Reynolds et al. 2018). Understanding the role of compositional versus configurational heterogeneity in this way is essential for promoting biodiversity in agricultural landscapes.

For ecology, we might a priori predict that changes in composition may have different effects on ecological patterns and processes than configuration. For instance, changing the composition of land uses may alter resource abundance and ultimately carrying capacity of species in a region of interest. In contrast, altering the

Libros de Estadística en PDF | Statistics Books in PDF

configuration of land uses may influence movement-related processes (Cushman et al. 2012) and changes in resource quality across space via changes in edge effects (Sisk et al. 1997; Ries et al. 2004; Pfeifer et al. 2017). For conservation, problems of composition focus on questions regarding "how much" and "what variety" (Fahrig 2001). In contrast, configuration focuses on questions of "where" and "under what context" (Lookingbill et al. 2010b).

Because composition focuses on "how much," its quantification is relatively straightforward. Compositional metrics, such as habitat amount or the proportion of a land use/land cover and the diversity of cover types, are easily calculated and interpreted (see Examples below). However, landscape configuration is much more challenging to quantify. Landscape and spatial ecologists have devoted a great deal of effort to the quantification of landscape configuration and there are now over one hundred metrics that have been developed (Cushman et al. 2008). It is not our goal to describe all of these composition and configuration metrics. Rather it may be useful to interpret what general aspects of configuration that these metrics attempt to quantify.

Configuration metrics capture several related concepts to varying degrees. These include contrast, aggregation, dispersion, interspersion, isolation, and subdivision (Table 3.1) (McGarigal et al. 2002). *Contrast* reflects configuration because it explicitly considers the adjacency of different land-cover types, such as edge contrast (Suarez et al. 1997; Fletcher and Koford 2003; Ries and Sisk 2010). *Aggregation* and its related elements—*dispersion*, *interspersion*, *isolation*, and *subdivision*—also focus on configuration because each attempts to quantify the context of cells or patches in a landscape, such as the proximity of land cover to similar or different land-cover types. Each of these issues has been shown to be important in metapopulation persistence, the outcome of species interactions, and other ecological processes (see Chaps. 10 and 11; Tilman and Lehman 1997; Ovaskainen et al. 2002).

3.2.4.2 Scale for Land-Cover Quantification

Land cover can be quantified in many different ways for ecological problems. Quantification of land-cover heterogeneity can occur at different scales or levels: at the cell (grain) level, the patch-level, the class-level, and the landscape-level (Fig. 3.3). In this way, metrics are typically organized as "patch-level," "class-level," or "landscape-level" metrics when quantifying pattern on land-cover and land-use maps (Cushman et al. 2008). These different levels can be applied to problems that focus on patches, neighborhoods, or the entire landscape (sometimes called the "scope" of the investigation) (McGarigal et al. 2002).

At the cell level, heterogeneity is quantified without respect to patches. The result is that each cell has a summary statistic reflecting land-cover heterogeneity. The most common example of this approach is the use of moving-window analyses, where quantification of the land cover at or surrounding each cell (typically based on a buffer or kernel; see Sect. 3.3.3.4) on the map is quantified, resulting in a new map

Libros de Estadística en PDF | Statistics Books in PDF

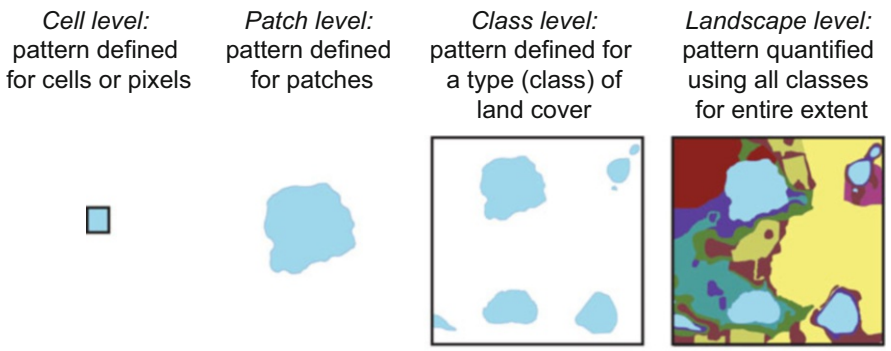


Fig. 3.3 Land-cover patterns are frequently quantified based on different scales of heterogeneity. Pattern can be summarized for each cell or pixel on a landscape (frequently through moving-window analysis), for individual patches (either characteristics of the patches themselves or their surroundings), for land-cover types or classes across the entire landscape, or through summaries using information on all classes in the landscape

where each cell has a unique summary value. Morphological imaging processing occurs at the cell level as well (see Advanced Topics) (Vogt et al. 2007).

At the patch level, the focus is on quantifying aspects of patches, such as their size, isolation, or the amount of perimeter (i.e., length of boundaries). This scale requires the formal delineation of patches in the landscape. Patch-level metrics have played an integral role in island biogeography (MacArthur and Wilson 1967), patch dynamics (Pickett and Thompson 1978; Wu and Loucks 1995), and metapopulation dynamics (Levins 1969; Hanski 1998). The island model and patch-matrix-corridor model largely focused on patch-level metrics to interpret spatial pattern.

The class-level focuses on summarizing variation in a cover type or "class" across the landscape. Thus, a "class" is simply a cover type or category on a map that describes a type of land cover or land use. It can include landscape-scale summaries of patch variation (e.g., mean patch size) or metrics that do not require patch delineation (e.g., the proportion of forest). In this situation, focal land-cover types are emphasized, but position relative to other land-cover types is sometimes considered (e.g., edge contrast). Class-level metrics have been instrumental in understanding effects of habitat loss and fragmentation (e.g., Villard et al. 1999).

The landscape-level focuses on quantifying land-cover variation for the entire landscape. In this situation, all land-cover types are typically considered and specific variation for each land-cover type (i.e., class-level heterogeneity) is usually ignored or pooled to quantify the overall pattern of the landscape. An example where landscape-level metrics have been important comes in understanding the role of heterogeneity on biodiversity in agricultural landscapes (Fahrig et al. 2011).

Several investigations have emphasized that many of the metrics that aim to quantify variation in landscape composition and configuration are highly correlated (McGarigal and McComb 1995, Fortin et al. 2003). Attempts have been made to identify which types of metrics provide important and non-redundant, or

Libros de Estadística en PDF | Statistics Books in PDF

complementary, information about patterns (Riitters et al. 1995; Neel et al. 2004; Wang et al. 2014). Cushman et al. (2008) provided one of the most thorough analyses of these correlations. They emphasized three important characteristics when using these types of metrics: strength, universality, and consistency. They attempted to identify components of landscape structure (i.e., components were combinations of correlated metrics describing land-cover pattern) that were derived from 103 landscape metrics. For these components, strength refers to the amount of variation explained by a component across classes and regions, universality is the percentage of classes or regions from which a component is found, while consistency describes the stability of the component interpretation across classes and regions. From these analyses, they identified seven major class-level components and eight major landscape-level components that explained land-cover variation (Cushman et al. 2008).

3.3 Examples in R

3.3.1 Packages in R

In R, there are a few libraries that can be used for land-cover quantification. We will start with some simple examples using the `raster` package (for raster data) (Hijmans and Van Etten 2012) and the `rgeos` package (for vector-based data) (Bivand and Rundall 2017). `SDMTools` (VanDerWal et al. 2010) provides some more advanced metrics for land-cover quantification that are largely based off of the program Fragstats (McGarigal et al. 2002). `lulcc` is a package in R that focuses on quantifying LULC change (Moulds et al. 2015). At the time of publication, `landscapemetrics` was released for calculating a wide variety of landscape metrics (Hesselbarth et al. 2018), which should also be considered, but it is not covered here.

3.3.2 The Data

To illustrate land-cover variation across landscapes, we return to the land-cover data used in Chap. 2 on scale. Land-cover data comes from the 2011 National Land Cover Database (NLCD) (Homer et al. 2015). We focus primarily on one landscape considered in Chap. 2 so that we can more readily visualize and interpret land-cover variation.

Libros de Estadística en PDF | Statistics Books in PDF

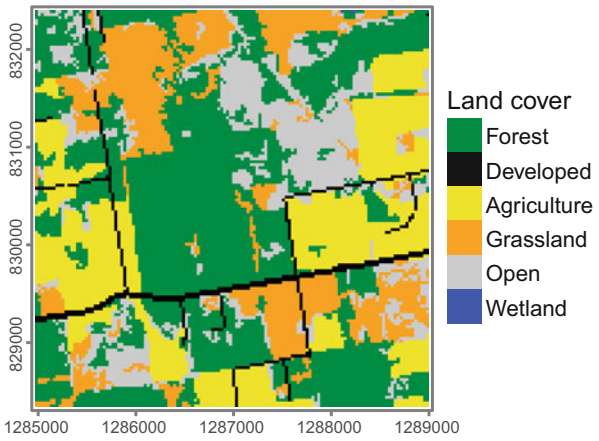


Fig. 3.4 The landscape being considered. Shown is the reclassified National Land Cover Database for a 4 × 4 km landscape in Alabama, which was sampled in Gottlieb et al. (2017)

3.3.3 Quantifying Land-Cover Variation at Different Scales

We first load the raster layer being considered and take a look at its attributes, including the thematic resolution being considered, the grain, and extent. In this situation, we reclassified the NLCD layer to simplify the thematic resolution down to six categories (Fig. 3.4): forest, developed, agriculture (rowcrops), grassland, open, and wetlands.

```
> library(raster)
> library(SDMTools)

> nlcd <- raster("nlcd2011gv2sr")

#grain and extent
> res(nlcd)
> extent(nlcd)

#nlcd thematic resolution
> levels(nlcd)
```

The resolution is 30 × 30 m and the extent covers approximately 4 × 4 km. With the `levels` function, we find that initially R did not treat the land-cover data as factors, so we reformat the raster layer to a factor.

```
#convert land-cover integers to factors
> nlcd <- as.factor(nlcd)
```

R is now treating the land-cover categories as factors, but they are only labeled as integer values. For mapping we may want to label the integers based on the land-

Libros de Estadística en PDF | Statistics Books in PDF

cover type classifications. We can provide labels and plot the map. To appropriately label the legend, the `rasterVis` package (Lamigueiro and Hijmans 2018) provides a straightforward approach with the `levelplot` function.

```
#add names of land-cover categories to raster
> land_cover <- levels(nlcd)[[1]]
> land_cover[,"landcover"] <- c("forest", "developed", "ag",
  "grass", "open", "wetland")
levels(nlcd) <- land_cover

#plot with custom color scheme
> library(rasterVis)
> land_col <- c("green", "orange", "yellow", "brown", "white",
  "blue")
> plot(nlcd, legend = T, col = land_col)
> levelplot(nlcd, col.regions = land_col, xlab = "", ylab = "")
```

3.3.3.1 Patch-Level Quantification

In ecology there has long been a focus on understanding variation among habitat patches. Classic examples include forest fragments (Whitcomb et al. 1976), grassland meadows that are interspersed by forest (Harrison 1991), or wetland patches (Naugle et al. 1999; Lookingbill et al. 2010a).

To quantify characteristics of patches, the first step is to delineate the patches themselves. This step is not trivial and can have important impacts on the conclusions regarding the effects of patch variation on ecological patterns and processes. For vector maps, typically patches are delineated by the user (e.g., hand digitizing aerial photographs). However, for raster-based maps, we typically automate patch delineation, using one of two common rules: the four-neighbor rule (also known as the "rook's rule") and the eight-neighbor rule (also known as the "Queen's rule") (Fig. 3.5). Using the four-neighbor rule will invariably result in a greater number of patches and smaller patches than using the eight-neighbor rule. Note in some situations we might want to use a 16-neighbor rule, such as if we would like for patch delineation to account for the potential for gap-crossing by organisms (Bowman and Fahrig 2002), although in practice this is rarely done.

We first summarize forest cover in the landscape from a patch perspective. Common patch-level metrics include patch size, perimeter–area ratio, core area, and patch isolation. Patch size and isolation are common metrics used in island biogeography, metapopulation biology, and metacommunity ecology (MacArthur and Wilson 1967; Hanski 1998; Holyoak et al. 2005). *Patch size* is important because it can predict local extinction probabilities in patches, it may be related to resource amount and variation, and it may influence immigration rates and habitat selection (Johnson and Igl 2001; Bowman et al. 2002), resulting in variation in

Libros de Estadística en PDF | Statistics Books in PDF

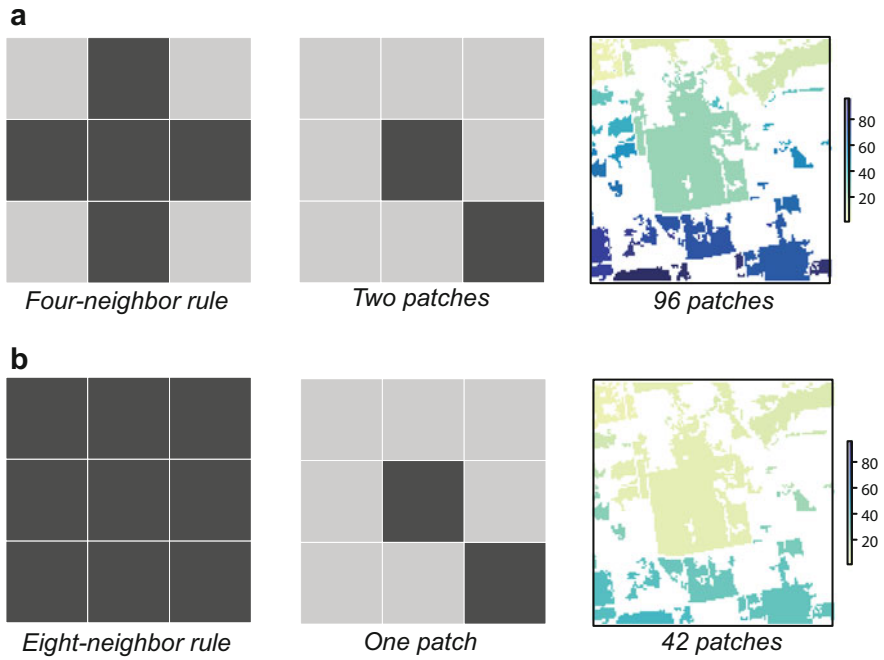


Fig. 3.5 Patch definition: four-neighbor and eight-neighbor rules. (**a**) Patches delineated in the landscape using the four-neighbor rule and (**b**) the eight-neighbor rule. In the landscape being considered, a four-neighbor rule leads to 96 patches being delineated, whereas an eight-neighbor rule results in 42 patches being delineated (right panels). Note that a 16-neighbor rule is occasionally used and allows for organisms to move 1 cell in intervening land-use while still considering it part of the same patch (such as crossing internal gaps)

occupancy, abundance, and diversity with patch size. *Patch isolation* is thought to be critical for colonization and dispersal rates, thereby influencing occupancy, abundance, and diversity (Moilanen and Hanski 2001). Note that it has been argued that patch isolation may be correlated with habitat amount at landscape scales, such that it can be difficult to interpret if such patterns with patch isolation are from isolation processes or through issues of habitat area at a landscape scale (Fahrig 2003, 2013).

Metrics of *patch shape* (e.g., perimeter–area ratio) and *core area* have long been used in conservation biology (Temple and Cary 1988; Laurance 1991; Ewers and Didham 2007) because they tell us something about the relative amounts of edge that may influence conditions in patches. For patch shape metrics, we only need to calculate the perimeter for each patch and we then take this value and divide by the area of the patch. *Core area* is defined as the area of a patch free from edge effects (Temple and Cary 1988). For that metric, we need to define the distance at which edge effects penetrate into patches (referred to at the "distance of edge influence") (Chen et al. 1992; Harper et al. 2005; Ries et al. 2017). With that value, we can then create buffers within patches to identify core areas (buffering inside rather than outside of patches, like what we did in Chap. 2).

Libros de Estadística en PDF | Statistics Books in PDF

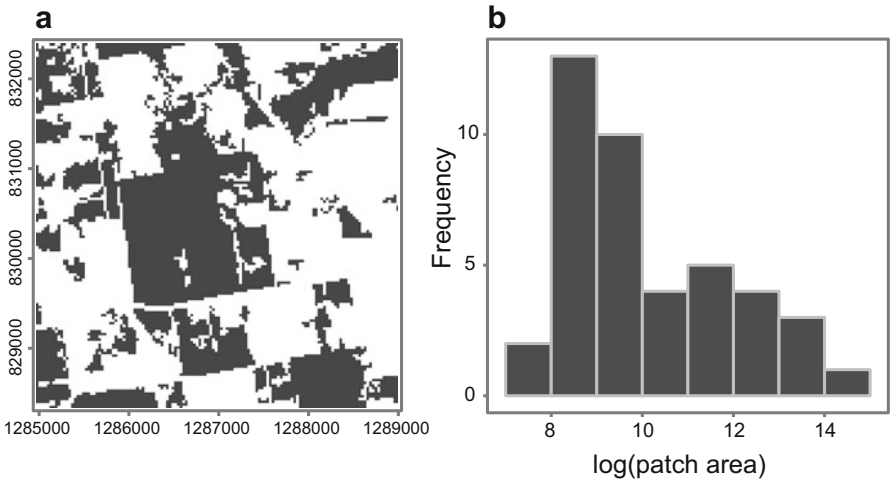


Fig. 3.6 (**a**) A binary representation of forest land cover for the landscape shown in Fig. 3.4, and (**b**) a summary metric of patch-level variation regarding the distribution of forest patch size (log (area)) in the landscape

We focus on identifying patches of forest and interpreting their variation. First, we delineate patches, then we use these delineations to calculate metrics for the patches that describe variation in patch structure. We reclassify the NLCD layer to create a binary layer of forest, akin to the island model of land-cover variation. To do so, we create a reclassification matrix, similar to the approach shown in Chap. 2. In this matrix, the first column is the original land-cover categories and the second column is the new categories.

```
#create a reclassification matrix
> nlcd.cat <- unique(nlcd)
> nlcd.cat.for <- c(1, 0, 0, 0, 0, 0)

> reclass.mat <- cbind(nlcd.cat, nlcd.cat.for)

#forest binary layer from reclassification matrix
> nlcd.forest <- reclassify(nlcd, reclass.mat)
> plot(nlcd.forest)
```

With this new binary forest layer (Fig. 3.6a) we can delineate forest patches. Both the `raster` and `SDMTools` packages have a means to do this, but currently the `SDMTools` package only allows patch delineation based on an eight-neighbor rule (Fig. 3.5b). Consequently, we will use the `raster` package to have more flexibility in accomplishing this task using the `clump` function:

Libros de Estadística en PDF | Statistics Books in PDF

Table 3.2 Types of *patch-level* metrics that `SDMTools` calculates

Metric type	Metric	Interpretation
Area	Number of cells/area	Metrics of patch size (differences related to the units used).
Edge	Perimeter	A measure of edge length, does not differentiate whether edge is interior or exterior.
	Number of edge cells (interior/exterior)	Metrics of edge length that distinguish interior edges (i.e., holes/perforations) versus exterior edges (patch boundaries).
	Core area/core area index	Metrics of core area based simply on "non-edge" cells, such that the depth of edge influence is implicitly considered to be the grain of the map.
Shape	Perimeter–area ratio	Metric of patch shape.
	Shape index	Metric of patch shape, where perimeter is divided by square root of patch area.
	Fractal dimension	Metric of shape complexity based on 2 × log perimeter to log area ratio.

```
#create patchIDs using clump from raster for 8-neighbor rule
> forest.patchID <- clump(nlcd.forest, directions = 8)
```

Note that, similar to cell IDs (see Chap. 2), this function labels patches based on integer values, starting in the northwestern (top left) portion of the map and working down (Fig. 3.5). With this new patch ID layer, we can calculate a variety of patch-based metrics using the `PatchStats` function in `SDMTools`:

```
> for.pstat <- PatchStat(forest.patchID, cellsize =
  res(nlcd.forest)[[1]])
```

In this function, we pass the length of cells into the `cellsize` argument to allow for proper calculation of area and length measurements. These calculations are in the units passed to the function; for instance, in the above code, we pass `cellsize` based on meters, such that area is in m^2 and edge is in m. This function automatically calculates many patch-based metrics (Table 3.2) and returns a data frame, where each row is a patch and each column is a metric.

```
> names(for.pstat)

[1] "patchID"           "n.cell"       "n.core.cell"     "n.edges.perimeter"
[5] "n.edges.internal"  "area"         "core.area"       "perimeter"
[9] "perim.area.ratio"  "shape.index" "frac.dim.index"  "core.area.index"
```

These metrics focus on area, edge, perimeter, shape, and core metrics. Edge is reported in two ways: the number of edge segments along the perimeter and the number of internal edges. For instance, a single cell or pixel that is an isolated patch would have 4 units for `n.edges.perimeter`. In contrast, `perimeter` takes `n.edges.perimeter` and multiplies it by the length (e.g., 4 × 200 in the

Libros de Estadística en PDF | Statistics Books in PDF

previous example). Core area (`core.area`) is simply based on the number of cells that are not adjacent to a patch *boundary*, such that this function does not allow explicit calculation of core area arising from greater distances from edge (see below for alternative approaches to do so). For shape, the function provides perimeter–area ratio, a shape index and the fractal dimension, which is another metric for describing patch shape. Perimeter–area ratio is an intuitive metric for patch shape, but it unfortunately varies with the size of the patch due to the scaling relationship of perimeter ($2\pi r$) to that of area (πr^2; e.g., compare a perimeter–area ratio for circles of different sizes). The shape index alleviates this problem by scaling perimeter to the square root of area. This metric takes on a value of 1 for regular shapes, such as a circle, and increases (unbounded) as patch irregularity increases. The fractal dimension also does not suffer from this scaling issue and some find this metric useful because it is straightforward to interpret this metric across scales. For this metric, the range is from 1 to 2, with regular shapes (e.g., circle) approach a value of 1, while highly irregular shapes approach a value of 2.

Summaries of patch metrics can be derived using functions on the data frame. For example, we calculate the number of patches on the map, mean of patch metrics and the standard deviation (SD) of those metrics with simple R commands:

```
#number of patches
> nrow(for.pstat)

##
[1] 42

#mean patch metrics
> for.pstat.mean <- colMeans(for.pstat[,2:ncol(for.pstat)])

#SD of patch metrics
> for.pstat.sd <- apply(for.pstat[,2:ncol(for.pstat)], 2, sd)
```

The `apply` function is very flexible in this way. Here it applies functions to the columns of the data based on the second argument in the function (2; note for applying calculations on rows of the data, one would pass 1). Similarly, we can visualize the variation or heterogeneity in metrics, such as the log of patch area. Patch area is often transformed to a log scale for practical reasons: biologically, we expect a change in 10 ha to be more important when contrasting a 5 ha to a 15 ha patch than when contrasting a 1000 ha to 1010 ha patch. For example, a histogram of the distribution of patch areas is straightforward to implement (Fig. 3.6b).

```
> hist(log(for.pstat$area))
```

While each of these patch-level metrics captures subtly different aspects of patch structure, many of these metrics are highly correlated (Fig. 3.7). Note that some summaries of patch metrics are also provided when doing a class-level analysis.

Libros de Estadística en PDF | Statistics Books in PDF

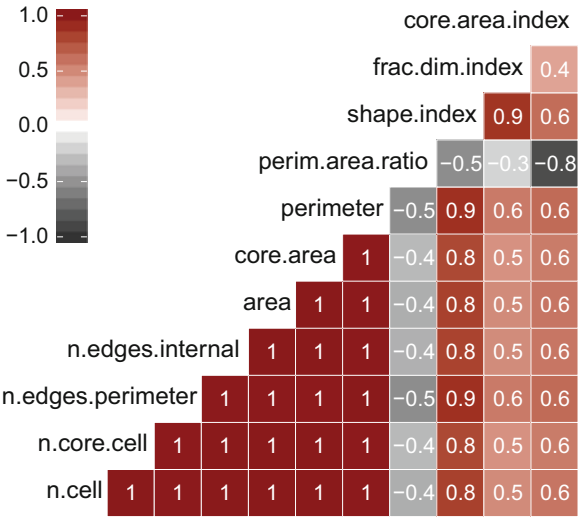


Fig. 3.7 Correlations among patch-level metrics for the landscape

3.3.3.2 Class-Level Quantification

We can also easily quantify patterns of land cover at the class-level. In this case, we can focus on metrics that do not require delineating patches, such as forest area, or we can get summary, patch-based metrics for the entire landscape (such as the standard deviation of patch sizes, as shown above). In any class-level metric, the metrics describe a focal land-cover class and most do not explicitly account for other land-cover types (see landscape-level metrics below). Some exceptions include class-level metrics that focus on edge contrast and interspersion, both of which account for variation in other land-cover types to quantify pattern of a focal land-cover type (see below).

To calculate class-based metrics, we use the `ClassStat` function from `SDMTools` in a similar way as we calculated patch-based metrics:

```
#calculation based on forest layer
> for.cstat <- ClassStat(nlcd.forest, cellsize =
  res(forest)[[1]])

#calculation based on nlcd layer (all land-cover types)
> nlcd.cstat <- ClassStat(nlcd, cellsize =
  res(nlcd)[[1]])
```

When looking over these metrics, several metrics provided are summary statistics for patch-level metrics (e.g., mean, minimum, maximum, and standard deviation of patch size), while others are unique to pattern at the class level (Table 3.3). Looking at the correlations of these metrics across classes in the landscape can help interpret

Libros de Estadística en PDF | Statistics Books in PDF

Table 3.3 Types of *class-level* metrics that `SDMTools` calculates

Metric type	Metric	Interpretation
Patch	Number of patches, patch density	Number of patches, and the number of patches relative to the study area extent. Common metrics of fragmentation.
	Patch area (mean, sd, min, max)	Summary statistics for the distribution of patch sizes.
	Largest patch index	The size of the largest patch on the map. Thought to be important for connectivity, based on percolation theory.
Area	Total area and proportion area	The area and its proportion on the map, ignoring patch boundaries/characteristics. Frequently used in landscape ecology and correlated with several metrics.
Edge	Total edge, edge density	Total length of edge for the class and edge scaled to the class area. Because total edge has a non-linear relationship with class area, edge density is often used.
	Core area (mean, sd, min, max)	Summary statistics for the distribution of patch core areas.
Shape	Perimeter–area, shape, fractal dimension (mean, sd, min, max)	Summary statistics for the distribution of patch shape statistics.
	Landscape shape index	Standardized measure of edge density that ignores patch boundaries. A class-level analog to the patch shape index.
Aggregation	Proportion of like adjacencies	Based on an adjacency matrix that relates cells and their neighbors. It is the proportion of links that are the same class type relative to the total number of links. Captures aggregation, but does not account for proportion of class area, which can make it misleading.
	Aggregation index	An area-weighted mean based on class proportion. Accounts for class area, such that maximum aggregation occurs when all land-cover occurs in one patch.
Contiguity	Patch cohesion index	A metric that uses a ratio of perimeter relative to the ratio of perimeter and area of patches, scaled to the map extent. Provides information on the physical connectedness of the land-cover and is related to the "clumpiness" of the class.
	Splitting index	The ratio of the total map area to the sum of the patch areas. Interpreted as the "effective mesh number."
	Effective mesh size	Related to the inverse of the splitting index and redundant with the landscape division index, but provides a measure of area (rather than probability). Provides the area-weighted mean patch size relative to the total map area.
	Landscape division index	Interpreted as the probability that two randomly chosen cells in the landscape are *not* situated in the same patch.

Libros de Estadística en PDF | Statistics Books in PDF

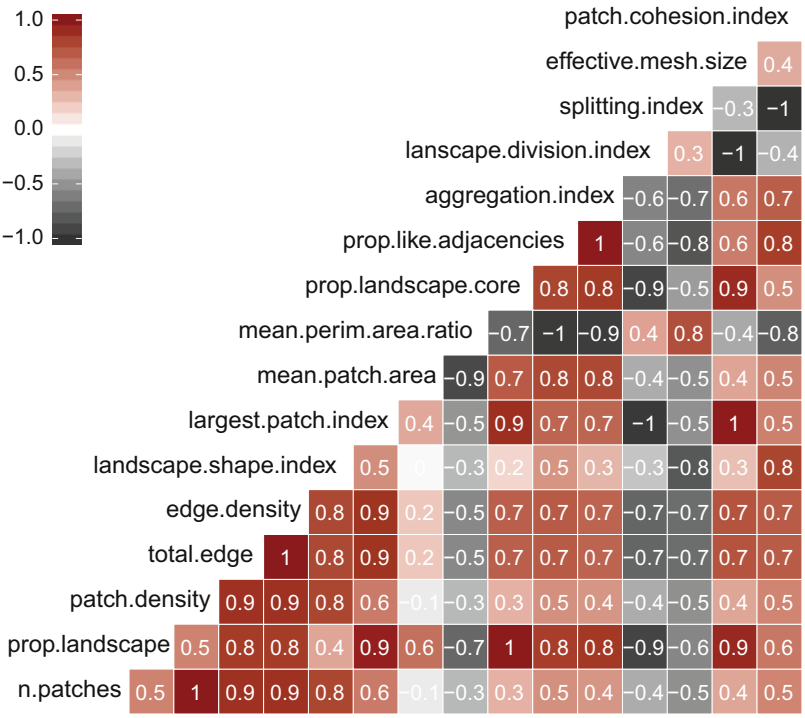


Fig. 3.8 Correlations among class-level metrics for the landscape

the extent to which metrics are capturing similar elements of class-level structure (Fig. 3.8).

We can check that these metrics are consistent with calculations based on the data frame provided with the `PatchStat` function:

```
#mean patch size
> for.cstat[for.cstat$class == 1, "mean.patch.area"]
> for.pstat.mean["area"]

#standard deviation of patch shape
> for.cstat[for.cstat$class == 1, "sd.shape.index"]
> for.pstat.sd["shape.index"]
```

The above calculations illustrate how some class-level metrics can be derived directly from patch-level metrics. In summary, the `SDMTools` package provides several metrics for quantifying land-cover patterns at the patch and class-level, similar to the popular program Fragstats (McGarigal et al. 2002).

What metrics does Fragstats calculate that `SDMTools` does not? At the patch- and class-levels, some general types of metrics that this package does not calculate include some important *functional metrics*, where we alter metrics based on species

Libros de Estadística en PDF | Statistics Books in PDF

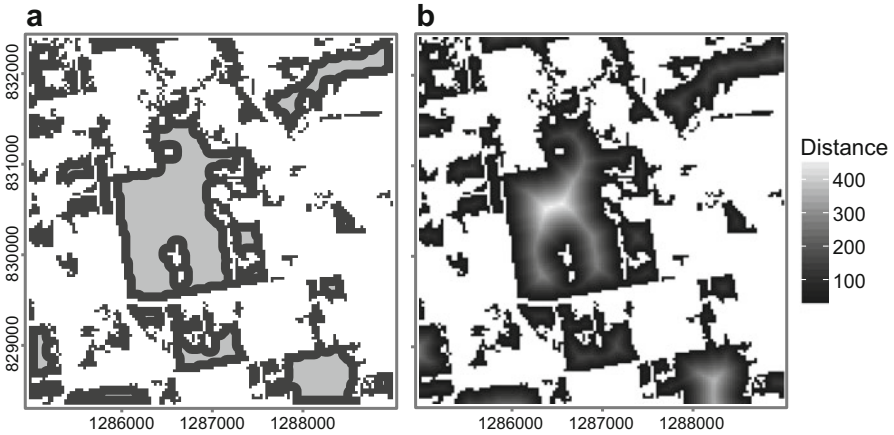


Fig. 3.9 (**a**) Polygon representation of forest patches on the landscape, where core area (>100 m from edge) is in light gray. (**b**) Raster representation of distance to forest edge (m), which can be used to delineate core areas based on different distances to edge and for implementing the effective area model

or process variability. These include core area metrics (where we adjust core area based on the distance of edge influence) and edge contrast, or the relative amount of dissimilarity between land-cover types at the edge (Ries et al. 2004; Watling and Orrock 2010). Other types of metrics not calculated include the radius of gyration (a measure of patch extent; Baker et al. 2015), and some patch-level isolation metrics, such as the proximity index and the nearest neighbor distance (Gustafson and Parker 1994; Moilanen and Nieminen 2002). Finally, note that Fragstats calculates area-weighted means of patch metrics in addition to simple mean of those patch metrics. The rationale for this approach is that area-weighted mean metrics provide more of a landscape-centric view because they, "reflect the average conditions of a pixel chosen at random or the conditions that an animal dropped at a random [pixel] the landscape would experience" (Jaeger 2000). In contrast, simple means only provide a patch-centric view of average patch characteristics (not that of conditions of an animal dropped randomly on the landscape because individuals landing in a specific patch would be contingent on patch size).

While `SDMTools` does not calculate core area and isolation metrics, such metrics can be calculated in a straightforward way using other packages in R. First, consider the calculating core areas based on a distance of 100 m. To approach this problem, we convert our raster map to a vector map with the `rasterToPolygons` function and then buffer within patch polygons using the `rgeos` package (Fig. 3.9a).

```
#create polygon layer
> library(rgeos)
> forest.poly <- rasterToPolygons(forest.patchID, dissolve = T)
```

Libros de Estadística en PDF | Statistics Books in PDF

```
#create core polygons and calculate their area
> core.poly <- gBuffer(forest.poly, width = −100, byid = T)
> core.area <- gArea(core.poly, byid = T)
```

This general approach is useful for delineating core areas of different sizes; however, it may be computationally slow for large landscapes. In addition, in many situations we may want to map more detail in edge effects, such as using the "effective area model" (Sisk et al. 1997). The effective area model maps edge response functions, such as variation in abundance as a function of distance from edge, for different edge types. Such models can be implemented by creating raster layers based on distances to boundaries. For instance, we can use the `distance` function in the `raster` package to create a new raster layer that shows the distance to edge. To implement this approach, we reformat our forest layer, such that forest is NA. This function will then calculate the nearest distance from each non-NA pixel to each NA pixel.

```
#re-format raster
> nlcd.forestNA <- nlcd.forest
> nlcd.forestNA[nlcd.forestNA == 1] <- NA

#create a distance to edge raster for forest land cover
> forest.dist <- raster::distance(nlcd.forestNA)
```

Much more information is provided with this new raster layer (Fig. 3.9b). Note that in the above code, we specified the `raster` package in the call of the `distance` function (`raster::distance`), because the `SDMTools` package also has a different distance function.

Isolation-related patch metrics can be quantified using distances based on patch centroids or edge–edge distances. Both of these types of distance metrics can be calculated with the `rgeos` package.

```
#centroids of polygons
> forest.centroid <- gCentroid(forest.poly, byid = T)

#edge-edge distance matrix
> edge.dist <- gDistance(forest.poly, byid = T)

#centroid-centroid distance matrix
> cent.dist <- gDistance(forest.centroid, byid = T)
```

With these distance matrices, we can use the `apply` function to calculate the nearest neighbor distances, or the minimum distance from one patch to any other patch in on the map. To do so, we first make the diagonal of the distance matrix NA, so that we ignore the diagonal (the focal patch, for which distance $= 0$) when summarizing information regarding other patches on the map. We can then use the `apply` function to identify the minimum distance to another patch.

Libros de Estadística en PDF | Statistics Books in PDF

```
#patch-level nearest-neighbor distance
> diag(cent.dist) <- NA
> diag(edge.dist) <- NA

> nnd.cent <- apply(cent.dist, 1, min, na.rm = T)
> nnd.edge <- apply(edge.dist, 1, min, na.rm = T)
```

Note here that using edge–edge (`nnd.edge`) versus centroid–centroid (`nnd.cent`) distances provides different results and these distances are not correlated ($r = 0.02$). The distance matrix can also be used to derive other patch-level and class-level summary statistics (e.g., mean distance, SD distance) in a straightforward way.

The proximity index incorporates both area and the distance matrix and is frequently used as a metric of patch isolation (Gustafson and Parker 1992, 1994). It is typically defined based on only considering patches within a neighborhood of a focal patch:

$$\text{prox}_i = \sum_{j=1}^{n} \frac{a_j}{d_{ij}^2} \tag{3.1}$$

where only the patches within the neighborhood of i are considered. This metric shares some similarity with metapopulation metrics for patch isolation (see Chap. 10). Note that some formulations of this metric, like that described above, only consider distances from the focal patch to all other patches in the neighborhood while others consider distances/linkages between non-focal patches within the neighborhood as well. We can calculate the proximity index by first creating a vector of patch area. Then we need to alter the distance matrix to only consider patches within a neighborhood of the patch, say 1000 m. Finally, we divide area by the distance with the `sweep` function and sum across all j patches to quantify the proximity index for patch i:

```
#patch area
> patch.area <- data.frame(id=for.pstat$patchID, area=for.pstat
  $area)

#neighborhood for proximity index to be calculated
> h <- edge.dist
> h2 <- 1 / h^2
> h2[edge.dist>1000] <- 0
> diag(h2) <- 0

#calculate proximity index
> patch.prox <- rowSums(sweep(h2, 2, patch.area$area, "*"))
```

Note that in this approach, we do not want the diagonal of the distance matrix to be NA. Rather, we use the `h2` matrix as an indicator matrix for only summing elements where distances are <1000 (not including the diagonal). With these results,

Libros de Estadística en PDF | Statistics Books in PDF

Table 3.4 Types of *landscape-level* metrics considered

Metric type	Metric	Interpretation
Patch	Number of patches, patch density	Number of patches, and the number of patches relative to the study area extent, summarized across all land-cover types.
	Largest patch index	The size of the largest patch on the map. Thought to be important for connectivity, based on percolation theory.
Edge	Total edge	Total length of edge across all classes.
	Edge density	Total length edge across all classes scaled to the class area.
Aggregation	Aggregation index	An area-weighted mean based on class proportion. Accounts for class area, such that maximum aggregation occurs when all land-cover occurs in one patch.
	Percentage of like adjacencies	Based on an adjacency matrix that relates cells and their neighbors. It is the proportion of links that are the same class type relative to the total number of links. Captures aggregation, but does not account for proportion of class area, which can make it misleading.
	Contagion	A measure of dispersion and interspersion, formally defined as the probability of finding a cell of type *i* next to a cell of type *j*.
Diversity	Land-cover richness	The number of land-cover types in the landscape.
	Shannon diversity	A metric based on information theory that weights both the number (richness) and evenness of land-cover types.
	Shannon evenness	Quantifies the distribution of relative areas among land-cover types. The complement of evenness is dominance.

we find that the patch proximity metric is weakly correlated with both patch area ($r = 0.09$) and nearest neighbor distance based on edge–edge distances ($r = -0.16$).

3.3.3.3 Landscape-Level Quantification

Landscape-level metrics are not considered in `SDMTools`, unfortunately. Here we provide code for some prominent landscape-level metrics (Table 3.4). Some common landscape-level metrics can be gleaned from summaries of class-level metrics in `SDMTools`, whereas others require writing new functions. We illustrate both approaches below.

Landscape metrics that can readily be derived from class-level metrics include the number of patches (NP), patch density (PD), largest patch index (LPI), total edge (TE), edge density (ED), and aggregation index (AI). At the landscape-level, these metrics are typically summing values for class-level metrics (e.g., NP, PD, TE), or taking the maximum value (LPI). Some examples include:

```
> land.NP <- sum(nlcd.cstat$n.patches)
> land.PD <- sum(nlcd.cstat$patch.density)
> land.LPI <- max(nlcd.cstat$largest.patch.index)
> land.TE <- sum(nlcd.cstat$total.edge) / 2
```

Libros de Estadística en PDF | Statistics Books in PDF

```
> land.ED <- sum(nlcd.cstat$edge.density) / 2
> land.AI <- sum(nlcd.cstat$prop.landscape *
  nlcd.cstat$aggregation.index)
```

Here, we divide the edge metrics by 2 because each edge segment will be counted twice when summing across class-level metrics (i.e., an edge will be counted once for each land-cover type in the adjacency). Also, the aggregation index is simply a weighted mean of the class-level metrics for aggregation. Note that `SDMTools` scales some of these metrics in a slightly different way than the Fragstats program when considering area/length, where `SDMTools` uses the square of the units provided in the layer (e.g., m^2), while Fragstats uses hectares (ha).

Land-cover richness and diversity are frequently considered. Land-cover richness is simply the number of land-cover types in an area of interest. So, if we are only interested in one or a few landscapes, then this is straightforward to calculate with simple output from the `raster` package (or with output from `SDMTools`). For example, we can calculate land cover richness for our landscape as:

```
> richness <- length(unique(values(nlcd)))
```

If we would like to calculate land-cover richness repeatedly for neighborhood, like when using a moving window analysis (see below), we can create a function to call for each neighborhood, x, such as:

```
> richness <- function(x) (length(unique(na.omit(x))))
```

Shannon's diversity, D, and evenness, E, indices are other popular measures, defined as:

$$D = -\sum_{i=1}^{n} P_i \ln(P_i) \tag{3.2}$$

And

$$E = \frac{-\sum_{i=1}^{n} P_i \ln(P_i)}{\ln(n)}. \tag{3.3}$$

For an entire landscape, it is straightforward to calculate D and E using output from the `table` function.

```
> table(values(nlcd)))

##
 1 2 3 4 5 6
7405 623 4010 3114 2935 3
```

Libros de Estadística en PDF | Statistics Books in PDF

This function returns the number of cells for each land-cover type on the map. We can then use this information to calculate diversity and evenness.

```
> C <- table(values(nlcd))
> P <- C / sum(C)
> D <- -sum(P * log(P))
> E <- D / log(length(C))
```

Note that in R, `log` defaults to calculating the natural log (i.e., $\ln(x)$).

Other landscape-level metrics that require new functions for their quantification (i.e., they cannot appropriately be summarized from class-level metrics) focus primarily on aggregation-related metrics. Aggregation-related metrics can capture several related concepts, including dispersion and interspersion. *Dispersion* indices focus on spatial mixing of a class type (ignoring other class types), while *interspersion* metrics focus on spatial mixing of different class types (ignoring dispersion of a specific class type) (Table 3.1). One prominent metric is *contagion*, which is an intuitive, landscape-level metric that captures both dispersion and interspersion. Contagion has been quantified in subtly different ways. A common formulation of contagion is (Li and Reynolds 1993, Riitters et al. 1996):

$$\text{Contagion} = 1 + \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left[P_{ij}\right] \ln \left[P_{ij}\right]}{2 \ln (n)} \tag{3.4}$$

where $P_{ij} = P_i P_{j/i}$, and

$$P_{j/i} = \frac{N_{ij}}{N_i}. \tag{3.5}$$

Here, n is the number of land-cover types (classes), P_i is the proportion of the landscape of land-cover type i, N_{ij} is the number of adjacencies between pixels of land-cover types i and j and N_i is the total number of adjacencies of land-cover type i and all land-cover types (including i). For this metric, we multiply the probability of a land-cover type by the conditional probability of that type being adjacent to a different land-cover type j and then sum this expression. Note the similarity of the contagion index to that of Shannon's Evenness index, E. The matrix $\mathbf{N}$ taken from the elements N_{ij} is a commonly used summary statistic in several landscape-level metrics (Turner and Gardner 2015). Some other relevant measures that can be derived from $\mathbf{N}$ include the *percentage of like adjacencies* and the *aggregation index* (Table 3.4). Note that calculating $\mathbf{N}$ requires using a patch-definition rule (e.g., Fragstats uses a four-neighbor rule).

One way to calculate this measure is to take advantage of the `adjacent` function in the `raster` package to calculate N_{ij}.

Libros de Estadística en PDF | Statistics Books in PDF

```
#identify adjacent cells
> adj <- adjacent(nlcd, 1:ncell(nlcd), directions = 4, pairs =
  T, include = T)
> head(adj, 2)

##
 from to
[1,] 1 1
[2,] 2 2
```

This function identifies all of the pairwise combinations of adjacencies on the map, including like adjacencies (i.e., two cells of the same land-cover type) with the term `include` = `T`. This information can be summarized to get **N** with the `table` function, which counts the values on the nlcd map based on the identified adjacencies:

```
> N <- table(nlcd[adj[,1]], nlcd[adj[,2]])

##
 1 2 3 4 5 6
 1 33155 410 584 884 1755 5
 2 410 1983 399 156 156 0
 3 584 399 18484 74 363 0
 4 884 156 74 13388 979 1
 5 1755 156 363 979 11363 1
 6 5 0 0 1 1 5
```

From there, the remaining terms are straightforward to calculate. A function for calculating contagion using the formula of Riitters et al. (1996) is:

```
> contagion <- function(r){
 adj <- adjacent(r, 1:ncell(r), directions = 4)
 Nij <- table(r[adj[,1]], r[adj[,2]])
 Nij <- unclass(Nij) #convert table format to matrix format

 Ni <- rowSums(Nij)
 Pj_i <- as.matrix(Nij / Ni)

 Pi <- as.vector(unclass(table(values(r))) / ncell(r))
 Pij <- Pi * Pj_i
 n <- length(Pi)

 #Ritters et al. 1996 formula
 contagion <- 1 + sum(Pij * log(Pij),na.rm = T)/(log(n^2 + n) -
 log(2))

 return(contagion)
}
```

Libros de Estadística en PDF | Statistics Books in PDF

The above function breaks the steps of calculating contagion into its parts. We first calculate the N_{ij}. Note that a rate-limiting step here is the construction of **N** using the `table` function. Scaling this function to larger landscapes would require using faster alternatives, such as the `data.table` function. Then P_i and $P_{j/i}$ are calculated. Finally, we put this together using the approach of Riitters et al. (1996), wherein a slight modification of the denominator is used in calculating contagion.

The general approach for calculating N_{ij} can be used to also calculate the percentage of like adjacencies, PLADJ, at the landscape-level. This metric quantifies the degree of dispersion of land-cover types. As this metric gets larger, the land-cover types are more aggregated. It is defined as:

$$\text{PLADJ} = \frac{\sum_{i=1}^{n} N_{ii}}{\sum_{i=1}^{n} \sum_{j=1}^{n} \left[N_{ij}\right]} \times 100 \tag{3.6}$$

This metric can be calculated in R with the following function:

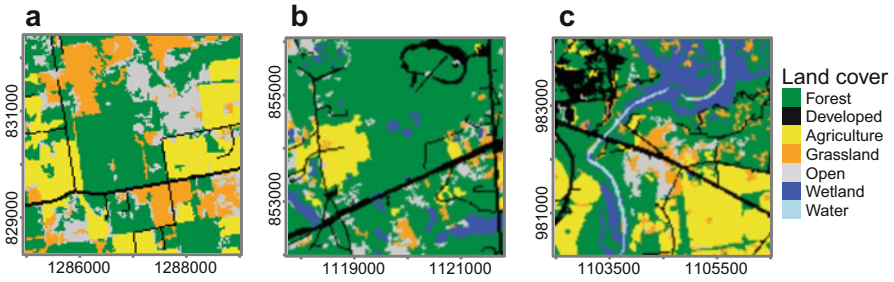
```
> PLADJ <- function(r){
adj <- adjacent(r, 1:ncell(r), directions = 4)
Nij <- table(r[adj[,1]], r[adj[,2]])
Nij <- unclass(Nij)

PLADJ <- sum(diag(Nij)) / sum(Nij) * 100
return(PLADJ)
}
```

To provide context for these landscape-level metrics, we contrast the landscape used so far (Fig. 3.4) with two other landscapes that were sampled in Chap. 2 (Fig. 3.10). For each landscape, we apply these functions to interpret landscape-level variation. One landscape is dominated by forest (Fig. 3.10b), whereas the other appears to be highly fragmented (Fig. 3.10c). It is notable that the forest-dominated landscape has generally similar landscape-level metric values to our original landscape except those related to landscape diversity and evenness. This similarity is driven by the fact that non-forest land cover is generally configured in small patches with a large proportion of edge. The landscape that appears fragmented (Fig. 3.10c) does have more patches, more edge, and less aggregation than the other landscapes. Note that these numbers can vary subtly with other programs, such as Fragstats, based largely on the underlying assumptions of the calculations (patch delineation rules, how boundaries are considered, etc.).

Taken together, these analyses illustrate how several landscape-level metrics can be calculated in R. It also illustrates how the use of landscape-level metrics can sometimes be more difficult to interpret than for patch or class-level metrics, because typically landscape-level metrics are pooling or summarizing information across all land-cover types on the map (compare metrics for landscapes a and b in Fig. 3.10). This pooling makes the metrics more difficult to interpret biologically than with other types of metrics. Nonetheless, in some situations, we expect biologically that

Libros de Estadística en PDF | Statistics Books in PDF



Metric type	Metric	Landscape a	Landscape b	Landscape c
Patch	Number of patches	166	174	232
	Largest patch index	0.189	0.439	0.108
Edge	Total edge	181081	180540	227220
	Edge density	0.011	0.011	0.014
Aggregation	Aggregation index	84.6	84.5	80.0
	Percentage of like adjacencies	83.9	83.9	79.2
	Contagion	0.337	0.471	0.282
Diversity	Land-cover richness	6	7	7
	Shannon diversity	1.41	1.17	1.16
	Shannon evenness	0.79	0.60	0.83

Fig. 3.10 Contrasting landscape-level metrics for three landscapes (**a**–**c**)

landscape-level metrics should better describe key issues of relevance to biodiversity, such as questions regarding the role of *landscape heterogeneity* in agricultural landscapes (Fahrig et al. 2011; Reynolds et al. 2018), and the importance of "countryside" biogeography (Brosi et al. 2008; Mendenhall et al. 2014), where interest lies in understanding the value of biodiversity across human dominated land uses.

3.3.3.4 Moving-Window Analysis

Each of the above approaches is often applied to replicated landscapes. In Chap. 2, for example, we calculated the proportion of forest cover surrounding different locations (a "patch-landscape," or "focal-patch" sampling design) (Fahrig 2003). Another approach to landscape quantification is using a moving-window analysis, which is akin to a "neighborhood" analysis and loosely captures some ideas of an "ecological neighborhood" (Addicott et al. 1987).

In a moving window analysis, for each pixel on a map we quantify land cover in a surrounding neighborhood. The result is a new map that visualizes the neighborhood variation in land-cover properties. These maps can then be used for sampling or for making predictive maps, such as maps of predicted species distribution (see Chap. 7).

The `raster` package provides a means to implement a moving-window analysis in a straightforward way with the `focal` function. Moving windows can be based

Libros de Estadística en PDF | Statistics Books in PDF

on different shaped windows, such as rectangles or circles. We first create a weight matrix with the `focalWeight` function that defines the window size and shape:

```
#focal buffer matrix for moving windows
> buffer.radius <- 100
> fw.100m <- focalWeight(nlcd, buffer.radius, type = 'circle')

#re-scale weight matrix to 1/0 for calculations
> fw.100m <- ifelse(fw.100m > 0, 1, 0)
> fw.100m

##
 [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,] 0 0 1 1 1 0 0
[2,] 0 1 1 1 1 1 0
[3,] 1 1 1 1 1 1 1
[4,] 1 1 1 1 1 1 1
[5,] 1 1 1 1 1 1 1
[6,] 0 1 1 1 1 1 0
[7,] 0 0 1 1 1 0 0
```

This is a square matrix where a circle is approximated based on the radius considered. Note that the cells in this matrix reflect the grain of the map being considered and `raster` creates the size of the matrix to match the length of the radius/grain. The `focalWeight` function can also be used to consider Gaussian kernels, as discussed in Chap. 2, by specifying `type = 'Gauss'` and setting the value for sigma (the smoothing parameter; see Fig. 2.10). For example, a Gaussian kernel with sigma = 50 would be quantified as:

```
> focalWeight(nlcd, c(50, 100), type = 'Gauss')

##
 [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,] 0.00 0.01 0.01 0.01 0.01 0.01 0.00
[2,] 0.01 0.01 0.02 0.03 0.02 0.01 0.01
[3,] 0.01 0.02 0.04 0.05 0.04 0.02 0.01
[4,] 0.01 0.03 0.05 0.06 0.05 0.03 0.01
[5,] 0.01 0.02 0.04 0.05 0.04 0.02 0.01
[6,] 0.01 0.01 0.02 0.03 0.02 0.01 0.01
[7,] 0.00 0.01 0.01 0.01 0.01 0.01 0.00
```

Here, the two numbers for *d* reflect sigma and the window size to be considered (100 m; same as above). The use of a Gaussian kernel allows for the weighting scheme to decline with distance (Fig. 2.10).

With this weight matrix, we can then use the `focal` function to run a moving window analysis. For each pixel, this function will multiple the `focalWeight` matrix by the raster. If the matrix is a series of 0's and 1's, in effect this will mask all values outside the neighborhood (by multiplying those values by 0).

Libros de Estadística en PDF | Statistics Books in PDF

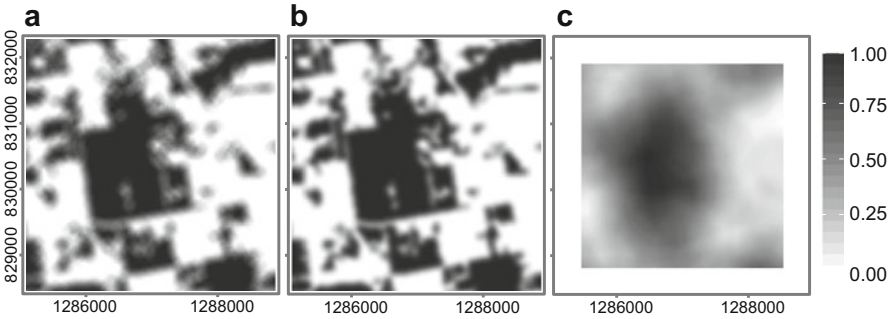


Fig. 3.11 Moving window analysis for neighborhood quantification around cells. (**a**) 100 m moving window (simple buffer with equal weighting) of the proportion of forest cover, (**b**) 100 m moving window of forest cover using a Gaussian kernel (sigma = 50), and (**c**) 500 m moving window using a simple buffer with equal weighting

We illustrate two examples. First, we calculate the proportion of forest cover surrounding each pixel. To do so, we use the `sum` function to sum the total forest cover within each window (Fig. 3.11). Note that below we illustrate this process in a couple of steps for clarity, but it could be streamlined by using a weighted average instead (see Chap. 2). Second, we can call our own defined functions of land-cover pattern. Here, we call our own function to calculate land-cover richness in surrounding each pixel to illustrate.

```
#forest cover moving window; number of cells
> forest.100m <- focal(nlcd.forest, w = fw.100m, fun = "sum",
  na.rm=T)

#proportion
> forest.prop.100m <- forest.100m / sum(fw.100m)

#richness moving window
> richness.100m <- focal(nlcd, fw.100m, fun = richness)
```

Some metrics at the neighborhood scale can be more difficult to calculate in an efficient way. For instance, calculating Shannon's diversity is less straightforward, because the calculation described above would take too long (the `table` function used in the above description is relatively slow). A much quicker way is to create individual maps that describe the proportion of each land cover category with a moving window and then use raster algebra across maps to derive a new map of diversity at the neighborhood scale. A function to accomplish this for Shannon's diversity is:

```
> diversity <- function(landcover, radius) {

 n <- length(unique(landcover))
```

Libros de Estadística en PDF | Statistics Books in PDF

```
#Create focal weights matrix
fw.i <- focalWeight(landcover, radius, "circle")

#create new layer for diversity
D <- landcover
values(D) <- 0

#function for log(p)*p
log.i <- function(x) ifelse(x == 0, 0, x * log(x))

#for each landcover category, create a moving window map and sum
for (i in 1:length(n)) {
focal.i <- focal(landcover == i, fw.i)
D <- D + calc(focal.i, log.i)
}

D <- D * -1
return(D)
}

> diversity.100m <- diversity(landcover = nlcd, radius = 100)
```

Overall, if we contrast this diversity map to that of land-cover richness, we find that these two metrics across the landscape are weakly correlated ($r = 0.24$).

3.3.4 *Simulating Land Cover: Neutral Landscapes*

Landscape and spatial ecologists frequently generate random or *neutral landscapes* to represent land-cover variation (Gardner et al. 1987; O'Neill et al. 1992; Neel et al. 2004; Etherington et al. 2015). These landscapes vary in the degree of complexity (Pe'er et al. 2013; Etherington et al. 2015). The application of these maps also varies considerably (With 1997; With and King 1997). Some map representations aim to capture a minimal amount of pattern and process to provide a "null" representation of landscapes (Gardner and Urban 2007).

The two most common landscape characteristics considered are: (1) the proportion of habitat or class of land-cover, p; and (2) the degree to which it is aggregated (or conversely, fragmented). Most approaches focus on a binary representation of the landscape (e.g., habitat v non-habitat), but some approaches extend this to several land-cover types (Saura and Martinez-Millan 2000).

We start with the simplest representation of a neutral landscape, sometimes referred to as a "simple random" landscape (Gardner et al. 1987). In this model, the only parameter considered is p. We can simulate a neutral landscape by making independent, random draws from probability distributions for each cell or pixel on the landscape. Frequently, either a uniform distribution is used (U~(0,1)), or a Bernoulli distribution is used (Binomial~(p, 1)). For the uniform distribution, if

Libros de Estadística en PDF | Statistics Books in PDF

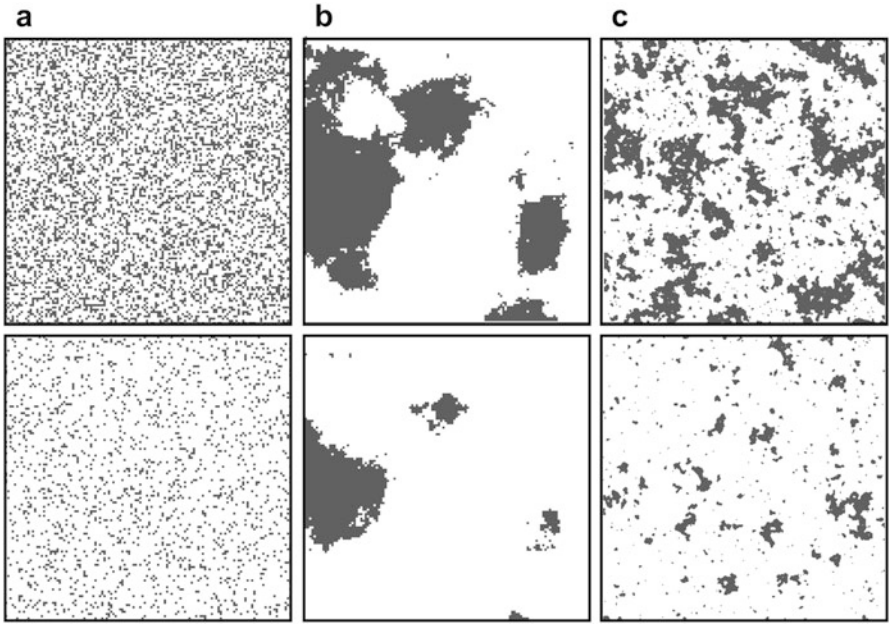


Fig. 3.12 Neutral landscapes can be generated in several ways. (**a**) Simple random landscapes with different proportions of habitat. (**b**) Fractal landscapes can be created that vary the degree of aggregation of habitat. Shown is $H = 0.7$. (**c**) Random landscapes generated with the modified random cluster algorithm can also create aggregated patterns. Shown is $p = 0.55$. Upper row of maps reflect approximately 30% habitat, while lower row is approximately 10% habitat

the draw is less than p, the cell is marked as habitat and non-habitat otherwise. A Bernoulli distribution is a type of binomial distribution, where only one "trial" is considered. It is often the distribution used when describing a (weighted) coin toss. Here we illustrate the use of the Bernoulli distribution (Fig. 3.12).

```
#landscape dimensions
> dimX <- 128
> dimY <- 128

#simple random with 30% habitat
> sr.30 <- raster(ncol = dimX, nrow = dimY, xmn = 0, xmx = dimX,
  ymn = 0, ymx = dimY)
> sr.30[] <- rbinom(ncell(sr.30), prob = 0.3, size = 1)

#simple random with 10% habitat
> sr.10 <- raster(ncol = dimX, nrow = dimY, xmn = 0, xmx = dimX,
  ymn = 0, ymx = dimY)
> sr.10[] <- rbinom(ncell(sr.10), prob=0.1, size=1)
```

Libros de Estadística en PDF | Statistics Books in PDF

Using a Bernoulli distribution is useful in the sense that it is an appropriate probability distribution for binary outcomes; however, it does not guarantee that a random landscape will have exactly p proportion of the landscape as habitat (or land cover). A uniform distribution provides more precision in this way because a quantile can be taken from the realizations of the uniform distribution across the map (see below). Overall, simple random landscapes are a useful starting point, but the patterns generated from these approaches are not similar to real-world patterns. Instead, they tend to generate patterns that resemble static white noise from old television sets.

Other common approaches that incorporate aggregation in addition to simply the amount of habitat include the use of Gaussian random fields (see Chap. 5), fractional Brownian motion (or fractal landscapes), and the use of various clustering algorithms (Keitt 2000; Saura and Martinez-Millan 2000; Chipperfield et al. 2011; Remmel and Fortin 2013). Gaussian random fields models use parameters that describe spatial dependence (from geostatistics; Chap. 5) to make predictions across a region of interest. These models are formally related to fractal models that have been widely used in ecology (Keitt 2000). We will illustrate these in Chap. 5 once spatial dependence is formally introduced.

Neutral models based on fractal algorithms have been widely applied in spatial ecology. The reason is that the degree of aggregation of habitat and the amount of habitat can be precisely and independently controlled. For most of these applications, the "mid-point displacement" algorithm has been used to generate fractal landscapes (Saupe 1988, p. 83–85). This is a relatively simple, recursive algorithm that takes square maps of power 2 (e.g., 32, 64, 128 cells in a linear dimension) and with each recursive partition, it breaks the line at its midpoint, adding some noise to the value at newly created point based on the degree of aggregation, H, termed the Hurst exponent. H is related to the fractal dimension (the precise relationship depends on the dimensions being considered, that is, 1D, 2D, or 3D). H ranges between 0 and 1; as H approaches 1, the map is highly aggregated (a high degree of spatial autocorrelation; see Chap. 5), while as H approaches 0, the map becomes much more fragmented, resembling a simple random map described above.

Fractal-like landscapes can be generated with a few different packages in R, including the `RandomFields`, `FieldSim`, `NLMR`, and `Voss` packages (Shitov and Moskalev 2005; Brouste et al. 2007; Schlather et al. 2015; Sciaini et al. 2018). We note that at the time of publication, the `NLMR` package was released, which offers a means to contrast several types of neutral landscapes with one R package, including the use of the midpoint displacement algorithm (Sciaini et al. 2018). Here, we illustrate the general idea with the fractal Brownian function, `Voss2d`, in the `Voss` package. This package uses a recursive method described by Voss (1985). This method is similar to the midpoint displacement algorithm where successive random additions of Gaussian noise, except that all of the points are modified with each recursive step and not just the newly created points, as in the midpoint displacement algorithm (Saupe 1988). For this package, we specify H and g, which specifies the dimensions of the landscape ($2^g \times 2^g$). These types of models will create a continuous, rugged surface that we then slice through to create a binary map of habitat/non-habitat for a given p.

Libros de Estadística en PDF | Statistics Books in PDF

```
> library(Voss)
> voss <- voss2d(g = 7, H = 0.7)
> str(voss)

##
List of 3
 $ x: num [1:129] 0 0.00781 0.01562 0.02344 0.03125 ...
 $ y: num [1:129] 0 0.00781 0.01562 0.02344 0.03125 ...
 $ z: num [1:129, 1:129] 0.0182 0.0348 0.0494 0.0526 0.0887 ...
```

The object created is a list with x–y coordinates and the value of the terrain, z, which defaults to being centered at 0. We can then create a binary representation of this map by quantifying the quantiles of the values generated and then truncating the fractal map based on p. Below we create maps for 10% and 30% habitat (Fig. 3.10).

```
#identify threshold
> voss1.thres <- quantile(voss$z, prob = 0.1)
> voss3.thres <- quantile(voss$z, prob = 0.3)

#truncate
> voss$z1 <- ifelse(voss$z < voss1.thres, 1, 0)
> voss$z3 <- ifelse(voss$z < voss3.thres, 1, 0)
```

Note that these maps can potentially be discretized into >2 categories, which has been used to reflect spatial variation in environmental gradients and habitat quality (With 1997).

An alternative approach to fractal landscapes is the modified random clusters (MRC) algorithm (Saura and Martinez-Millan 2000). This algorithm uses a series of successive steps to generate clustered land-cover. First, a simple random map is generated, similar to that described above. Second, clusters are determined, which is functionally identical to a four-neighbor patch delineation rule described earlier (Fig. 3.4). Third, clusters are assigned to the focal land-cover type being considered, based on an expected total proportional area, A, assigned for one land-cover type (e.g., forest). Note that "expected" is used here, because in practice, this expectation may not be possible, depending on the cluster size distribution of the map. Finally, the remaining land-cover types are filled in using the same steps. This function uses two parameters: p and A. In this setting, p controls the degree of fragmentation (p does not have the same meaning as used above!): it is highly non-linear and related to the "percolation threshold" in simple random neutral landscapes (Gardner et al. 1987). The percolation threshold is the point at which habitat becomes fully connected on a landscape, such that there is one cluster: for a simple random landscape, like that described above, it is approximately 0.59 (Gardner et al. 1987). Because the MRC algorithm focuses on cluster assignment from a simple random mapping process (steps 1 and 2 above), it is sensitive to this percolation threshold. That is not necessarily a problem, but it means that most of the interesting variability in pattern generated from this algorithm occurs when p is close to the percolation threshold. It also means that near or above this threshold, it is likely not

Libros de Estadística en PDF | Statistics Books in PDF

possible to generate landscapes where the observed *A* matches the expected *A* we are attempting to capture.

A basic form of this model can be implemented in the `secr` package (Efford 2018) with the `randomHabitat` function. We will illustrate the use of this function by focusing on examples shown in Saura and Martinez-Millan (2000). We first create a "mask," which is simply the extent or shape of the area of interest (note we could also create a mask in this package using an irregular polygon). We then provide values of *p* and *A* to generate the random maps.

```
> library(secr)
> tempmask <- make.mask(nx = dimX, ny = dimX, spacing = 1)

> p55A3 <- randomHabitat(tempmask, p = 0.58, A = 0.3)
> p55A1 <- randomHabitat(tempmask, p = 0.55, A = 0.1)

> plot(p55A3, dots = FALSE, col = "green")
```

Taken together, the wide variety of neutral landscape maps that have been developed provide a means of interpreting spatial pattern under conditions of limited biological process. These types of maps have been used to address a wide variety of issues and several generalizable insights have emerged from them (Turner and Gardner 2015). For example, these types of models have illustrated that the extent of the landscape under consideration can greatly affect pattern metrics due to the truncation of extents. Neutral models have revealed potential thresholds in connectivity and how connectivity of habitats can vary greatly with habitat amount (see Chap. 9). Neutral landscapes have also been used to interpret whether observed land-cover patterns are potentially significant (Remmel and Fortin 2013, 2017).

3.4 Next Steps and Advanced Issues

3.4.1 Testing for Pattern Differences Between Landscapes

A critical question that often arises in land-cover analysis is whether or not the observed patterns are meaningful, unusual, or significantly altered due to some sort of environmental change. For example, Tinker et al. (2003) were interested in understanding whether the current patterns of forest structure in a national forest were different than those in a nearby national park (Yellowstone), where different management practices occurred. In such situations, it is often important to have an understanding of the historic range of variability in landscape pattern through, for example, a long-term time series of land cover (Gustafson 1998).

With information on land-cover maps either across space or over time, how do we test for significant pattern on maps? The `PatternClass` package provides one means to do so (Remmel and Fortin 2013, 2017). This package takes binary

Libros de Estadística en PDF | Statistics Books in PDF

landscapes and estimates the magnitude of spatial dependence in the observed landscape using geostatistical techniques (see Chap. 5). With this estimate, it simulates neutral landscapes with the same magnitude of spatial dependence. Neutral landscapes can then be compared with the observed land-cover pattern to infer significance.

3.4.2 Land-Cover Quantification via Image Processing

An alternative approach to land-cover quantification comes from the application of morphological image processing techniques to land-cover maps (Vogt et al. 2007; Riitters et al. 2009). This approach focuses on the geometric/configuration property of each cell in the landscape and classifies each of them into a specific category, such as edge, core, perforated, and patch. In doing so, this approach provides an intuitive means to visualize landscape patterns by providing a map of this cell classification. Note that much of the requirements for this type of quantification is effectively done in `SDMTools` (e.g., through the use of delineating internal and external edges), but this package summarizes these values across the map rather than assigning the values to the cells on the map for visualization.

3.4.3 Categorical Versus Continuous Metrics

A common criticism of land-use and land-cover analyses is that quantification is based on categorical maps, where in reality most of these maps are derived from underlying remote sensing imagery that is inherently continuous. Not only is this criticism relevant to the data being processed, but it is also relevant for spatial ecology concepts. For instance, using categorical maps for quantifying spatial pattern implicitly focuses on the patch-matrix-corridor and landscape mosaic paradigms, which may be misleading in situations where underlying resources vary continuously. This has led to developments of other conceptual paradigms, such as the continuum concept (Fischer and Lindenmayer 2006) and the related gradient paradigm (McGarigal et al. 2009).

Surface metrics have been advanced to quantify continuous environmental variation (McGarigal et al. 2009; Hoechstetter et al. 2011). These metrics capture patterns of continuous variation on maps, such as digital elevation maps. Metrics can focus on change (e.g., slope), aggregation (~patches), etc. McGarigal et al. (2009) introduced surface metrics to landscape ecology. They emphasized some parallels with types of metrics for categorical maps. Some metrics are compositional (non-spatial), emphasizing variation in magnitudes of variables (e.g., amplitude) and the rates of change (e.g., slope). Other metrics focus on spatial (horizontal) variability, similar to configuration metrics. Surface metrics have been increasingly used under a variety of contexts (Moniem and Holland 2013; Frazier 2016). There are

Libros de Estadística en PDF | Statistics Books in PDF

currently no R packages dedicated to surface metrics, but some simple metrics can be calculated manually within the `raster` package (e.g., see Chap. 6).

3.5 Conclusions

Land-use and land-cover change are primary issues influencing biodiversity and ecosystem services across the planet (Lawler et al. 2014; Newbold et al. 2015). The clearing and degradation of habitat can reduce biodiversity and increase extinction risks for many species (Wilcove et al. 1998; Brooks et al. 2002). Land-use intensification for agricultural production can lower biodiversity substantially (Tscharntke et al. 2005). Urbanization is occurring in many biodiversity hotspots, impacting biodiversity and ecosystem services (Miller and Hobbs 2002). Effects of land-use and land-cover change can also interact with other environmental changes, such as climate change (Laurance and Useche 2009; Cote et al. 2016).

Quantifying these changes has emerged as a staple of much of spatial ecology, geography, and conservation science (O'Neill et al. 1999; Malanson et al. 2006; Kupfer 2012). Several conceptual frameworks have been advanced for interpreting these changes and hundreds of metrics have been derived for quantification of these patterns. Yet many of these metrics are related (Riitters et al. 1995; Fortin et al. 2003; Cushman et al. 2008; Wang et al. 2014). These metrics capture both variation in composition and configuration occurring at different scales. Understanding these relationships and how metrics capture the spatial scales of composition and configuration of land use and land cover is essential for appropriate applications. We recommend that emphasis is placed on functional, rather than structural, metrics when possible and that conceptual frameworks should be chosen based on the species or process under investigation. In addition, more mechanistic data and modeling can help to illuminate the effects of land-cover and land-use change on biological processes, which can be challenging when using pattern data alone due to the correlated nature of such data (e.g., Figs. 3.7 and 3.8). Such understanding may improve predictions of the effects of land-cover and land-use change on biodiversity and help guide more effective conservation strategies.

References

Addicott JF, Aho JM, Antolin MF, Padilla DK, Richardson JS, Soluk DA (1987) Ecological neighborhoods: scaling environmental patterns. Oikos 49(3):340–346. https://doi.org/10.2307/3565770

Baker CM, Hughes BD, Landman KA (2015) Length-based connectivity metrics and their ecological interpretation. Ecol Indic 58:192–198. https://doi.org/10.1016/j.ecolind.2015.05.046

Benton TG, Vickery JA, Wilson JD (2003) Farmland biodiversity: is habitat heterogeneity the key? Trends Ecol Evol 18(4):182–188. https://doi.org/10.1016/s0169-5347(03)00011-9

Libros de Estadística en PDF | Statistics Books in PDF

Bivand R, Rundall C (2017) rgeos: interface to geometry engine - open source (GEOS). R package version 0.3-26

Blair RB (1996) Land use and avian species diversity along an urban gradient. Ecol Appl 6 (2):506–519. https://doi.org/10.2307/2269387

Bowman J, Fahrig L (2002) Gap crossing by chipmunks: an experimental test of landscape connectivity. Can J Zool 80(9):1556–1561. https://doi.org/10.1139/z02-161

Bowman J, Cappuccino N, Fahrig L (2002) Patch size and population density: the effect of immigration behavior. Conserv Ecol 6(1):9

Brockerhoff EG, Jactel H, Parrotta JA, Quine CP, Sayer J (2008) Plantation forests and biodiversity: oxymoron or opportunity? Biodivers Conserv 17(5):925–951. https://doi.org/10.1007/s10531-008-9380-x

Brooks TM, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Rylands AB, Konstant WR, Flick P, Pilgrim J, Oldfield S, Magin G, Hilton-Taylor C (2002) Habitat loss and extinction in the hotspots of biodiversity. Conserv Biol 16(4):909–923. https://doi.org/10.1046/j.1523-1739.2002.00530.x

Brosi BJ, Daily GC, Shih TM, Oviedo F, Durán G (2008) The effects of forest fragmentation on bee communities in tropical countryside. J Appl Ecol 45(3):773–783. https://doi.org/10.1111/j.1365-2664.2007.01412.x

Brouste A, Istas J, Lambert-Lacroix S (2007) On fractional Gaussian random fields simulations. J Stat Softw 23(1):1–23

Brudvig LA, Leroux SJ, Albert CH, Bruna EM, Davies KF, Ewers RM, Levey DJ, Pardini R, Resasco J (2017) Evaluating conceptual models of landscape change. Ecography 40(1):74–84. https://doi.org/10.1111/ecog.02543

Chen JQ, Franklin JF, Spies TA (1992) Vegetation responses to edge environments in old-growth douglas-fir forests. Ecol Appl 2(4):387–396. https://doi.org/10.2307/1941873

Chipperfield JD, Dytham C, Hovestadt T (2011) An updated algorithm for the generation of neutral landscapes by spectral synthesis. PLoS One 6(2):e17040. https://doi.org/10.1371/journal.pone.0017040

Collinge SK, Forman RTT (1998) A conceptual model of land conversion processes: predictions and evidence from a microlandscape experiment with grassland insects. Oikos 82(1):66–84. https://doi.org/10.2307/3546918

Cote IM, Darling ES, Brown CJ (2016) Interactions among ecosystem stressors and their importance in conservation. Proc R Soc B 283(1824). https://doi.org/10.1098/rspb.2015.2592

Cushman SA, McGarigal K, Neel MC (2008) Parsimony in landscape metrics: strength, universality, and consistency. Ecol Indic 8(5):691–703. https://doi.org/10.1016/j.ecolind.2007.12.002

Cushman SA, Gutzwiller KJ, Evans JS, McGarigal K (2010) The gradient paradigm: a conceptual and analytical framework for landscape ecology. In: Cushman SA, Huettman F (eds) Spatial complexity, informatics, and wildlife conservation. Springer, Tokyo

Cushman SA, Shirk A, Landguth EL (2012) Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. Landsc Ecol 27 (3):369–380. https://doi.org/10.1007/s10980-011-9693-0

Diamond JM (1975) The island dilemma: lessons of modern biogeographic studies for the design of natural reserves. Biol Conserv 7(2):129–146. https://doi.org/10.1016/0006-3207(75)90052-x

Didham RK, Kapos V, Ewers RM (2012) Rethinking the conceptual foundations of habitat fragmentation research. Oikos 121(2):161–170. https://doi.org/10.1111/j.1600-0706.2011.20273.x

Efford MG (2018) secr: spatially explicit capture-recapture models. R package version 3.1.6

Etherington TR, Holland EP, O'Sullivan D (2015) NLMpy: a PYTHON software package for the creation of neutral landscape models within a general numerical framework. Methods Ecol Evol 6(2):164–168. https://doi.org/10.1111/2041-210x.12308

Ewers RM, Didham RK (2006) Confounding factors in the detection of species responses to habitat fragmentation. Biol Rev 81(1):117–142. https://doi.org/10.1017/s1464793105006949

Libros de Estadística en PDF | Statistics Books in PDF

Ewers RM, Didham RK (2007) The effect of fragment shape and species' sensitivity to habitat edges on animal population size. Conserv Biol 21(4):926–936. https://doi.org/10.1111/j.1523-1739.2007.00720.x

Ewers RM, Scharlemann JPW, Balmford A, Green RE (2009) Do increases in agricultural yield spare land for nature? Glob Chang Biol 15(7):1716–1726. https://doi.org/10.1111/j.1365-2486.2009.01849.x

Fahrig L (2001) How much habitat is enough? Biol Conserv 100(1):65–74

Fahrig L (2003) Effects of habitat fragmentation on biodiversity. Annu Rev Ecol Evol Syst 34:487–515. https://doi.org/10.1146/annurev.ecolsys.34.011802.132419

Fahrig L (2013) Rethinking patch size and isolation effects: the habitat amount hypothesis. J Biogeogr 40(9):1649–1663. https://doi.org/10.1111/jbi.12130

Fahrig L (2017) Ecological responses to habitat fragmentation *per se*. Annu Rev Ecol Evol Syst 48:1–23

Fahrig L, Baudry J, Brotons L, Burel FG, Crist TO, Fuller RJ, Sirami C, Siriwardena GM, Martin J-L (2011) Functional landscape heterogeneity and animal biodiversity in agricultural landscapes. Ecol Lett 14(2):101–112. https://doi.org/10.1111/j.1461-0248.2010.01559.x

Fischer J, Lindenmayer DB (2002) The conservation value of paddock trees for birds in a variegated landscape in southern New South Wales. 2. Paddock trees as stepping stones. Biodivers Conserv 11(5):833–849. https://doi.org/10.1023/a:1015318328007

Fischer J, Lindenmayer DB (2006) Beyond fragmentation: the continuum model for fauna research and conservation in human-modified landscapes. Oikos 112(2):473–480. https://doi.org/10.1111/j.0030-1299.2006.14148.x

Fischer J, Lindenmayer DB (2007) Landscape modification and habitat fragmentation: a synthesis. Glob Ecol Biogeogr 16(3):265–280. https://doi.org/10.1111/j.1466-8238.2007.00287

Fischer J, Lindenmayer DB, Fazey I (2004) Appreciating ecological complexity: habitat contours as a conceptual landscape model. Conserv Biol 18(5):1245–1253. https://doi.org/10.1111/j.1523-1739.2004.00263.x

Flather CH, Bevers M (2002) Patchy reaction-diffusion and population abundance: the relative importance of habitat amount and arrangement. Am Nat 159(1):40–56

Fletcher RJ Jr, Koford RR (2003) Spatial responses of Bobolinks (*Dolichonyx oryzivorus*) near different types of edges in northern Iowa. Auk 120(3):799–810

Fletcher RJ Jr, Didham RK, Banks-Leite C, Barlow J, Ewers RM, Rosindell J, Holt RD, Gonzalez A, Pardini R, Damschen EI, Melo FPL, Ries L, Prevedello JA, Tscharntke T, Laurance WF, Lovejoy T, Haddad NM (2018) Is habitat fragmentation good for biodiversity? Biol Conserv 226:9–15

Forman RTT (1995a) Land mosaics: the ecology of landscapes and regions. Cambridge University Press, Cambridge

Forman RTT (1995b) Some general principles of landscape and regional ecology. Landsc Ecol 10(3):133–142. https://doi.org/10.1007/bf00133027

Forman RTT, Godron M (1981) Patches and structural components for a landscape ecology. Bioscience 31(10):733–740. https://doi.org/10.2307/1308780

Fortin MJ, Boots B, Csillag F, Remmel TK (2003) On the role of spatial stochastic models in understanding landscape indices in ecology. Oikos 102(1):203–212. https://doi.org/10.1034/j.1600-0706.2003.12447.x

Frazier AE (2016) Surface metrics: scaling relationships and downscaling behavior. Landsc Ecol 31(2):351–363. https://doi.org/10.1007/s10980-015-0248-7

Gardner RH, Urban DL (2007) Neutral models for testing landscape hypotheses. Landsc Ecol 22(1):15–29. https://doi.org/10.1007/s10980-006-9011-4

Gardner RH, Milne BT, Turner MG, O'Neill RV (1987) Neutral models for the analysis of broad-scale landscape pattern. Landsc Ecol 1(1):19–28. https://doi.org/10.1007/bf02275262

Gottlieb IGW, Fletcher RJ, Nunez-Regueiro MM, Ober H, Smith L, Brosi BJ (2017) Alternative biomass strategies for bioenergy: implications for bird communities across the southeastern United States. Glob Change Biol Bioenergy 9(11):1606–1617. https://doi.org/10.1111/gcbb.12453

Libros de Estadística en PDF | Statistics Books in PDF

Grassini P, Cassman KG (2012) High-yield maize with large net energy yield and small global warming intensity. Proc Natl Acad Sci U S A 109(4):1074–1079. https://doi.org/10.1073/pnas.1116364109
Grimm NB, Faeth SH, Golubiewski NE, Redman CL, Wu JG, Bai XM, Briggs JM (2008) Global change and the ecology of cities. Science 319(5864):756–760. https://doi.org/10.1126/science.1150195
Gustafson EJ (1998) Quantifying landscape spatial pattern: what is the state of the art? Ecosystems 1(2):143–156. https://doi.org/10.1007/s100219900011
Gustafson EJ, Parker GR (1992) Relationships between landcover proportion and indexes of landscape spatial pattern. Landsc Ecol 7(2):101–110. https://doi.org/10.1007/bf02418941
Gustafson EJ, Parker GR (1994) Using an index of habitat patch proximity for landscape design. Landsc Urban Plan 29(2–3):117–130. https://doi.org/10.1016/0169-2046(94)90022-1
Haddad NM, Brudvig LA, Clobert J, Davies KF, Gonzalez A et al (2015) Habitat fragmentation and its lasting impact on Earth. Sci Adv 1:e1500052
Hadley AS, Betts MG (2016) Refocusing habitat fragmentation research using lessons from the last decade. Curr Landsc Ecol Rep 1:55–66. https://doi.org/10.1007/s40823-016-0007-8
Haila Y (2002) A conceptual genealogy of fragmentation research: from island biogeography to landscape ecology. Ecol Appl 12(2):321–334. https://doi.org/10.2307/3060944
Hall LS, Krausman PR, Morrison ML (1997) The habitat concept and a plea for standard terminology. Wildl Soc Bull 25(1):173–182
Hanski I (1998) Metapopulation dynamics. Nature 396(6706):41–49
Hanski I (1999) Metapopulation ecology. Oxford University Press, Oxford
Harper KA, Macdonald SE, Burton PJ, Chen JQ, Brosofske KD, Saunders SC, Euskirchen ES, Roberts D, Jaiteh MS, Esseen PA (2005) Edge influence on forest structure and composition in fragmented landscapes. Conserv Biol 19(3):768–782
Harrison S (1991) Local extinction in a metapopulation context: an empirical evaluation. Biol J Linn Soc 42(1–2):73–88
Haynes KJ, Dillemuth FP, Anderson BJ, Hakes AS, Jackson HB, Jackson SE, Cronin JT (2007) Landscape context outweighs local habitat quality in its effects on herbivore dispersal and distribution. Oecologia 151(3):431–441. https://doi.org/10.1007/s00442-006-0600-3
Hesselbarth MHK, Sciaini M, Nowosad J (2018) landscapemetrics: landscape metrics for categorical map patterns. R package version 0.1.1
Hijmans RJ, Van Etten J (2012) raster: geographic analysis and modeling with raster data. R package version 2.5-8
Hill MF, Caswell H (1999) Habitat fragmentation and extinction thresholds on fractal landscapes. Ecol Lett 2(2):121–127
Hoechstetter S, Walz U, Thinh NX (2011) Adapting lacunarity techniques for gradient-based analyses of landscape surfaces. Ecol Complex 8(3):229–238. https://doi.org/10.1016/j.ecocom.2011.01.001
Holyoak M, Leibold MA, Holt RD (2005) Metacommunities: spatial dynamics and ecological communities. University of Chicago Press, Chicago
Homer C, Dewitz J, Yang LM, Jin S, Danielson P, Xian G, Coulston J, Herold N, Wickham J, Megown K (2015) Completion of the 2011 National Land Cover Database for the Conterminous United States - representing a decade of land cover change information. Photogramm Eng Remote Sens 81(5):345–354. https://doi.org/10.14358/pers.81.5.345
Huffaker CB (1958) Experimental studies on predation: dispersion factors and predator-prey oscillations. Hilgardia 27:343–383
Jaeger JAG (2000) Landscape division, splitting index, and effective mesh size: new measures of landscape fragmentation. Landsc Ecol 15(2):115–130. https://doi.org/10.1023/a:1008129329289
Jetz W, Wilcove DS, Dobson AP (2007) Projected impacts of climate and land-use change on the global diversity of birds. PLoS Biol 5(6):1211–1219. https://doi.org/10.1371/journal.pbio.0050157

Libros de Estadística en PDF | Statistics Books in PDF

Johnson DH, Igl LD (2001) Area requirements of grassland birds: a regional perspective. Auk 118 (1):24–34
Keitt TH (2000) Spectral representation of neutral landscapes. Landsc Ecol 15(5):479–493. https://doi.org/10.1023/a:1008193015770
Kotliar NB, Wiens JA (1990) Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity. Oikos 59(2):253–260
Kupfer JA (2012) Landscape ecology and biogeography: rethinking landscape metrics in a post-FRAGSTATS landscape. Prog Phys Geogr 36(3):400–420. https://doi.org/10.1177/0309133312439594
Kupfer JA, Malanson GP, Franklin SB (2006) Not seeing the ocean for the islands: the mediating influence of matrix-based processes on forest fragmentation effects. Glob Ecol Biogeogr 15 (1):8–20. https://doi.org/10.1111/j.1466-822x.2006.00204.x
Lambin EF, Turner BL, Geist HJ, Agbola SB, Angelsen A, Bruce JW, Coomes OT, Dirzo R, Fischer G, Folke C, George PS, Homewood K, Imbernon J, Leemans R, Li XB, Moran EF, Mortimore M, Ramakrishnan PS, Richards JF, Skanes H, Steffen W, Stone GD, Svedin U, Veldkamp TA, Vogel C, Xu JC (2001) The causes of land-use and land-cover change: moving beyond the myths. Global Environ Change 11(4):261–269
Lamigueiro OP, Hijmans R (2018) rasterVis. R package version 0.44
Laurance WF (1991) Edge effects in tropical forest fragments: application of a model for the design of nature reserves. Biol Conserv 57(2):205–219. https://doi.org/10.1016/0006-3207(91)90139-z
Laurance WF, Useche DC (2009) Environmental synergisms and extinctions of tropical species. Conserv Biol 23(6):1427–1437. https://doi.org/10.1111/j.1523-1739.2009.01336.x
Lawler JJ, Lewis DJ, Nelson E, Plantinga AJ, Polasky S, Withey JC, Helmers DP, Martinuzzi S, Pennington D, Radeloff VC (2014) Projected land-use change impacts on ecosystem services in the United States. Proc Natl Acad Sci U S A 111(20):7492–7497. https://doi.org/10.1073/pnas.1405557111
Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. Bull Entomol Soc Am 15:237–240
Li HB, Reynolds JF (1993) A new contagion index to quantify spatial patterns of landscapes. Landsc Ecol 8(3):155–162. https://doi.org/10.1007/bf00125347
Li H, Reynolds JF (1995) On definition and quantification of heterogeneity. Oikos 73(2):280–284
Lindenmayer DB, Fischer J (2007) Habitat fragmentation and landscape change: an ecological and conservation synthesis. Island Press, Washington, DC
Lookingbill TR, Elmore AJ, Engelhardt KAM, Churchill JB, Gates JE, Johnson JB (2010a) Influence of wetland networks on bat activity in mixed-use landscapes. Biol Conserv 143 (4):974–983. https://doi.org/10.1016/j.biocon.2010.01.011
Lookingbill TR, Gardner RH, Ferrari JR, Keller CE (2010b) Combining a dispersal model with network theory to assess habitat connectivity. Ecol Appl 20(2):427–441
MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press, Princeton, NJ
Macedo MN, DeFries RS, Morton DC, Stickler CM, Galford GL, Shimabukuro YE (2012) Decoupling of deforestation and soy production in the southern Amazon during the late 2000s. Proc Natl Acad Sci U S A 109(4):1341–1346. https://doi.org/10.1073/pnas.1111374109
Malanson GP, Zeng Y, Walsh SJ (2006) Landscape frontiers, geography frontiers: lessons to be learned. Prof Geogr 58(4):383–396. https://doi.org/10.1111/j.1467-9272.2006.00576.x
McGarigal K, McComb WC (1995) Relationships between landscape structure and breeding birds in the Oregon Coast range. Ecol Monogr 65(3):235–260. https://doi.org/10.2307/2937059
McGarigal K, Cushman SA, Neel MC, Ene E (2002) FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: http://www.umass.edu/landeco/research/fragstats/fragstats.html

Libros de Estadística en PDF | Statistics Books in PDF

McGarigal K, Tagil S, Cushman SA (2009) Surface metrics: an alternative to patch metrics for the quantification of landscape structure. Landsc Ecol 24(3):433–450. https://doi.org/10.1007/s10980-009-9327-y

McIntyre S, Barrett GW (1992) Habitat variegation, an alternative to fragmentation. Conserv Biol 6 (1):146–147. https://doi.org/10.1046/j.1523-1739.1992.610146.x

McIntyre S, Hobbs R (1999) A framework for conceptualizing human effects on landscapes and its relevance to management and research models. Conserv Biol 13(6):1282–1292. https://doi.org/10.1046/j.1523-1739.1999.97509.x

Mendenhall CD, Karp DS, Meyer CFJ, Hadly EA, Daily GC (2014) Predicting biodiversity change and averting collapse in agricultural landscapes. Nature 509(7499):213. https://doi.org/10.1038/nature13139

Miller JR, Hobbs RJ (2002) Conservation where people live and work. Conserv Biol 16 (2):330–337. https://doi.org/10.1046/j.1523-1739.2002.00420.x

Moilanen A, Hanski I (2001) On the use of connectivity measures in spatial ecology. Oikos 95 (1):147–151

Moilanen A, Nieminen M (2002) Simple connectivity measures in spatial ecology. Ecology 83 (4):1131–1145

Moniem H, Holland JD (2013) Habitat connectivity for pollinator beetles using surface metrics. Landsc Ecol 28(7):1251–1267. https://doi.org/10.1007/s10980-013-9886-9

Moulds S, Buytaert W, Mijic A (2015) An open and extensible framework for spatially explicit land use change modelling: the lulcc R package. Geosci Model Dev 8(10):3215–3229. https://doi.org/10.5194/gmd-8-3215-2015

Naugle DE, Higgins KF, Nusser SM, Johnson WC (1999) Scale-dependent habitat use in three species of prairie wetland birds. Landsc Ecol 14(3):267–276. https://doi.org/10.1023/a:1008088429081

Neel MC, McGarigal K, Cushman SA (2004) Behavior of class-level landscape metrics across gradients of class aggregation and area. Landsc Ecol 19(4):435–455. https://doi.org/10.1023/B:LAND.0000030521.19856.cb

Newbold T, Hudson LN, Hill SLL, Contu S, Lysenko I, Senior RA, Borger L, Bennett DJ, Choimes A, Collen B, Day J, De Palma A, Diaz S, Echeverria-Londono S, Edgar MJ, Feldman A, Garon M, Harrison MLK, Alhusseini T, Ingram DJ, Itescu Y, Kattge J, Kemp V, Kirkpatrick L, Kleyer M, Correia DLP, Martin CD, Meiri S, Novosolov M, Pan Y, Phillips HRP, Purves DW, Robinson A, Simpson J, Tuck SL, Weiher E, White HJ, Ewers RM, Mace GM, Scharlemann JPW, Purvis A (2015) Global effects of land use on local terrestrial biodiversity. Nature 520(7545):45. https://doi.org/10.1038/nature14324

O'Neill RV, Gardner RH, Turner MG (1992) A hierarchical neutral model for landscape analysis. Landsc Ecol 7(1):55–61. https://doi.org/10.1007/bf02573957

O'Neill RV, Riitters KH, Wickham JD, Jones KB (1999) Landscape pattern metrics and regional assessment. Ecosyst Health 5(4):225–233. https://doi.org/10.1046/j.1526-0992.1999.09942.x

Oliver T, Roy DB, Hill JK, Brereton T, Thomas CD (2010) Heterogeneous landscapes promote population stability. Ecol Lett 13(4):473–484. https://doi.org/10.1111/j.1461-0248.2010.01441.x

Ovaskainen O, Sato K, Bascompte J, Hanski I (2002) Metapopulation models for extinction threshold in spatially correlated landscapes. J Theor Biol 215(1):95–108. https://doi.org/10.1006/jtbi.2001.2502

Pe'er G, Zurita GA, Schober L, Bellocq MI, Strer M, Muller M, Putz S (2013) Simple process-based simulators for generating spatial patterns of habitat loss and fragmentation: a review and introduction to the G-RaFFe model. PLoS One 8(5):e64968. https://doi.org/10.1371/journal.pone.0064968

Pettorelli N, Vik JO, Mysterud A, Gaillard JM, Tucker CJ, Stenseth NC (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends Ecol Evol 20 (9):503–510. https://doi.org/10.1016/j.tree.2005.05.011

Pfeifer M, Lefebvre V, Peres CA, Banks-Leite C, Wearn OR, Marsh CJ, Butchart SHM, Arroyo-Rodríguez V, Barlow J, Cerezo A, Cisneros L, D'Cruze N, Faria D, Hadley A, Harris SM,

Libros de Estadística en PDF | Statistics Books in PDF

Klingbeil BT, Kormann U, Lens L, Medina-Rangel GF, Morante-Filho JC, Olivier P, Peters SL, Pidgeon A, Ribeiro DB, Scherber C, Schneider-Maunoury L, Struebig M, Urbina-Cardona N, Watling JI, Willig MR, Wood EM, Ewers RM (2017) Creation of forest edges has a global impact on forest vertebrates. Nature 551:187. https://doi.org/10.1038/nature24457 https://www.nature.com/articles/nature24457#supplementary-information

Phalan B, Onial M, Balmford A, Green RE (2011) Reconciling food production and biodiversity conservation: land sharing and land sparing compared. Science 333(6047):1289–1291. https://doi.org/10.1126/science.1208742

Pickett STA, Thompson JN (1978) Patch dynamics and design of nature reserves. Biol Conserv 13 (1):27–37. https://doi.org/10.1016/0006-3207(78)90016-2

Pinto-Correia T, Mascarenhas J (1999) Contribution to the extensification/intensification debate: new trends in the Portuguese montado. Landsc Urban Plan 46(1–3):125–131. https://doi.org/10.1016/s0169-2046(99)00036-5

Polasky S, Nelson E, Camm J, Csuti B, Fackler P, Lonsdorf E, Montgomery C, White D, Arthur J, Garber-Yonts B, Haight R, Kagan J, Starfield A, Tobalske C (2008) Where to put things? Spatial land management to sustain biodiversity and economic returns. Biol Conserv 141 (6):1505–1524. https://doi.org/10.1016/j.biocon.2008.03.022

Prugh LR, Hodges KE, Sinclair ARE, Brashares JS (2008) Effect of habitat area and isolation on fragmented animal populations. Proc Natl Acad Sci U S A 105(52):20770–20775. https://doi.org/10.1073/pnas.0806080105

Remmel TK, Fortin MJ (2013) Categorical, class-focused map patterns: characterization and comparison. Landsc Ecol 28(8):1587–1599. https://doi.org/10.1007/s10980-013-9905-x

Remmel TK, Fortin MJ (2017) What constitutes a signficant difference in landscape pattern? (using R). In: Gergel SE, Turner MG (eds) Learning landscape ecology: concepts and techniques for a sustainable world, 2nd edn. Springer, New York

Reynolds C, Fletcher RJ, Carneiro CM, Jennings N, Ke A, LaScaleia MC, Lukhele MB, Mamba ML, Sibiya MD, Austin JD, Magagula CN, Mahlaba T, Monadjem A, Wisely SM, McCleery RA (2018) Inconsistent effects of landscape heterogeneity and land-use on animal diversity in an agricultural mosaic: a multi-scale and multi-taxon investigation. Landsc Ecol 33(2):241–255. https://doi.org/10.1007/s10980-017-0595-7

Ries L, Sisk TD (2010) What is an edge species? The implications of sensitivity to habitat edges. Oikos 119(10):1636–1642. https://doi.org/10.1111/j.1600-0706.2010.18414.x

Ries L, Fletcher RJ, Battin J, Sisk TD (2004) Ecological responses to habitat edges: mechanisms, models, and variability explained. Annu Rev Ecol Evol Syst 35:491–522. https://doi.org/10.1146/annurev.ecolsys.35.112202.130148

Ries L, Murphy SM, Wimp GM, Fletcher RJ Jr (2017) Closing persistent gaps in knowledge about edge ecology. Curr Landsc Ecol Rep 2(1):30–41

Riitters KH, O'Neill RV, Hunsaker CT, Wickham JD, Yankee DH, Timmins SP, Jones KB, Jackson BL (1995) A factor-analysis of landscape pattern and structure metrics. Landsc Ecol 10(1):23–39. https://doi.org/10.1007/bf00158551

Riitters KH, O'Neill RV, Wickham JD, Jones KB (1996) A note on contagion indices for landscape analysis. Landsc Ecol 11(4):197–202. https://doi.org/10.1007/bf02071810

Riitters K, Vogt P, Soille P, Estreguil C (2009) Landscape patterns from mathematical morphology on maps with contagion. Landsc Ecol 24(5):699–709. https://doi.org/10.1007/s10980-009-9344-x

Rindfuss RR, Walsh SJ, Turner BL, Fox J, Mishra V (2004) Developing a science of land change: challenges and methodological issues. Proc Natl Acad Sci U S A 101(39):13976–13981. https://doi.org/10.1073/pnas.0401545101

Roff DA (1974) Analysis of a population model demonstrating importance of dispersal in a heterogeneous environment. Oecologia 15(3):259–275. https://doi.org/10.1007/bf00345182

Saupe D (1988) Algorithms for random fractals. In: Petigen HO, Saupe D (eds) The science of fractal images. Springer, New York, pp 71–113

Libros de Estadística en PDF | Statistics Books in PDF

Saura S, Martinez-Millan J (2000) Landscape patterns simulation with a modified random clusters method. Landsc Ecol 15(7):661–678. https://doi.org/10.1023/a:1008107902848
Schlather M, Malinowski A, Menck PJ, Oesting M, Strokorb K (2015) Analysis, simulation and prediction of multivariate random fields with package random fields. J Stat Softw 63(8):1–25
Sciaini M, Fritsch M, Scherer C, Simpkins CE (2018) NLMR and landscapetools: an integrated environment for simulating and modifying neutral landscape models in R. Methods Ecol Evol. https://doi.org/10.1101/307306
Shitov VV, Moskalev PV (2005) Modification of the Voss algorithm for simulation of the internal structure of a porous medium. Tech Phys 50(2):141–145. https://doi.org/10.1134/1.1866426
Sisk TD, Haddad NM, Ehrlich PR (1997) Bird assemblages in patchy woodlands: modeling the effects of edge and matrix habitats. Ecol Appl 7(4):1170–1180
Sodhi NS, Koh LP, Clements R, Wanger TC, Hill JK, Hamer KC, Clough Y, Tscharntke T, Posa MRC, Lee TM (2010) Conserving Southeast Asian forest biodiversity in human-modified landscapes. Biol Conserv 143(10):2375–2384. https://doi.org/10.1016/j.biocon.2009.12.029
Suarez AV, Pfennig KS, Robinson SK (1997) Nesting success of a disturbance-dependent songbird on different kinds of edges. Conserv Biol 11(4):928–935
Temple SA, Cary JR (1988) Modeling dynamics of habitat-interior bird populations in fragmented landscapes. Conserv Biol 2(4):340–347. https://doi.org/10.1111/j.1523-1739.1988.tb00198.x
Thornton DH, Wirsing AJ, Roth JD, Murray DL (2013) Habitat quality and population density drive occupancy dynamics of snowshoe hare in variegated landscapes. Ecography 36 (5):610–621. https://doi.org/10.1111/j.1600-0587.2012.07737.x
Tilman D, Lehman CL (1997) Habitat destruction and species extinctions. In: Tilman D, Kareiva P (eds) Spatial ecology: the role of space in population dynamics and interspecific interactions. Princeton University Press, Princeton, NJ
Tilman D, Balzer C, Hill J, Befort BL (2011) Global food demand and the sustainable intensification of agriculture. Proc Natl Acad Sci U S A 108(50):20260–20264. https://doi.org/10.1073/pnas.1116437108
Tilman D, Clark M, Williams DR, Kimmel K, Polasky S, Packer C (2017) Future threats to biodiversity and pathways to their prevention. Nature 546(7656):73–81. https://doi.org/10.1038/nature22900
Tinker DB, Romme WH, Despain DG (2003) Historic range of variability in landscape structure in subalpine forests of the Greater Yellowstone Area, USA. Landsc Ecol 18(4):427–439
Tscharntke T, Klein AM, Kruess A, Steffan-Dewenter I, Thies C (2005) Landscape perspectives on agricultural intensification and biodiversity - ecosystem service management. Ecol Lett 8 (8):857–874. https://doi.org/10.1111/j.1461-0248.2005.00782.x
Tscharntke T, Tylianakis JM, Rand TA, Didham RK, Fahrig L, Batary P, Bengtsson J, Clough Y, Crist TO, Dormann CF, Ewers RM, Frund J, Holt RD, Holzschuh A, Klein AM, Kleijn D, Kremen C, Landis DA, Laurance W, Lindenmayer D, Scherber C, Sodhi N, Steffan-Dewenter I, Thies C, van der Putten WH, Westphal C (2012) Landscape moderation of biodiversity patterns and processes - eight hypotheses. Biol Rev 87(3):661–685. https://doi.org/10.1111/j.1469-185X.2011.00216.x
Turner MG (1989) Landscape ecology: the effect of pattern on process. Annu Rev Ecol Syst 20:171–197. https://doi.org/10.1146/annurev.ecolsys.20.1.171
Turner MG, Gardner RH (2015) Landscape ecology in theory and practice, 2nd edn. Springer, New York
Turner BL, Lambin EF, Reenberg A (2007) The emergence of land change science for global environmental change and sustainability. Proc Natl Acad Sci U S A 104(52):20666–20671. https://doi.org/10.1073/pnas.0704119104
VanDerWal J, Shoo L, Januchowski S (2010) SDMTools: Species Distribution Modelling Tools: tools for processing data associated with species distribution modelling exercises. R package version 1.1
Vergara PM, Meneses LO, Grez AA, Quiroz MS, Soto GE, Perez-Hernandez CG, Diaz PA, Hahn IJ, Fierro A (2017) Occupancy pattern of a long-horned beetle in a variegated forest landscape:

Libros de Estadística en PDF | Statistics Books in PDF

linkages between tree quality and forest cover across spatial scales. Landsc Ecol 32(2):279–293. https://doi.org/10.1007/s10980-016-0443-1
Villard MA, Metzger JP (2014) Beyond the fragmentation debate: a conceptual model to predict when habitat configuration really matters. J Appl Ecol 51(2):309–318. https://doi.org/10.1111/1365-2664.12190
Villard MA, Trzcinski KM, Merriam G (1999) Fragmentation effects on forest birds: relative influence of woodland cover and configuration on landscape occupancy. Conserv Biol 13 (4):774–783
Vitousek PM (1994) Beyond global warming: ecology and global change. Ecology 75 (7):1861–1876. https://doi.org/10.2307/1941591
Vogt P, Riitters KH, Estreguil C, Kozak J, Wade TG (2007) Mapping spatial patterns with morphological image processing. Landsc Ecol 22(2):171–177. https://doi.org/10.1007/s10980-006-9013-2
Voss RF (1985) Random fractal forgeries. In: Earnshaw RA (ed) Fundamental algorithms in computer graphics. Springer, New York, pp 805–883
Walz U, Hoechstetter S, Dragut L, Blaschke T (2016) Integrating time and the third spatial dimension in landscape structure analysis. Landsc Res 41(3):279–293. https://doi.org/10.1080/01426397.2015.1078455
Wang XL, Blanchet FG, Koper N (2014) Measuring habitat fragmentation: an evaluation of landscape pattern metrics. Methods Ecol Evol 5(7):634–646. https://doi.org/10.1111/2041-210x.12198
Watling JI, Orrock JL (2010) Measuring edge contrast using biotic criteria helps define edge effects on the density of an invasive plant. Landsc Ecol 25(1):69–78. https://doi.org/10.1007/s10980-009-9416-y
Whitcomb RF, Lynch JF, Opler PA, Robbins CS (1976) Island biogeography and conservation: strategy and limitations. Science 193(4257):1030–1032
Wiens JA (1976) Population responses to patchy environments. Annu Rev Ecol Syst 7:81–120
Wiens JA (1995) Landscape mosaics and ecological theory. In: Hansson L, Fahrig L, Merriam G (eds) Mosaic landscapes and ecological processes. Chapman and Hall, London, pp 1–26
Wilcove DS (1985) Nest predation in forest tracts and the decline of migratory songbirds. Ecology 66(4):1211–1214. https://doi.org/10.2307/1939174
Wilcove DS, Rothstein D, Dubow J, Phillips A, Losos E (1998) Quantifying threats to imperiled species in the United States. Bioscience 48(8):607–615
With KA (1997) The application of neutral landscape models in conservation biology. Conserv Biol 11(5):1069–1080. https://doi.org/10.1046/j.1523-1739.1997.96210.x
With KA, King AW (1997) The use and misuse of neutral landscape models in ecology. Oikos 79 (2):219–229. https://doi.org/10.2307/3546007
Wright CK, Larson B, Lark TJ, Gibbs HK (2017) Recent grassland losses are concentrated around US ethanol refineries. Environ Res Lett 12(4). https://doi.org/10.1088/1748-9326/aa6446
Wu JG, Loucks OL (1995) From balance of nature to hierarchical patch dynamics: a paradigm shift in ecology. Q Rev Biol 70(4):439–466. https://doi.org/10.1086/419172

Libros de Estadística en PDF | Statistics Books in PDF

Chapter 4
Spatial Dispersion and Point Data



4.1 Introduction

Points of information located in space can describe a variety of ecological processes and conservation problems, ranging from GPS locations of species occurrence to the origin of spread of an invasive species. Point data, or data that describe distinct locations in space, might reflect the locations of individual trees (Condit et al. 2000), nests of birds (Bayard and Elphick 2010), patches of habitat (Lancaster et al. 2003), or patchy disturbances (e.g., burrows; Schooley and Wiens 2001). Often the focus of point pattern analysis is on quantifying spatial dispersion (aggregated, uniform, or random distributions; Fig. 4.1), determining if and how dispersion varies with spatial and temporal scale, and understanding the causes of these patterns (Illian et al. 2008; Wiegand and Moloney 2014; Velazquez et al. 2016).

Why are point patterns important? Understanding such patterns lies at the heart of interpreting territoriality, interference competition, social behavior, etc. Community ecologists have long hypothesized that intraspecific *aggregation* can promote community coexistence (Ives and May 1985), and the spatial patterns of individuals can help explain why aggregation occurs (Melles et al. 2009; Lara-Romero et al. 2016). Spatial point patterns can also provide insight into species interactions (Andersen 1992; Rodriguez-Perez et al. 2012) and mechanisms of coexistence (Brown et al. 2011). From a conservation perspective, the aggregation of patches, which can be viewed as points across a landscape, is predicted to reduce extinction rates in populations (Ovaskainen et al. 2002). When using models, such as species–area models, to predict the biodiversity consequences of both habitat loss and other forms of disturbance, whether changes occur in a spatially random or in an aggregated manner can greatly alter conclusions (Seabloom et al. 2002; Kallimanis et al. 2005). Furthermore, understanding the spatial scale of these patterns can give insight into processes that drive species distributions (Wiegand et al. 2009), disturbance regimes (Yang et al. 2008), species invasion (Kelly and Meentemeyer 2002; Deckers et al.

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_4

Libros de Estadística en PDF | Statistics Books in PDF

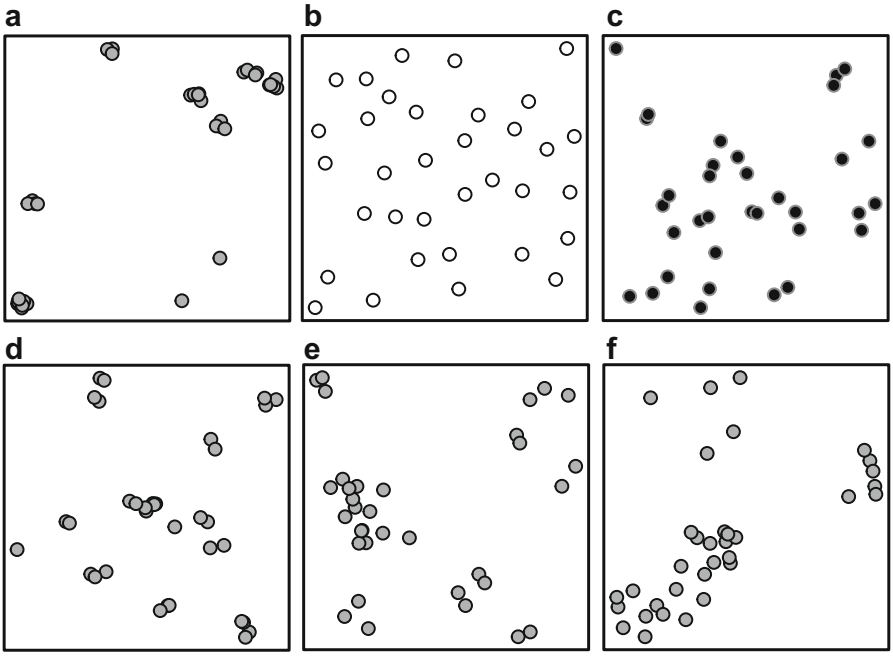


Fig. 4.1 The spatial dispersion of points and scale dependence in spatial dispersion. (**a**) Aggregated, (**b**) uniform (regular), and (**c**) random distributions of points. (**d–f**) Aggregation occurring at increasingly large spatial scales. Aggregation based on a Matérn cluster process, while regular distribution based on a Matérn inhibition process

2005; Maheu-Giroux and de Blois 2007), and population persistence (Adler and Nuernberger 1994).

Here, we provide an introduction and overview on the use of spatial point pattern analysis to address ecological and conservation questions related to the spatial dispersion of species. Spatial point pattern analysis focuses on examining patterns of points to establish whether there are regularities in the process they represent (are they clustered, randomly spread, or evenly distributed?), and the spatial scales (extents) at which these patterns occur. We first describe common characteristics of point data and related point patterns. We then provide a brief summary of different types of statistical models used to identify spatial point patterns and the scale(s) at which they occur. We illustrate these models with data on plant distributions and show how point data can be simulated to better interpret why point patterns occur in nature.

Libros de Estadística en PDF | Statistics Books in PDF

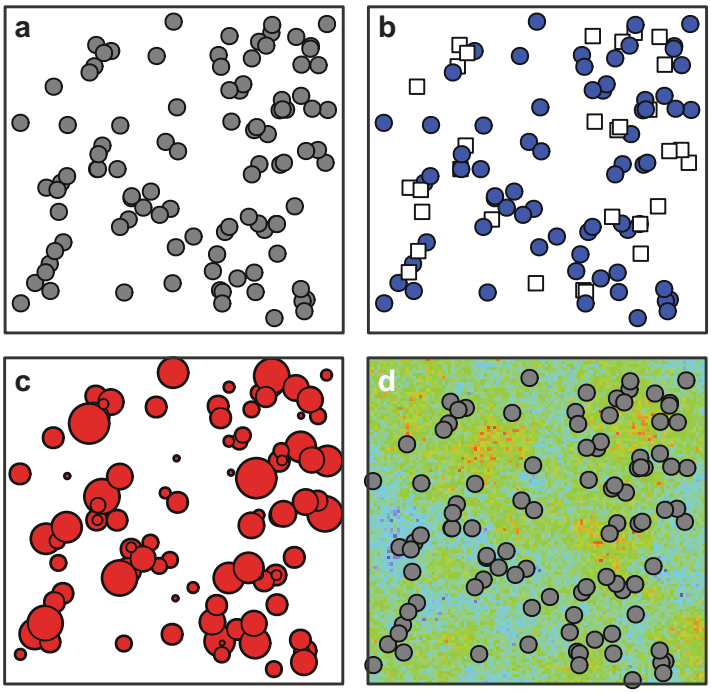


Fig. 4.2 Some common characteristics of point patterns. (**a**) A point pattern measured in a square plot (boundary). (**b**) Bivariate marks, such as those describing the presence and absence of a species on potential habitat. (**c**) An example of a continuous mark, in this case the area of the point. (**d**) A point pattern with an underlying gradient of interest (with variation in the gradient indicated by different colors) that may help explain a point pattern

4.2 Key Concepts and Approaches

4.2.1 Characteristics of Point Patterns

Point data come in a variety of forms. For the approaches considered in this chapter, all point data consist of *x*–*y* coordinates of point locations (Fig. 4.2). An inherent assumption is that the spatial pattern of the points was generated by a unique *point process* over the entire study area. Hence, the delineation of the study area in terms of size and location is important and will affect the point pattern observed. The shape of the study area can be regular (e.g., rectangular) or can be based on highly irregular polygons. Locations of points can have attributes, which can be categorical (e.g., species names) or quantitative (biomass, height, etc.). These attributes are often referred to as *marks*. In some situations, we may also want to integrate other sources of information or covariates (e.g., elevation) into our interpretation of point patterns.

Point patterns arise from *point processes* (Table 4.1) (Illian et al. 2008). Point processes are random processes that result in realizations (i.e., observations) of point

Libros de Estadística en PDF | Statistics Books in PDF

Table 4.1 Common terms for the spatial dispersion of point data in ecology

Term	Description
Aggregation	A pattern of dispersion where the units are closer together than what would be expected if they were randomly dispersed. Can operate at different scales.
Edge effect	An effect in point pattern analysis that arises near plot boundaries due to a lack of knowledge regarding points occurring outside of plots, which can cause bias in the interpretation of point patterns.
First-order statistics	Global statistics summarizing point patterns across an entire study area.
Homogeneous point process	When the intensity (~density) of points is constant (i.e., does not change) over space.
Inhomogeneous point process	When the intensity (~density) of points changes with location.
Isotropy	When the point process does not change with direction.
Marks	Attributes of points. Can be qualitative or quantitative.
Point process	Random processes that result in realizations of point patterns across space and/or time.
Second-order statistics	Local statistics that summarize point patterns in a neighborhood around a location.
Stationarity	When the point process does not change over space (or time), which is referred to as a homogeneous point process.
Thinned point process	A sample from an underlying point process.

patterns across space and/or time. The simplest point process that is frequently used is the Poisson point process. A Poisson point process is based on the idea that points are independently distributed in space (Diggle 2003). Such a distribution can be described based on the intensity (~or density) of points, within a specified region (see below).

Point processes are frequently described as being *homogeneous* or *stationary* versus being inhomogeneous or non-stationary (Diggle 2003). When point processes are homogeneous, the intensity does not vary over space (or time when considering spatiotemporal data). Conversely, *inhomogeneous* point processes have spatially (or temporally) varying intensities over the study area. In such situations, a common objective is trying to understand what factors can explain these trends.

Thinned point processes occur when the observed point locations come from a sample of the underlying point process (Diggle 2003; Illian et al. 2008). This sampling, or "thinning," can lead to inhomogeneity in the intensity function independent of the underlying point process, which can be due to biological processes, such differential mortality, or through study design issues. For example, presence-only data frequently used in modeling species distributions can be viewed as thinned point data (Warton and Shepherd 2010), which are often biased near roads and such bias can influence predictions of models (Elith et al. 2006; Phillips et al. 2009). In such cases, the thinning process can be estimated to account for such biased sampling (Fithian et al. 2015). Here, we focus on data that do not come from a thinned point process but rather are complete censuses in a region of interest.

Libros de Estadística en PDF | Statistics Books in PDF

However, we will come back to this issue when analyzing opportunistic data of species distributions in Chap. 7 (Renner et al. 2015).

4.2.2 Summary Statistics for Point Patterns

To interpret spatial point patterns, we can measure a variety of properties of the spatial pattern. Illian et al. (2008) categorized the various ways to summarize point patterns based on two dimensions: (1) whether summaries are numerical or functional; and (2) point or location focused (see also Wiegand and Moloney 2014). Numerical summaries focus on "global" statistics in the sense that they consider patterns (aggregation, randomness, or regular distributions) over the entire study area based on count data from contiguous sampling units (e.g., quadrats). These statistics have a long history in ecology (Clark and Evans 1954; Lloyd 1967; Velazquez et al. 2016). They focus on summary statistics taken from all of the points, such as the mean and variance of counts of points taken from sampling units. For instance, evidence for aggregation or regular dispersion has traditionally come from interpreting the simple mean–variance ratio wherein the expected absence of spatial pattern (randomness) follows a Poisson distribution. If the mean $\gg$ variance, there tends to be evidence for regularity, if the mean $\cong$ variance, there tends to be evidence for randomness, and if the mean $\ll$ variance, there tends to be evidence for aggregation (Dale and Fortin 2014). Yet numerical statistics do not provide information regarding the scale or explicit spatial patterning of the points. They are simply contingent on how the data are collected (e.g., counts in quadrats). We will not focus on these statistics because they provide less information on spatial aspects of point patterns.

Functional statistics use a variety of approaches to understand the local spatial pattern and to estimate the spatial scale of such patterns (Diggle 2003; Illian et al. 2008). Functional statistics can potentially capture the concept of ecological neighborhoods for point data (Addicott et al. 1987), as described in Chap. 2, and can help identify the spatial scale(s) at which pattern occurs (Gustafson 1998). Here, we focus primarily on functional statistics and how they can be used in spatial ecology.

Point-focused statistics summarize patterns from the perspective of individual focal points. For example, the pattern of points can be described based on nearest neighbor patterns, where distances from a focal point to the next closest point are measured. Location-focused statistics, in contrast, summarize patterns from (typically all) locations within the study area where points may or may not occur.

A third, related dimension regarding summary statistics relates to how point data are summarized across points. *First-order statistics* focus on summarizing point data based individual points without explicit focus on relations to other points. This is typically accomplished with the intensity function, $\lambda(x)$, which quantifies the relative density of points at or around location x. This is a location-based, first-order statistic because it summarizes information at location x without using explicit relational data regarding inter-point information. *Second-order statistics* focus on statistical

Libros de Estadística en PDF | Statistics Books in PDF

relationships estimated from information contained in pairs of points. For instance, quantifying the number of points within a certain distance r of a focal point would be a functional, point-based, second-order statistic, because this statistic uses information on pairs of points (not locations) and can be calculated across a range of scales. There are also higher order statistics (e.g., using information from sets of three points) (Wiegand and Moloney 2014), but in practice, the overwhelming majority of spatial point pattern analysis focuses on first-order and second-order statistics.

4.2.2.1 Null Models

To assess the significance of the observed point pattern, null models are typically used. In this way, the observed point pattern is compared to point patterns generated from null models (Baddeley et al. 2014). The Complete Spatial Randomness (CSR, i.e., a homogeneous Poisson process) null model is commonly used (Wiegand and Moloney 2004). Under CSR, it is assumed that: (1) the number of points (n) in a region follows a Poisson distribution with a mean, λ (sometimes called the intensity; $\lambda = n/A$, where A is the area of the study region); and (2) given the number of points in the region, the points are an independent random sample having the same probability of occurring anywhere, such that points come from the uniform distribution across the study region (Diggle 2003). While it is likely that CSR rarely occurs in nature, it is the simplest null model to contrast with the observed pattern, which can help determine whether or not the observed pattern is non-random.

To generate a CSR pattern within a regularly shaped study area, a Poisson point process is used. In a nutshell, for a given number of points (n), we can simulate a CSR pattern many times. These will provide confidence envelopes for CSR (Baddeley et al. 2014). This general approach is typically referred to as Monte Carlo simulations (Manly 2006). If the observed data fall outside of the confidence envelopes, there is evidence for significant spatial patterns that differ from CSR.

While Monte Carlo methods are often used to explore significance under CSR, other null models can be considered. For instance, certain point processes assume aggregation processes (e.g., Thomas or Matérn processes) while others can capture inhibition processes (e.g., Gibbs or Strauss processes; Illian et al. 2008). These processes that assume non-random patterns can also be considered null models for point patterns (see Sect. 4.3.8).

Alternatively, we might know of an underlying resource gradient that a species uses. We could develop a null model that uses this underlying gradient as a null. For instance, point locations can be driven, in part, by environmental gradients, such as elevation or precipitation. With information on these gradients, we can account for effects of gradients while asking whether points are aggregated or regularly dispersed (e.g., Melles et al. 2009). Such a test would ask whether the species shows characteristic spatial patterns above and beyond the underlying resource gradient. These types of tests are based on inhomogeneous point processes, where the inhomogeneity is driven by environmental gradients. See Sect. 4.3.7 for an example.

Libros de Estadística en PDF | Statistics Books in PDF

4.2.2.2 Inhomogeneous Point Process Models

A powerful advance in point pattern analysis is the development of inhomogeneous point process models. An *inhomogeneous point process model* is similar to a generalized linear model for point data, where we are modeling the intensity of points in the study area (Illian et al. 2008, 2013; Renner et al. 2015). Typically, this is formulated as an inhomogeneous Poisson point process (as opposed to other processes, such as Thomas processes) which is very similar to Poisson regression. This modeling framework is inhomogeneous because it accounts for covariates that may influence the intensity of points, as well as the potential for non-stationarity in point patterns (see below). The point process model can be defined as:

$$\lambda(s) = \exp(\alpha + \beta x(s) + \ldots), \tag{4.1}$$

where λ is the intensity of the point process at locations s across the study area, x is a covariate at location s, and α and β are parameters to be estimated. Inhomogeneous point process models can be used in a variety of ways, both in terms of exploratory data analysis, for inference, and for prediction and spatial mapping (see below and Chaps. 7 and 8).

4.2.3 Common Statistical Models for Point Patterns

A variety of exploratory data analysis approaches have been derived for interpreting point patterns. For second-order statistics, some common approaches include measuring the degree of spatial aggregation between neighboring points within a circular area of a given radius or calculating the number of points at a specific distance from a focal point (Fig. 4.3). Each of these measurements asks subtly different questions. Wiegand et al. (2013) illustrated how different point processes can lead to similar patterns in some types of statistical summaries and argued that several statistics should be used to fully capture spatial point patterns. Below we illustrate some of the most common approaches.

4.2.3.1 Ripley's *K* (and *L*)

We start with describing the Ripley's K function (Ripley 1976) because it is probably the most common second-order statistic for point patterns used in ecology. Ripley's K calculates the degree of spatial aggregation of points within a circle of radius r and contrasts the observed pattern to that expected under CSR (Fig. 4.3a). Ripley's K is defined as:

Libros de Estadística en PDF | Statistics Books in PDF

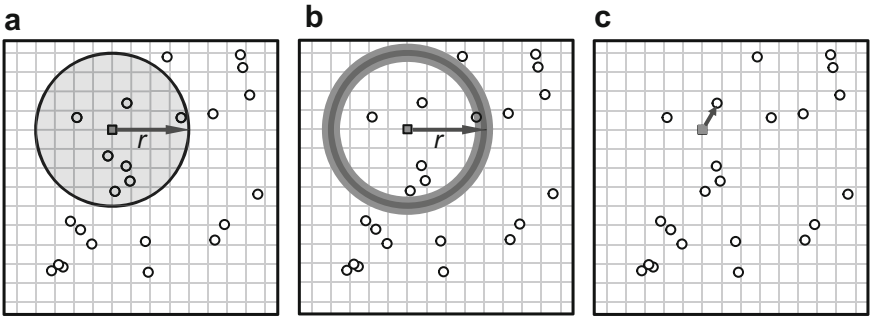


Fig. 4.3 Second-order statistics for interpreting spatial point patterns capture spatial information of points in different ways. (**a**) The K (or L) function uses information from all points within a specified area (buffer of radius, r) around a focal point (square). (**b**) The g function (pair correlation function) uses point information at a specified radius r from a focal point or location. This distance is often smoothed with a kernel bandwidth, indicated by dark shading around radius r represented by the grey line. In this situation, the bandwidth is relatively wide such that it overlaps with one point. (**c**) The G function uses the nearest neighbor distances between points (arrow denotes the nearest neighbor from a focal point)

$$K(r) = \frac{E}{\lambda}, \tag{4.2}$$

where $\lambda = n/A$, and $E = \#$ points within radius r of an arbitrarily chosen point (not including that point). Thus, if $K(r)$ is larger than expected for a radius r, the observed number of events shows evidence of clustering, whereas if $K(r)$ is smaller than expected for a radius r there is evidence for regularity. Given that the area of a circle is πr^2, the expected number of points in a circle of radius r under CSR is:

$$E_{\text{CSR}} = \lambda \pi r^2. \tag{4.3}$$

Now, if we rearrange and substitute, we get that the expected value of the K function under CSR is:

$$K(r)_{\text{CSR}} = \pi r^2. \tag{4.4}$$

This expectation regarding CSR results in an exponential increase in expected values due to r^2, which has some undesirable properties. Consequently, we can "linearize" this function, such that the expected value under CSR equals r (Ripley 1979):

$$\widehat{L}(r) = \sqrt{\frac{K(r)}{\pi}} = r. \tag{4.5}$$

Libros de Estadística en PDF | Statistics Books in PDF

Graphically, plots of $L(r)$ as a function of r will result in the 1:1 line being the expected value under CSR. In this situation, if $L(r) > r$, there is evidence for spatial aggregation, whereas if $L(r) < r$ there is evidence for regularity. Another popular linearization is to set the expected value = 0 under CSR:

$$\widehat{L}(r) = \sqrt{\frac{K(r)}{\pi}} - r. \tag{4.6}$$

In this case, if $L(r) > 0$, there is evidence for aggregation, whereas if $L(r) < 0$ there is evidence for regularity. When using the linearized version of the K function (i.e., the L function), it is important to know whether it is linearized to r or 0.

In practice, Ripley's K is calculated as follows:

$$\widehat{K}(r) = A \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{w_{ij} I_r(d_{ij} < r)}{n^2}, \tag{4.7}$$

where d_{ij} is the distance between events i and j, I_r is an indicator function taking a value of 1 if $d_{ij} < r$ or zero otherwise, and w_{ij} is a weighting factor that corrects for study area *edge effects*. If we ignore edge correction issues (we will expand on these below), this can simplify to:

$$\widehat{K}(r) = \lambda^{-1} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{I_r(d_{ij} < r)}{n}. \tag{4.8}$$

Note Eq. (4.8) is very similar to how we initially defined K. In a nutshell, this statistic is calculating the mean number of points within a distance r, scaled to the average intensity, λ, and we then compare it to what is expected under CSR (or other null models).

4.2.3.2 Pair Correlation Function

While the K and L functions have a long history of use, the interpretation of the scales at which patterns occur can be unclear. For instance, if there is a strong aggregated pattern at approximately a 5 m radius but no effect at larger scales, Ripley's K functions may suggest that aggregated patterns occur at larger scales, because $r > 5$ m still uses data from $r < 5$ m. That is, the K (and L) function is a cumulative function, where all points less than r are used. Consequently, distance-based analyses (Fig. 4.3b) are often recommended either to complement or in lieu of area-based analyses like Ripley's K (Illian et al. 2008). The most common statistic for a distance-based analysis is the pair correlation function, g, where:

Libros de Estadística en PDF | Statistics Books in PDF

$$g(r) = \frac{\frac{dK(r)}{dr}}{2\pi r}. \tag{4.9}$$

Essentially, the g function calculates the slope of the K function at radius r divided by the circumference of radius r. The value expected for this statistic under CSR is 1, if $g(r) > 1$, there is evidence for aggregation, whereas if $g(r) < 1$ there is evidence for regularity. Note that in practice, this idea has sometimes been implemented by defining a ring width (Δr) or bin size rather than using the derivative in Eq. (4.9) (Wiegand and Moloney 2004). If bands are too narrow, data sparseness will cause spurious irregularities in the function. If the bands are too wide, they start to approach the K function. Rather than manually determining the ring width, we can use a smoothing kernel to address this issue (Penttinen et al. 1992) (see below; Chap. 6). Wiegand and Moloney (2004) also discussed the O-ring statistic, which is related to the pair correlation function, where $O(r) = \lambda g(r)$.

While the pair correlation function is not used as frequently as the K function in ecology, several scientists have argued that it is more informative than the K function because it can better isolate the relevant spatial scales of point patterns (Illian et al. 2008). In fact, the K function can be viewed as simply a cumulative version of the pair correlation function (Baddeley 2007; Wiegand and Moloney 2014; Velazquez et al. 2016).

4.2.3.3 Distances Between Neighbors: The *G*-Function

The G-function estimates the cumulative distribution of the nearest neighbor distances for a given point, sometimes referred to as the "event-to-event" distribution (Fig. 4.3c). This function is useful for interpreting the probability of finding the nearest neighbor for a given distance r. It can also be used when considering nearest neighbor distance metrics in studies of habitat fragmentation (see Chap. 3) by providing a means to interpret the spatial scaling of patch isolation. This function can be written as:

$$\widehat{G}(r) = \frac{1}{N} \sum_{i=1}^{n} 1(d_i \leq r), \tag{4.10}$$

where N is the number of points and d_i is the distance between observed nearest neighbors. The expected value of the G-function under CSR is:

$$G(r)_{\mathrm{CSR}} = 1 - \exp(-\lambda \pi r^2). \tag{4.11}$$

When the observed $G(r)$ is greater than $G_{\mathrm{CSR}}(r)$, distances are shorter than expected from a Poisson process, indicative of spatial aggregation. In contrast, when $G(r) < G_{\mathrm{CSR}}(r)$, distances are longer than what is expected from a Poisson

Libros de Estadística en PDF | Statistics Books in PDF

process, indicative of a uniform dispersion. This function has an intuitive interpretation; however, it ignores all point information beyond that of the nearest neighbor (unlike K, L, or g). As a consequence, it tends to capture smaller scale heterogeneity in point patterns relative to other summary statistics (Wiegand and Moloney 2014). There are other related functions that focus on location rather than points (e.g., the empty space function, $F(r)$), but we will not elaborate on those here.

4.2.3.4 Bivariate and Multivariate Marks

The above functions have been extended for interpreting the spatial pattern between two types of points (Andersen 1992). This extension opens up several problems that can be tackled by spatial point pattern analysis, such as the role of interspecific competition on species distribution, resource selection by animals, and how human activities may impact oviposition behavior. This idea can be more generally framed as an analysis of marked points. *Marks* are simply information about the events (points) and can be categorical (e.g., predator versus prey) or continuous (e.g., dbh of trees).

For two categorical points (1 and 2), the sample bivariate (or cross) Ripley's K function can be described as:

$$\widehat{K_{12}}(r) = A \sum_{\substack{i=1 \\ i \neq j}}^{n_1} \sum_{\substack{j=1 \\ j \neq i}}^{n_2} \frac{I_r\left(d_{ij} < r\right)}{n_1} n_2. \tag{4.12}$$

This is then repeated for K_{21} and a weighted average is calculated. The expectation for the function is contingent on the underlying question (see below for more details). This general approach can be extended to several types of points (Condit et al. 2000).

When using continuous marks, the mark correlation function is often used (Penttinen et al. 1992):

$$\widehat{K_m}(r) = \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} I_r m_i m_j, \tag{4.13}$$

where m_i is the quantitative mark at location i. The expected value is then:

$$K_m(r)_{\mathrm{CSR}} = \sqrt{\frac{\widehat{K_m}(r)}{\pi \mu^2}}, \tag{4.14}$$

where μ is the mean value for the mark. Note that if you replace $m_i m_j$ with $(m_i - m_j)^2$ in Eq. (4.13), the function becomes nearly equivalent to the empirical variogram (see

Libros de Estadística en PDF | Statistics Books in PDF

Chap. 5). Thus, the mark correlation function asks whether quantitative values of marks tend to be positively or negatively associated (e.g., larger trees tend to be close to other large trees).

4.2.3.5 Edge Effect Correction

The above statistics ignore the problem of edge effects. In point pattern analysis, *edge effects* arise near the boundary of the study area of interest where there is no information on points located outside the study area. Consequently, when considering a radius r from a point near a boundary, the number of observed points is likely lower than the true number of points if points could occur outside of study area. Statistics that do not account for edge effects tend to result in biased estimates of spatial dispersion. It is therefore recommended to account for edge effects. Note that under some situations, such as if the boundary is a hard (true) boundary for points, it may make sense to not account for edge effects (Lancaster and Downes 2004).

Each of the above statistics (K, L, pair correlation g, etc.) can be adjusted for edge effects. Some corrections use weights for points, some use buffers, and some use some sort of wrapping (e.g., a torus) to adjust for boundary effects. A simple correction based on weights that Ripley (1988) suggested is sometimes referred to as an "isotropic correction." In this case, if the radius r is completely inside the plot, the weight, w, equals 1. If part of the circle is outside plot, this correction uses the inverse of the fraction of circumference lying inside the plot (e.g., if ½ of circumference is inside the plot, the $w = 2$ for points inside). This correction is intuitive but does not work well for irregular complex boundaries. The "translate correction" is a common correction that works on all plot/window geometries (also known as toroidal correction). This correction does not use weights but rather wraps points around the plot like a torus.

4.2.3.6 General Assumptions

The above analyses all assume that the point process is *stationary* (homogeneous). The analyses also assume therefore that the process is *isotropic*. Yet the above functions can be extended to accommodate *inhomogenous point processes*, where intensity varies across space.

These models also assume a complete census of the point locations, not a sample. It has been argued that if the observed points are a random sample of the true distribution that these analyses should still be valid (because a Poisson point process that is thinned randomly/independently is still a Poisson point process). However, if point locations are a biased sample, then such bias could alter the conclusions about the point pattern.

Libros de Estadística en PDF | Statistics Books in PDF

4.3 Examples in R

4.3.1 Packages in R

In R, there are a few libraries that can be used for spatial point pattern analysis. `Spatial` (Venables and Ripley 2002) allows for limited point-pattern analysis, whereas `spatstat` is a more flexible and comprehensive library. We will focus on the use of `spatstat` (Baddeley and Turner 2005).

In `spatstat`, the basic data types are Point Patterns (`ppp`), Windows (`owin`), and Pixel Images (`im`). A point pattern is a dataset recording the spatial locations of all "events" or "individuals" observed in a certain region. A window is a region in two-dimensional space. It usually represents the boundaries of the study area. A pixel image is an array of values for each grid point in a rectangular grid inside the window. It may contain covariate data (e.g., taken from a raster grid) or it may be the result of calculations (such as kernel density smoothing function).

4.3.2 The Data

As an example of interpreting point data and its spatial patterns, we analyze plant distribution data collected at the Ordway-Swisher Biological Station, a core NEON (National Ecological Observatory Network) site for the southeastern United States (Kampe et al. 2010; Kao et al. 2012). Prickly pear cactus (*Opuntia humifusa*) is a common plant found in old fields (Fig. 4.4a) and other upland areas that have limited canopy cover. Prickly pear cactus is of considerable interest for three reasons. First, it is a common resource for several insects and vertebrates that use it for foraging and breeding (Sauby et al. 2012; Grunwaldt et al. 2015; Lavelle et al. 2015). Second, some species of *Opuntia* are grown as agricultural crops in some regions of the world (Lopez 1995; Cruz-Rodriguez et al. 2016). Third, *Opuntia* has invaded some areas where it is not native, becoming a problematic species that can dominate an area (e.g., some areas of Australia and South Africa) (Freeman 1992; Novoa et al. 2016).

This system is well-suited for exploring the possibilities of how spatial point pattern analysis can be applied to address applied questions in ecology. We use data of *O. humifusa* locations in a 50×50 m plot (Fig. 4.3b), which were mapped using a high-resolution GPS (~30 cm error) as part of a larger investigation on habiat loss and fragmentation (Fletcher et al. 2018). Each cactus location includes information on the size of cactus, as well as the presence of an insect herbivore, *Chelinidea vittiger*, that specializes on *Opuntia* and has been used as a biological control for invasive *Opuntia* in some regions of the world (DeVol and Goeden 1973). Cactus size can be characterized as a continuous mark, whereas the presence–absence of *C. vittiger* can be characterized as a binary (or qualitative) mark. Our goal is to interpret whether there is any evidence of spatial aggregation of cactus and the biological control agent, along with identifying the spatial scales of pattern.

Libros de Estadística en PDF | Statistics Books in PDF

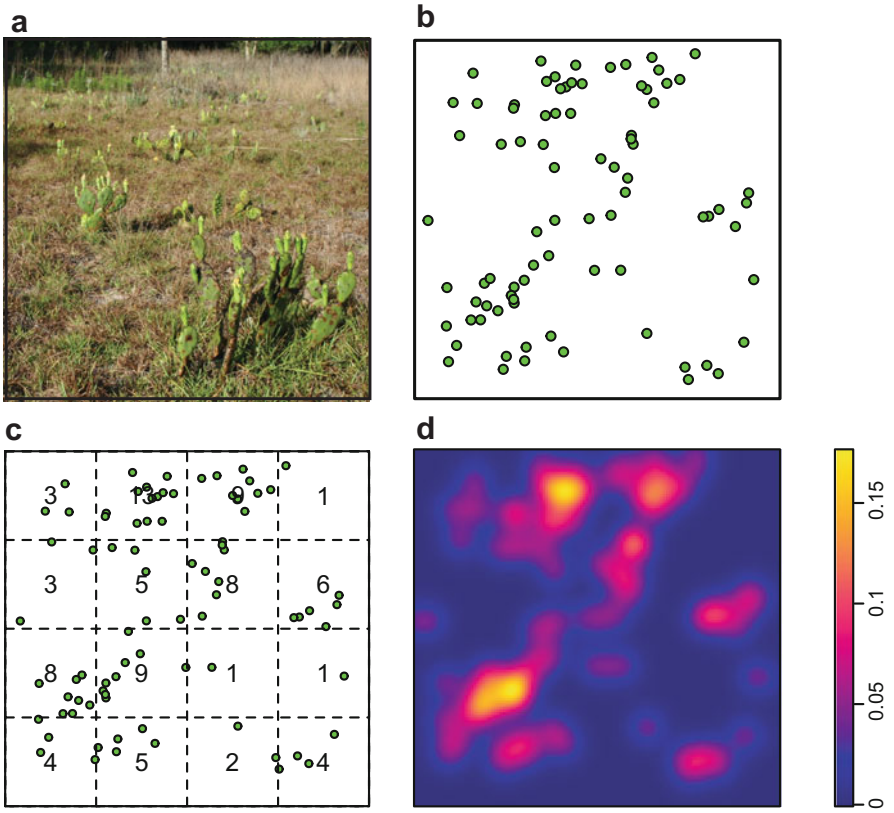


Fig. 4.4 *Opuntia humifusa* at the Ordway-Swisher Biological Station. (**a**) *O. humifusa* shows a patchy distribution. (**b**) The observed point pattern of cactus locations in a 50 × 50 m plot. (**c**) Quadrat counts of *Opuntia humifusa* and (**d**) the intensity mapped using a kernel plot

4.3.3 Creating Point Pattern Data and Visualizing It

First, we load the `spatstat` package and data (`cactus.csv`), and we create relevant `spatstat` objects. Because this study area is a square (50 × 50 m), we can provide `spatstat` with the corners of the study area (`cactus_boundary.csv`) to delineate the window size (e.g., latitude/longitude or UTMs).

```
> library(spatstat)
> cactus <- read.csv('cactus.csv', header = T)
> boundary <- read.csv("cactus_boundary.csv", header = T)
> ppp.window <- owin(xrange = c(boundary$Xmin, boundary$Xmax),
  Yrange = c(boundary$Ymin, boundary$Ymax))

> ppp.cactus <- ppp(cactus$East, cactus$North, window =
  ppp.window)
```

Libros de Estadística en PDF | Statistics Books in PDF

spatstat can also take polygon files (e.g., shp files) for delineating the study area. For example, we can load an .shp file of a polygon of the plot extent with the rgdal package and create a win object from that in a straightforward manner:

```
> library(rgdal)
> boundary.poly <- readOGR("cactus_boundary.shp")
> ppp.window.poly <- as.owin(boundary.poly)
```

Once the data are in spatstat format, there are several exploratory graphs and summary statistics that spatstat can provide. Below are a few examples (Fig. 4.4c).

```
> plot(ppp.cactus) #graph of point locations
> plot(density(ppp.cactus, 1)) #density/intensity plot

> summary(ppp.cactus)

##
Planar point pattern: 82 points
Average intensity 0.0262668 points per square unit

Coordinates are given to 1 decimal place
i.e. rounded to the nearest multiple of 0.1 units

Window: rectangle = [403368, 403424.6] x [3285673, 3285728] units
Window area = 3121.81 square units
```

This summary shows that there are 82 points (cactus patches) and provides the observed intensity, λ. The density plot (Fig. 4.4d) can be a helpful visualization of intensity of points across the plot. By plotting the spatial intensity in this way, we can get an idea of whether or not there may be spatial trends in the point occurrences that may violate the assumption of a homogeneous point process.

We can also make tallies of counts of point locations based on quadrats overlaid on the plot (Fig. 4.4c). To determine whether these quadrat counts conform to CSR (i.e., a homogeneous Poisson process), we can use a simple Chi-square test statistic.

```
> Q <- quadratcount(ppp.cactus, nx = 4, ny = 4)#12.5x12.5m quadrats
> plot(ppp.cactus)
> plot(Q, add = TRUE)
> quadrat.test(ppp.cactus, nx = 4, ny = 4,method = "Chisq")

##
Chi-squared test of CSR using quadrat counts
 Pearson X2 statistic

data: ppp.cactus
X2 = 35.463, df = 15, p-value = 0.004223
alternative hypothesis: two.sided

Quadrats: 4 by 4 grid of tiles
```

Libros de Estadística en PDF | Statistics Books in PDF

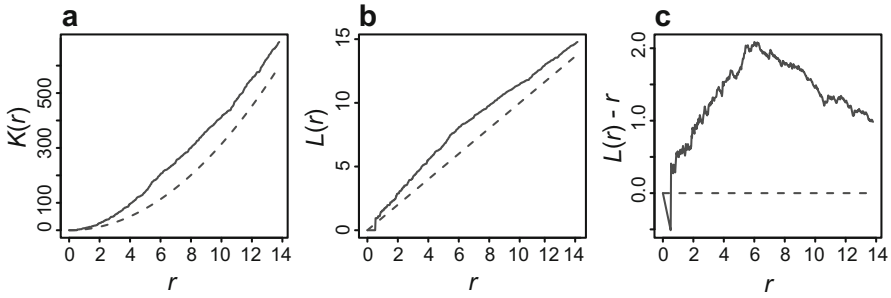


Fig. 4.5 (**a**) Ripley's *K* and the linearized *L* functions, where (**b**) is scaled such that the expectation is *r*, while (**c**) is scaled such that the expectation is zero. For each, the solid line shows the observed value while the dashed line shows the expectation under complete spatial randomness (a Poisson point process). In this figure, there are no corrections for edge effects

This test statistic suggests highly a non-random point pattern at the scale of the quadrat that we defined. Note that this test is more akin to a first-order point pattern analysis because it is based on the dispersion of points among sampling quadrats.

4.3.4 Univariate Point Patterns

Second-order point pattern analyses can readily be implemented in `spatstat`. Below illustrates Ripley's *K* and the standardized *L* functions (Fig. 4.5), initially ignoring edge effects (`correction="none"`).

```
> Knone <- Kest(ppp.cactus, correction = "none")
> plot(Knone)

> Lnone <- Lest(ppp.cactus, correction = "none")
> plot(Lnone) #standardized to a 1:1 expectation
> plot(Lnone, . - r ~ r) #standardized to a zero expectation
```

You will notice that for these functions, two lines are drawn. The "L_{pois}" line is a dashed line that represents the expected (theoretical) value based on a Poisson process (CSR). The way that `spatstat` calculates *L* is to linearize *K* such that the expected value is *r* (or the radius). The other solid line represents the estimated *L* (linearized *K*), when we ignore edge effects.

The above analysis ignores the problem of edge effects. `spatstat` provides a variety of edge corrections. We contrast an isotropic and translate correction for adjusting for boundary effects (Fig. 4.6). The isotropic correction uses a simple weighting scheme for the area sampled near the plot boundary (Ripley 1988), while

Libros de Estadística en PDF | Statistics Books in PDF

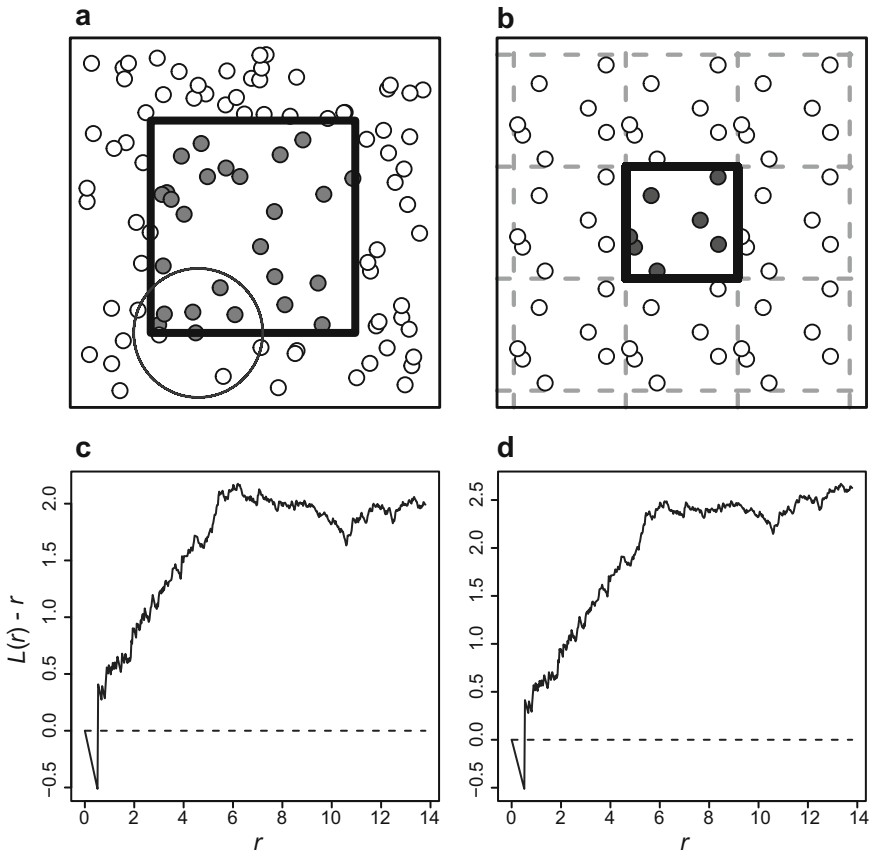


Fig. 4.6 The effect of edge correction on inferences from point patterns. (**a**) To correct for edge effects, weights can be used to offset the number of points observed within a radius inside the plot (gray points), depending on the amount of the radius that falls outside the plot. (**b**) Alternatively, the point pattern can be shifted (e.g., a toroidal shift shown here) to account for boundary corrections. (**c**) Point pattern analysis based on the *L* function that uses a weighting function (isotropic), and (**d**) using the translate correction (toroidal shift). For **c** and **d**, the solid line shows observed value, while the dashed line shows expected value under CSR. Contrast (**c**) and (**d**) to Fig. 4.5c

the translate correction uses a toroidal shift. We adjust for potential boundary effects by typing:

```
> Liso <- Lest(ppp.cactus, correction = "isotropic")
> plot(Liso, . - r~r)

> Ltrans <- Lest(ppp.cactus, correction = "translate")
> plot(Ltrans, . - r ~ r)
```

Libros de Estadística en PDF | Statistics Books in PDF

When comparing the L function that ignores boundaries to those above that account for boundaries, notice that patterns change at larger distances—we expect that the L function at larger distances should potentially be more biased than at smaller distances (because larger radii will naturally overlap more with the boundary of the study area). When we ignore edge effects, we are in effect counting fewer points within the radius r near the boundary, so the observed value for L or K should have an artifact of decreasing as r increases.

The analyses so far are exploratory. While the observed statistics (K, L) appear different than the expectation, it is unclear if these are substantially (or significantly) different. To conduct formal inference regarding if the point pattern follows CSR, we can use Monte Carlo simulations to calculate a confidence envelope under CSR with the `envelope` function. This function can be applied to several point pattern statistics.

```
> Lcsr <- envelope(ppp.cactus, Lest, nsim = 99, rank = 1,
  correction = "translate", global = FALSE)
> plot(Lcsr, . - r ~ r, shade=c("hi", "lo"), legend = F)
```

In the `envelope` function, rank specifies the alpha for the simulations. For a `rank=1`, the max and min are used for envelopes, such that for 99 simulations, $\alpha = 0.01$ while for 19 simulations, $\alpha = 0.05$. Also note that we used `global = FALSE`. This means that these are "pointwise envelopes." These envelopes work better for L than K because of variance stabilizing properties.

Plots of pointwise envelopes show the stated upper and lower quantiles of simulated patterns for any distance r (Fig. 4.7). Because such analyses are calculating envelopes for many distances, pointwise envelopes with a specified α should not be used to reject a null model at that level (because of the multiple tests). Consequently, there are alternative global tests that can be used in this way. While global tests are under active development (Baddeley et al. 2014; Wiegand et al. 2016), `spatstat` does provide one option for a global test (using `global = T` in the

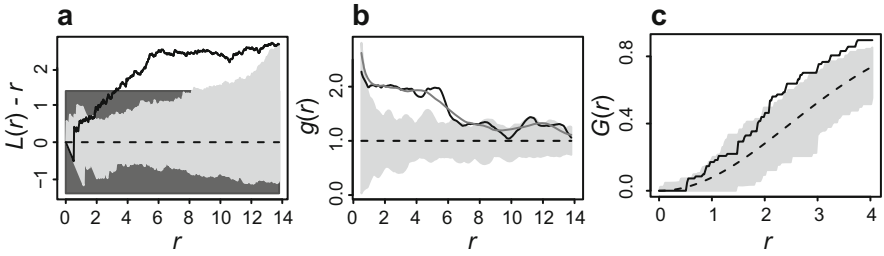


Fig. 4.7 Confidence envelopes for complete spatial randomness of the (**a**) L function, (**b**) pair correlation function, and the (**c**) G-function. For (**a**), both 99% pointwise and global envelopes are shown (global in dark gray). For (**b**), two bandwidths around the radius r are shown (the default, 0.15, in black; 0.4 in gray). For each, the solid line shows observed value while the dashed line represents expectation under complete spatial randomness (a Poisson point process)

Libros de Estadística en PDF | Statistics Books in PDF

above model). This approach estimates the maximum deviation from the Poisson point process across all r (i.e., $D = \max|K(r) - K\text{pois}(r)|$). This approach is referred to as a simultaneous envelope (or critical band) rather than a pointwise envelope (Fig. 4.7a). If the observed line falls outside the simultaneous envelope at any point on r, we would reject the null hypothesis.

Now, say we are more interested in estimating the distance at which spatial patterns arise, such that using a "ring" rather than a circle (as in Ripley's K) is more appropriate. To estimate the pair correlation function, g, most of the arguments are similar to above. The main exception is that instead of calling `Lest`, we call `pcf` (pair correlation function; Fig. 4.7b):

```
> Ptrans <- pcf(ppp.cactus, correction = "translate")
> plot(Ptrans)

> Penv <- envelope(ppp.cactus, pcf, nsim = 99, rank = 1, correction =
  "translate", global = FALSE)
> plot(Penv, shade = c("hi", "lo"), legend = FALSE)
```

The `pcf` function uses a smoothing kernel such that distance bins are not needed. The default bandwidth coefficient (related to sigma in a Gaussian kernel; see Chap. 2) for the smoothing kernel is set to 0.15 (Stoyan and Stoyan 1994). We can adjust the smoothing on the pair correlation function using the `stoyan` command in the `pcf` function. Increasing the value of the bandwidth coefficient (e.g., `stoyan = 0.4`) results in a less wiggly g function (Fig. 4.7b).

Finally, we can use similar arguments for the G-function to estimate the probability of finding a nearest neighbor as a function of distance (Fig. 4.7c). `spatstat` uses a similar approach as above with the `Gest` function. Note that for `Gest`, there are subtly different ways to account for edge effects relative to above. Below we use `rs`, the reduced sample correction. We can check the observed G-function calculated by `spatstat` to the cumulative distribution function of the empirical data (with the `ecdf` function):

```
> Gtrans <- Gest(ppp.cactus, correction = "rs")
> plot(Gtrans, legend = F)

> Genv <- envelope(ppp.cactus, Gest, nsim = 99, rank = 1,
  correction = "rs", global = FALSE)
> plot(Genv, shade = c("hi", "lo"), legend = FALSE)

#nearest neighbor distance for each point
> nn.dist <- nndist(ppp.cactus)
> plot(ecdf(nn.dist), add = T)
```

Note that the radius considered for the G-function is much smaller than for the L-function or the pair correlation function. This makes sense, because the nearest neighbor distances will emphasize the shortest distances between points.

Libros de Estadística en PDF | Statistics Books in PDF

Taken together, the analyses using the *L*, *g*, and *G*-functions provide complementary insights regarding the spatial pattern of *Opuntia*. Using pointwise envelopes, the *L* function suggests an aggregated pattern occurring at scales of approximately 2–13 m, while the pair correlation function suggests that most of the observed effect in the *L* function is generated from shorter distances, on the order of 2–6 m. Similarly, the *G*-function suggests that nearest neighbor distances are random at very small scales (<2 m), while distances are closer than expected at larger scales, consistent with aggregation.

4.3.5 Marked Point Patterns

Many of the above univariate analyses can be extended to ask interesting and important questions based on marked point patterns. First, consider the issue of resource use versus availability in the context of marked point patterns (Lancaster and Downes 2004). There are several insect herbivores that use *Opuntia* cactus. We may be interested in interpreting the spatial dispersion of these herbivores. We want to know the distribution of herbivores, given the underlying distribution of cactus. For this and other complexities, such as interpreting spatial covariance between species, we need to use a marked point pattern analysis.

In this situation, we will interpret the spatial distribution of an insect herbivore, *Chelinidea vittiger*, on cactus. *C. vittiger* is a pest insect that has been used as a biological control for *Opuntia* cactus in Australia (although with limited effectiveness) (DeVol and Goeden 1973). It specializes on *Opuntia*, where it feeds and breeds on cactus segments. It is a poor disperser (Fletcher et al. 2011) and tends to show aggregated distribution patterns on cactus (Miller et al. 2013).

To interpret *C. vittiger* distribution, we can use a randomization procedure where we relabel marks (shuffle the locations of used versus unused cacti, termed "random labeling") to interpret the observed pattern of the herbivore with the `rlabel` function. That is, we are interested in insect dispersion, conditional on cactus distribution. In other circumstances, we might be interested in the joint distribution of two marked processes (e.g., competition between two species). For such situations, we can use the `rshift` function in a similar way as using `rlabel`. Rather than shuffling labels, the `rshift` function performs a toroidal shift of the point pattern of one mark while leaving the other marked point pattern constant.

First, we need to make a new `spatstat` object that includes the marks for presence–absence data. In the data provided, there is information from six surveys conducted at each patch regarding the number of *C. vittiger* detected per patch. Here, we truncate the data to produce a bivariate mark of presence–absence of *C. vittiger*:

```
> cactus$CheliPA <-as.factor(ifelse(cactus$chelinidea > 0,
  "presence", "absence"))
```

Libros de Estadística en PDF | Statistics Books in PDF

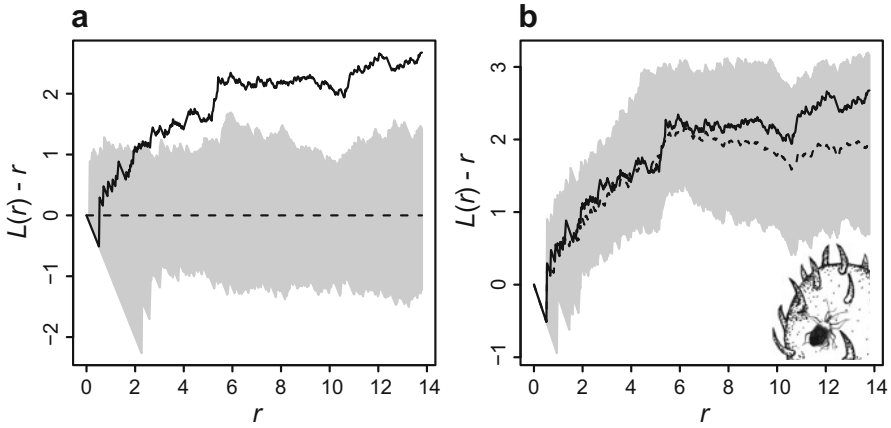


Fig. 4.8 Bivariate *L* function to distinguish spatial use versus availability. (**a**) Spatial point pattern analysis for *Chelinidea vittiger* occurrence use without conditioning on cactus pattern (availability) shows evidence for aggregation. (**b**) Bivariate analysis, conditioning on cactus distribution (using a random labeling procedure), shows that herbivores are distributed randomly. The solid line is the observed value, while the dotted line is the expected value under CSR

With this new variable, `CheliPA`, we can create a new `spatstat` object.

```
> ppp.PA <- ppp(cactus$East, cactus$North, window = ppp.window,
  marks = cactus$CheliPA)
> split(ppp.PA) #summary statistics by mark
> plot(split(ppp.PA)) #separate plots
```

We first interpret the spatial pattern of *C. vittiger* distribution, ignoring the underlying distribution of cactus (Fig. 4.8a).

```
> cheli.data <- subset(cactus, chelinidea > 0)
> ppp.bug <- ppp(cheli.data$East, cheli.data$North, window =
  ppp.window)

> Lbug <- envelope(ppp.bug, Lest, nsim = 99, rank = 1, i =
  "presence", global = F)
```

Then, we contrast these results to those based on a bivariate *K* function (or a bivariate pair correlation function) with a random-labeling simulation to interpret the spatial pattern of marks (Fig. 4.8b).

```
> Lmulti <- envelope(ppp.PA, Lcross, nsim = 99,rank = 1, I =
  "presence", global = FALSE, simulate =
  expression(rlabel(ppp.PA)))
```

Libros de Estadística en PDF | Statistics Books in PDF

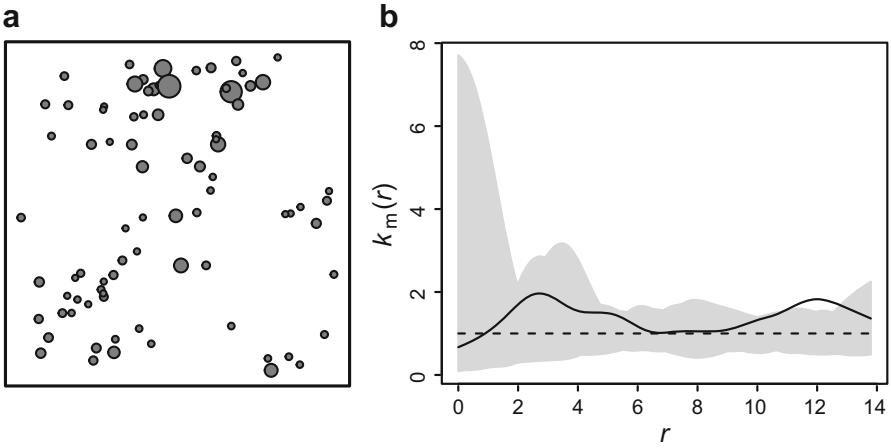


Fig. 4.9 (**a**) Patch area (~plant area) and (**b**) mark correlation analysis of cactus patch area, along with 99% pointwise envelopes. Values <1 and outside the envelope indicate negative relationships of patch area with patches *r* distance away, while values >1 indicate positive relationships

Taken together, if we only did an analysis that ignored the distribution of cactus, we would have very different conclusions regarding the spatial pattern (Fig. 4.8), where we would conclude that *C. vittiger* has an aggregated distribution. However, this pattern is driven by the pattern of cactus, which constrains the distribution of *C. vittiger*.

Finally, we can consider continuous marks through the use of the mark correlation function. For this example, we consider cactus area as a continuous mark. Consequently, we ask whether cactus tends to be aggregated by size or if there is an inhibition process where larger cacti tend to be near smaller ones. This can be done by creating a `spatstat` object where we use cactus size as the quantitative mark and then use the `markcorr` function (Fig. 4.9):

```
> ppp.area <- ppp(cactus$East, cactus$North, windo w=
  ppp.window, marks = cactus$Area)
> mcf.area <- markcorr(ppp.area)
> MCFenv <- envelope(ppp.area, markcorr, nsim = 99, correction =
  "iso", global = FALSE)
> plot(MCFenv, shade = c("hi", "lo"), legend = F)
```

Note that in this case, if the observed value is above 1, there is evidence for a positive mark correlation, such that large cacti tend to be near other large cacti. If the value is <1, large cacti tend to be near smaller cacti. This analysis suggests that there is no strong spatial pattern of cactus size across the plot (i.e., large cacti are not aggregated).

Libros de Estadística en PDF | Statistics Books in PDF

4.3.6 *Inhomogeneous Point Processes and Point Process Models*

Point process models allow for understanding and accounting for inhomogeneous point processes: when point processes vary by location, such as across environmental gradients. An inhomogeneous point process model is similar to a generalized linear model (GLM) for point data (very similar to a Poisson regression; more on GLMs in Chap. 6), where we are modeling the intensity of points in the study area (Renner et al. 2015).

We could use a variety of covariates to account for inhomogeneous point processes. We will consider two types of covariates. First, we start with simply accounting for spatial trend based on *x*–*y* coordinates (see Ch. 6 for more on spatial trend). We can fit different point process models and inspect the model fit. Second, we will import a raster layer that quantifies herbaceous vegetation height in the plot (see Chap. 5 for more on these data and their interpretation). Surrounding vegetation height may be relevant for interpreting *Opuntia* distribution due to light limitation (Hicks and Mauchamp 2000) or indirect effects from variation in herbivory (Burger and Louda 1994). To fit a point process model, we use the `ppm` function.

```
#simple intercept and trend models based on x,y coordinates
> pp.int <- ppm(ppp.cactus, ~ 1) #no trend(homogeneous)
> pp.xy <- ppm(ppp.cactus, ~ x + y) #linear trend
> pp.xy2 <- ppm(ppp.cactus, ~ polynom(x, y, 2)) #quadratic trend
```

Adding *x*-*y* coordinates in a point process model can sometimes cause difficulty for model convergence, such that it may require rescaling coordinates. In the above models, we manually centered the `ppp` objects (window and point coordinates by subtracting the mean of the *x* and *y* coordinates; code not shown) to insure convergence, but the `rescale` function in the `spatstat` package could also be used. To use a raster layer, we must convert the raster (Fig. 4.10a) to a matrix and then an image file that `spatstat` can interpret. We can then fit the model and contrast models with AIC.

```
#model based on covariates from a raster layer
> library(raster)
> veg.height <- raster('cactus_matrix')

#raster into an image covariate that spatstat can read
> veg.height <- data.frame(rasterToPoints(veg.height))
> veg.height <- veg.height[order(veg.height$x, veg.height$y), ]#sort
> veg.height.mat <- matrix(NA, nrow=length(unique(veg.height$x)),
  ncol=length(unique(veg.height$y)))
> veg.height.mat[] <- veg.height$Height
> cov.veg <- im(mat = veg.height.mat,
  Xrange = c(boundary$Xmin, boundary$Xmax),
  Yrange = c(boundary$Ymin, boundary$Ymax))
```

Libros de Estadística en PDF | Statistics Books in PDF

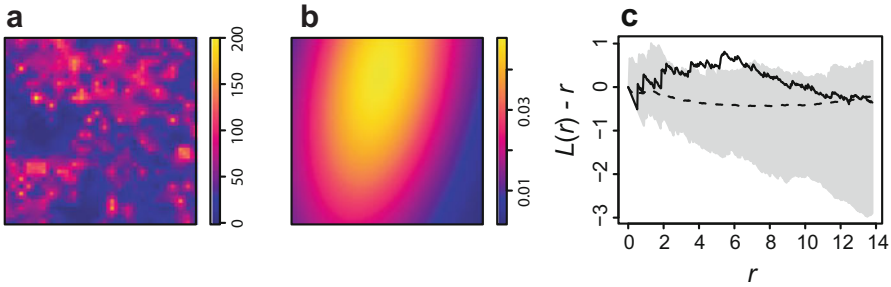


Fig. 4.10 Inhomogeneous *L* function analysis based on spatial trend in points. (**a**) Vegetation height covariate; (**b**) Plot of best-fitting spatial trend based on kernel density function. (**c**) Inhomogeneous *L* function that accounts for spatial trend

```
#point process model based on vegetation covariate
> pp.veg <- ppm(ppp.cactus,~ veg, covariates =
  list(veg=cov.veg))

#model selection with AIC
> data.frame(model = c("int", "xy", "xy^2", "veg"),
  AIC = c(AIC(pp.int), AIC(pp.xy), AIC(pp.xy2), AIC(pp.veg)))

##
model AIC
1 int 762.8697
2 xy 761.0581
3 xy^2 753.3127
4 veg 754.8890
```

Based on the model selection criteria (AIC), there is some evidence of spatial heterogeneity in the intensity of points, suggesting that this could be considered an inhomogeneous point process. The most supported model is one with a quadratic trend, but including the raster image of vegetation height also had some support. We can plot this estimated process with the `predict` function:

```
#plot the point process models
> plot(predict(pp.xy2, type = "trend"))
> plot(ppp.cactus, add = T)
```

We can also adjust for this pattern in the *K* function. To do so, we first need to make an image object of the predicted point process model that `spatstat` can use (similar to a raster map of the covariate of interest). We then use the `Linhom` function to account for this heterogeneity when quantifying spatial dispersion (Fig. 4.10c).

Libros de Estadística en PDF | Statistics Books in PDF

```
> pp.xy.pred <- predict.ppm(pp.xy2, type = "trend")
> Lxycsr <- envelope(ppp.cactus, Linhom, nsim = 99, rank = 1,
  correction = "translate", simulate =
  expression(rpoispp(pp.xy.pred)), global = F)
> plot(Lxycsr, . - r ~ r, shade = c("hi", "lo"), legend = F)
```

Note that the clustering previously observed is largely (but not entirely) accounted for when considering this inhomogeneity. This general approach can be used to control for factors that may influence patterns to ask whether clustering or uniform distributions occur after controlling for these effects.

4.3.7 Alternative Null Models

While CSR is a useful starting point as a null model, in some situations we may be interested in using alternative null models. Some null models can be derived from a Poisson cluster process. Two common Poisson cluster processes considered in ecology are Matérn cluster processes and Thomas cluster processes (Velazquez et al. 2016). In a Matérn cluster process, there are two types of points. The first are "parent" points, which have a Poisson distribution. Second, for each parent point, there are "offspring" points, which are independently and uniformly distributed around the parent points within a radius r (Fig. 4.11a). Consequently, these "offspring" points generate an underlying aggregated pattern. Similarly, with a Thomas process, "offspring" points are generated with parents but with an isotropic Gaussian distribution (similar to a Gaussian kernel described in Chap. 2). Such a process could reflect biological phenomena such as seed dispersal from parent plants.

We can use these alternative null models in `spatstat`, with the above functions (K, L, pair correlation g, etc.). For example, a K function with a Thomas process as a null model can be quantified as:

```
> Kthomas <- kppm(ppp.cactus, ~ 1, "Thomas")
```

To interpret this model, `summary(Kthomas)` provides a wealth of information regarding the fitted model. For example, it provides an estimate of the mean cluster size (4.6 points), as well as the best fit scale for that size (3.7 m). Note that in the above `kppm` function, we can also account for covariates (the "~1" states to not consider covariates and only include an intercept in the model; `~polynom(x, y, 2)` would account for the trend shown above).

We can use the `envelope` function here as well to interpret the point pattern.

```
> Kthomas.env <- envelope(Kthomas, Lest, nsim = 99, rank = 1,
  global = F)
> plot(Kthomas.env, . - r ~ r, shade=c("hi", "lo"), legend = F)
```

Libros de Estadística en PDF | Statistics Books in PDF

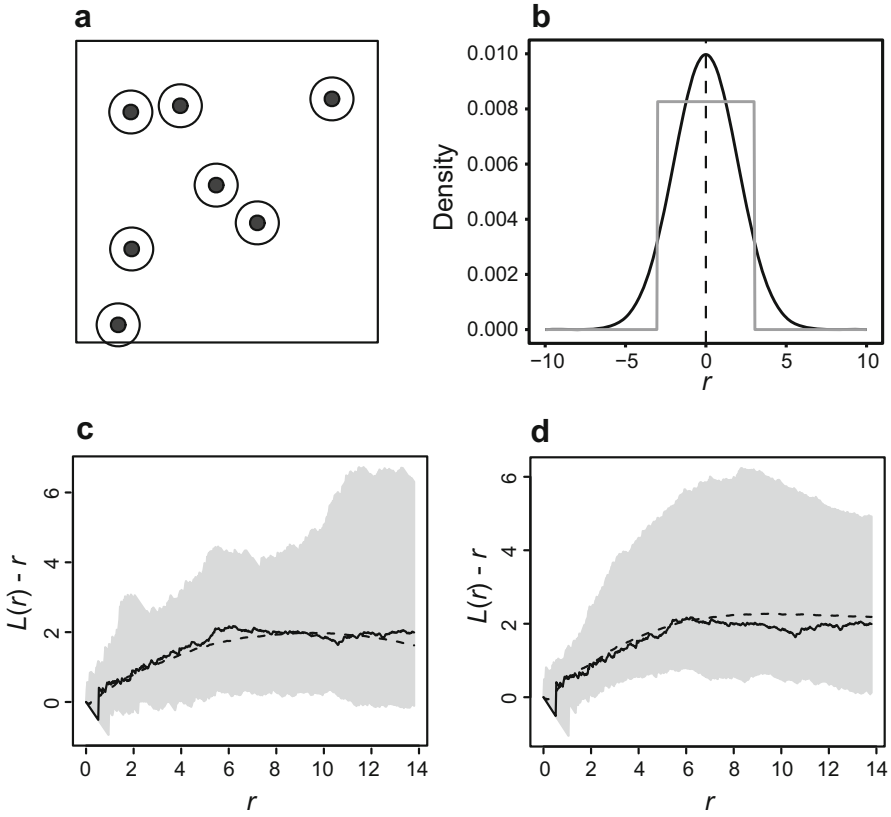


Fig. 4.11 Alternative models that assume a clustered point process. (**a**) Most models start with assuming that "parent" points come from an underlying Poisson point process, with "offspring" points that are clustered within a radius or buffer from the parent points. Shown is a Poisson process with a buffer radius in which offspring points may cluster. (**b**) Clustering of offspring points can occur based on different underlying distributions. Shown are cross-sections of bivariate kernels for a Matérn cluster process and a Thomas process, where a Matérn assumes that offspring points are clustered around parents based on a uniform distribution (gray line), while a Thomas process assumes clustering of offspring based on an isotropic normal distribution (black line). (**c**, **d**) Resulting *L* functions based on these processes fit to the cactus data, where the solid line is the observed value, while the dashed line is the expected values under the (**c**) Thomas and (**d**) Matérn processes

These models suggest that the observed point pattern is not substantially different from what would be expected from a Thomas process (Fig. 4.11d) or a Matérn process (Fig. 4.11c). This result is consistent with the previous results, which illustrated the aggregative pattern.

Libros de Estadística en PDF | Statistics Books in PDF

4.3.8 Simulating Point Processes

It is frequently useful to be able to simulate patterns. Simulations provide a means to create a known "truth" that we can then use to evaluate potential hypotheses or models that might be of interest to spatial ecology and conservation. Indeed, simulations are increasingly used in ecology to interpret the accuracy and sensitivity of models (Kery and Royle 2016). We will consider the problem of simulating spatial data throughout this book.

We can simulate point processes. In fact, for some problems mentioned above, we were simulating data when using Monte Carlo simulations to infer significance of point patterns. To illustrate, we can simulate a homogeneous point process with the same intensity, λ, as in our empirical data, using the `rpoispp` function:

```
> sim.pp <- rpoispp(lambda=intensity(ppp.cactus), nsim = 4,
  win=ppp.window)

#access the x-y coordinates for points in simulation 1
> sim.pp[[1]]$x
> sim.pp[[1]]$y
```

In this case, we passed λ of the empirical data using the command `intensity`, and requested four simulations (realizations) based on a Poisson point process with the observed intensity. This function can be adjusted for inhomogeneous point processes by passing a functional relationship that describes the inhomogeneous process on the x–y plane, rather than a mean intensity:

```
#make a function based on ppm model coefficients
> pp2.fun <- function(x, y) {exp(pp.xy2$coef[1] + pp.xy2$coef[2]
  * x + pp.xy2$coef[3] * y + pp.xy2$coef[4] * I(x^2) +
  pp.xy2$coef[5] * x * y + pp.xy2$coef[6] * I(y^2))

#or using expectation from xy^2 model
> pp2.sim <- rpoispp(pp.xy.pred, nsim = 4)

#simulate inhomogeneous point process from ppm
> pp2.sim <- rpoispp(pp2.fun, nsim = 4, win = ppp.window)
```

Finally, we note that other processes can be simulated, such as Matérn and Thomas processes mentioned above, using a similar approach.

Libros de Estadística en PDF | Statistics Books in PDF

4.4 Next Steps and Advanced Issues

4.4.1 Space-Time Analysis

Space-time analysis of point patterns is an active and developing area of research (Cressie and Wikle 2011). In the simplest case, one could consider time of an event (e.g., Julian date of predation on a nest) as a continuous mark to interpret whether events are spatially clustered in time (i.e., predation occurs locally nearby in time, such as when a predator alters movement paths upon patch discovery to be more sinusoidal). In that situation, using a mark correlation function may provide some insight.

More broadly, the above models for inhomogeneous K and g functions have been extended to space-time point processes (Gabriel and Diggle 2009). These functions can be used to identify spatiotemporal regularity or aggregation of point processes. While `spatstat` provides some functionality for space-time analyses, for more on space-time point pattern analysis, including simulating spatiotemporal data, see the `stpp` package (Gabriel et al. 2013).

4.4.2 Replicated Point Patterns

The problems and approaches illustrated in this chapter reflect situations with one landscape, region or plot. In some situations, we may have replicated plots and there may be interest in making general conclusions regarding these replicated point patterns.

There are at least two ecological scenarios where multiple plots may occur (Diggle 2003). In the first, we may have several sites where point patterns are observed. If we can assume that points could potentially disperse across sites, one way to approach this is to consider all sites simultaneously within our study `ppp` window, with areas between sites masked out of the window. With this `ppp` window, analysis can proceed in a way similar to the analyses described above. A second scenario occurs when we have replicated sites where we do not expect dispersal between them and that these sites can be considered replicates of a treatment (e.g., in an experiment), a land-use condition, etc. In this case, we may want to make conclusions across replicates. Using replicated plots can be beneficial for several reasons, including understanding sparse point patterns (e.g., point patterns of rare species) and spatial variation in point processes (Bagchi and Illian 2015). This scenario has been commonly approached by separately analyzing each site and then combining summary statistics across sites. For a detailed overview, see Wiegand and Moloney (2014).

`spatstat` has recently been extended to accommodate this latter scenario, where data from replicate plots are stored as a `hyperframe` (a generalization of the data frame object used in R). Some of the above functions, such as `ppm`, have

Libros de Estadística en PDF | Statistics Books in PDF

been extended to replicate point patterns. This is an area of active development (Baddeley et al. 2015; Bagchi and Illian 2015).

4.5 Conclusions

The analysis of spatial point patterns has rapidly emerged as a frontier in spatial ecology. There have been several developments in this field over the past 15 years. With this increase in methods, there are now a variety of ways to approach the problem of spatial point pattern analysis (Wiegand and Moloney 2014; Velazquez et al. 2016).

Several issues should be considered when conducting spatial point pattern analysis (Velazquez et al. 2016). First, it is often useful to consider more than one second-order statistic. Each of the statistics we considered (K, L, g, G) captures different elements of spatial point patterns, such that they provide complementary information. In most situations, analyses should adjust for edge (boundary) effects. The potential for inhomogeneous point patterns needs to also be considered in most situations. Finally, the use of pointwise envelopes should be treated carefully and the use of global envelopes should be considered (Baddeley et al. 2014; Wiegand et al. 2016).

While much of the ecological work on spatial point pattern analysis has focused on patterns of plant distributions, there are also several opportunities to apply these approaches to answer critical questions in animal ecology (e.g., spatial patterns of nest predation) and conservation biology (e.g., the spatial scales of effects of point source pollution). We expect that the scope of spatial point pattern analysis will continue to increase in the coming years to help address ecological and conservation problems.

References

Addicott JF, Aho JM, Antolin MF, Padilla DK, Richardson JS, Soluk DA (1987) Ecological neighborhoods: scaling environmental patterns. Oikos 49(3):340–346. https://doi.org/10.2307/3565770

Adler FR, Nuernberger B (1994) Persistence in patchy irregular landscapes. Theor Popul Biol 45 (1):41–75

Andersen M (1992) Spatial analysis of two species interactions. Oecologia 91(1):134–140

Baddeley A (2007) Spatial point processes and their applications. In: Baddeley A, Barany I, Schneider R, Weil W (eds) Stochastic geometry, Lecture notes in mathematics, vol 1892. Springer, Berlin, pp 1–75

Baddeley A, Turner R (2005) spatstat: an R package for analyzing spatial point patterns. J Stat Softw 12(6):1–42

Baddeley A, Diggle PJ, Hardegen A, Lawrence T, Milne RK, Nair G (2014) On tests of spatial pattern based on simulation envelopes. Ecol Monogr 84(3):477–489. https://doi.org/10.1890/13-2042.1

Libros de Estadística en PDF | Statistics Books in PDF

Baddeley A, Rubak E, Turner R (2015) Spatial point patterns: methodology and applications with R. CRC Press, Boca Raton, FL

Bagchi R, Illian JB (2015) A method for analysing replicated point patterns in ecology. Methods Ecol Evol 6(4):482–490. https://doi.org/10.1111/2041-210x.12335

Bayard TS, Elphick CS (2010) Using spatial point-pattern assessment to understand the social and environmental mechanisms that drive avian habitat selection. Auk 127(3):485–494. https://doi.org/10.1525/auk.2010.09089

Brown C, Law R, Illian JB, Burslem D (2011) Linking ecological processes with spatial and non-spatial patterns in plant communities. J Ecol 99(6):1402–1414. https://doi.org/10.1111/j.1365-2745.2011.01877.x

Burger JC, Louda SM (1994) Indirect versus direct effects of grasses on growth of a cactus (opuntia-fragilis) - insect herbivory versus competition. Oecologia 99(1–2):79–87. https://doi.org/10.1007/bf00317086

Clark PJ, Evans FC (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology 35:445–453

Condit R, Ashton PS, Baker P, Bunyavejchewin S, Gunatilleke S, Gunatilleke N, Hubbell SP, Foster RB, Itoh A, LaFrankie JV, Lee HS, Losos E, Manokaran N, Sukumar R, Yamakura T (2000) Spatial patterns in the distribution of tropical tree species. Science 288 (5470):1414–1418. https://doi.org/10.1126/science.288.5470.1414

Cressie N, Wikle CK (2011) Statistics for spatio-temporal data. Wiley, Chichester

Cruz-Rodriguez JA, Gonzalez-Machorro E, Gonzalez AAV, Ramirez MLR, Lara FM (2016) Autonomous biological control of Dactylopius opuntiae (Hemiptera: Dactyliiopidae) in a prickly pear plantation with ecological management. Environ Entomol 45(3):642–648. https://doi.org/10.1093/ee/nvw023

Dale MRT, Fortin MJ (2014) Spatial analysis: a guide for ecologists, 2nd edn. Cambridge University Press, Cambridge

Deckers B, Verheyen K, Hermy M, Muys B (2005) Effects of landscape structure on the invasive spread of black cherry Prunus serotina in an agricultural landscape in Flanders, Belgium. Ecography 28(1):99–109. https://doi.org/10.1111/j.0906-7590.2005.04054.x

DeVol JE, Goeden RD (1973) Biology of Chelinidea vittiger with notes on its host-plant relationships and value in biological weed control. Environ Entomol 2:231–240

Diggle PJ (2003) Statistical analysis of spatial point patterns, 2nd edn. Arnold Press, London

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29 (2):129–151

Fithian W, Elith J, Hastie T, Keith DA (2015) Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods Ecol Evol 6(4):424–438. https://doi.org/10.1111/2041-210x.12242

Fletcher RJ, Reichert BE, Holmes K (2018) The negative effects of habitat fragmentation operate at the scale of dispersal. Ecology 99(10):2176–2186

Fletcher RJ Jr, Acevedo MA, Reichert BE, Pias KE, Kitchens WM (2011) Social network models predict movement and connectivity in ecological landscapes. Proc Natl Acad Sci U S A 108:19282–19287

Freeman DB (1992) Prickly pear menace in eastern Australia 1880–1940. Geogr Rev 82 (4):413–429. https://doi.org/10.2307/215199

Gabriel E, Diggle PJ (2009) Second-order analysis of inhomogeneous spatio-temporal point process data. Statistica Neerlandica 63(1):43–51. https://doi.org/10.1111/j.1467-9574.2008.00407.x

Gabriel E, Rowlingson B, Diggle PJ (2013) stpp: an R package for plotting, simulating and analyzing spatio-temporal point patterns. J Stat Softw 53(2):1–29

Libros de Estadística en PDF | Statistics Books in PDF

Grunwaldt JM, Guevara JC, Grunwaldt EG (2015) Review of scientific and technical bibliography on the use of Opuntia spp. as forage and its animal validation. J Prof Assoc Cactus Dev 17:13–32

Gustafson EJ (1998) Quantifying landscape spatial pattern: what is the state of the art? Ecosystems 1(2):143–156. https://doi.org/10.1007/s100219900011

Hicks DJ, Mauchamp A (2000) Population structure and growth patterns of Opuntia echios var. gigantea along an elevational gradient in the Galapagos Islands. Biotropica 32(2):235–243. https://doi.org/10.1111/j.1744-7429.2000.tb00466.x

Illian J, Penttinen A, Stoyan H, Stoyan D (2008) Statistical analysis and modelling of spatial point patterns. Wiley, Chichester

Illian JB, Martino S, Sorbye SH, Gallego-Fernandez JB, Zunzunegui M, Esquivias MP, Travis JMJ (2013) Fitting complex ecological point process models with integrated nested Laplace approximation. Methods Ecol Evol 4(4):305–315. https://doi.org/10.1111/2041-210x.12017

Ives AR, May RM (1985) Competition within and between species in a patchy environment: relations between microscopic and macroscopic models. J Theor Biol 115(1):65–92

Kallimanis AS, Kunin WE, Halley JM, Sgardelis SP (2005) Metapopulation extinction risk under spatially autocorrelated disturbance. Conserv Biol 19(2):534–546

Kampe TU, Johnson BR, Kuester M, Keller M (2010) NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. J Appl Remote Sens 4. https://doi.org/10.1117/1.3361375

Kao RH, Gibson CM, Gallery RE, Meier CL, Barnett DT, Docherty KM, Blevins KK, Travers PD, Azuaje E, Springer YP, Thibault KM, McKenzie VJ, Keller M, Alves LF, Hinckley ELS, Parnell J, Schimel D (2012) NEON terrestrial field observations: designing continental-scale, standardized sampling. Ecosphere 3(12):1–17. https://doi.org/10.1890/es12-00196.1

Kelly M, Meentemeyer RK (2002) Landscape dynamics of the spread of sudden oak death. Photogramm Eng Remote Sens 68(10):1001–1009

Kery M, Royle JA (2016) Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS. Academic, San Diego

Lancaster J, Downes BJ (2004) Spatial point pattern analysis of available and exploited resources. Ecography 27(1):94–102. https://doi.org/10.1111/j.0906-7590.2004.03694.x

Lancaster J, Downes BJ, Reich P (2003) Linking landscape patterns of resource distribution with models of aggregation in ovipositing stream insects. J Anim Ecol 72(6):969–978. https://doi.org/10.1046/j.1365-2656.2003.00764.x

Lara-Romero C, de la Cruz M, Escribano-Avila G, Garcia-Fernandez A, Iriondo JM (2016) What causes conspecific plant aggregation? Disentangling the role of dispersal, habitat heterogeneity and plant-plant interactions. Oikos 125(9):1304–1313. https://doi.org/10.1111/oik.03099

Lavelle MJ, Blass CR, Fischer JW, Hygnstrom SE, Hewitt DG, VerCauteren KC (2015) Food habits of adult male white-tailed deer determined by camera collars. Wildl Soc Bull 39 (3):651–657. https://doi.org/10.1002/wsb.556

Lloyd M (1967) Mean crowding. J Anim Ecol 36(1):1–30. https://doi.org/10.2307/3012

Lopez AD (1995) Review: Use of the fruits and stems of the prickly pear cactus (Opuntia spp) into human food. Food Sci Technol Int 1(2–3):65–74

Maheu-Giroux M, de Blois S (2007) Landscape ecology of Phragmites australis invasion in networks of linear wetlands. Landsc Ecol 22(2):285–301. https://doi.org/10.1007/s10980-006-9024-z

Manly BFJ (2006) Randomization, bootstrap and Monte Carlo methods in biology, 3rd edn. CRC Press, Boca Raton, FL

Melles SJ, Badzinski D, Fortin MJ, Csillag F, Lindsay K (2009) Disentangling habitat and social drivers of nesting patterns in songbirds. Landsc Ecol 24(4):519–531. https://doi.org/10.1007/s10980-009-9329-9

Miller CW, Fletcher RJ Jr, Gillespie SR (2013) Conspecific and heterospecific cues override resource quality to influence offspring production. PLoS One 8(7):e70268. https://doi.org/10.1371/journal.pone.0070268

Libros de Estadística en PDF | Statistics Books in PDF

Novoa A, Kaplan H, Wilson JRU, Richardson DM (2016) Resolving a prickly situation: involving stakeholders in invasive cactus management in South Africa. Environ Manag 57(5):998–1008. https://doi.org/10.1007/s00267-015-0645-3

Ovaskainen O, Sato K, Bascompte J, Hanski I (2002) Metapopulation models for extinction threshold in spatially correlated landscapes. J Theor Biol 215(1):95–108. https://doi.org/10.1006/jtbi.2001.2502

Penttinen A, Stoyan D, Henttonen HM (1992) Marked point processes in forest statistics. For Sci 38(4):806–824

Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl 19(1):181–197. https://doi.org/10.1890/07-2153.1

Renner IW, Elith J, Baddeley A, Fithian W, Hastie T, Phillips SJ, Popovic G, Warton DI (2015) Point process models for presence-only analysis. Methods Ecol Evol 6(4):366–379. https://doi.org/10.1111/2041-210x.12352

Ripley BD (1976) Second-order analysis of stationary point processes. J Appl Probab 13(2):255–266. https://doi.org/10.2307/3212829

Ripley BD (1979) Tests of randomness for spatial point patterns. J R Stat Soc Series B Methodol 41(3):368–374

Ripley BD (1988) Statistical inference for spatial processes. Cambridge University Press, Cambridge

Rodriguez-Perez J, Wiegand T, Traveset A (2012) Adult proximity and frugivore's activity structure the spatial pattern in an endangered plant. Funct Ecol 26(5):1221–1229. https://doi.org/10.1111/j.1365-2435.2012.02044.x

Sauby KE, Marsico TD, Ervin GN, Brooks CP (2012) The role of host identify in determining the distribution of the invasive moth Cactoblastis cactorum (Lepidoptera: Pyralidae) in Florida. Fla Entomol 95(3):561–568

Schooley RL, Wiens JA (2001) Dispersion of kangaroo rat mounds at multiple scales in New Mexico, USA. Landsc Ecol 16(3):267–277. https://doi.org/10.1023/a:1011122218548

Seabloom EW, Dobson AP, Stoms DM (2002) Extinction rates under nonrandom patterns of habitat loss. Proc Natl Acad Sci U S A 99(17):11229–11234. https://doi.org/10.1073/pnas.162064899

Stoyan D, Stoyan H (1994) Fractals, random shapes and point fields: Methods of -geometrical statistics. New York: Wiley

Velazquez E, Martinez I, Getzin S, Moloney KA, Wiegand T (2016) An evaluation of the state of spatial point pattern analysis in ecology. Ecography 39(11):1042–1055. https://doi.org/10.1111/ecog.01579

Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York

Warton DI, Shepherd LC (2010) Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. Ann Appl Stat 4(3):1383–1402. https://doi.org/10.1214/10-aoas331

Wiegand T, Moloney KA (2004) Rings, circles, and null-models for point pattern analysis in ecology. Oikos 104(2):209–229. https://doi.org/10.1111/j.0030-1299.2004.12497.x

Wiegand T, Moloney K (2014) Handbook of spatial point-pattern analysis in ecology. Chapman & Hall, CRC Applied Environmental Statistics, Boca Raton, FL

Wiegand T, Martinez I, Huth A (2009) Recruitment in tropical tree species: revealing complex spatial patterns. Am Nat 174(4):E106–E140. https://doi.org/10.1086/605368

Wiegand T, He F, Hubbell SP (2013) A systematic comparison of summary characteristics for quantifying point patterns in ecology. Ecography 36(1):92–103. https://doi.org/10.1111/j.1600-0587.2012.07361.x

Wiegand T, Grabarnik P, Stoyan D (2016) Envelope tests for spatial point patterns with and without simulation. Ecosphere 7(6). https://doi.org/10.1002/ecs2.1365

Yang J, He HS, Shifley SR (2008) Spatial controls of occurrence and spread of wildfires in the Missouri Ozark Highlands. Ecol Appl 18(5):1212–1225. https://doi.org/10.1890/07-0825.1

Libros de Estadística en PDF | Statistics Books in PDF

Chapter 5
Spatial Dependence and Autocorrelation



5.1 Introduction

Spatial patterns are omnipresent in both environmental and ecological data (Wagner and Fortin 2005). In Chap. 4, we focused on point patterns to characterize the degree and spatial scale of aggregation or regular dispersion based on the *x*–*y* coordinates of point events (e.g., tree locations, nest locations). In Chap. 5, we are interested in interpreting spatial patterns of quantitative measures taken from sampling the environment.

The spatial analysis of such continuous variables falls under the realm of *spatial statistics*, traditionally aimed at quantifying spatial pattern and its statistical significance, and *geostatistics*, traditionally aimed at quantifying spatial variance and using this information to spatially interpolate data (Oliver and Webster 1991; Cressie 1993; Haining 2003; Dale and Fortin 2014). Spatial statistics can identify the spatial scale of patterns (i.e., the characteristic scale(s); Chap. 2). These approaches largely focus on quantifying and interpreting *spatial dependence*, or the similarity of a variable as a function of spatial location and/or geographic distance.

Tobler's first law of geography lies at the foundation of these issues, where "everything is related to everything else, but near things are more related than distant things" (Tobler 1970). To operationalize this fundamental idea, spatial statistics and geostatistics estimate spatial variance or covariance/correlation as a function of geographic distance. Tobler's law implies that at close distances, covariance or correlation of measurements should be high, yet as distances increase, the covariance or correlation should decline.

Spatial dependence can suggest key processes driving ecological patterns as a function of scale, such as spatial interactions among organisms as well as organism responses to environmental gradients that contain spatial dependence (Wagner and Fortin 2005) (Fig. 5.1). Yet spatial dependence can also be a nuisance for statistical inferences because data containing spatial dependence do not fulfill the common

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_5

Libros de Estadística en PDF | Statistics Books in PDF

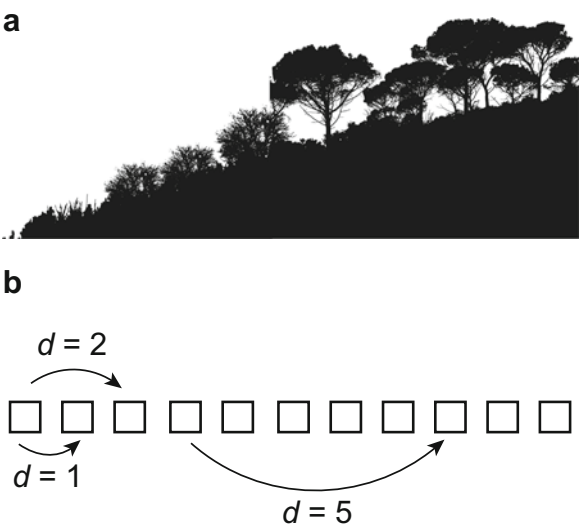


Fig. 5.1 The problem of spatial dependence. (**a**) When considering the environment across space, environmental measures such as elevation or canopy cover tend to be more similar at close locations and similarity declines with distance. (**b**) When sampling these gradients, we can interpret this spatial dependence based on lag distances between measurement locations. Shown are three lag distances ($d = 1, 3, 5$)

statistical assumption of independence assumed in many traditional statistics (Legendre 1993).

Here, our goals are to provide the key concepts needed to: (1) estimate the degree of spatial autocorrelation in data and potential spatial scale of the pattern; (2) understand how the estimated spatial variance can be used to interpolate and simulate spatial patterns using kriging and related approaches; and (3) identify the characteristic spatial scale(s) in the data using multiscale analysis (e.g., wavelet and spectral decomposition). We illustrate these concepts by extending our example from Chap. 4 on *Opuntia* cactus.

5.2 Key Concepts and Approaches

5.2.1 The Causes of Spatial Dependence

The terms *spatial dependence* and *spatial autocorrelation* are often used interchangeably, yet each term has a subtle different meaning based on why similarity of measurements in space occur (Table 5.1). To understand these differences, it is useful to make the distinction of whether spatial pattern is driven by *endogeneous* or *exogeneous* mechanisms (Bolker 2003). Endogeneous mechanisms are those that directly occur from the organism or processes being considered, which result in spatial pattern. Some common examples include localized dispersal generating spatial clustering of organisms, or social or grouping behavior (e.g., schooling fish or herds of ungulates). Exogeneous mechanisms, in contrast, are those that occur outside of the organism or process being measured, such as spatial aggregation of

Libros de Estadística en PDF | Statistics Books in PDF

Table 5.1 Common terms for spatial dependence in ecology

Term	Description
Anisotropy	When data have properties that are direction-dependent. Contrast to isotropy.
Correlogram	A plot of autocorrelation as a function of lag distance.
Endogenous process	A process that directly arises from the organism or response variable being considered, which results in patterns of spatial dependence.
Exogeneous process	A process that arises from outside of the organism or response variable being measured, such as spatial aggregation of resources or environmental gradients used by the organism of interest. Sometimes referred to as induced spatial dependence.
Isotropy	When data are uniform in all directions, which is frequently assumed in the analysis of spatial dependence.
Kriging	A method of interpolation for which the interpolated values are modeled via spatial covariance functions derived from variograms.
Scalogram	A plot of the wavelet variance as a function of a scaling factor related to distance.
Stationarity	When spatial pattern does not change over space or time (i.e., there is no trend in spatial dependence), which is frequently assumed in the analysis of spatial dependence.
Spatial autocorrelation	In a narrow sense, spatial dependence that arises from endogeneous processes.
Spatial dependence	Similarity in a response variable as a function of spatial location/distance, which can be driven by endogeneous or exogeneous processes.
Variogram	A plot of the spatial covariance as a function of lag distance. Different quantities are sometimes plotted, with the most common being semivariance.

resources or environmental gradients used by the organism of interest, which is sometimes referred to as "indirect" mechanisms and *induced spatial dependence* (Peres-Neto and Legendre 2010). In this context, spatial dependence is often considered a broad term for statistical spatial covariance that can be driven by both exogenous and endogenous processes. Note that this has also been referred to as *spatial legacy* (Peres-Neto and Legendre 2010). In contrast, spatial autocorrelation is sometimes considered as a certain type of spatial dependence that is driven by endogenous processes alone (Dale and Fortin 2014).

5.2.2 Why Spatial Dependence Matters

Given that spatial dependence is widespread in nature, why might we care? There are several answers to this question. First, there are practical reasons: when spatial dependence occurs, sampling locations within the range of dependence are no longer independent from each other. This issue is particularly troublesome, given that many common statistical tests assume that samples are independent. For instance, in a linear regression model, we often write:

Libros de Estadística en PDF | Statistics Books in PDF

$$y_i = \alpha + \beta x_i + \varepsilon_i, \tag{5.1}$$

where y is our response variable at location i, α is the intercept, β is the deterministic slope of the relationship of x with y, and ε is the error term. Where is independence assumed? In the error term of the model, we assume that errors are normally distributed with a mean of zero and a variance, σ^2, which is assumed to be iid—independent and identically distributed. This assumption means that each residual i (i.e., the difference between the observed and predicted value for i) is not dependent on other residuals and each comes from the same underlying distribution (see Chap. 6 for more details). Consequently, the problem of spatial dependence arises in our assumptions of the error in the model.

What is the problem if we violate this assumption? When spatial dependence occurs in our data and we ignore it, it often leads to type I error, where we infer significant patterns in the data that may in fact not exist. This is in contrast to a type II error, where we fail to conclude a significant pattern occurs when in fact exists. The issue of type I error arises because we are implicitly assuming that we have a larger sample size (and thus greater degrees of freedom) than we actually do, sometimes referred to a pseudo-replication (Hurlbert 1984). This assumption leads to artificially small estimates of uncertainty (or artificially high precision), such as standard errors (SEs) or confidence intervals (CIs) for parameter estimates, such as the SE of β in Eq. (5.1). Spatial dependence is thus thought to primarily bias our interpretation of the precision, not point estimates (e.g., we might adequately estimate β but not the SE or CI of β). Consequently, accounting for spatial dependence in statistical models, such as linear regression, may be necessary in some cases (see Chap. 6 for examples on how to do so). Alternatively, by identifying the scale(s) at which spatial dependence occurs, we may better design investigations to minimize problems of spatial dependence (Oliver and Webster 1991), such as spacing sampling locations at distances (i.e., lag distances) greater than the expected range of spatial dependence in the data.

The second reason why we might care about spatial dependence is that describing spatial dependence in our data may provide insights toward understanding key biological processes that generate the spatial patterns we are observing. For instance, when spatial dependence arises, is this pattern revealing the scale of social behavior, environmental variation in key resources, or dispersal (Brown et al. 1995; Koenig 1998; Fletcher and Sieving 2010; Cohen et al. 2016)? While quantifying spatial dependence alone may not provide rigorous answers to such questions, it may generate hypotheses or further predictions to help isolate the causes of spatial dependence.

Finally, spatial dependence can alter conclusions regarding conservation threats for many species and conservation strategies (Carroll and Pearson 2000; Landeiro and Magnusson 2011; Yoo and Ready 2016). For example, Koenig and Liebhold (2016) illustrated that there has been increasing spatial synchrony (one form of spatial dependence; see Chap. 10) in wintering birds across North America with warming temperatures over a 50 year time period. They emphasized that such

Libros de Estadística en PDF | Statistics Books in PDF

synchrony may have detrimental effects of population persistence through a reduction in demographic rescue (i.e., when dispersal reduces the probability of extinction of local populations).

5.2.3 *Quantifying Spatial Dependence*

There is a variety of ways to quantify spatial dependence. Here, we focus on the use of correlograms and semivariograms, which are complementary approaches frequently used in ecology and spatial statistics.

5.2.3.1 Correlograms

To understand how spatial statistics estimate spatial autocorrelation, it is useful to recall formulas for correlations, variances, and covariances. The spatial statistics we present emerge clearly from these classical statistics.

Recall the formula for a simple Pearson linear correlation, r, for two variables, z_1 and z_2:

$$r(z_1, z_2) = \frac{\sum_{i=1}^{n} (z_{1i} - \bar{z}_1)(z_{2i} - \bar{z}_2)}{\sqrt{\sum_{i=1}^{n} (z_{1i} - \bar{z}_1)^2 \sum_{i=1}^{n} (z_{2i} - \bar{z}_2)^2}} = \frac{\mathrm{Cov}(z_1, z_2)}{\sqrt{\mathrm{Var}(z_1)\mathrm{Var}(z_2)}}, \tag{5.2}$$

where $r(z_1, z_2)$ ranges from -1 to 1. The key is to extend this idea over space.

Moran's I test statistic extends the standard Pearson correlation over space (increasing lag distances) to estimate the degree of spatial autocorrelation for a quantitative variable, z, as:

$$I = \frac{n}{\mathbf{W}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{n} (z_i - \bar{z})^2}, \tag{5.3}$$

where $\mathbf{W}$ is a weight matrix that describes the dependency between locations i and j. Typically, this is a neighborhood indicator matrix, where $w_{ij} = 1$ if i and j are adjacent and 0 otherwise. Note this matrix is often row standardized, such that $\sum_j w_{ij} = 1$. This statistic can also be calculated for different distance categories, or bins, to interpret spatial dependence as a function of distance as:

Libros de Estadística en PDF | Statistics Books in PDF

$$I(d) = \frac{n}{\mathbf{W}(\mathbf{d})} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(d)\left(z_i - \bar{z}\right)\left(z_j - \bar{z}\right)}{\sum_{i=1}^{n} \left(z_i - \bar{z}\right)^2}. \tag{5.4}$$

Notice how similar Moran's I and Pearson correlation coefficients are: in essence Moran's $I(d)$ is a Pearson's coefficient computed for one variable against itself according to increasing distances among sampling locations (d) (Fig. 5.1). The plot of $I(d)$ as a function of distance class is called a spatial *correlogram*: its shape helps to interpret how the spatial pattern varies with distance and to estimate the spatial scale of the pattern. When z is normally distributed and that there are enough pairs of sampling locations per distance class (usually more than 20 pairs), the $I(d)$ will vary between +1 (where positive values indicate positive spatial autocorrelation) and -1 (where negative values indicate negative spatial autocorrelation), while values close to 0 indicates the absence of spatial pattern. Thus, Moran's I behaves as a Pearson correlation coefficient, and it is frequently used by ecologists because of its intuitive interpretation. Yet, when $I(d)$ is computed with less than 20 pairs, its value can be greater than 1 or smaller than -1. To avoid this known "boundary" or "edge effect" (Chap. 4), correlograms are often computed only up ½ or ⅔ of the maximum distance between the sampling locations to ensure adequate sample size for each distance bin (Dale and Fortin 2014). Note that Eq. (5.3) provides a common, global test for spatial dependence, while Eq. (5.4) is typically only used for the generation of correlograms. We focus on the use of correlograms because they provide much richer and intuitive information regarding spatial dependence.

Moran's I is an isotropic (i.e., pooled in all directions) averaged value of spatial autocorrelation per distance class for the entire extent of a study area. To detect the potential for *anisotropy* (i.e., spatial autocorrelation that varies in different directions) in the spatial pattern, the estimation of spatial autocorrelation can be computed using both distance and angle classes (i.e., different directions).

As Moran's I is a dimensionless number, it can be compared across different variables. One limitation of Moran's I is that it is sensitive to outliers (e.g., one or a few points can generate significant, erroneous autocorrelation). This is why some researchers transform the data (e.g., log-transformation of the response variable) to reduce the impacts of outliers. Because of this sensitivity, a similar statistic, Geary's c, has been developed. Geary's c values range from 0 (positive spatial autocorrelation) to 2 (negative spatial autocorrelation) and 1 indicates the absence of spatial autocorrelation. Yet Geary's c is also somewhat sensitive to outliers. As Geary's c is in essence the standardized equivalent of the semivariance presented below, we will not focus on Geary's here (but see Dale and Fortin 2014 for details).

Significance for each Moran's I coefficient can be based on Monte Carlo randomizations or through normal approximations. If significance is assessed using normal approximations then the assumption of *stationarity* needs to be valid. Stationarity is a term that describes a situation where the process that generated the spatial pattern does not vary in across a study area (e.g., mean and variance are similar throughout the region of interest) (Haining 2003). As the same data are used to compute $I(d)$ at increasing distances, the $I(d)$ values are not independent. This is

Libros de Estadística en PDF | Statistics Books in PDF

the same statistical issue encountered previously regarding spatial point pattern analysis (Chap. 4), which requires the use of multiple comparison corrections. Therefore, a Bonferroni adjustment correction (or something similar) should be applied that accounts for the number of distance classes computed, k, to adjust the significance level (Brunsdon and Comber 2015). For instance, using a Bonferroni correction, for a $I(d)$ to be statistically significant, its probability needs to be smaller or equal to $0.05/k$ (e.g., for $k = 15$, the adjusted probability to be significant based on a Bonferroni correction is $0.05/25 = 0.003$).

5.2.3.2 Variograms

Geostatistics comes at the same goal of estimating spatial dependency through a slightly different means (Cressie 1993). Instead of starting with a correlation coefficient (i.e., standardized covariance) such as the Moran's I, geostatistics stem from the sample variance and covariance instead:

$$\mathrm{Var}(z) = \frac{1}{n-1}\sum_{i=1}^{n}\left(z_i - \bar{z}\right)^2, \tag{5.5}$$

$$\mathrm{Cov}(z_1, z_2) = \frac{1}{n-1}\sum_{i=1}^{n}\left(z_{1,i} - \bar{z}_1\right)\left(z_{2,i} - \bar{z}_2\right). \tag{5.6}$$

The semivariance, γ, is calculated as:

$$\gamma(d) = \frac{1}{2n(d)}\sum_{i}^{n(d)}\left[\left(z(x_i) - z(x_i + d)\right)\right]^2, \tag{5.7}$$

where z is the value of the variable at location x_i, and $n(d)$ is the number of pairs of sampling locations at distance class d. Note the similarities with the variance equation.

The term "semi" comes from the fact that we divide by 2 (it helps to stabilize the statistical properties of the metric). Again, plotting γ as a function of d produces a semivariogram, often simply referred to as a *variogram*. Note that semivariance is on the same units as the data (e.g., km). Unlike Moran's I, but like variance, $\gamma(d) \geq 0$ and there is no upper bound. For interpreting the shape of the semivariogram, small values (closest to 0) indicate strong spatial covariance (i.e., strong spatial pattern), whereas larger values indicate less spatial covariance (i.e., weak or no spatial pattern). It is a rule of thumb to only interpret $^2/_3$ of the total distance (extent) considered, similar to that for Moran's I (Cressie 1993; Dale and Fortin 2014); for larger distances, the $n(d)$ is typically too small for reliable inference.

Semivariance computed from observed data are called "empirical," "experimental," or "observed" variograms. Empirical variograms simply plot the semivariances

Libros de Estadística en PDF | Statistics Books in PDF

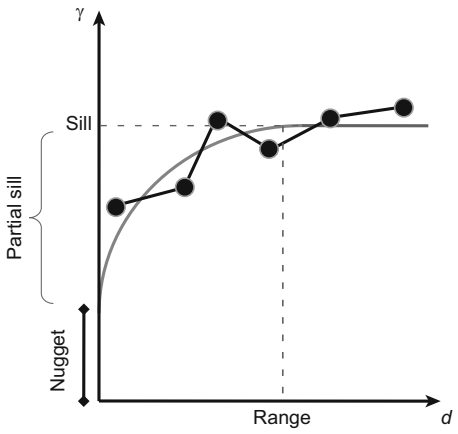


Fig. 5.2 Empirical and theoretical variograms, including the parameters of the variogram model. The empirical variogram shown as black points/line, while the theoretical variogram shown as a curved, asymptotic gray line

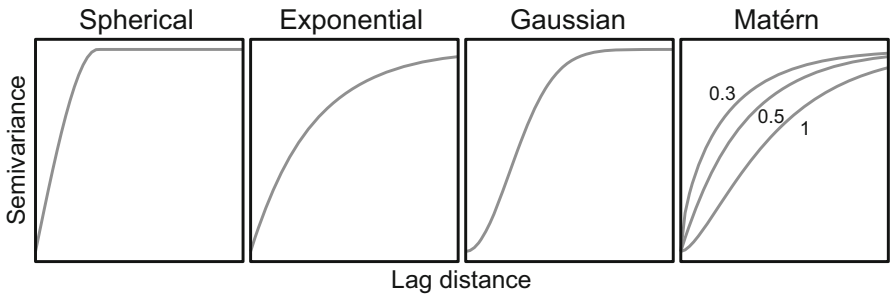


Fig. 5.3 Some common theoretical variogram models. For the Matérn variogram, different levels of kappa, a parameter in this model, are shown (note that when kappa = 0.5, the Matérn variogram becomes equivalent to the exponential variogram)

as a function of distance class, *d* (the spatial lag distance). Theoretical (or model-based) variograms can be fitted to an empirical variogram to spatially interpolate data at unsampled locations and to formally estimate the spatial scale of the pattern. In the presence of spatial pattern, three parameters relevant to interpreting the semivariance can be estimated from the fitted theoretical variogram: the range, sill, and nugget (Fig. 5.2). The *nugget* is the *y*-intercept at the origin that is greater than zero. It explains the variability in the data that occurs at very short distances, which could occur from measurement errors, sampling bias, or other random factors. The *range* indicates the distance up which the spatial dependence occurs, such that beyond the range, the data are no longer spatially autocorrelated. The *sill* is the value of semivariance beyond the estimated range, i.e., the variability that cannot be attributed to spatial autocorrelation. Note that some theoretical models assume that there is no sill (e.g., an exponential model; Fig. 5.3) while others assume that there is no nugget (i.e., the intercept = 0) (see Dale and Fortin 2014). If our interest is in spatial interpolation (which historically was the goal of variogram analysis and

Libros de Estadística en PDF | Statistics Books in PDF

geostatistics), we need to estimate parameters of model-based semivariograms and determine their relative fit to the data, using, for example, model-selection approaches (Burnham and Anderson 1998).

5.2.3.3 Kriging

To spatially interpolate across a region (e.g., make a predictive map of the response variable) based on the degree of spatial dependence, kriging is often used. *Kriging* is essentially a weighted moving average technique that uses estimates from a semivariogram (range, nugget, and sill) to perform spatial interpolation. More specifically, it is a set of linear regressions that determine the best combination of weights to interpolate across a region of interest by minimizing the variance from the spatial covariance in the data, where weights are derived from the estimates from the variogram (Dale and Fortin 2014; Oliver and Webster 2014).

The general form of the kriging model can be described as (Brunsdon and Comber 2015):

$$z = f(x_i) + v(x_i) + \varepsilon_i, \tag{5.8}$$

where $f(x_i)$ is a deterministic trend function (e.g., the response is non-stationary and may change with latitude or longitude), $v(x_i)$ describes the spatial dependence based on variogram parameters, and ε_i is the error. When there is no deterministic trend, *ordinary kriging* is used to interpolate based solely on the variogram parameters. In contrast, *universal kriging* assumes a large-scale, deterministic trend in the data, $f(x_i)$ (non-stationarity). This component is sometimes referred to as trend-surface analysis, which will be discussed in Chap. 6. Mathematical details about the different types of kriging algorithms can be found in Cressie (1993) and Haining (2003). Oliver and Webster (2014) provided a useful, practical tutorial on kriging.

In general, kriging is preferred for spatial interpolation in contrast to other simpler approaches. For example, a common, intuitive approach is inverse distance weighting (IDW) interpolation. IDW interpolates based on estimates that provide greater weight from nearby locations rather than distant ones. However, unlike kriging, this approach does not provide an objective means to determine the magnitude of distance-based weighting or the extent (maximum distance/limiting radius) for weighting. IDW also cannot provide SEs or other measures of uncertainty for predictions. Kriging, in contrast, has been shown to provide the best linear unbiased prediction for unsampled locations and can provide SEs for predictions. Reliable use of kriging requires proper estimation of the variogram model (Oliver and Webster 2014).

Libros de Estadística en PDF | Statistics Books in PDF

5.2.3.4 Some Extensions

For binary data, semivariance can be calculated with indicator functions by replacing $z(x_i)$ in Eq. (5.7) with an indicator function (Rossi et al. 1992). Monte Carlo randomizations are typically used to infer significance in such situations.

Both Moran's $I(d)$ and the semivariance γ function can be extended to address spatial correlations between two variables, termed "cross-correlograms" and "cross-semivariograms," respectively (Goovaerts 1994; Wackernagel 2003). For instance, a cross-variogram between variables u and v can be defined as:

$$\gamma_{uv}(d) = \frac{1}{2n(d)} \sum_{i}^{n(d)} \left[\left(z_u(x_i) - z_u(x_i + d) \right) \right] \left[\left(z_v(x_i) - z_v(x_i + d) \right) \right], \tag{5.9}$$

Semivariance, like Moran's I, is a "global" statistic. These models have also been extended to estimate variation in local intensity of spatial dependencies, referred to as LISA, or Local Indicators of Spatial Association (Anselin 1995; Boots 2002). These local measures are sometimes used to identify hotspots of intensity across landscapes (Nelson and Boots 2008).

5.2.3.5 Statistical Nuisance

Finally, spatial dependence can often be more of a statistical nuisance issue for ecological and conservation problems. In these cases, we might want to adjust conventional analyses (e.g., linear regression) to deal with dependence. This can be done in a variety of ways (Keitt et al. 2002; Beale et al. 2010). For example, conventional generalized linear models (GLMs) have been extended to adjust for spatial dependence by directly modeling the covariance of the residuals (generalized least square regression, GLS). Another common practice is to assess whether or not spatial dependence remains after a conventional analysis. This is frequently done by calculating Moran's I on the residuals from a model (Dormann et al. 2007). If there is evidence for autocorrelation of the residuals, then the conventional analysis should be replaced with one that formally adjusts for spatial dependence. We will consider these approaches in detail in Chap. 6.

5.3 Examples in R

5.3.1 Packages in R

There are a few libraries to choose from for quantifying spatial dependence. We will focus on using `geoR` (Ribeiro and Diggle 2016), `spdep` (Bivand and Piras 2015), `gstat` (Pebesma 2004), `pgirmess` (Giraudoux 2018), and `ncf` (Bjørnstad and Falck 2001). `Spatial` (Venables and Ripley 2002) allows for limited geostatistical

Libros de Estadística en PDF | Statistics Books in PDF

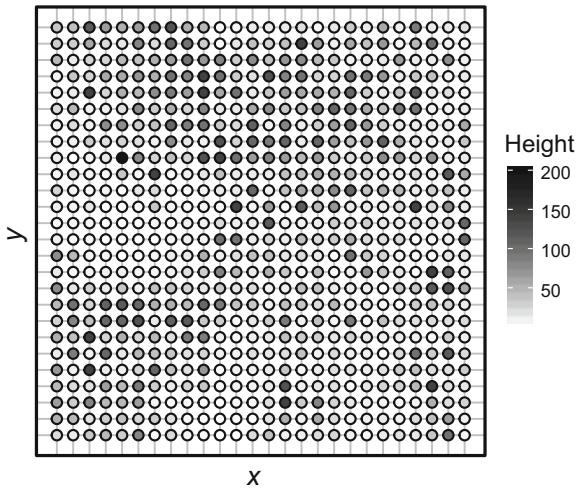


Fig. 5.4 A map of vegetation height measurements (cm) taken every 2 m on the plot

analysis (empirical correlograms and variograms), which comes with the VR Bundle when installing R. `spdep` has more options for correlograms and other spatial features (Bivand 2006). `geoR` provides a model-based variogram analysis based on maximum likelihood, while `gstat` has several geostatistics features, including the use of cross-variograms. We will also use the `ncf` package, which can fit spline (smoothed/non-parametric) correlograms and can provide a bootstrap approach for assessing statistical significance. We will implement kriging in with `geoR` and `gstat`.

5.3.2 The Data

As an example of interpreting spatial dependence, we return to the system considered in Chap. 4: old fields and prickly pear cactus (*Opuntia humifusa*) at the Ordway-Swisher Biological Station. In Chap. 4, we focused on data of *O. humifusa* locations in a 50 × 50 m plot, which were mapped using a high-resolution GPS (~30 cm error). Here, we focus on data from the surrounding matrix: samples of vegetation height taken systematically through the plot across a grid of sampling points spaced 2-m apart (Fig. 5.4) as part of a larger study on habitat loss and fragmentation (Fletcher et al. 2018). This information is relevant to movement of a pest insect considered in Chap. 4, *Chelinidea vittiger* (Schooley and Wiens 2004; Fletcher et al. 2014; Acevedo and Fletcher 2017) and we can use these measurements to first interpret spatial dependence of vegetation in the matrix and then create a map of vegetation height (via kriging) for understanding connectivity between *O. humifusa* patches (connectivity is covered in Chap. 9).

Our goals are to first interpret spatial dependence with the use of Moran's *I* and correlograms. We then use variograms to interpret the scale of spatial dependence and illustrate how model-based variograms can be used in kriging. Next, we

Libros de Estadística en PDF | Statistics Books in PDF

illustrate how kriging-related approaches can be used for generating spatial maps, similar in concept to what we illustrated in Chap. 3 with neutral landscape models. We end by introducing approaches to interpret multiscale spatial dependence.

5.3.3 *Correlograms*

We first import and visualize our data (`'cactus_matrix.csv'`). We will use a couple of different packages for calculating Moran's *I* and correlograms, contrasting what each can provide. We contrast these options because each uses different methods to infer statistical significance of potential spatial dependence, each varies in the complexity of coding required, and each can be helpful under different circumstances.

```
# load the matrix data into R:
> matrix <- read.csv(cactus_matrix.csv', header = T)
> head(matrix, 3)

##
 x y Height
1 0 0 35
2 0 2 65
3 0 4 75
```

With the data loaded, we can plot the data in several ways to interpret it. For example, we plot variation in vegetation height (`Height`) based on *x*–*y* coordinates, using a gray scale (with 12 breaks using the `cut` function) to fill points (using `pch=21`, which allows the fill of points to differ) to visualize variation in the matrix (Fig. 5.4).

```
> plot(matrix[,"y"] ~ matrix[,"x"],
  pch =21, bg = gray.colors(12)[cut(matrix[,3], breaks = 12)])
```

In correlogram (and variogram) analyses, we should truncate the range of lag distances at which we consider spatial dependence to approximately ½ to ⅔ the total distance observed. We can determine this distance by creating a pairwise distance matrix from the sampling locations. Because of the small spatial scale at which this plot occurs, we do not need to worry about projections for this calculation.

```
#calculate a distance matrix
> coords <- cbind(matrix$x, matrix$y)
> colnames(coords) <- c("x", "y")
> distmat <- as.matrix(dist(coords))

#maximum distance to consider in correlogram/variogram
> maxdist <- 2/3 * max(distmat)
```

Libros de Estadística en PDF | Statistics Books in PDF

To interpret spatial dependence with Moran's *I*, we start with the simplest approach and package and work up to less simple, but more flexible approaches. The first is the `pgirmess` package, which is a wrapper package for the `spdep` package. The `spdep` package has several useful spatial analysis functions, but it is less user-friendly than some other common spatial packages. The `pgirmess` package is more user-friendly in this way (but less flexible). We will use `pgirmess` and then contrast it to the `ncf` and `spdep` packages for generating different types of correlograms. In this package, we use the `correlog` function, specifying the coordinates for each sample and the measurement (i.e., height). We also specify that we want to use `method = "Moran"` (this package can also calculate Geary's *c*), the number of distance classes to consider, and we ask for the test to be two-sided (i.e., testing for both the potential of positive and negative spatial dependence).

```
> library(pgirmess)

#correlog from pgirmess
> correlog.pgirmess <- correlog(coords, matrix$Height, method =
  "Moran", nbclass = 14, alternative = "two.sided")

#summary
> head(round(correlog.pgirmess, 2))

##
 dist.class coef p.value n
[1,] 4.45 0.19 0.00 21692
[2,] 9.36 0.08 0.00 37708
[3,] 14.27 −0.01 0.22 51132
[4,] 19.18 −0.04 0.00 55500
[5,] 24.09 −0.02 0.00 61012
[6,] 28.99 −0.01 0.12 58540
```

In the above code, we find that the `correlog` function creates a matrix that contains each distance class considered (with `dist.class` reflecting the center of each bin), the Moran coefficient for that distance, the *p*-value, and the sample size (number of pairs of locations used) for that distance. This package uses normal approximations to test for the significance of spatial autocorrelation (i.e., it assumes the response variable is normally distributed and uses asymptotic theory to derive *p*-values). This approximation can be fast and relatively easy to implement, but it makes some key assumptions (e.g., normality in the residuals of the response data). We can then plot the correlogram

```
#correlogram plot
> plot(correlog.pgirmess)
> abline(h = 0)
```

The plot provides a visualization of Moran's *I* as a function of distance, with distances of significant spatial dependence shown in red (Fig. 5.5a is a generalized plot that contrasts this approach with those described below). This analysis suggests that positive spatial dependence is significant out to approximately 10 m, with some

Libros de Estadística en PDF | Statistics Books in PDF

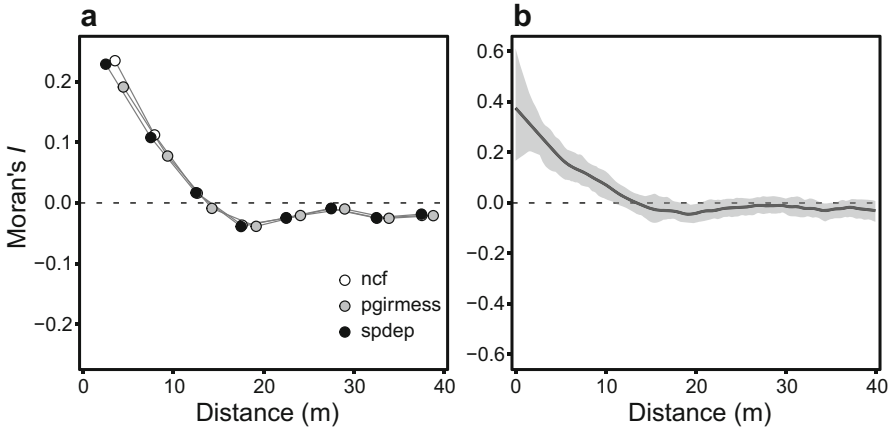


Fig. 5.5 (**a**) Contrasting distance-binned correlograms and (**b**) a spline correlogram. For (**a**), three approaches taken from different R packages are shown. For (**b**), the bootstrap confidence envelopes are shown. `Pgrimess` uses normal approximations to infer significance, `ncf` uses Monto Carlo permutations, while `spdep` is the most flexible package and can use either normal approximations or permutations to infer significance

evidence of negative spatial dependence at moderate distances. Note that this approach defaults to calculating Moran's *I* for distances up to the maximum distance in the observed data; however, we should ignore distances beyond ½ to ⅔ of the maximum distance.

An alternative approach is to use the `ncf` package. This package can provide non-parametric tests of significance for correlograms. It can also provide spline correlograms (Bjørnstad and Falck 2001). In spline correlograms, Moran's *I* is estimated with cubic splines that provide a smooth relationship across a variable of interest (more on this in Chap. 6) such that binning of distances is not required. This aspect is one benefit of using this package. This package has two approaches for interpreting potential significance of the correlogram. The first is a bootstrap approach (Efron 1979) to generate pointwise confidence intervals for the correlogram, such that evidence for spatial dependence is inferred when the confidence intervals do not overlap zero. Bootstrapping is a resampling technique used for inferring uncertainty in sample estimates and/or statistical significance in data. Bootstrapping involves resampling the data with replacement many times, where for each sample the variable of interest is calculated (in this case, Moran's *I*). The distribution of values of the estimate can then be used to approximate confidence intervals. A second approach `ncf` uses is the use of Monte Carlo permutations to generate a null envelope for spatial dependence, analogous to what we used in Chap. 4 for point patterns.

To use the `ncf` package, we need to either detach the `pgrimess` package or call the relevant function in `ncf` differently. This is because one of the functions we will use in the `ncf` package, `correlog`, has the same name as the one used above for `pgrimess`. If we do not want to detach `pgirmess` we can call the function from `ncf` as `ncf::correlog`.

Libros de Estadística en PDF | Statistics Books in PDF

```
> library(ncf)

#correlogram with Monte Carlo test
> correlog.ncf <- ncf::correlog(x = matrix$x, y = matrix$y, z =
  matrix$Height, increment = 5, resamp = 99)
> plot(correlog.ncf)
> abline(h = 0)
```

With this approach, we find similar evidence for spatial dependence based on the Monte Carlo permutations (Fig. 5.5a). However, in this case we may interpret that positive spatial dependence occurs at slightly greater distances than observed when using normal approximations with `pgrimess`. Note that in this function, the entire distance range is also considered although we should ignore distances beyond ½ to ⅔ of the maximum distance for inferences.

We can contrast these results with the use of spline correlograms with the `spline.correlog` function. We request a bootstrapping approach to infer significance in this situation.

```
#spline correlogram with 95% pointwise bootstrap CIs
> spline.corr <- spline.correlog(x = matrix$x, y = matrix$y, z =
  matrix$Height, xmax = maxdist, resamp = 100, type = "boot")

#plot with point-wise 95% CIs from bootstrap
> plot (spline.corr)
```

This correlogram suggests that the slight negative spatial dependence identified with the above approaches at moderate distances is too weak to infer a statistical pattern (i.e., the bootstrap confidence intervals overlap zero) (Fig. 5.5b).

Finally, we illustrate using the `spdep` package for correlograms. This package provides greater flexibility in the development of correlograms than other packages. For example, we can generate indicator correlograms for binary (0, 1) response data (see above). With this package, we will first calculate a general Moran's *I* that is sometimes used as an overall test of spatial dependence in data (Bivand et al. 2013). We then show how to create a customized correlogram using a similar approach.

To interpret spatial dependence with the `spdep` package, we must manually create the spatial weights matrix, **W**, shown in Eqs. (5.3) and (5.4). Note that `spdep` actually stores **W** in a list format rather than matrix format, because in many cases the former can be more compact and take up less storage computationally. Calculating **W** can be accomplished using the `knearneigh`, the `dnearneigh`, or the `cell2nb` functions. Here, we use the `dnearneigh` function, which creates a list, where each element is a vector for the neighbor IDs for each sample. Neighbors are identified based on distances specified with `dnearneigh`. The `knearneigh` finds the *k* closest neighbors, which could vary in distance with some sampling designs, while the `cell2nb` identifies data on regular grids, like that used here, but it is less generalizable so we do not focus on this function. Below we specify `d1 = 0` (minimum distance) and `d2 = 3` (maximum distance), which with these data results in an eight-neighbor function (i.e., queen's rule) (Fig. 5.6).

Libros de Estadística en PDF | Statistics Books in PDF

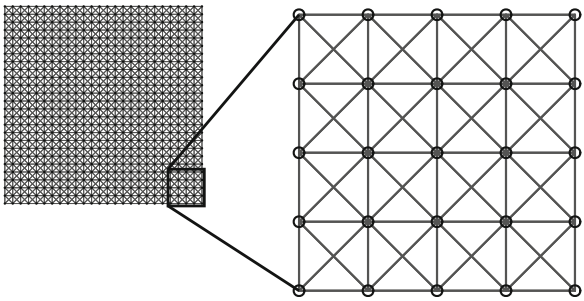


Fig. 5.6 Neighborhood matrix used for calculating Moran's I from $d = 0$–3. This matrix identifies neighbors based on an eight-neighbor rule (Queen's rule; see Chap. 3)

```
> library(spdep)

#make a neighborhood list:
> neigh <- dnearneigh(x = coords, d1 = 0, d2 = 3, longlat = F)

#plot the neighborhood
> plot(neigh, coordinates(coords))
```

To then calculate Moran's I, we convert the `neigh` object to a spatial weights list. As part of this procedure, we specify `style='W'`, which means that we will create a row-standardized **W**:

```
> wts <- nb2listw(neighbours = neigh, style = 'W', zero.policy = T)
```

With these spatial weights, we can now calculate Moran's I. `spdep` allows for inferring significance through normal approximations using `moran.test` (similar to `pgirmess`) or through Monte Carlo permutations using `moran.mc` (similar to `ncf`):

```
> mor.mc <- moran.mc(x = matrix$Height, listw = wts,
  nsim = 999, zero.policy = T)
> mor.norm <- moran.test(x = matrix$Height, listw = wts,
  randomisation = F, zero.policy = T)
> mor.mc

##
 Monte-Carlo simulation of Moran I

data: matrix$Height
weights: wts
number of simulations + 1: 1000

statistic = 0.27366, observed rank = 1000, p-value = 0.001
alternative hypothesis: greater

> mor.norm
```

Libros de Estadística en PDF | Statistics Books in PDF

```
##
 Moran I test under normality

data: matrix$Height
weights: wts

Moran I statistic standard deviate = 13.819, p-value < 2.2e-16
alternative hypothesis: greater
sample estimates:
Moran I statistic Expectation Variance
 0.2736595356 −0.0014814815 0.0003964261
```

In this case, both approaches yield identical estimates of Moran's *I* (0.274) and both provide a global test suggesting that spatial dependence is statistically significant.

Now we take the above approach calculate Moran's *I* for specific lag distance categories, generate a permutation value for each category, and then put together for a correlogram. We first create a data frame for storing the output and then provide a `for` loop that repeats the above process for each lag distance.

```
#first, create a df for storing data
> correlog.sp <- data.frame(dist = seq(5, 0.5 * max(distmat), by
  = 5), MoransI = NA, Null.LCL = NA, Null.UCL = NA, Pvalue = NA)

#Calculate Moran's I for lag distances
> for (i in 1:nrow(correlog.sp)){

 d.start <- correlog.sp[i, "dist"] − 5
 d.end <- correlog.sp[i, "dist"]
 neigh <- dnearneigh(x = coords, d1 = d.start, d2 = d.end,
 longlat = F)
 wts <- nb2listw(neighbours = neigh, style = 'W', zero.policy
 = T)
 mor.i <- moran.mc(x = matrix$Height, listw = wts, nsim = 99,
zero.policy = T)

 #summarize results from spdep
 correlog.sp[i, "dist"] <- (d.end + d.start)/2
 correlog.sp[i, "MoransI"] <- mor.i$statistic
 correlog.sp[i, "Null.LCL"] <- quantile(mor.i$res, p = 0.025)
 correlog.sp[i, "Null.UCL"] <- quantile(mor.i$res, p = 0.975)
 correlog.sp[i, "Pvalue"] <- mor.i$p.value
 }
> plot(y = correlog.sp$MoransI, x = correlog.sp$dist)
> abline(h = 0)
> lines(correlog.sp$dist, correlog.sp$Null.LCL, col = "red")
> lines(correlog.sp$dist, correlog.sp$Null.UCL, col = "red")
```

We have now seen several ways to calculate correlograms, each of which provides different benefits and limitations. In this case, the correlograms showed

Libros de Estadística en PDF | Statistics Books in PDF

generally similar patterns (Fig. 5.6). The use of normal approximations to interpret the significance of spatial autocorrelation (`pgirmess` and `spdep` packages) can be helpful with large data sets, where Monte Carlo tests can prove computationally expensive. However, Monte Carlo tests can be helpful when data are not normally distributed. The `ncf` package provides a means to not resort to binning of lag distances, which can be helpful and provides a straightforward bootstrapping procedure to infer significance. The `spdep` package allows great flexibility for calculating correlograms, but is less user-friendly.

5.3.4 Variograms

To illustrate empirical and model-based semivariograms, we use both the `geoR` and `gstat` packages. We primarily focus on the `geoR` package because it enables likelihood-based comparisons (e.g., AIC) between model-based variograms, which is useful for identifying the best variogram model for inferences and interpolation (Oliver and Webster 2014), and it provides an interesting Monte Carlo approach. The `gstat` package provides more options for different types of model-based variograms and can calculate cross-variograms so we briefly illustrate its use as well. We first create a `geoR` object that consists of the *x*–*y* coordinates and the value at each coordinate, which in this case is vegetation height. We refer to the measurements at sampling locations as *z*.

```
#load packages
> library(geoR)
> library(gstat)

#create a geoR object
> geo.veg <- as.geodata(matrix)
```

The `geoR` package provides a useful scheme for visualizing the raw data: `plot(geo.veg)` provides a four-panel plot. The first panel shows the sampling locations, where the measurements, *z* (vegetation height in this example), are shown as a color ramp, with low values being blue and high values red. The second and third panels show *z* values as a function of *x* and *y* coordinates. These panels can help for visually interpreting whether there is potential anisotropy in the data (directionality or trend in *z* as a function of *x*–*y* locations). The final panel provides a histogram (and density plot) of the *z* values.

We can calculate the empirical variogram for the data using the `variog` function in the `geoR` package. We will set the maximum distance considered based on our above code. Note that to calculate an empirical variogram, we bin lag distances. `geoR` will automatically do that for us, but we can also manually define the break points in the lag distance categories used for the semivariogram (Fig. 5.7):

Libros de Estadística en PDF | Statistics Books in PDF

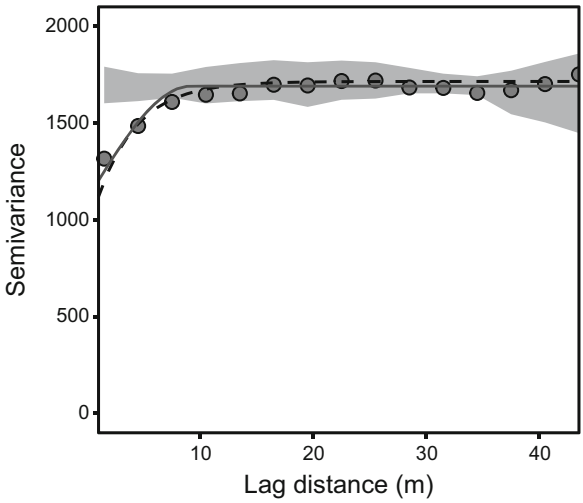


Fig. 5.7 Empirical and theoretical (exponential, black dashed, and spherical, gray solid, models) variograms for interpreting spatial dependence of vegetation height. Also shown are the 99% null pointwise envelopes (shaded region)

```
#Empirical semivariogram
> emp.geoR <- variog(geo.veg, max.dist = maxdist)
> plot(emp.geoR)

#standardize break points to a minimum 3-m lag distance
> emp.geoR <- variog(geo.veg, max.dist = maxdist,
  breaks = c(seq(0, maxdist, by = 3)))
> plot(emp.geoR)
```

In `gstat`, we can create empirical variograms by first creating an object that `gstat` can read (specifying the coordinates of the data) and then using the `variogram` function in `gstat`:

```
> gstat.veg <- matrix
> coordinates(gstat.veg) <- ~x + y
> emp.gstat <- variogram(Height ~ 1, cutoff = maxdist, width =
  3, gstat.veg)
> plot(emp.gstat)
```

Comparing the two packages illustrates that they provide essentially identical empirical variograms.

The above variograms assumed *isotropy*—no directionality in spatial dependence. We can subset our data based on direction to visually consider whether there might be evidence for anisotropy in spatial dependence using the `variog4` function in `geoR` or by adding the `alpha` argument to the `variogram` function in `gstat`. In both cases, data are subset such that four varigrams are calculated for the 0°, 45°, 90°, 135° directions (Fig. 5.8a), where 0° covers the range from −22.5° to 22.5°, 45° covers 22.5° to 67.5°, etc.:

Libros de Estadística en PDF | Statistics Books in PDF

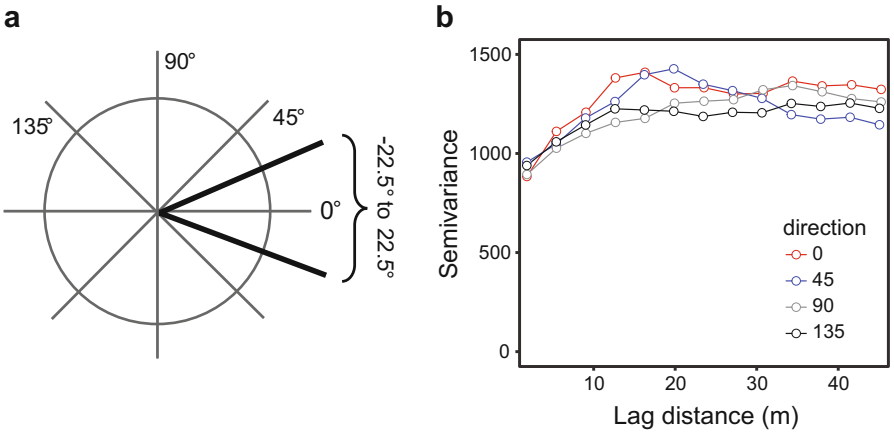


Fig. 5.8 Directional variography subsets the data to interpret variation in spatial dependence in different directions. (**a**) 0°, 45°, 90°, and 135° are typically considered (with windows ±22.5°). Larger values (between 180° and 360°) provide the same patterns because the semivariance formula is symmetric. (**b**) Directional variograms for vegetation height on the plot

```
#in geoR, variogram in each direction
> emp4.geoR <- variog4(geo.veg, max.dist = maxdist)
> plot(emp4.geoR)

#in gstat
> emp4.gstat <- variogram(Height ~ 1, cutoff = maxdist, alpha =
  c(0, 45, 90, 135), gstat.veg)
> plot(emp4.gstat)
```

In this plot (Fig. 5.8b), strong differences in the empirical variograms would suggest that anisotropy might be occurring in the data. Note, however, that to calculate these directional variograms, `geoR` and `gstat` are subsetting the data into four subsets, such that less data are used in each individual variogram. Consequently, the directional variograms may bounce around more than the variogram considered with all of the data. Why are only 0°, 45°, 90°, 135° considered? If directions between 180° and 360° were considered, for example, it would result in the same variogram patterns, because the calculation of the variogram is symmetric (the squared term in Eq. (5.7); $[z(x_i) - z(x_i + d)]^2 = [z(x_i + d) - z(x_i)]^2$).

We can fit theoretical variograms to the data using maximum likelihood techniques with the `likfit` function in `geoR` and contrast different variogram models using model selection criteria (e.g., Akaike's Information Criterion, AIC) (Oliver and Webster 2014). To do so, we must provide initial values for the partial sill (i.e., sill—nugget; Fig. 5.2) and the range, for which we can make an educated guess based on the empirical variogram. To fit exponential and spherical variogram models (Fig. 5.3):

Libros de Estadística en PDF | Statistics Books in PDF

```
#exponential variogram
> mlexp <- likfit(geo.veg, cov.model = "exp", ini = c(700, 10))

#spherical variogram
> mlsph <- likfit(geo.veg, cov.model = "sph", ini = c(700, 10))
> summary(mlexp)

##
Summary of the parameter estimation
-----------------------------------
Estimation method: maximum likelihood

Parameters of the mean component (trend):
 beta
43.0708

Parameters of the spatial component:
 correlation function: exponential
 (estimated) variance parameter sigmasq (partial sill) = 504.7
 (estimated) cor. fct. parameter phi (range parameter) = 5.884
 anisotropy parameters:
 (fixed) anisotropy angle = 0 ( 0 degrees )
 (fixed) anisotropy ratio = 1

Parameter of the error component:
 (estimated) nugget = 732

Transformation parameter:
 (fixed) Box-Cox parameter = 1 (no transformation)

Practical Range with cor=0.05 for asymptotic range: 17.62812

Maximised Likelihood:
 log.L n.params AIC BIC
 "-3298" "4" "6603" "6621"

non spatial model:
 log.L n.params AIC BIC
 "-3368" "2" "6739" "6748"

Call:
likfit(geodata = geoR.veg, ini.cov.pars = c(500, 15), cov.model = "exp")

> AIC(mlexp, mlsph)

##
 df AIC
mlexp 4 6603.375
mlsph 4 6603.830
```

Libros de Estadística en PDF | Statistics Books in PDF

The output from these models provides several key insights. For our purposes, we will focus on two types of important output. First, for each model, the log-likelihood, AIC and BIC (Bayesian Information Criterion) are provided for interpreting model fit and model selection. These values are provided for the spatial model considered, as well as a "non-spatial" model, which would assume a constant variance (i.e., variance does not change with lag distance). The output also provides estimates of the range, nugget, and partial sill for the model under consideration. For some theoretical variograms, the "practical range" is also provided. The practical range uses an approximation (which varies, depending on the theoretical variogram model) to determine the effective range distance when the variogram function shows a smoothed asymptotic relationship to the sill (e.g., exponential models; Fig. 5.3). For instance, in an exponential variogram it is typically defined as the distance where the variance reaches 95% of the estimated sill. In this example, the exponential variogram fits the data slightly better than the spherical variogram, based on AIC. Both of these models fit the data substantially better than a non-spatial model.

We can fit an exponential variogram in `gstat` as:

```
> exp.gstat <- fit.variogram(emp.gstat, vgm("Exp"))
```

Note that while `gstat` does not implement model selection based on likelihood techniques, it does provide a wider variety of model-based variograms that `geoR`. These alternatives can be perused with the `vgm()` and `show.vgm()` functions.

Finally, we can overlay plots of the theoretical variograms with the empirical variograms (Fig. 5.7):

```
> plot(emp.geoR)
> lines(mlexp, col = "blue")
> lines(mlsph, col = "red")
```

We can use model selection to contrast spatial and non-spatial models provided in the output of the `likfit` function. Another useful approach is to determine confidence envelopes of spatial randomness (analogous to envelopes calculated in Chap. 4). Null envelopes can then be overlaid with the empirical and theoretical variograms. In `geoR`, we can obtain null envelopes Monte Carlo permutations. The code below executes 99 permutations, where vegetation height is shuffled among x–y coordinates and plots the maximum and minimum values at each distance lag relative to the empirical variogram:

```
> emp.env <- variog.mc.env(geo.veg, obj.var = emp.geoR)
> plot(emp, envelope = emp.env)
> lines(mlexp, col = "blue")
```

These envelopes describe the variance as a function of lag distance under spatial randomness, given the underlying data. Consequently, when our observed variogram falls outside of this envelope, there is some signature of significant spatial

Libros de Estadística en PDF | Statistics Books in PDF

dependence in the data. In this case, we observe that the variogram only falls outside of the null envelope at distances <10 m, loosely similar to our conclusions on spatial dependence using correlograms (Fig. 5.5).

5.3.5 *Kriging*

With our theoretical variogram model, we can create an interpolated map using *kriging*. We could potentially make kriged predictions onto our observed sampling locations or onto a grid that covers the entire plot. We will illustrate the latter, where we use the `expand.grid` function to create a new set of locations. Note that distances between this expanded grid will provide the resolution of the map that we create. First, we krige with `geoR` (Fig. 5.9a).

```
#grid with 1-unit intervals (1-m)
> new.grid.1m <- expand.grid(0:max(matrix$x), 0:max(matrix$y))

#kriging: krige.control, cov.pars: partial sill, range
> krig.geoR.exp <- krige.conv(geoR.veg, locations = new.grid.1m,
  krige = krige.control(cov.pars = c(mlexp$cov.pars[1],
  mlexp$cov.pars[2]), nugget = mlexp$nugget,
  cov.model = "exp", type.krige = "OK"))

#get the prediction values for the kriged surface
> image(krig.geoR.exp, main = "kriged estimates")
```

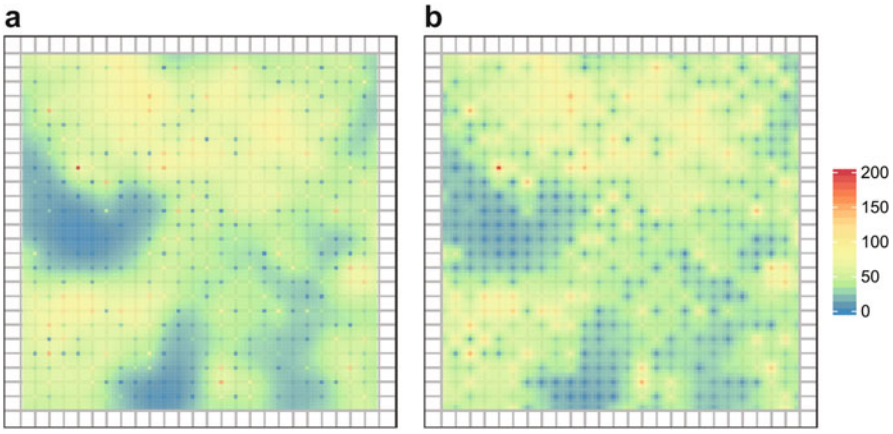


Fig. 5.9 Kriged map of vegetation height based on the exponential model (**a**). Also shown is the background sampling grid (2×2 m) for reference. (**b**) Inverse distance weighting interpolation (IDW)

Libros de Estadística en PDF | Statistics Books in PDF

In the above code, we take the estimates from the exponential variogram and use them in ordinary kriging (`type.krige = "OK"`) for spatial interpolation. The output includes predictions that we can use for mapping, as well as uncertainty in those predictions. In our case, the initial sampling grid was detailed at a high resolution, such that the uncertainty is very low. Note also that the kriged image uses our z values at the observed samples and only makes predictions for unsampled areas on our new grid. We can also plot the uncertainty in the predictions:

```
> image(krig.geoR.exp, val = sqrt(krig.geoR.exp $krige.var),
  main = "kriging SE")
```

Here, the model does not estimate variance for the sampled points; it fixes the variance to zero for those locations. Consequently, one could remove those sampled points from the mapping of the uncertainty in kriged predictions. This kriged map could then be used as a raster image for other purposes.

We can also implement kriging in `gstat` with the `krige` function.

```
> new.grid.1m <- expand.grid(x = 0:max(matrix$x), y =
  0:max(matrix$y))
> gridded(new.grid.1m) <- ~x + y
> krig.gstat <- krige(Height ~ 1, gstat.veg, new.grid.1m, model
  = exp.gstat)

#plot
> image(krig.gstat, main = "kriging-gstat")
```

In `gstat`, we need to have labels for the x–y coordinates in the new grid (unlike `geoR`). As an aside, inverse distance weighting interpolation is also straightforward in `gstat` with the `idw` function (Fig. 5.9b):

```
> idw.gstat <- idw(Height ~ 1, gstat.veg, new.grid.1m)
```

We can check the similarity in the kriged predictions from `geoR` and `gstat` and inverse distance weighting by calculating the correlation between predictions as:

```
> cor(cbind(geoR.exp = krig.geoR.exp$predict,
  gstat.exp = krig.gstat$var1.pred,
  gstat.idw = idw.gstat$var1.pred))

##
geoR.exp gstat.exp gstat.idw
geoR.exp 1.000 1.000 0.984
gstat.exp 1.000 1.000 0.984
gstat.idw 0.984 0.984 1.000
```

The two packages provide identical predictions based on kriging. In this case, the inverse distance weighting also provides nearly identical predictions to kriging. This

Libros de Estadística en PDF | Statistics Books in PDF

is not surprising, given the dense, regular sampling in the plot. With sparse and/or irregularly spaced sampling, we might expect these approaches to be less correlated.

5.3.6 Simulating Spatially Autocorrelated Data

Once the parameters of a theoretical variogram have been estimated, one can use these values to generate simulated spatially autocorrelated data having the same statistical properties of the observed spatial pattern using an annealing algorithm (Cressie 1993) or a Gaussian random fields algorithm, both of which are stochastic distribution functions (Lantuéjoul 2002). These procedures, and others, are often used to generate null reference distributions to test significance of observed spatial patterns in ecological data (e.g., Remmel and Fortin 2013). Note that when simulating Gaussian random fields, the simulations by default have a mean of zero.

We can use the `gstat` or `RandomFields` package (Schlather et al. 2015) to simulate spatial patterns based on the variogram parameters. Here, we show the use of `RandomFields`, which has more flexibility in this regard than `gstat`. We will also provide include the mean value of observed vegetation height (otherwise, the mean value of the simulated random field would be approximately zero).

```
#variogram models to simulate
> library(RandomFields)
> model.exp <- RMexp(var = mlexp$cov.pars[1], scale =
  mlexp$cov.pars[2]) + RMnugget(mlexp$nugget) + RMtrend(mean =
  mean(matrix$Height))

> dimx <- 1:50
> dimy <- 1:50

#simulate
> sim.exp <- RFsimulate(model = model.exp, x = dimx, y = dimy)
> data.sim <- as.matrix(sim.exp)

#plot with image
> image(dimx, dimy, data.sim,xlab = "x", ylab = "y")

#plot with raster package
> library(raster)
> RMexp.grid <- raster(data.sim)
> plot(RMexp.grid)
```

These simulated maps (Fig. 5.10) are called "unconditional Gaussian random fields." If we provide sample values for mapping (as in kriging), then the maps would be considered "conditional Gaussian random fields." Note that even when adjusting for the mean value of vegetation height with the `RMtrend` function, this

Libros de Estadística en PDF | Statistics Books in PDF

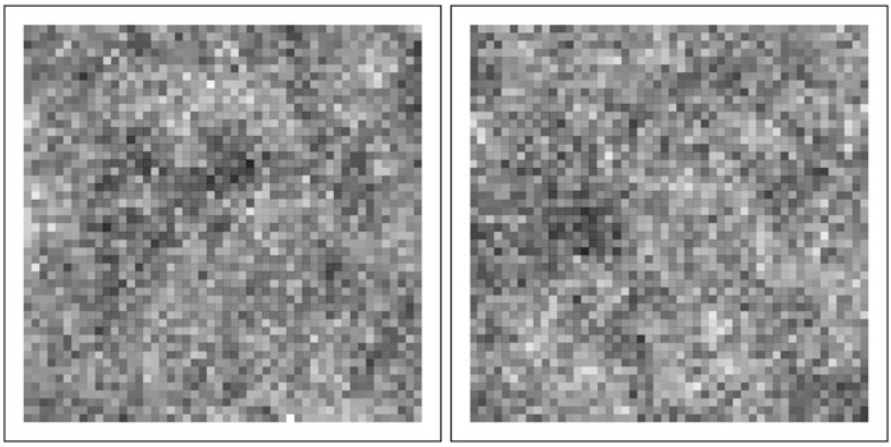


Fig. 5.10 Two realizations of simulating spatial dependence under the exponential variogram model fit to the vegetation data using unconditional Gaussian random fields

approach does end up simulating some values less than zero, which is not biologically plausible, because it is a Gaussian (normally distributed) random field. The `RandomFields` package has options for simulating fields that can circumvent this problem, but it is beyond the scope of this book.

The approach above can also be used to more generally create spatially autocorrelated maps of different degrees, similar in function to the neutral landscape approaches described in Chap. 3. For example, we illustrate how altering the partial sill and range parameters can generate different types of neutral landscape maps (Fig. 5.11) with the following alternative scenarios:

```
> model.exp.ps2r5 <- RMexp(var = 20, scale = 5) + RMnugget(var = 2)
> model.exp.ps8r5 <- RMexp(var = 80, scale = 5) + RMnugget(var = 2)
> model.exp.ps2r20 <- RMexp(var = 20, scale = 20) + RMnugget(var = 2)
```

The above scenarios take a base model, where the range = 5, the nugget = 2 and the partial sill = 20 and then increased the partial sill 4× and the range 4×. When plotting realizations of these models (similar to above), it is clear that changes in the partial sill increases the magnitude of variation and increases in the range parameter makes the map smoother. When truncating these maps similar to a neutral landscape scenario where we alter the proportion of habitat or land cover on the landscape (Chap. 3), however, changing the partial sill has negligible effects on the map while increasing the range leads to much greater aggregation of the land cover (Fig. 5.11).

Libros de Estadística en PDF | Statistics Books in PDF

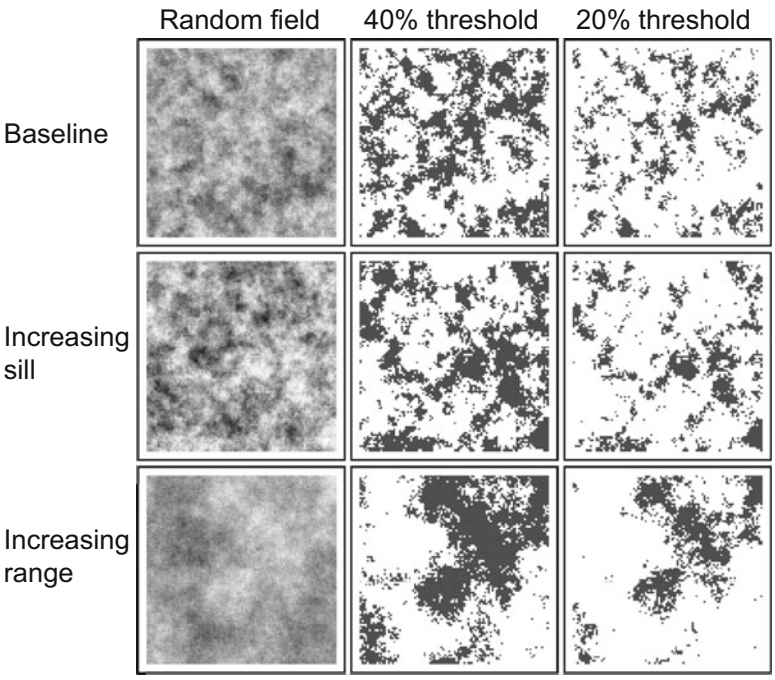


Fig. 5.11 Using unconditional Gaussian random fields to simulate neutral landscapes with varying spatial dependence based on an exponential variogram model. The first row is taken from a model with partial sill = 20, range = 5, and nugget = 2. The second row increases the partial sill 4×, while the third row increases the range 4×. Shown is the continuous random field and two thresholds of this map, similar to the use of neutral landscape models described in Chap. 3

5.3.7 Multiscale Analysis

With the availability of remotely sensed data and increasingly large databases that span broad extents, areas under investigation are usually large enough to include the effects of several process acting at various spatial scales that generate observed spatial patterns. With the potential of such multiscale effects, the first step consists therefore to identify the key spatial scales of the patterns. There are two multiscale analysis methods that can be used to decompose the key spatial scales from remotely sensed data or other data from a study area: Fourier spectral decomposition and hierarchical wavelet decomposition analysis (Keitt and Urban 2005).

5.3.7.1 Wavelets and Fourier Series

Fourier's technique and wavelets are related (Dale et al. 2002). Fourier's technique assumes the data have been generated by stationary processes that occur across the entire area of interest. With this approach, processes are envisioned as a series of sine and cosine waves operating at different scales that sum together to drive observed

Libros de Estadística en PDF | Statistics Books in PDF

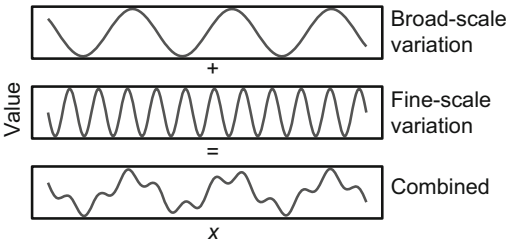


Fig. 5.12 How Fourier transforms work. Fourier transforms assume that observed variation comes from multiple processes operating at different scales. This approach uses sine/cosine waves of different amplitudes and periods to capture such variation. These broad and fine-scale waves are combined to interpret observed variation. This approach assumes stationarity, where waves operate across the entire extent of interest, unlike wavelet transforms

variation (Fig. 5.12). This formulation can be helpful in many situations when broad extents are considered; indeed, linear trends across broad extents are unlikely in many situations (Austin 2002).

When the stationary assumption is not fulfilled, a wavelet discrete transform can be used, as we illustrate in the example below. Wavelets transforms use a similar approach to Fourier decomposition but with two key differences. First, wavelet transforms come in a variety of shapes (the "haar," "Mexican hat," etc.; Dale et al. 2002), such that users need not assume only sine/cosine functions to describe spatial variation. Second, wavelet transforms do not assume stationarity, but instead allow for local variation in wavelet templates at different resolutions to be fit to observed data.

These techniques tend to be applied to dyadic grids of data, where the dimensions of the grid are of a power 2 (e.g., 32×32, 128×128), similar to our use of fractal algorithms in Chap. 3. The reason for this constraint is that it allows us to recursively decompose the spatial variation on the map. For instance, if we have a map of dimensions 64×64, this map can be broken into successive blocks representing different spatial resolutions, such as four blocks of 32×32 units, to interpret spatial variation. Note that wavelet analysis has been extended to work with maps that do not comply with this constraint, but that is beyond the scope of our application here.

Wavelets can be discrete or continuous. Here, we focus on the simplest, discrete transform, the Haar (Fig. 5.13). Using observed data, the discrete wavelet transform can be computed for each sampling location, or pixel, for a series of wavelet template scales that are as a power of two. Then, the wavelet values can be mapped and all scales analyzed. The plot of wavelet variance against the scaling factor is called a *scalogram* (Dale and Fortin 2014). The highest wavelet variance values indicate the spatial scales that fit the data best.

We calculate wavelets using the `waveslim` package (Whitcher 2015). We use the Haar wavelet, which is a common type of wavelet used in spatial analysis (Fig. 5.13). To calculate wavelets, we need to pass the maximum scale being considered, which should be a power of 2 (8, 16, 32, 64, etc.). First, we reformat

Libros de Estadística en PDF | Statistics Books in PDF

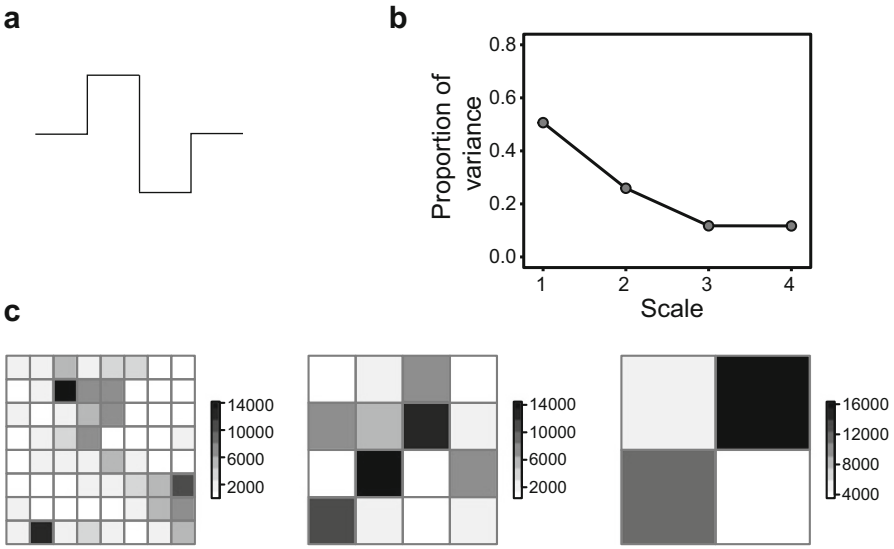


Fig. 5.13 Wavelets and scalograms. (**a**) The Haar template, which is a common discrete wavelet template. (**b**) The scalogram from the fitting the Haar template to a portion of the observed data, where variance is plotted as a function of scale. (**c**) Maps of the spatial variation identified at each scale (scales 1–3) from wavelet analysis

the data into a square matrix using the `acast` function from the `reshape2` package (Wickham 2007):

```
> library(reshape2)
> matrix.mat <- acast(matrix, x ~ y, value.var = "Height")
> dim(matrix.mat)

##
[1] 26 26
```

This reformatting emphasizes that the sampling grid used is a 26 × 26 grid. For the purposes here, we will subset the grid to become a dyadic grid that is 16 × 16.

```
> max.scale <- 4

#DWT: Discrete Wavelet Transform
> library(waveslim)
> x.dwt <- dwt.2d(matrix.mat[1:16, 1:16], 'haar', J = max.scale)
```

This function creates new matrices that describe the wavelet variance at different scales. To do so, it creates three bands of variation, labeled LH, HL, and HH. We sum the squared values of these bands to quantify a total measure of wavelet

Libros de Estadística en PDF | Statistics Books in PDF

variance. We can then calculate the proportion of variance at each scale and plot the scalogram (Fig. 5.13b).

```
#Sum the wavelet spectrums
> t.var <- (sum(x.dwt$LH1^2 + x.dwt$HL1^2 + x.dwt$HH1^2)
  + sum(x.dwt$LH2^2 + x.dwt$HL2^2 + x.dwt$HH2^2)
  + sum(x.dwt$LH3^2 + x.dwt$HL3^2 + x.dwt$HH3^2)
  + sum(x.dwt$LH4^2 + x.dwt$HL4^2 + x.dwt$HH4^2))

#proportional variance
> x.lev.1 <- (sum(x.dwt$LH1^2 + x.dwt$HL1^2 + x.dwt$HH1^2)) / t.var
> x.lev.2 <- (sum(x.dwt$LH2^2 + x.dwt$HL2^2 + x.dwt$HH2^2)) / t.var
> x.lev.3 <- (sum(x.dwt$LH3^2 + x.dwt$HL3^2 + x.dwt$HH3^2)) / t.var
> x.lev.4 <- (sum(x.dwt$LH4^2 + x.dwt$HL4^2 + x.dwt$HH4^2)) / t.var

> var.all.dwt <- c(x.lev.1, x.lev.2, x.lev.3, x.lev.4)
> sum(var.all.dwt)

#Scalogram: plotting global Wavelet spectrum profiles
> plot(var.all.dwt, pch = 21, type = "b", lwd = 1, ylab = "Average
  Variance", xlab = "Scale")
```

The scalogram suggests that most of the spatial variation occurs at the finest distance considered. Finally, we can plot the wavelet images (Fig. 5.13c) at each scale using the `raster` package. Below we show an example of scale 1.

```
#Map Wavelet values according to scales
> wave.raster1 <- raster((x.dwt$LH1^2 + x.dwt$HL1^2 +
  x.dwt$HH1^2))
> plot(wave.raster1)
```

These measures of variation at different scales can then be used as predictor variables in regression or related analyses to account for spatial dependence arising at different scales (Keitt and Urban 2005). The key to do so is to link the submatrices to the appropriate response data in a hierarchical way.

5.3.7.2 Eigenvector Spectral Decomposition

When data are sampled from an irregular grid or layout in a contiguous fashion, an eigenvector spectral decomposition can be used to identify the key scales that match the data. Here, we use principal coordinates of neighborhood matrices, PCNM, which is a special case of the generalized Moran's Eigenvector Map (Dray et al. 2006, 2012). Unlike the wavelet analysis that is performed on the data given the spatial layout of contiguous pixels, the PCNM multiscale analysis is performed on the *x*–*y* coordinates of the sampling locations. In this approach, there are potentially as many PCNM spatial scales as there are sampling locations. PCNM uses a

Libros de Estadística en PDF | Statistics Books in PDF

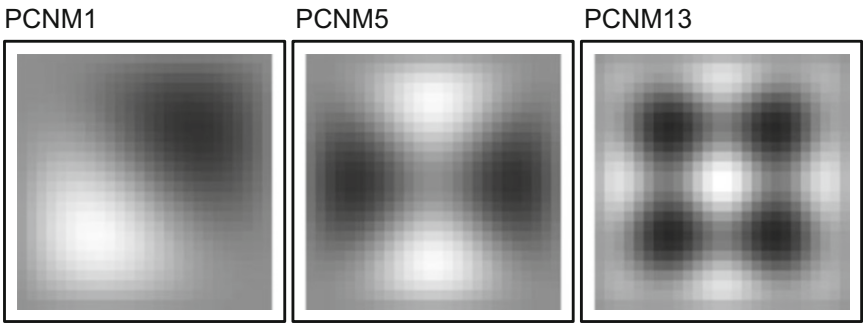


Fig. 5.14 Eigenvector mapping captures spatial variation at different scales. The first few eigenvectors capture broad-scale spatial variation, similar to what would be captured if a linear or polynomial term of x–y coordinates was considered (e.g., in trend surface analysis) whereas the latter eigenvectors capture increasingly fine scale variation. Shown are the first three eigenvectors selected with forward, stepwise regression on the observed vegetation data, where dark pixels indicate higher values of the eigenvectors

Principal Coordinates Analysis (PCoA; Gower 1966), which is also referred to as metric multidimensional scaling or classical scaling (Legendre and Legendre 2012). PCoA shares some similarities to the more common principal components analysis (PCA), but it focuses on using distance or similarity matrices (rather than the original data, as in PCA) to position objects in a space of lower dimensionality than the original data, with a focus on Euclidean distance space. PCNM provides eigenvectors that can capture spatial structure: as the eigenvectors increase, they capture increasingly finer scales of space that resemble sin waves of smaller and smaller periods (Fig. 5.14), analogous to Fourier's technique.

Once the PCNMs are computed they can be used as spatial predictors either in a multiple regression (Dormann et al. 2007) or other analyses, such as redundancy analysis for community data (see Chap. 11). Because there as many PCNM eigenvectors as the number of locations, model selection is necessary to reduce the number of eigenvectors considered. Narrowing the number of spatial scales can also be guided using knowledge about the scales of processes that may have generated the data; for example large-scale trend, intermediate patchiness, and small-scale patchiness. If we were to only consider the first few eigenvectors, this may be functionally similar to what has been referred to a trend surface analysis, where x or y coordinates are used as predictors (and potentially their polynomial terms, such as x^2 or x^3) in regression models to allow for large-scale variation in spatial dependence (see Chap. 6). In contrast, the larger eigenvectors capture fine-scale variation in spatial dependence.

We illustrate the eigenvector method with the grid used for kriging above. We can use the `vegan` package (Oksanen et al. 2018) to first determine the PCNM based on a distance matrix calculated from the sample locations. Eigenvectors for each sampling location are calculated.

Libros de Estadística en PDF | Statistics Books in PDF

```
> library(vegan)

#PCNM on distance matrix based on coords
> xypcnm <- pcnm(dist(coords))

#eigenvectors:
> xypcnm$vectors
```

We can visualize the eigenvectors in a variety of ways. Here we create a raster of the eigenvectors with the `raster` package and plot (Fig. 5.14).

```
#create raster
pcnm1.raster <- rasterFromXYZ(data.frame(x = matrix$x, y =
matrix$y, z = xypcnm$vectors[,1]))
plot(pcnm1.raster)
```

Once we have generated the eigenvectors, we can use them as predictors for vegetation height. Given the large number of eigenvectors that are generated, a common approach is to use a procedure to select a subset of the eigenvectors for further inclusion (Bauman et al. 2018). One approach is forward selection procedure to determine the eigenvectors that best explain the response variable (Dray et al. 2006; Blanchet et al. 2008). Other options are possible, such as using the reduction in spatial autocorrelation in the residuals of models (see Chap. 6; Dray et al. 2006; Dormann et al. 2007). Here, we simply focus on a forward selection procedure suggested by Blanchet et al. (2008), which has been shown to reliably capture multiscale spatial dependence in some situations (Bauman et al. 2018). This approach can be implemented in the `adespatial` package (Dray et al. 2018). In this approach, we first fit a full (global) model with all eigenvectors as covariates. From this model, we extract the adjusted R^2 value, which provides a measure of the variation explained from the eigenvectors (adjusted for the number of variables in the model). Blanchet et al. (2008) then proposed a forward selection approach with a "double-stopping" rule; that is, the forward selection terminates either when the current model reaches the adjusted R^2 of the full model, or when new eigenvectors are no longer significant, based on a prespecified α. Below, we implement this approach, using a conservative $\alpha = 0.005$, given the large number of potential covariates included.

```
> library(adespatial)
> height <- matrix$Height
> xypcnm.df <- data.frame(xypcnm$vectors)

#fit full model
> xypcnm.full <- lm(height ~ ., data = xypcnm.df)
> R2adj <- summary(xypcnm.full)$adj.r.squared
```

Libros de Estadística en PDF | Statistics Books in PDF

```
#forward selection with adespatial
> xypcnm.for <- forward.sel(height, xypcnm$vectors, adjR2thresh =
  R2adj, alpha = 0.005, nperm = 999)
```

In this case, we find that ten eigenvectors are retained in the linear regression model to explain the spatial variation in vegetation height (the total number retained might vary slightly due to the stochastic nature of the permutation test used; see Fig. 5.14 for plots of three of the selected eigenvectors). In Chap. 6, we dive deeper into spatial regression and how this method and others can account for spatial dependence while interpreting environmental relationships.

5.4 Next Steps and Advanced Issues

5.4.1 Local Spatial Dependence

Throughout this chapter, we have focused on "global" spatial statistics that assume stationarity. However, the intensity of spatial dependence can often vary across a region. When the stationary assumption is not valid, other methods can be used. There are two general ways in which uncovering local spatial dependence can occur (Brunsdon and Comber 2015). First, some approaches take global indices and use decomposition methods to understand the role of individual locations on the global statistic. For example, local indicators of spatial association, or LISA (Anselin 1995; Boots 2002), take indices like Moran's *I* to determine the contribution of each observation to the global statistic, which can subsequently be mapped to understand spatial variation in spatial dependence. The examples of using Moran's *I* in this chapter can be extended to interpret local Moran's *I* in both the `ncf` and `spdep` packages. Second, moving window analyses can be done to understand spatial dependencies in the chosen windows. In this approach, global statistics are applied, but only to the neighborhood (window) under consideration. See Chap. 3 for examples of moving window analyses.

5.4.2 Multivariate Spatial Dependence

In ecology, data are frequently multivariate. For instance, in community ecology we often work with matrices of species occurrence or abundance. In these cases, we may be interested in understand and accounting for multivariate spatial dependence (Dray et al. 2012). Many of the methods described in this chapter can be extended to multivariate data (Wackernagel 2003). When two types of data occur, cross-correlograms and cross-variograms can be used to understand the spatial dependence between variables (Wagner 2003). See Chap. 11 for some discussion of these approaches in the context of spatially structured communities.

Libros de Estadística en PDF | Statistics Books in PDF

5.5 Conclusions

The detection, characterization, and significance testing of spatial pattern is the first step to understand spatial ecological data and the processes that generated them. Spatial dependence commonly occurs in ecological data and it is often argued that failure to account for spatial dependence can impact inferences in ecology (Legendre 1993; Dormann et al. 2007; Beale et al. 2010). Here, we illustrate ways in which spatial dependence can be diagnosed in ecological data. This type of diagnosis can be useful to provide insights into why patterns occur in data and whether spatial dependence can be problematic for inferences on ecological patterns and processes.

Both correlograms and variograms provide useful insights for interpreting the magnitude and extent of spatial dependence in data. These approaches provide much richer information than single tests of spatial dependence, such as using a single Moran's *I* test statistic (Eq. 5.3). Correlograms have the benefit of providing a standardized metric (i.e., correlation coefficient) that can be compared across variables, while variograms provide a means to formally estimate the scale of spatial dependence through the use of model-based variograms and estimating the spatial range. Variogram modeling can also be used for interpolating spatial data via kriging, providing a formal means for predicting ecological patterns across space. A variety of techniques have been used for inferring the significance of spatial dependence, with Monte Carlo permutations providing perhaps the most flexibility in their applications. Multiscale analyses can also provide useful insights, in particular in situations where questions and/or data come from broad extents where multiple scales of spatial dependence may operate.

References

Acevedo MA, Fletcher RJ Jr (2017) The proximate causes of asymmetric movement across heterogeneous landscapes. Landsc Ecol 32:1285–1297

Anselin L (1995) Local indicators of spatial association: LISA. Geogr Anal 27(2):93–115

Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol Model 157(2–3):101–118

Bauman D, Drouet T, Dray S, Vleminckx J (2018) Disentangling good from bad practices in the selection of spatial or phylogenetic eigenvectors. Ecography 41:1–12

Beale CM, Lennon JJ, Yearsley JM, Brewer MJ, Elston DA (2010) Regression analysis of spatial data. Ecol Lett 13(2):246–264. https://doi.org/10.1111/j.1461-0248.2009.01422.x

Bivand R (2006) Implementing spatial data analysis software tools in R. Geogr Anal 38(1):23–40. https://doi.org/10.1111/j.0016-7363.2005.00672.x

Bivand R, Piras G (2015) Comparing implementations of estimation methods for spatial econometrics. J Stat Softw 63(18):1–36

Bivand RS, Pebesma EJ, Gomez-Rubio V (2013) Applied spatial data analysis with R. Use R! 2nd edn. Springer, New York

Bjørnstad ON, Falck W (2001) Nonparametric spatial covariance functions: estimation and testing. Environ Ecol Stat 8(1):53–70. https://doi.org/10.1023/a:1009601932481

Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables. Ecology 89(9):2623–2632. https://doi.org/10.1890/07-0986.1

Libros de Estadística en PDF | Statistics Books in PDF

Bolker BM (2003) Combining endogenous and exogenous spatial variability in analytical population models. Theor Popul Biol 64(3):255–270. https://doi.org/10.1016/s0040-5809(03)00090-x
Boots B (2002) Local measures of spatial association. Ecoscience 9(2):168–176
Brown JH, Mehlman DW, Stevens GC (1995) Spatial variation in abundance. Ecology 76 (7):2028–2043. https://doi.org/10.2307/1941678
Brunsdon C, Comber L (2015) An introduction to R for spatial analysis and mapping. Sage Publications, Inc, London
Burnham KP, Anderson DR (1998) Model selection and inference: a practical information-theoretic approach. Springer, New York
Carroll SS, Pearson DL (2000) Detecting and modeling spatial and temporal dependence in conservation biology. Conserv Biol 14(6):1893–1897. https://doi.org/10.1046/j.1523-1739.2000.99432.x
Cohen JM, Civitello DJ, Brace AJ, Feichtinger EM, Ortega CN, Richardson JC, Sauer EL, Liu X, Rohr JR (2016) Spatial scale modulates the strength of ecological processes driving disease distributions. Proc Natl Acad Sci U S A 113(24):E3359–E3364. https://doi.org/10.1073/pnas.1521657113
Cressie NAC (1993) Statistics for spatial data. Wiley, Chichester
Dale MRT, Fortin MJ (2014) Spatial analysis: a guide for ecologists, 2nd edn. Cambridge University Press, Cambridge
Dale MRT, Dixon P, Fortin MJ, Legendre P, Myers DE, Rosenberg MS (2002) Conceptual and mathematical relationships among methods for spatial analysis. Ecography 25(5):558–577. https://doi.org/10.1034/j.1600-0587.2002.250506.x
Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kuehn I, Ohlemueller R, Peres-Neto PR, Reineking B, Schroeder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30(5):609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x
Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol Model 196(3–4):483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015
Dray S, Pelissier R, Couteron P, Fortin MJ, Legendre P, Peres-Neto PR, Bellier E, Bivand R, Blanchet FG, De Caceres M, Dufour AB, Heegaard E, Jombart T, Munoz F, Oksanen J, Thioulouse J, Wagner HH (2012) Community ecology in the age of multivariate multiscale spatial analysis. Ecol Monogr 82(3):257–275. https://doi.org/10.1890/11-1183.1
Dray S, Bauman D, Blanchet G, Borcard D, Clappe S, Guenard G, Jombart T, Larocque G, Legendre P, Madi N, Wagner HH (2018) adespatial: multivariate spatial analysis. R package version 0.2-0
Efron B (1979) Bootstrap methods - another look at the jackknife. Ann Stat 7(1):1–26. https://doi.org/10.1214/aos/1176344552
Fletcher RJ Jr, Sieving KE (2010) Social-information use in heterogeneous landscapes: a prospectus. Condor 112:225–234
Fletcher RJ Jr, Acevedo MA, Robertson EP (2014) The matrix alters the role of path redundancy on patch colonization rates. Ecology 95(6):1444–1450
Fletcher RJ, Reichert BE, Holmes K (2018) The negative effects of habitat fragmentation operate at the scale of dispersal. Ecology 99(10):2176–2186
Giraudoux P (2018) pgirmess: spatial analysis and data mining for field ecologists. R package version 1.6.9
Goovaerts P (1994) Study of spatial relationships between two sets of variables using multivariate geostatistics. Geoderma 62(1–3):93–107. https://doi.org/10.1016/0016-7061(94)90030-2
Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53(3–4):325–338
Haining R (2003) Spatial data analysis: theory and practice. Cambridge University Press, Cambridge
Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. Ecol Monogr 54(2):187–211. https://doi.org/10.2307/1942661

Libros de Estadística en PDF | Statistics Books in PDF

Keitt TH, Urban DL (2005) Scale-specific inference using wavelets. Ecology 86(9):2497–2504. https://doi.org/10.1890/04-1016
Keitt TH, Bjørnstad ON, Dixon PM, Citron-Pousty S (2002) Accounting for spatial pattern when modeling organism-environment interactions. Ecography 25(5):616–625
Koenig WD (1998) Spatial autocorrelation in California land birds. Conserv Biol 12(3):612–620. https://doi.org/10.1046/j.1523-1739.1998.97034.x
Koenig WD, Liebhold AM (2016) Temporally increasing spatial synchrony of North American temperature and bird populations. Nat Clim Chang 6(6):614. https://doi.org/10.1038/nclimate2933
Landeiro VL, Magnusson WE (2011) The geometry of spatial analyses: implications for conservation biologists. Natureza & Conservacao 9(1):7–19. https://doi.org/10.4322/natcon.2011.002
Lantuéjoul C (2002) Geostatistical simulation, models, and algorithms. Springer, Berlin
Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? Ecology 74(6):1659–1673
Legendre P, Legendre L (2012) Numerical Ecology, 3rd Edition. Elsevier, Amsterdam
Nelson TA, Boots B (2008) Detecting spatial hot spots in landscape ecology. Ecography 31(5):556–566. https://doi.org/10.1111/j.0906-7590.2008.05548.x
Oksanen J, Guillaume B, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens HH, Szoecs E, Wagner H (2018) Vegan: community ecology package. R version 2.4-6
Oliver MA, Webster R (1991) How geostatistics can help you. Soil Use Manag 7(4):206–217. https://doi.org/10.1111/j.1475-2743.1991.tb00876.x
Oliver MA, Webster R (2014) A tutorial guide to geostatistics: computing and modelling variograms and kriging. Catena 113:56–69. https://doi.org/10.1016/j.catena.2013.09.006
Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. Comput Geosci 30(7):683–691. https://doi.org/10.1016/j.cargo.2004.03.012
Peres-Neto PR, Legendre P (2010) Estimating and controlling for spatial structure in the study of ecological communities. Glob Ecol Biogeogr 19(2):174–184. https://doi.org/10.1111/j.1466-8238.2009.00506.x
Remmel TK, Fortin MJ (2013) Categorical, class-focused map patterns: characterization and comparison. Landsc Ecol 28(8):1587–1599. https://doi.org/10.1007/s10980-013-9905-x
Ribeiro PJ, Jr., Diggle PJ (2016) geoR: analysis of geostatistical data. vol R package version 1.7-5.2
Rossi RE, Mulla DJ, Journel AG, Franz EH (1992) Geostatistical tools for modeling and interpreting ecological spatial dependence. Ecol Monogr 62(2):277–314. https://doi.org/10.2307/2937096
Schlather M, Malinowski A, Menck PJ, Oesting M, Strokorb K (2015) Analysis, simulation and prediction of multivariate random fields with package random fields. J Stat Softw 63(8):1–25
Schooley RL, Wiens JA (2004) Movements of cactus bugs: patch transfers, matrix resistance, and edge permeability. Landsc Ecol 19(7):801–810
Tobler WR (1970) Computer movie simulating urban growth in Detroit region. Econ Geogr 46(2):234–240. https://doi.org/10.2307/143141
Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York
Wackernagel H (2003) Multivariate geostatistics: an introduction with applications, 3rd edn. Springer, Berlin
Wagner HH (2003) Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. Ecology 84(4):1045–1057. https://doi.org/10.1890/0012-9658(2003)084[1045:scipci]2.0.co;2
Wagner HH, Fortin MJ (2005) Spatial analysis of landscapes: concepts and statistics. Ecology 86(8):1975–1987. https://doi.org/10.1890/04-0914
Whitcher B (2015) waveslim: basic wavelet routines for one-, two- and three-dimensional signal processing. R package version 1.7.5
Wickham H (2007) Reshaping data with the reshape package. J Stat Softw 21(12):1–20
Yoo J, Ready R (2016) The impact of agricultural conservation easement on nearby house prices: incorporating spatial autocorrelation and spatial heterogeneity. J For Econ 25:78–93. https://doi.org/10.1016/j.jfe.2016.09.001

Libros de Estadística en PDF | Statistics Books in PDF

Chapter 6
Accounting for Spatial Dependence in Ecological Data



6.1 Introduction

Inference and prediction are fundamental to all aspects of ecology and conservation. Yet the presence of dependency in the data due to either phylogeny, space, or time can impair the statistical inference and subsequent ecological interpretation of the pattern(s) observed (Sokal and Oden 1978; Swihart and Slade 1985; Garland et al. 1992; Lennon 2000; Miller 2012). In this chapter, we will focus specifically on how the presence of spatial dependency complicates our ability to make statistical inferences and prediction (Legendre 1993), as the principles due to space are analogous to those due to time and phylogeny (Bauman et al. 2018). It is important to understand how statistical biases due to spatially structured data can affect answering a wide array of ecological questions ranging from species–environment relationships to predicting the spread of invasive species. Consequently, there is an increasing emphasis on formally accounting for spatial dependence in inferential problems in ecology and conservation (Segurado et al. 2006; Dormann et al. 2007; Hooten et al. 2007; Carroll and Johnson 2008; Beale et al. 2010; Crase et al. 2014).

Accounting for spatial dependence in modeling is, however, very challenging. This challenge arises because spatial dependence in data can emerge for a variety of reasons (see Chap. 5). In particular, when modeling spatial data, spatial dependence can occur simply due to model mis-specification, such an important covariate not being included in the model or that its functional relationship is mis-specified (e.g., effects may be non-linear). Spatial dependence could also occur through processes such as localized dispersal or social behavior (Koenig 1999). In these cases, adding environmental covariates will likely not be sufficient for appropriate inferences.

Here, we provide an overview regarding several ways in which space has been addressed in regression-like models of species–environment relationships. Regression models are frequently used in ecology and conservation to address a variety of problems, ranging from interpreting habitat suitability to forecasting the effects of climate change (Guisan and Zimmermann 2000; Algar et al. 2009). Our overview is

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_6

Libros de Estadística en PDF | Statistics Books in PDF

largely guided by some comprehensive reviews and syntheses on the topic (Keitt et al. 2002; Dormann et al. 2007; Miller et al. 2007; Diniz et al. 2009; Bini et al. 2009; Beale et al. 2010), but we update these syntheses with more recent advances (Crase et al. 2012; Rousset and Ferdy 2014; Bardos et al. 2015; Blangiardo and Cameletti 2015; Ver Hoef et al. 2018). Our goals are threefold. We first describe the problem of spatial dependence on inferences in ecology and conservation. Then, we discuss how to diagnose problems of spatial dependence. Finally, we illustrate common ways to address these statistical problems using a variety of approaches aimed at accounting for spatial dependence in statistical analyses.

6.2 Key Concepts and Approaches

6.2.1 The Problem of Spatial Dependence in Ecology and Conservation

Bivand (1980) was one of the first to explore the importance of spatial dependence on statistical inference from correlation coefficients, a problem that Legendre (1993) later clearly illustrated for ecology. These articles highlight how spatial correlations may create spurious inference and ecological interpretation when spatial dependency of the data is ignored (Fig. 6.1). Depending of the magnitude of spatial autocorrelation (see Chap. 5), parameter estimation can be erroneous and hence our subsequent understanding of ecological patterns and processes: at small values of spatial autocorrelation (e.g., <0.2) the effect tends to be negligible, whereas when the value of spatial autocorrelation is high (e.g., >0.2) then the effect tends to be important and will affect statistical inferences (Bivand 1980). The reason for this problem generally lies in the estimation of uncertainty, where standard errors and confidence intervals around point estimates of correlation coefficients (and other parameters) tend to be artificially narrow. This issue can be considered from the point of degrees of freedom (df), where one df is counted for each independent observation. Yet spatial dependence causes observations to not be independent, such that each observation should not be counted as one df. In effect, this issue essentially leads to "pseudo-replication" in space, a well-known problem for ecology (Hurlbert 1984).

This problem has several practical consequences for conservation. For example, Crase et al. (2014) illustrated that ignoring spatial dependence in forecasts of species response to climate change leads to greater estimated effects of climate change. Ignoring spatial dependence has also been shown to affect conservation planning and understanding habitat suitability for a wide range of species of conservation concern (Carroll and Johnson 2008; Lichstein et al. 2002; Carroll et al. 2010).

Several approaches have been proposed to account for spatial dependence in statistical analyses and modeling (Keitt et al. 2002; Dray et al. 2006; Carl and Kuhn 2010). In the simplest approaches, we might subset data such that sample points are

Libros de Estadística en PDF | Statistics Books in PDF

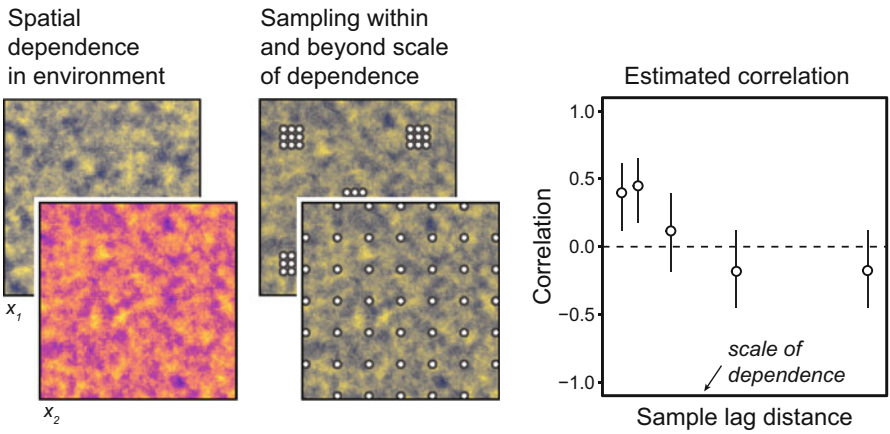


Fig. 6.1 The problem of spatial dependence for ecological inferences. When spatial dependence occurs and is ignored, type I error rates increase. Shown are two, independently derived, environmental variables, x_1 and x_2, that have spatial dependence (generated from a Gaussian random field; see Chap. 5). If sampling occurs within the range of spatial dependence, spurious inferences can occur when such dependence is ignored. In contrast, if sampling is implemented beyond the range of spatial dependence, reliable inference is obtained. Shown are results from Pearson correlation coefficient between the environmental variables taken from five sampling designs that vary in their spatial distribution based on lag distance (each has the same number of samples). Correlations are high when sampling is implemented within the range of spatial dependence, but declines as the lag distance between samples increases

greater than the range of estimated spatial autocorrelation (Chap. 5) (Hawkins et al. 2007), or perhaps just adjust α levels in statistical tests to be more conservative (Dale and Fortin 2014). Some of the most common approaches focus on extending linear regression models to accommodate spatial dependence by either using *autocovariate* variables (Table 6.1) to account for spatial dependence (Augustin et al. 1996; Wagner and Fortin 2005; Betts et al. 2006; Melles et al. 2011) or geostatistical models (see Chap. 5; Cressie 1993). Ordination techniques for community data can also be used to account for spatial structure in the data (see Chap. 11; Wagner 2003, 2004; Dray et al. 2012). Below we explain some of the most common approaches in detail. To do so, we first reintroduce the generalized linear model, which was briefly described in Chap. 2, and use this model framework to build from for accounting for spatial dependence.

6.2.2 The Generalized Linear Model and Its Extensions

Before jumping into approaches aimed at dealing with spatial dependence, we briefly discuss some critical background material. As a reminder, linear regression

Libros de Estadística en PDF | Statistics Books in PDF

Table 6.1 Common terms for spatial regression analysis in ecology

Term	Description
Aerial data	Spatial polygon data that are typically exhaustive tessellations of an area.
Autocovariate	A predictor variable that quantifies the frequency (or related metrics) of the response variable in the surrounding neighborhood.
Autoregressive model	Models that use information on the neighborhood matrix to account for spatial dependence based on deviations from the expected values.
Fixed effect	Deterministic effects that are constant across samples.
Lattice data	Spatial data indexed over a regularly spaced set of points.
Multilevel model	A type of mixed model, where random effects are used to capture hierarchies in the system.
Neighborhood matrix	A square matrix (dimensions are the number of sample points) that quantifies relationships between sampling points, such as binary neighbor connections or distance-weighted linkages.
Random effect	Effects that come from a distribution and vary across samples.
Residual	The difference between the observed value of the dependent variable and the predicted value.
Spatial filtering	When fixed effect covariates in a regression are added that attempt capture the spatial signal through the inclusion of functions of x–y coordinates or related distance metrics.
Tessellation	An arrangement of polygons closely fitted together without gaps or overlapping boundaries.
Trend surface analysis	Analyses where variation in the response variable is expressed as a function of the geographic coordinates of the sampling locations.

and ANOVA are types of linear models (Nelder and Wedderburn 1972). A linear model can be described as:

$$y_i = \alpha + \beta_1 x_i + \varepsilon_i, \tag{6.1}$$

where y_i is the response variable for sampling unit i (e.g., density of a species at a location), α is the intercept, β_1 is the slope (coefficient), x_i is the explanatory variable measured at i, and ε_i is the error, which is assumed to come from a normal distribution and be iid = independent and identically distributed. That is, each residual i is not dependent on other residuals and each comes from the same underlying distribution. This error distribution is assumed to come from a normal distribution with a mean of zero and an unknown finite variance, written as $\varepsilon_i \sim N(0, \sigma^2)$. Plotting the residuals of the model, or the deviation of the predictions to the observed data for a given value of x (Fig. 6.2), helps understand whether this assumption is met. Note that the equivalence of linear regression and ANOVA in this framework can be seen by considering categorical treatments (x_i) in an ANOVA as "dummy" variables (e.g., 0, 1 variables) in a regression model.

Linear models can be extended in two very useful ways. The first major extension, the *generalized linear model* (GLM), allows for alternative distributions for the response variable other than the normal distribution. These other distributions specifically come from the exponential family of distributions, which includes

Libros de Estadística en PDF | Statistics Books in PDF

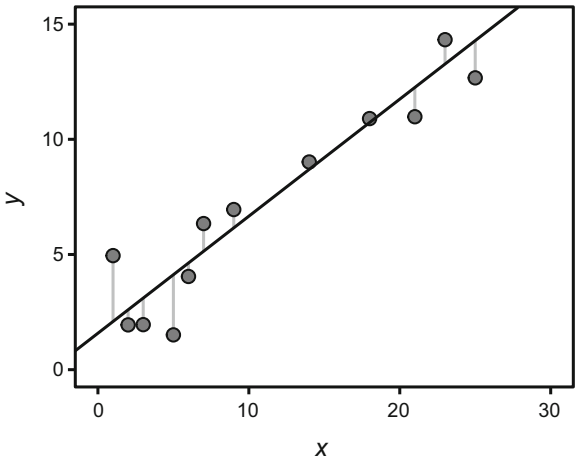


Fig. 6.2 A linear regression model and the residuals from that model. In standard regression techniques, residuals (the difference of the observed value from the predicted value of the response variable for a given value of the explanatory variable) are assumed to be independent and identically distributed. When spatial autocorrelation occurs in the residuals of models, such autocorrelation can impact inference if ignored. Dots represent observed data, black line is the prediction from the linear model, and the vertical gray lines represent residuals

distributions such as the Poisson, binomial, Bernoulli, and gamma distributions. This extension greatly increases the flexibility of these models, allowing for responses such as the presence/absence of a species at a sampling location (a Bernoulli distribution). The classic text for generalized linear models is McCullagh and Nelder (1989). In GLMs, we specify a link function and a distribution for the errors (ε).

Perhaps the two most common types of GLMs are logistic regression and Poisson (or log-linear) regression. For logistic regression we have:

$$\text{logit}(p_i) = \alpha + \beta_1 x_i, \tag{6.2}$$

where p_i is the expected probability of a "success" and a "logit" link function is used (i.e., $\log(p_i/(1 - p_i))$). In this case, we assume a binomial error distribution. A binomial distribution can be thought of as a distribution that arises from a series of coin tosses. If there is only one toss, it is called a Bernoulli distribution; if there is more than one toss (sometimes referred to as "trials"), then the distribution is called a binomial distribution. In the latter case, we are interested in the frequency or proportion of "successes" out of the total number of trials.

For a Poisson regression, we have:

$$\log(\lambda_i) = \alpha + \beta_1 x_i, \tag{6.3}$$

where λ_i is the expected count for sample i and we use a "log" link function and assume a Poisson error distribution. The Poisson distribution is a discrete

Libros de Estadística en PDF | Statistics Books in PDF

distribution were values are integers greater than or equal to zero (i.e., negative values are not allowed). The Poisson distribution assumes that the mean equals the variance, which is often a restrictive assumption. A related distribution that relaxes this assumption in the negative binomial distribution. There are several other types of GLMs; however, we will focus on only a few in this book. Interested readers should see Bolker (2008) and Bolker et al. (2009) for the use of GLMs in ecology.

The second major extension of a linear model is to allow for random effects, what is frequently termed a random-effects model, or if fixed effects are considered alongside random effects, a mixed model. *Random effects* can be contrasted with *fixed effects* (the β above) in several ways. Random effects are extremely flexible in how they can accommodate complex data structures and provide inference unattainable with fixed effects. Some uses for random effects include: (1) conditional inference-when you would like to make inferences on a particular sampling unit, location, etc. (e.g., a particular watershed contained within the study area); (2) accommodating block, split-plot, Latin-square, and other treatment structures in experiments; (3) more generally accounting for both temporal and spatial dependencies in data, such as temporal repeated measures or spatial autocorrelation; (4) when one thinks treatment effects may vary in space or time (similar to including an "interaction" term in a linear model); and (5) "broad-sense" inference: making inferences for an entire region/population from a sample (in contrast to "narrow-sense" inference, where we make inferences only for the specific samples or locations being considered) (Littell et al. 2006; Gelman and Hill 2007; Zuur et al. 2009).

There has been some confusion in ecology regarding when an effect should be considered random versus fixed, and how inferences may change depending on whether a variable is considered random or fixed. Gelman and Hill (2007) discussed how random effects have been loosely described and used in the literature, and the resulting problems that have arisen. We do not focus on this issue; rather we will simply consider mixed models as one means to accommodate spatial dependence.

We can formally describe a linear mixed model as:

$$y_i = \alpha + \beta_1 x_i + \gamma + \varepsilon_i, \tag{6.4}$$

where γ is a random effect and is typically assumed to be distributed $\sim N(0, \sigma^2)$. When we put these two extensions together, we have *generalized linear mixed models* (GLMMs), which are very powerful models that are seeing increasing use in ecology, evolution, and conservation (Bolker et al. 2009; Thorson and Minto 2015). Note that we can also model the variance, σ^2, as a variance–covariance matrix, which is how we specifically extend this model to explicitly account for spatial dependence, as we will see below.

Libros de Estadística en PDF | Statistics Books in PDF

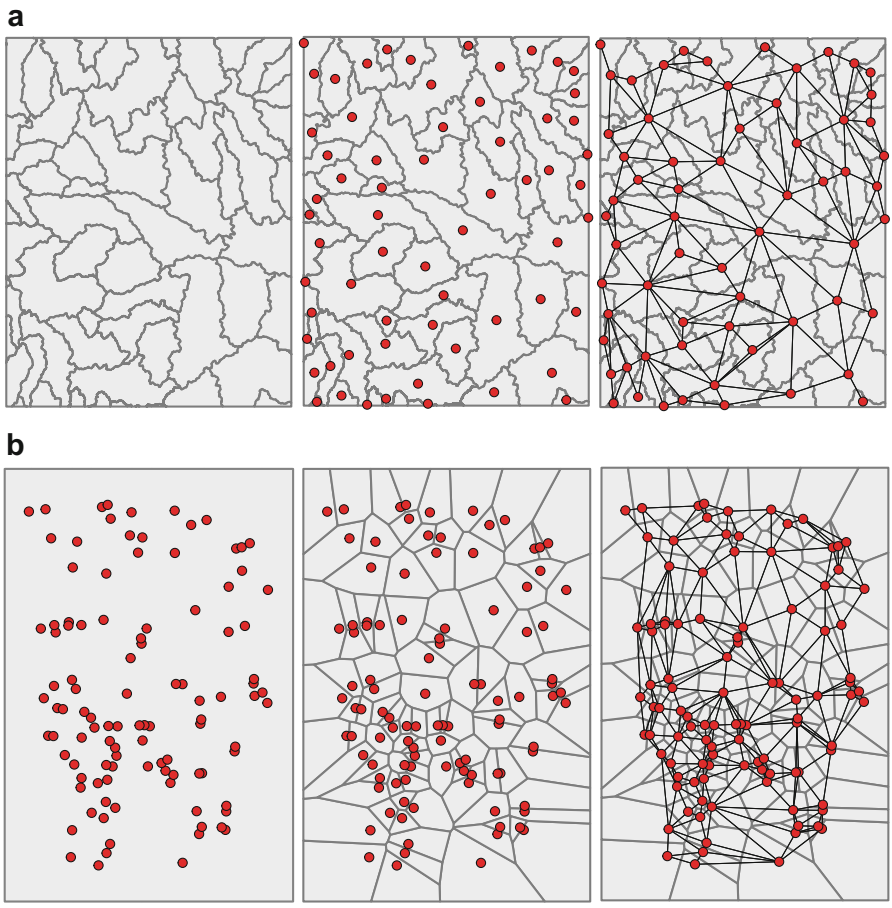


Fig. 6.3 Examples of aerial data used in spatial modeling. Aerial data can come from (**a**) polygon-based information (e.g., maps of counties, watersheds, etc.) or can be generated (**b**) from point or line data using Voronoi tessellation. In either approach, we can describe spatial dependence through the links among locations (right panel) with a spatial neighborhood (weights) matrix

6.2.3 General Types of Spatial Models

The vast diversity of spatial regression-like models can be organized in several ways. Three important properties include: (1) the type of response data (quantitative, count, presence–absence); (2) whether samples are irregularly spaced samples across continuous space or lattice/gridded data that are discrete in nature (Fig. 6.3); and (3) the way in which spatial dependence is considered.

The type of response data used will ultimately affect the type of regression model being fit. Different types of response data lend themselves to different distributions

Libros de Estadística en PDF | Statistics Books in PDF

used in GLM-like models. Overall, most of the approaches to spatial dependence have been better developed for normally distributed response variables than non-normally distributed response variables (Beale et al. 2010). Dealing with non-normal response data is generally more challenging than normally distributed data. For instance, data such as presence–absence data (0/1 data) have much less information content than normally distributed response variables, which impact the ability to identify, interpret, and account for spatial dependence in models.

Samples of data frequently come from *aerial data* (or *lattice data*) where neighborhoods are considered, such as samples arising from counties or watersheds. In such cases, spatial dependence is frequently considered based on neighboring polygons or related neighbors through the use of a *neighborhood matrix* (or spatial weights matrix). In contrast to aerial data, samples can also come from points across a study region. In this case, information on *x*–*y* coordinates are used either directly (e.g., using an *x*-coordinate as a predictor) or indirectly (e.g., by calculating distances between pairs of points).

Models can also be categorized based on how spatial structure is considered. For some models, often referred to as *spatial filtering* models (Getis and Griffith 2002), space is considered as predictor variables in a regression, where we attempt to "filter out" the spatial signal through the inclusion of functions of *x*–*y* coordinates or related distance metrics. In these cases, spatial dependence is thought to be largely dominated by exogenous drivers such as spatial dependence in environmental gradients, and often (but not always) occurs at relatively large scales (Fortin et al. 2012). In contrast, other models focus specifically on accounting for spatial dependence in the error terms of regression models. These models frequently assume spatial dependence is more localized and dominated mostly by endogenous processes (e.g., localized dispersal, species interactions) (Fortin et al. 2012; Teng et al. 2018).

6.2.4 Common Models that Account for Spatial Dependence

6.2.4.1 Trend Surface Analyses

Trend surface analyses use *x*–*y* coordinates in an attempt to capture large-scale spatial dependence in a region. There have been two common ways in which coordinates are added to regression models: polynomial regression (Haining 2003) and generalized additive models (GAMs) (Zuur et al. 2009).

The idea of trend surface analysis with polynomial regression is simply to include *x*–*y* coordinates and their polynomials (e.g., x^2, x^3, etc.) in the regression as covariates (Legendre 1993). Incorporating coordinates in this way is thought to be useful to deal with large-scale dependencies arising from exogenous processes (e.g., climate gradients across a geographic range), but it may be more limited in accounting of local autocorrelation. Legendre (1993) suggested simply adding quadratic and cubic terms for *x*–*y* coordinates to the regression model (Fig. 6.4). Adding quadratic and cubic terms allows for some potential non-linear responses across geographic

Libros de Estadística en PDF | Statistics Books in PDF

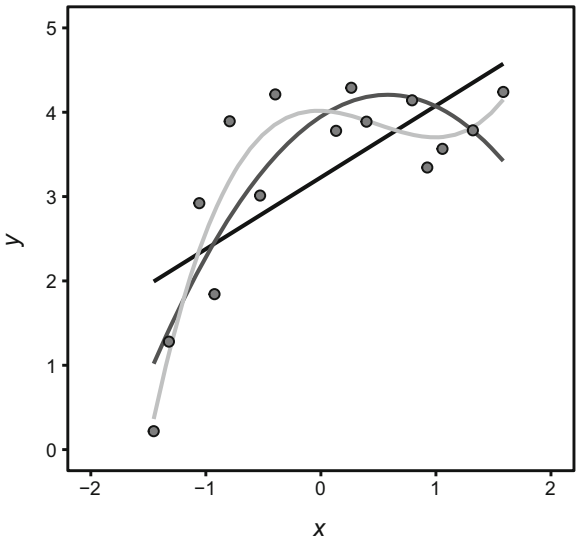


Fig. 6.4 Incorporating polynomial terms into a regression model to account for non-linearity in environmental relationships. Shown is an example of a linear model, contrasted with a model that adds a quadratic term, and a model that includes both a quadratic and cubic term

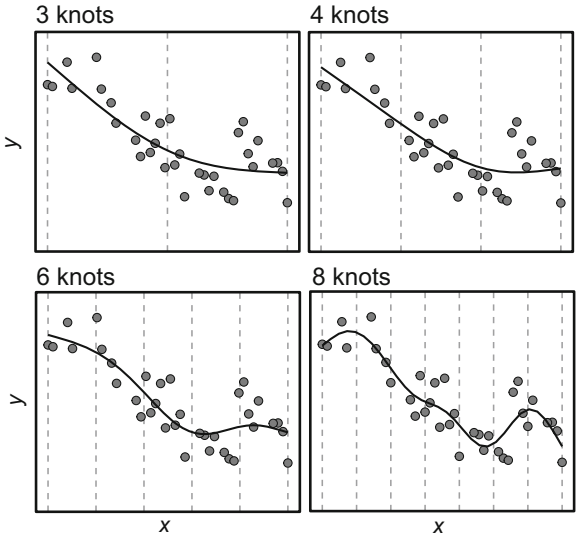


Fig. 6.5 The generalized additive model (GAM) and the concept of smoothers. Shown are GAMs fit to the data based on different numbers of knots (vertical lines; ranging from 3 to 8 knots). Within each knot, a simple spline (e.g., a cubic spline; see Fig. 6.4) is fit, with the constraint that splines must connect at the knots. As the number of knots increases, the complexity of the smoother function increases. Modified from Zuur et al. (2009)

space. Note that trend surface analysis will not formally adjust estimates of fixed effects for uncertainty due to spatial dependence (unlike mixed models, see below), but they may account for dependence in model residuals.

Generalized additive models (GAMs) (Hastie and Tibshirani 1986; Wood 2006) can be used in a similar way to trend surface analysis based on polynomial regression. GAMs use a class of equations called "smoothers" that attempt to generalize data into smooth curves by local fitting to subsections of the data (Fig. 6.5). This approach allows for more flexibility in capturing non-linearity in responses across

Libros de Estadística en PDF | Statistics Books in PDF

geographic space and GAMs have frequently been used in species distribution modeling more broadly (see Chap. 7). The simplest example of a smoother that is likely to be familiar to scientists is the running average, where one calculates the average value of data in a "window" across values of a covariate. While the running average is an example of a smoother, much more efficient smoothers have been developed. LOWESS (i.e., locally weighted regression; Cleveland 1979) is one example of a more efficient smoother used in some GAMs. The idea is to plot the value of the dependent variables (e.g., occurrences) along a single environmental variable, and then to calculate a smooth curve that fits the data as closely as possible while being parsimonious based on some sort of criterion. The algorithm fits a smooth curve to each variable and then combines the results additively. The approach generally employed with GAMs is to divide the data into some number of sections, using "knots" at the ends of the sections. Then a low-order polynomial or spline function (a spline is a function of polynomials relationships stitched together) is fit to the data in the section, with the added constraint that the second derivative of the function at the knots must be the same for both sections sharing that knot. This latter criterion eliminates kinks in the curve, and ensures that it is smooth and continuous (Fig. 6.5).

6.2.4.2 Eigenvector Mapping

Eigenvector mapping extends the general eigenvector approach described in Chap. 5 by using eigenvectors that describe different scales of spatial structure as predictors in regression models (Dray et al. 2006; Griffith and Peres-Neto 2006). In effect, this is somewhat similar to a trend surface model, but where eigenvector values, rather than x–y coordinates, are used as predictors. The ability of this approach to capture multiple scales of potential spatial structure is a relatively unique benefit in contrast to other approaches. Because each eigenvector captures spatial patterns at different scales, the combined use of several eigenvectors can potentially address problems of anisotropy and non-stationarity in spatial autocorrelated data. However, this approach and related techniques can sometimes lead to bias in coefficients of fixed effects and may not improve Type I error rates (Beale et al. 2010; Emerson et al. 2015).

Spatial eigenvectors are derived from a distance matrix from sample points, typically through the use of principal coordinates analysis (PCoA) on distance matrices (see Chap. 5; Dray et al. 2006). In this approach, a pairwise distance matrix is first calculated between all sampling points. This distance matrix is converted to a binary connectivity (or weights) matrix based on some distance threshold that allows for a minimum representation of connectivity among all points. For instance, a "minimum spanning tree," which is the minimum set of links that ensures all points being considered are connected, is often used as a parsimonious way to guarantee connectivity among all points considered (see below). With this binary connectivity matrix, PCoA (also known as classic multidimensional scaling) is performed. PCoA generates new components that capture the variation in the distance matrix, which

Libros de Estadística en PDF | Statistics Books in PDF

are summarized with eigenvalues and eigenvectors, similar to Principal Components Analysis (Legendre and Legendre 1998). The set of eigenvectors that reduce or eliminate spatial autocorrelation in the residuals of the models is then identified. This can be assessed through the use of Moran's *I* on the residuals of models that include eigenvectors as predictors (Dray et al. 2006). Those eigenvectors that reduce autocorrelation the most are then used as predictors in a standard regression model to "filter out" spatial dependence.

6.2.4.3 Autocovariate Models

In these and related models, we typically work with "areal" or "lattice" data, rather than point-based samples. Autocovariate regression is similar to linear regression, but an *autocovariate* is included into the regression model. This autocovariate can be defined in various ways, such as a weighted mean of the response variable in surrounding locations (Augustin et al. 1996):

$$\text{auto}_i = \frac{\sum_{j=1}^{k_i} w_i y_i}{\sum_{j=1}^{k_i} w_i}, \tag{6.5}$$

where auto_i is the spatially weighted mean of the response variable, y, in the neighborhood (with a neighbor set k_i, reflecting the size of neighborhood considered) around sample i. This autocovariate is frequently calculated based on first-order neighbors (e.g., adjacent polygons or surrounding eight cells in a lattice), but the idea can be extended to account for further away points, typically weighting points based on the inverse of the distance (samples farther away get less weight than those closer to the sample). This approach can be used in a generalized linear model context; for instance, applying autocovariates in logistic regression, termed autologistic regression, is a common approach in ecology (Augustin et al. 1996; Wintle and Bardos 2006).

In effect, this approach assumes that if nearby locations are occupied, there should be a greater likelihood that the focal point is occupied. This is a relatively simple approach, although in practice, it was shown to not perform well because it can cause bias in coefficients of fixed effects for environmental predictor variables (Dormann et al. 2007; Beale et al. 2010). In these cases, autocovariate models tended to de-emphasize the effect of the environmental covariates, while overemphasizing the effects of autocovariates, leading to Type II error in inferences on environmental covariates. This issue is at least partly driven by the fact that the autocovariate is calculated on the raw data before fitting the explanatory variables, even though explanatory variables may contain spatial dependence that can reduce spatial dependence in the residuals of models (Crase et al. 2012). There are also difficulties with using these models to interpolate (or extrapolate) to new locations (see below).

Crase et al. (2012) developed a related approach in which autocovariates are quantified from the residuals of models, rather than the raw data, termed the residual

Libros de Estadística en PDF | Statistics Books in PDF

autocorrelation approach (RAC). This approach replaces the use of the raw data (y_i in Eq. (6.5)), with $y_i - q_i$, where q_i is the fitted value from an environment-only model that ignores autocorrelation. This leads to an autocovariate that captures only the variance not explained by explanatory variables. Crase et al. (2012) argued that this approach better captures spatial dependence than using standard autocovariates because explanatory variables are fitted first to the data.

Bardos et al. (2015) raised concerns regarding the validity of prior analyses (Dormann et al. 2007; Beale et al. 2010) that emphasized bias in auto-models. They show that a weighting scheme based on weighted means (Eq. 6.5) is not valid for autocovariate models. Rather a weighted sums scheme should be used instead:

$$\text{auto}_i = \sum_{j=1}^{k_i} w_i y_i. \tag{6.6}$$

This weighting scheme has not been evaluated as thoroughly as a weighted means approach described above, but Bardos et al. (2015) illustrated that it may perform better, in terms of capturing autocorrelation and providing unbiased estimates of fixed effects.

6.2.4.4 Autoregressive Models

Autoregressive models work with aerial or lattice data, similar to autocovariate models. The difference lies in how spatial dependence is captured with these model formulations. Two common *autoregressive models* are simultaneous autoregressive models (SAR) and conditional autoregressive models (CAR) (Lichstein et al. 2002; Ver Hoef et al. 2018). In both SAR and CAR, spatial dependence is captured through the use of a spatial neighborhood weights matrix akin to autocovariate models, but dependence is described based on deviations from the expected value given the covariates (Keitt et al. 2002).

SAR and CAR models share several similar features. In practice, a primary difference is that SAR can accommodate anisotropic spatial dependence, while CAR cannot. Nonetheless, the CAR is often used. Also, note that some work suggests that both CAR and SAR perform well on regular lattices, but suffer diminished performance on irregular lattices (e.g., county or watershed data) (Wall 2004). Both of these models use a spatial weights matrix, $\mathbf{W}$, that captures the neighborhood surrounding sampling locations. Typically, $\mathbf{W}$ is a binary matrix that identifies neighbors, but it could also include non-binary data.

The general CAR model can be written in matrix notation as:

$$\boldsymbol{y} = \boldsymbol{\beta}\mathbf{X} + \rho\mathbf{W}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}) + \varepsilon, \tag{6.7}$$

Libros de Estadística en PDF | Statistics Books in PDF

where ρ is the first-order autocorrelation between neighbors, $\boldsymbol{\beta}$ is a vector of coefficients (i.e., slopes) related to the explanatory variables $\mathbf{X}$ described through the "design matrix" (i.e., a $N \times K$ matrix of explanatory variable values for each sample of data used in model fitting, where N is the total number of samples and K is the total number of explanatory variables). In this equation, $\boldsymbol{\beta}\mathbf{X}$ is the same as a standard regression (Eq. 6.1) written in matrix form (i.e., Eq. 6.1 can be rewritten in matrix form as $\mathbf{y} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon}$), such that the only difference in this equation and a standard GLM is that the ε in the standard GLM is now broken into $\rho\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \varepsilon$. The $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ captures the deviation of the observed data from that expected from the covariates and this is multiplied by the correlation for the neighbors ($\rho\mathbf{W}$**;** note that this only captures the neighbors because $\mathbf{W}$ is 0 for all non-neighbors).

There are several types of SAR models that capture different kinds of spatial dependence, which assume that the dependence occurs in the response variable, predictor variables, or the error (Dormann et al. 2007). The general SAR model can be written in matrix notation as:

$$\boldsymbol{y} = \boldsymbol{\beta}\mathbf{X} + \rho\mathbf{W}\boldsymbol{y} + \varepsilon. \tag{6.8}$$

While there are several types of SAR models, Ver Hoef et al. (2018) did not recommend the use of certain specifications of SAR models for ecological data, such as the use of "lag" or "SAR mixed models." See Kissling and Carl (2008), Dale and Fortin (2014), and Ver Hoef et al. (2018) for more details.

6.2.4.5 Multilevel Models

The effects of potential spatial dependence can be also handled by using "multilevel" or "hierarchical" modeling. This type of modeling is a natural extension of generalized linear models, where we specify random effects to account for dependencies (correlations and hierarchical structure) in the data. Thus, multilevel models can be considered one type GLMM. An excellent text on this approach is Gelman and Hill (2007). Keitt et al. (2002) also touched on this approach when they contrasted "blocking" with other approaches to addressing spatial dependence.

Multilevel models are relevant when there is a natural hierarchical structure to the data being used (Fortin et al. 2012). For example, point samples may be collected in a grid or along a transect (with replicate grids or transects across a region), samples may be nested with counties or watersheds nested within larger regions (e.g., states), etc. In the absence of such sampling structure, this framework may not be helpful for accounting for spatial dependences. Some reasons to consider multilevel models with spatial data: (1) it can accommodate using all the data to perform inferences when some groups or blocks have small sample size; (2) it provides more efficient inference for regression parameters; (3) it can appropriately include predictors at >1 level in a hierarchy (e.g., within patch, patch, and landscape predictors); and (4) it can provide correct estimates of uncertainty (standard error, confidence interval, etc.) (Gelman and Hill 2007). For example, if we collect multiple samples within patches

Libros de Estadística en PDF | Statistics Books in PDF

and sample across different landscapes or region, we could specify a multilevel regression as follows:

$$y_{i,p,l} = \alpha + \beta_1 x_i + \gamma_p + \delta_l + \epsilon_i, \tag{6.9}$$

where γ_p is a random effect of a patch, and δ_l is a random effect of the landscape or region. In doing so, this formulation acknowledges that observations within each region have some correlation/similarity.

6.2.4.6 Generalized Least Squares and Spatial Mixed Models

Generalized least squares models (GLS) and spatial mixed models are similar in scope to a multilevel model. The main conceptual difference is that we specify spatial correlation structures explicitly in the random effects (GLMMs) or residuals (GLS) by modeling the variance–covariance matrix over space.

In a GLS spatial model, we take a typical regression, $y_i = \alpha + \beta_1 x_i + \epsilon_i$, where ϵ is $\sim N(0, \sigma^2)$ and replace the variance on the error term with a variance–covariance matrix: $\varepsilon \sim N(0, \Sigma)$ (Keitt et al. 2002). In a GLMM spatial model, a similar approach is taken, but a variance–covariance matrix is added for the random effect: $\gamma \sim N(0, \Sigma)$ rather than the residuals (Littell et al. 2006). In both cases, parametric correlation functions are fit to explain the variance–covariance matrix by specifying model-based correlation structures, akin to model-based variogram structures we described in Chap. 5. These correlation structures are sometimes referred to as Gaussian random fields (Thorson and Minto 2015). For example, in the GLS we will consider below, we will fit a spatial exponential covariance (see Chap. 5):

$$\sum = \sigma^2 \begin{bmatrix} 1 & \exp\left(-\frac{d_{ij}}{\alpha}\right) \\ \exp\left(-\frac{d_{ij}}{\alpha}\right) & 1 \end{bmatrix}, \tag{6.10}$$

where σ^2 is the non-spatial variance estimated, d_{ij} is the distance between two observations i and j, and α is a parameter to be estimated (related to the range). With mixed effects, one can specify models that only account for spatial autocorrelation within the regions/groups specified by the random effect. For instance, Fletcher (2005) used this general approach to account for within-patch spatial dependence of species distribution while assuming that among-patch dependence was negligible. Similar to CAR and SAR, GLS has a strong foundation for normally distributed response variables, but the application of these models to non-normal data is more challenging (Rousset and Ferdy 2014). Note the utility of GLS may depend upon the scale of environmental relationships being considered. For instance, Diniz et al. (2003) found that GLS tended to de-emphasize covariates

Libros de Estadística en PDF | Statistics Books in PDF

operating across large spatial scales while overemphasizing covariates operating at more local scales.

6.2.5 *Inference Versus Prediction*

An implicit but pervasive issue regarding spatial regression and other modeling approaches considered in this book regards whether the goal of the work is for inference or prediction. When our goal is inference, we are interested in estimating factors influencing response variables (Stephens et al. 2007). In contrast, if our goal is prediction, we aim to build models that can make accurate predictions or projections across space and time (Boyce et al. 2002), including both interpolating between sample locations and predicting to new areas (i.e., model transferability; see Chap. 7 for more). Ecologists and conservation biologists often use models in both ways, but ultimately these are very different goals and model building and evaluation will be (or should be) different depending on the goal.

Spatial regression models can be helpful in problems of inference, where we are interested in understanding spatial or environmental relationships, such as factors related to species distribution and abundance. These approaches can potentially provide more reliable inference in regard to parameter estimates and their uncertainty, as well as more reliable statistical hypothesis tests. However, the use of these models for prediction, projection, or interpolation can sometimes be difficult, depending on the type of model considered. For example, with autocovariate models, prediction requires information on the response variable (e.g., occurrence) across the region being predicted, because the regression model includes this information in the form of the autocovariate (Augustin et al. 1996). In contrast, trend surface and related spatial filtering models are straightforward to use in prediction because only the physical locations are used as predictors. In some cases, spatial regression models are used for prediction where the dependence term is ignored (e.g., using only the fixed effects from a mixed effects model). Depending on the goal of spatial modeling, the utility of the above approaches may vary.

6.3 Examples in R

6.3.1 *Packages in R*

In R, there are a few libraries that can be used for spatial regression models. We use the `mgcv` package for fitting generalized additive models (Wood 2006), `lme4` for fitting multilevel models (Bates et al. 2015), and `vegan` (Oksanen et al. 2018) and `spdep` (Bivand and Piras 2015) for fitting eigenvector maps. We use the `spdep` package for models requiring lattice data (autocovariate, SAR, CAR) and interpreting autocorrelation in the residuals of models. We fit spatial GLS and mixed models with

Libros de Estadística en PDF | Statistics Books in PDF

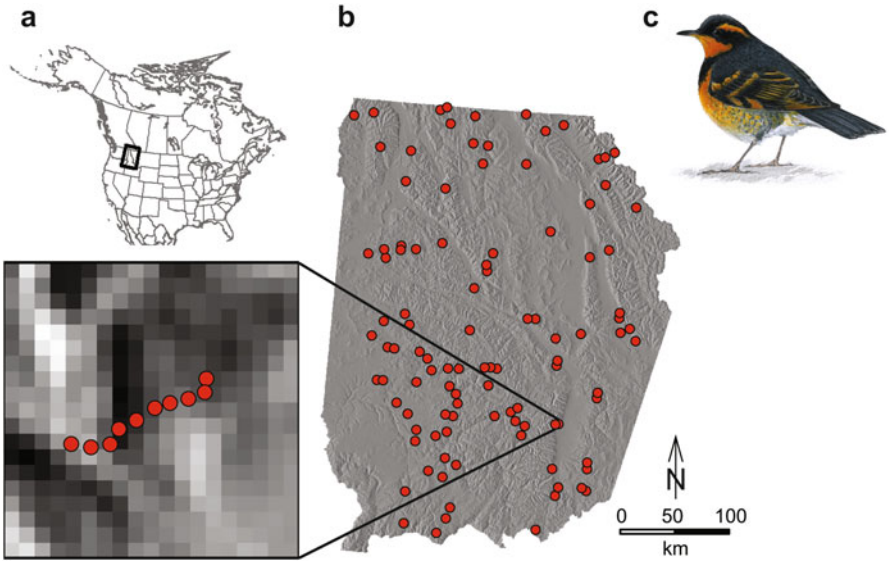


Fig. 6.6 The Northern Region Landbird Monitoring Program applies a hierarchical sampling design for surveying bird communities. This monitoring program covered (**a**) northern Idaho and western Montana, where (**b**) transects were distributed across different watersheds, with typically 10 points per transect. Here we focus on the occurrence of (**c**) varied thrush (picture courtesy of Matthew Dodder at http://www.birdguy.net/)

`MASS` (Venables and Ripley 2002) and `spaMM` (Rousset and Ferdy 2014), but other packages can be used, particularly Bayesian modeling packages (e.g., `spBayes`) (Finley et al. 2015).

6.3.2 The Data

Monitoring programs are often hierarchically structured and filled with spatio-temporal dependence in the data. The Northern Region Landbird Monitoring Program is one such example (Hutto and Young 2002). Sampling locations consisted of point counts (100-m radius), along a transect (typically 10 points/transect; transects are approximately 3 km long), with transects randomly selected within USFS Forest Regions across Montana and Idaho (Fig. 6.6). Ten-minute point counts were conducted by trained observers, where all birds seen or heard were recorded. Here we only consider birds detected within 100-m of the point. These points were also resampled over time, although we will not consider these temporal repeated measures here.

To interpret spatial regression models, we consider a simple environmental relationship of species occurrence along an elevation gradient. Elevation is

Libros de Estadística en PDF | Statistics Books in PDF

frequently considered to be an important, though often indirect, factor correlated with species distribution. We focus on the occurrence of the varied thrush (*Ixoreus naevius*) (Fig. 6.6), a migratory bird that breeds in the western USA. Varied thrush have declined in the western USA over the past several decades, based on Breeding Bird Survey data (Sauer et al. 2017), with the annual decline of approximately 2–3% per year (1966–2015: −2.47, 95% CI: −3.19, −1.79; 2005–2015: −3.32, 95% CI: −5.14, −1.56). Furthermore, they are often considered an old-growth, interior species (Brand and George 2001; Betts et al. 2018). Consequently, this species has been of some interest for conservation.

We fit logistic regression models and their spatial extensions to infer and predict the distribution of varied thrush as a function of elevation in this mountainous region. Here we focus on modeling detection/non-detection of thrushes (0/1 data). Elevation was derived from a 30-m resolution Digital Elevation Model (DEM). Prior to analysis, all GIS layers were aggregated to a common 200-m resolution, reflecting the grain of the sampling unit (100-m-radius point counts).

With this sampling design, there are likely observation errors in detecting varied thrushes, such that models that explicitly account for imperfect detection would be useful (McCarthy et al. 2012). Rota et al. (2011) estimated that detection probabilities of varied thrushes with this sampling design was relatively high ($p = 0.87$/count), which is likely driven by their distinctive and loud song. We do not consider that sampling error here to focus specifically on the problem of spatial dependence. See sect. 6.4 for further discussion on sampling errors.

6.3.3 *Models that Ignore Spatial Dependence*

To begin, we import a raster layer of elevation with the `raster` package and use this layer to also derive other key variables related to elevation, such as slope and aspect (Fig. 6.7).

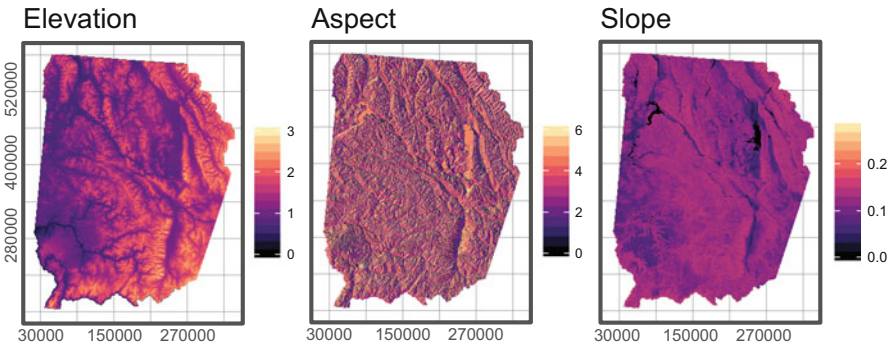


Fig. 6.7 The raster data considered come from a digital elevation model, including elevation (in km), aspect (in radians), and slope. Note that slope is double square-root transformed for visualization

Libros de Estadística en PDF | Statistics Books in PDF

Table 6.2 Terrain metrics that the `raster` package can calculate based on elevation data

Metric	Description
Aspect[a]	The compass direction that a slope faces.
Flow direction	The direction of the greatest drop in elevation (or smallest rise) from focal cell and its eight neighboring cells, coded as integer values in powers of 2 (starting east of focal cell and moving in a clockwise direction; 1, 2, 4, 8, 16, 32, 64, and 128).
Roughness	Difference between the maximum and minimum value of the focal cell and its eight neighboring cells.
Slope	The change in elevation, described as the difference in elevation between two points divided by the distance.
Terrain ruggedness index	Absolute difference between the value of a cell and the value of its eight neighboring cells (eight-neighbor rule).
Topographic position index	Difference between the value of a cell and the mean value of its eight neighboring cells (eight-neighbor rule).

[a]Can be measured in degrees or radians

```
> library(raster)
> elev <- raster("elev.gri")

#create aspect and slope layers from the elevation layer
> elev.terr <- terrain(elev, opt = c("slope", "aspect"), unit =
  "radians")
```

The `terrain` function in the `raster` package takes an elevation layer (e.g., DEM) and returns raster layers that are calculated from elevation, including slope, aspect, topographic position index, terrain ruggedness index (TRI), roughness, and flow direction (Wilson et al. 2007) (Table 6.2). Here, we just calculate slope and aspect (Fig. 6.7). Note that for this function, the projection must be set on the raster layer for implementation. This function defaults to aspect being calculated in radians, using the algorithm in Horn (1981).

We can merge the slope and aspect layers into a single raster stack that holds all raster layers. We create a single object that holds all of the data with the `stack` function:

```
#makes a multilayered file for extraction
> layers <- stack(elev, elev.terr)
> names(layers) <- c("elev", "slope", "aspect")
```

We first consider a non-spatial logistic regression model. To do so, we use the `extract` function to grab covariate values from `layers` at the sample locations from the survey data and we then merge the covariates with our data on thrush occurrence using `cbind`.

```
> point.data <- read.csv("vath_2004.csv", header=T)
> coords <- cbind(point.data$EASTING, point.data$NORTHING)
> land.cov <- extract(x = layers, y = coords)
> point.data <- cbind(point.data, land.cov)
```

Libros de Estadística en PDF | Statistics Books in PDF

We consider a simple set of logistic regression models. We expect that elevation may help explain varied thrush occurrence, where thrushes may be most likely to occur at either low or moderate elevations. Consequently, we consider quadratic terms in the logistic regression model to account for potential non-linearities (Fig. 6.4) in occurrence as a function of elevation. We also consider slope and aspect as proxies for local variation in environmental conditions. First, we transform the explanatory variables to a mean of 0 and a variance of 1 (sometimes referred to as a z-transformation or "centering and scaling"). Centering and scaling can help improve model convergence and facilitates comparing coefficients for different parameters.

```
> point.data$elevs <- scale(point.data$elev, center = T, scale = T)
> point.data$slopes <- scale(point.data$slope, center = T, scale = T)
> point.data$aspects <- scale(point.data$aspect, center = T, scale = T)
```

Note that the default for the `scale` function is to both center and scale, but we explicitly request this here to illustrate. Now we can fit logistic regression models of varying complexity.

```
> VATH.elev <- glm(VATH ~ elevs, family = "binomial", data =
  point.data)
> VATH.all <- glm(VATH ~ elevs + slopes + aspects, family =
  "binomial", data = point.data)
> VATH.elev2 <- glm(VATH ~ elev + I(elev^2), family = "binomial", data
  = point.data)
```

Note that to specify a quadratic term in R, we write `I(elev^2)`. This could also be accomplished through the `poly()` (see below). We can contrast model fit using AIC:

```
> round(AIC(VATH.elev, VATH.all, VATH.elev2), 2)

##
df AIC
VATH.elev 2 583.10
VATH.all 4 584.84
VATH.elev2 3 566.54

> summary(VATH.elev2)

##
Call:
glm(formula = VATH ~ elev + I(elev^2), family = "binomial", data = point.
data)

Deviance Residuals:
 Min 1Q Median 3Q Max
-0.6088 -0.5787 -0.5032 -0.3231 3.0804
```

Libros de Estadística en PDF | Statistics Books in PDF

```
Coefficients:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.984 1.990 -4.012 6.01e-05 ***
elev 10.698 3.227 3.316 0.000915 ***
I(elev^2) -4.476 1.281 -3.494 0.000475 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 584.34 on 804 degrees of freedom
Residual deviance: 560.54 on 802 degrees of freedom
AIC: 566.54

Number of Fisher Scoring iterations: 6
```

For each of these models, we can use the `summary` function to view the coefficients estimated from the model and related diagnostics. While this small set of candidate models is far from a complete set, from this comparison there is some evidence of thrush occurrence increasing at moderate elevations. This can be concluded because the linear elevation term is positive while the quadratic term is negative (both of which are significant based on *p*-values, or `Pr(>|z|)`, which will cause a humped-shaped relationship with elevation. We can plot this relationship by first generating a new data set to predict onto and then use the `predict` function (Fig. 6.8):

```
> elev <- seq(min(point.data$elev), max(point.data$elev), length = 15)
> newdata <- data.frame(elev = elev)

> glm.pred <- predict(VATH.elev2, newdata = newdata, type = "link", se = T)
> ucl <- glm.pred$fit + 1.96*glm.pred$se.fit
> lcl <- glm.pred$fit - 1.96*glm.pred$se.fit
```

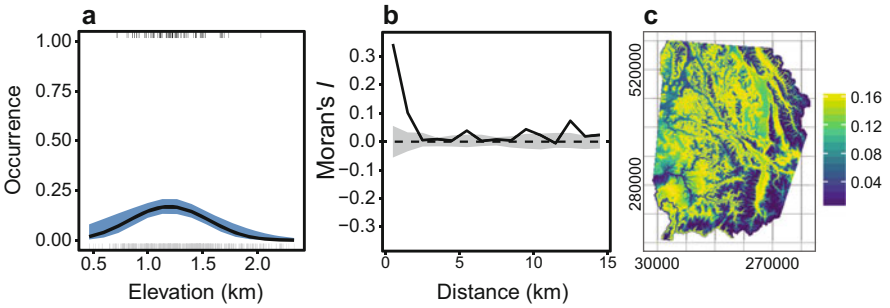


Fig. 6.8 (**a**) Predicted relationship (with 95% prediction intervals) of varied thrush occurrence with elevation based on a standard logistic regression model. (**b**) Correlogram using the raw response data, where gray region shows the 99% null envelope from a permutation test. (**c**) Mapping predictions from model

```
#create data frame and back-transform to probability scale
> glm.newdata <- data.frame(newdata, pred = plogis(glm.pred$fit),
  lcl = plogis(lcl),ucl = plogis(ucl))

> plot(glm.newdata$elev, glm.newdata$pred, ylim = c(0,0.5))
> lines(glm.newdata$elev, glm.newdata$lcl)
> lines(glm.newdata$elev, glm.newdata$ucl)
```

We can also plot predictions of this model across the study region by predicting onto the raster stack layer. To do so, the `raster` package defaults to making predictions on the link scale, but we can then back-transform the predictions on the raster to the probability scale (Fig. 6.8c).

```
> glm.raster <- predict(model = VATH.elev2, object = layers)
> glm.raster <- exp(glm.raster) / (1 + exp(glm.raster))
> plot(glm.raster, xlab = "Longitude", ylab = "Latitude")
```

In this model and subsequent models, we will generally focus on two issues. First, is there evidence for spatial autocorrelation in the residuals of the models? Second, how do estimated relationships, that is, the coefficients and standard errors (SEs), change depending on the model?

We can determine if spatial dependence might be problematic for inferences by considering if there is evidence for spatial dependence in the residuals of the model (Dormann et al. 2007; Beale et al. 2010). First, we consider if there is spatial autocorrelation in the response variable. For interpreting spatial autocorrelation, we will use the correlogram function described in Chap. 5 when we used the `spdep` package. The function in Chap. 5 was altered to allow specification of different bins for lag distances and the maximum distance considered. This function is useful because it can be readily used for both binary data (0/1 response data) and for other response variable distributions (e.g., residuals), although other functions, such as the `correlog` function in the `ncf` package (Bjørnstad and Falck 2001) could also do the trick. We call this function `icorrelogram` and add it to our data frame with the `source` function. We then plot the resulting the correlogram (Fig. 6.8b).

```
#import function
> source('icorrelogram.r')
```

To inspect this function, simply type:

```
> icorrelogram

##
function(locations,z, binsize, maxdist){

distbin <- seq(0,maxdist,by=binsize)
Nbin <- length(distbin)-1
```

```
moran.results <- data.frame(dist = rep(NA,Nbin),
Morans.i =NA,null.lcl=NA, "null.ucl"=NA)

for (i in 1:Nbin){
  d.start <- distbin[i]
  d.end <- distbin[i+1]
  neigh <- dnearneigh(x=locations, d1=d.start, d.end, longlat=F)
  wts <- nb2listw(neighbours=neigh, style='B', zero.policy=T)
  mor.i <- moran.mc(x=z, listw=wts, nsim=200, alternative="greater",
                    zero.policy=T)

  moran.results[i, "dist"]<-(d.end+d.start)/2
  moran.results[i, "Morans.i"]<-mor.i$statistic
  moran.results[i, "null.lcl"]<-quantile(mor.i$res, probs = 0.025,na.
  rm = T)
  moran.results[i, "null.ucl"]<-quantile(mor.i$res, probs = 0.975,na.
  rm = T)
}
return(moran.results)
}
```

This function identifies neighbors between points using the `dnearneigh` function for different distance classes. It then takes the object created, reformats it to a list of relevance to the **W** spatial neighbor matrix, and uses a `moran.mc` function to run a permutation-based Moran's *I*. The distance classes, Moran's *I* and the null envelope from the permutations are then stored in a data frame. We can run the function on the observed data and plot (Fig. 6.8):

```
#run correlogram function
> VATH.cor <- icorrelogram(locations = coords, z
  = point.data$VATH, binsize = 1000, maxdist = 15000)
> head(VATH.cor)

##
Dist Morans.i Null.lcl Null.ucl
1 500 0.34 -0.06 0.06
2 1500 0.10 -0.03 0.03
3 2500 0.01 -0.02 0.03

#plot correlogram
> plot(VATH.cor$Dist, VATH.cor$Morans.i, ylim = c(-0.5, 0.5))
> abline(h=0, lty = "dashed")
> lines(VATH.cor$Dist, VATH.cor$Null.lcl)
> lines(VATH.cor$Dist, VATH.cor$Null.ucl)
```

Now we consider if there is spatial autocorrelation in the residuals of the logistic regression model.

```
#residuals from quadratic elevation model
> VATH.elev2.res <- residuals(VATH.elev2, type = "deviance")
```

Note that we request the deviance-based residuals. For GLM-type models, there are several related residuals that could be calculated, the default being a deviance-based residual. For a binomial or Bernoulli GLM, this type of residual is calculated as:

$$d_i = s_i \sqrt{-2\big(y_i \log\big(\widehat{y}_i\big) + (1 - y_i)\log\big(1 - \widehat{y}_i\big)\big)}, \tag{6.11}$$

where d_i is the deviance of observation i, y_i is the observation, $\widehat{y}$ is the predicted value, and $s_i = 1$ if $y_i = 1$ and -1 if $y_i = 0$. The deviance residuals are potentially more useful in GLMs in comparison to others because they are directly related to the overall deviance (and likelihood) of the model, where the sum of the deviance residuals equals the deviance of the model (-2log-likelihood). We can visualize spatially the residuals by mapping them. More formally, we can assess this using the `icorrelogram` function:

```
#correlogram on residuals
> corr.res <- icorrelogram(locations = coords, z =
  VATH.elev2.res, binsize = 1000, maxdist = 15000)
```

Here, we find evidence for spatial autocorrelation in the residuals of the model (Fig. 6.9). Note that rather than using correlograms, we could have used semivariograms on the residuals to interpret spatial autocorrelation in the residuals (Beguin et al. 2012).

It is important to understand the interpretation of the use of residuals in this analysis in comparison to the raw data. For instance, if we fit an intercept-only (mean) model and contrast correlograms from the raw data and the residuals of the mean model:

```
> VATH.int <- glm(VATH ~ 1,family = "binomial", data = point.data)
> VATH.int.res <- residuals(VATH.int, type = "deviance")

> corr.int.res <- icorrelogram(locations = coords, z =
  VATH.int.res, binsize = 1000, maxdist = 15000)
> cor(VATH.cor$Morans.i, corr.int.res$Morans.i)

##
[1] 1
```

We find that the Moran's I is identical ($r = 1$). This illustrates the equivalence of considering residuals from regression models in correlograms when no predictors are considered to that of the raw data (Bivand et al. 2013).

Because of the spatial dependence in the residuals, we consider either subsetting the data based on the approximate range of spatial autocorrelation or regression-like models that attempt to account for spatial autocorrelation. First, we subset the data.

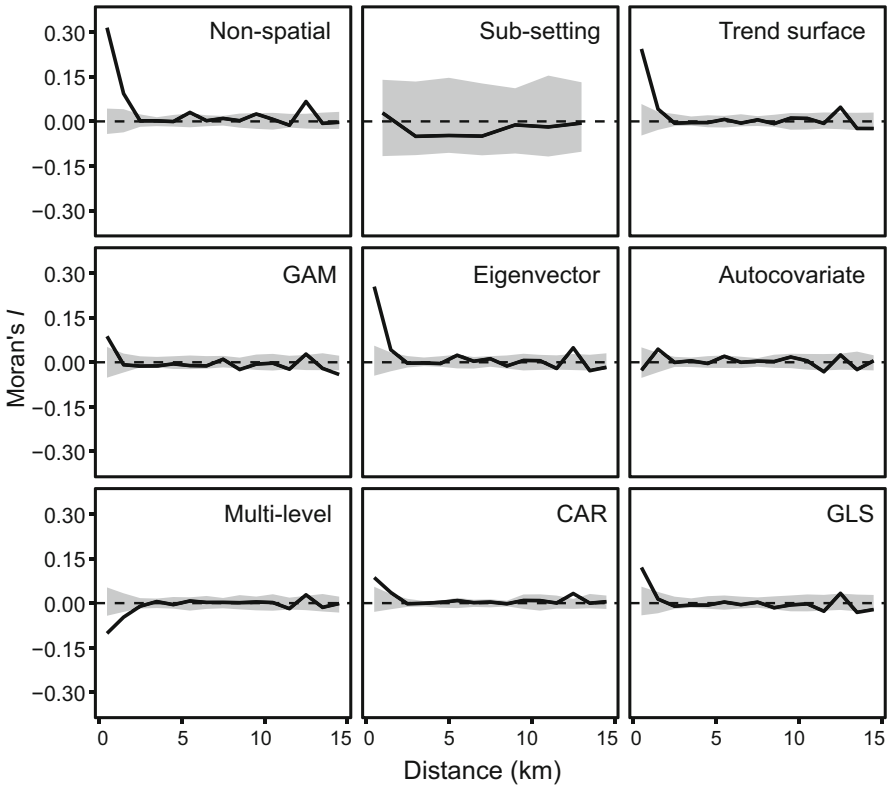


Fig. 6.9 Correlograms on the residuals of the models considered. Note that for subsetting the data, correlograms were calculated for wider lag distance bins because of less data being used

Given the sampling design and the shape of the correlogram (Fig. 6.8b), it would be natural to only consider one point per transect. Note we could also potentially pool across all points on each transect, however, such an approach would increase the spatial grain of the analysis, which might not be ideal. Below we use a function to pick one random point from each transect.

```
#shuffle points on transects
> rand.vector <- with(point.data, ave(POINT, as.factor(TRANSECT),
  FUN=function(x) sample(length(x))))

#pick one random point on transect and remove rest
> point.datasub <- point.data[rand.vector == 1,]

#coordinates from subset data
> coords.sub <- cbind(point.datasub$EASTING, point.datasub$NORTHING)
```

With this data subset, we then refit the logistic regression model.

```
> VATH.sub <- glm(VATH ~ elev + I(elev^2), family = "binomial", data =
  point.datasub)
> summary(VATH.sub)

##
Call:
glm(formula = VATH ~ elev + I(elev^2), family = "binomial", data = point.
datasub)

Deviance Residuals:
Min 1Q Median 3Q Max
-0.5673 -0.5408 -0.4677 -0.3022 2.6507

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.158 4.498 -1.369 0.171
elev 8.254 7.519 1.098 0.272
I(elev^2) -3.860 3.076 -1.255 0.209

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 109.89 on 166 degrees of freedom
Residual deviance: 105.18 on 164 degrees of freedom
AIC: 111.18

Number of Fisher Scoring iterations: 6
```

When we subset the data, our sample size decreases substantially, from 805 to 167 points. Not surprisingly, the SEs on the parameter estimates increase substantially and there is no longer strong evidence for an elevation effect. We can interpret whether this subsetting removed the spatial autocorrelations in the residuals of the model. Note that for this subset, we need to use a larger lag distance than 1-km because we no longer have data points <1 km (or alternatively, one could just increase the first few bin sizes). We calculate the correlogram using a 2-km lag distance.

```
> VATH.sub.res <- residuals(VATH.sub)

#correlogram on residuals
> corr.sub.res <- icorrelogram(locations = coords.sub, z =
  VATH.sub.res, binsize = 2000, maxdist = 15000)
```

This subsetting suggests that spatial autocorrelation is no longer problematic (Fig. 6.9), but there is a cost in terms of reduced power. Regression models that use all of the data but account for spatial dependence might be a useful alternative in this case.

6.3.4 Models that Account for Spatial Dependence

We consider several types of models that account for spatial dependence. These include: trend surface models, eigenvector-based models, autocovariate models (autologistic regression), autoregressive models (a CAR model), a multilevel model, generalized least squares, and spatial GLMMs.

6.3.4.1 Trend Surface Models

We consider two types of trend surface models. In the first model, we simply extend our logistic regression model to include *x*–*y* coordinates, along with their quadratic and cubic polynomial terms with the `I()` function.

```
> VATH.trend <- glm(VATH ~ elev + I(elev^2) + EASTING + NORTHING +
  I(EASTING^2) + I(EASTING^3) + I(NORTHING^2) + I(NORTHING^3),
  family = "binomial", data = point.data)
> summary(VATH.trend)

##
Call:
glm(formula = VATH ~ elev + I(elev^2) + EASTING + NORTHING +
I(EASTING^2) + I(EASTING^3) + I(NORTHING^2) + I(NORTHING^3),
family = "binomial", data = point.data)

Deviance Residuals:
Min 1Q Median 3Q Max
-1.0301 -0.5425 -0.2959 -0.1743 2.8718

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.861e+00 6.332e+00 -1.557 0.11943
elev 8.795e+00 3.138e+00 2.803 0.00507 **
I(elev^2) -3.195e+00 1.248e+00 -2.559 0.01049 *
EASTING 2.208e-04 4.769e-05 4.631 3.65e-06 ***
NORTHING -8.018e-05 5.076e-05 -1.580 0.11420
I(EASTING^2) -1.263e-09 2.806e-10 -4.502 6.72e-06 ***
I(EASTING^3) 2.090e-15 5.122e-16 4.081 4.48e-05 ***
I(NORTHING^2) 2.296e-10 1.366e-10 1.681 0.09275 .
I(NORTHING^3) -2.049e-16 1.179e-16 -1.738 0.08216 .
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 584.34 on 804 degrees of freedom
Residual deviance: 486.89 on 796 degrees of freedom
AIC: 504.89

Number of Fisher Scoring iterations: 6
```

In the above model, we manually added quadratic and cubic terms. A more automated way to do this is with the `poly` function, where specifying `poly (EASTING,3)` would add the linear, quadratic and cubic terms. Note that the `poly` function also standardizes polynomials to be orthogonal, removing the correlation between terms (which in many situations would be preferred). While the above model is straightforward to implement, it is limited in the spatial variation in can capture. An alternative to this model is to consider a generalized additive model (GAM), where we allow spline functions to capture spatial variation. The `mgcv` package provides a means to automate the selection of spline variation through the use of generalized cross-validation procedures. This model can be run as:

```
> library(mgcv)
> VATH.gam <- gam(VATH ~ elev + I(elev^2) + s(EASTING, NORTHING),
  family = "binomial", data = point.data)
```

In this model formulation, elevation is considered in a similar way as above but splines are considered for both Easting (x) and Northing (y) coordinates with the `s` command. This syntax defaults to automated selection of the number of knots being considered. We can manually adjust the number of knots (Fig. 6.5) by adding some syntax to the `s` command. We will look at GAMs in more detail in Chap. 7. In this case, the use of the gam formulation reduces spatial autocorrelation in the residuals (Fig. 6.9); however, it does not appear to fully remove the spatial dependence.

6.3.4.2 Eigenvector Mapping

To account for spatial dependence with eigenvector mapping, there are three steps. First, we create a neighborhood weights matrix with the `spdep` package. We can do this in several ways. Here we calculate a neighborhood weights matrix by using the maximum distance needed for a minimum spanning tree—the minimum set of connections needed to fully connect points across the landscape. The distance needed for a minimum spanning tree can be determined with the `vegan` package using the `spantree` function (note: this distance could also be determined using the `pcnm` function and finding the threshold, as discussed in Chap. 5).

```
> library(vegan)
> spantree.em <- spantree(dist(coords), toolong = 0)
> max(spantree.em$dist)

##
[1] 41351.09
```

We then identify neighborhoods with the `dnearneigh` function using this distance. With these neighbors, we extract the distances between neighbors with the `nbdists` function. Finally, we transform distances as suggested in Dormann et al. (2007) with the `lapply` function (because `nbdists` object is in list form),

and then create a list in the format relevant to the **W** matrix with the `nb2listw` function:

```
> dnn <- dnearneigh(coords, 0, max(spantree.em$dist))
> dnn_dists <- nbdists(dnn, coords)
> dnn_sims <- lapply(dnn_dists, function(x) (1 - ((x / 4)^2)))
> ME.weight <- nb2listw(dnn, glist = dnn_sims, style = "B",
  zero.policy = T)
```

With this **W** matrix, we use the `ME` function in the `spdep` package to identify the most important eigenvectors that reduce spatial dependence, based on a permutation bootstrap test on Moran's *I* for the residuals (Griffith and Peres-Neto 2006). In this function, we include the relevant covariates in the model formula, but we also add the neighborhood matrix (in list form):

```
> VATH.ME <- ME(VATH ~ elev + I(elev^2), family = "binomial", listw =
  ME.weight, data = point.data)
> VATH.ME$selection

##
Eigenvector ZI pr(ZI)
0 NA NA 0.01
1 796 NA 0.01
2 804 NA 0.03
3 805 NA 0.20

> head(fitted(VATH.ME),2)

##
 vec796 vec804 vec805
[1,] 0.003042641 -0.008629250 0.01187249
[2,] 0.003088077 -0.008737222 0.01196633
```

The `ME` function provides output regarding the eigenvectors selected but we need to then refit the logistic regression model with this eigenvectors included as covariates.

```
#new glm with ME covariates
> VATH.evm <- glm(VATH ~ elev + I(elev^2) + fitted(VATH.ME), family =
  "binomial", data = point.data)
> summary(VATH.evm)

##
Call:
glm(formula = VATH ~ elev + I(elev^2) + fitted(VATH.ME), family =
"binomial",
 data = point.data)
```

```
Deviance Residuals:
Min 1Q Median 3Q Max
-1.5359 -0.5175 -0.4027 -0.1416 2.7454

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.401 1.948 -4.312 1.62e-05 ***
elev 8.776 3.029 2.898 0.00376 **
I(elev^2) -3.112 1.168 -2.664 0.00773 **
fitted(VATH.ME)vec796 -14.742 3.198 -4.610 4.03e-06 ***
fitted(VATH.ME)vec804 -8.644 3.242 -2.666 0.00767 **
fitted(VATH.ME)vec805 38.110 8.789 4.336 1.45e-05 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 584.34 on 804 degrees of freedom
Residual deviance: 499.73 on 799 degrees of freedom
AIC: 511.73

Number of Fisher Scoring iterations: 7
```

In this case, the approach identifies three eigenvectors to include, each of which explains occurrence to some degree. However, the inclusion of these eigenvectors does not remove spatial autocorrelation in the residuals of the model (Fig. 6.9). Overall, the main difference in this approach relative to the trend surface model described above is the creation of the eigenvector covariates and determining which of these covariates to include in the final logistic regression model.

6.3.4.3 Autocovariate Models

To fit autocovariate models, we calculate new autocovariates and then use these covariates in a standard logistic regression model. We will calculate these autocovariates with the `autocov_dist` function in the `spdep` package. Because most of the significant autocorrelation in the residuals occurs <1 km (Fig. 6.8b), we will calculate the autocovariates at this scale.

```
> auto1km <- autocov_dist(point.data$VATH, coords, nbs = 1000, type =
  "one", style = "B", zero.policy = T)
```

The `type=` provides the weighting scheme. When inverse is specified, points are weighted by the inverse of the distance between the focal point and the neighboring point. If `"one"` is specified, all points within the distance (`nbs`) are given equal weight. `style` describes how the covariate will be calculated, with `"B"` reflecting a binary coding. Bardos et al. (2015) stated that using `style = "B"` provides a valid weighting scheme for autocovariate models.

We then fit standard logistic regression models with these covariates included.

```
> VATH.auto1km <- glm(VATH ~ elev + I(elev^2) + auto1km, family =
  "binomial", data = point.data)

> summary(VATH.auto1km)

##
Call:
glm(formula = VATH ~ elev + I(elev^2) + auto1km, family = "binomial",
data = point.data)

Deviance Residuals:
Min 1Q Median 3Q Max
-2.0314 -0.4131 -0.3809 -0.2902 2.9077

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.6518 1.9660 -3.383 0.000716 ***
elev 6.9046 3.1222 2.211 0.027006 *
I(elev^2) -2.8061 1.2106 -2.318 0.020450 *
auto1km 0.8665 0.1008 8.596 < 2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 584.34 on 804 degrees of freedom
Residual deviance: 470.99 on 801 degrees of freedom
AIC: 478.99

Number of Fisher Scoring iterations: 6
```

In this case, the inclusion of the autocovariate in the model is very significant, while the coefficients for the elevation effect decrease. In addition, the inclusion of the autocovariate removes the spatial autocorrelation in the residuals (Fig. 6.9).

6.3.4.4 Autoregressive Models

Fitting autoregressive models to non-normal data is challenging. One approach is to use Bayesian modeling. While there are some packages for fitting spatial autoregressive models with Bayesian modeling (e.g., see the `spBayes` package; Finley et al. 2015), using Bayesian methods for spatial regression is often computationally demanding. A new alternative is using "Integrated Nested Laplace Approximation" or INLA (Blangiardo and Cameletti 2015). The value of this approach is that it greatly reduces the computational demands of Bayesian modeling. However, it does only apply to certain types of Bayesian models. For example,

INLA can be used to fit CAR models for binary data. To do so, we need to create a neighborhood weights matrix of the "dgTMatrix" form, which is a type of a sparse matrix (sparse matrices are those that have very few observations and are largely filled with zeros. There are efficient ways to store and manipulate these types of matrices in R). We first create a neighborhood matrix by creating Thiessen polygons from the point data with the deldir and dismo packages (Fig. 6.3). Thiessen polygons, also known as Voronoi polygons, are based on Delaunay triangulation. These polygons partition a region into convex polygons such that each polygon contains exactly one point.

```
> library(INLA)
> library(deldir)
> library(dismo)

> thiessen <- voronoi(coords)

#plot thiessen polygons
> plot(thiessen)
> points(coords, col = "red")

> point.poly <- poly2nb(thiessen)

#plot neighborhood matrix
> plot(point.poly, coords, col = "red", add = T)

#format neighborhood matrix
> adj <- nb2mat(point.poly, style = "B")
> adj <- as(adj, "dgTMatrix")
```

With this neighborhood matrix, we can then fit the CAR model. To do so, for INLA we need to first specify the type of the model fitting, including the covariates being considered. For the CAR model, we add an observation-level covariate to the data frame (id) and then specify "besag" for the CAR model. We then fit the model with the inla function:

```
> point.data$id <- 1:nrow(point.data)
> VATH.inla <- inla(VATH ~ elev + I(elev^2) + f(id, model = "besag",
  graph = adj), family = "binomial", data = point.data,
  control.predictor = list(compute = TRUE))

> summary(VATH.inla)

##
Call:
c("inla(formula = form, family = \"binomial\", data = point.data, ", "
control.predictor = list(compute = TRUE))")
```

```
Time used:
 Pre-processing Running inla Post-processing Total
 2.8085 3.2775 0.5343 6.6203

Fixed effects:
 mean sd 0.025quant 0.5quant 0.975quant mode kld
(Intercept) -8.0537 1.9683 -12.1721 -7.9648 -4.4301 -7.7824 0
elev 10.8093 3.1908 4.9396 10.6618 17.4881 10.3574 0
I(elev^2) -4.5148 1.2666 -7.1800 -4.4519 -2.1971 -4.3222 0

Random effects:
Name Model
 ID Besags ICAR model

Model hyperparameters:
 mean sd 0.025quant 0.5quant 0.975quant mode
Precision for ID 18537.90 18336.86 1248.75 13131.81 66833.34 3386.31

Expected number of effective parameters(std dev): 2.993(0.0029)
Number of equivalent replicates : 268.99

Marginal log-Likelihood: -899.24
Posterior marginals for linear predictor and fitted values computed
```

This approach allows for a binomial CAR model (note that if our response data were normally distributed, we could use the `spautolm` function in the `spdep` package). The `spaMM` package can also fit CAR models to binomial data, but the above model in that package takes >50× longer to fit than with INLA. Also, the INLA approach is computationally much faster than using other Bayesian modeling approaches, which is a major benefit of this package. With the `inla` package, we must manually calculate residuals to interpret spatial autocorrelation:

```
#manual deviance residual calculation:
> VATH.inla.fit <- VATH.inla$summary.fitted.values$mean
> si <- ifelse(point.data$VATH==1, 1, -1)
> VATH.inla.res <- si * (-2 * (point.data$VATH * log(VATH.inla.fit) + (1 -
  point.data$VATH) * log(1 - VATH.inla.fit)))^0.5

#correlogram on residuals
> cor.inla.res <- icorrelogram(locations = coords, z = VATH.inla.res,
  binsize = 1000, maxdist = 15000)
```

In this case, we find that the CAR model removes most, but not all, of the autocorrelation in the residuals (Fig. 6.9). This may be due to the fact that the CAR model is only using first-order neighbors, such that only dependence between neighboring points (~300 m apart) is captured. As the observed spatial dependence in the residuals extends out to 1–2 km (Fig. 6.8b), this smaller scale is not sufficient in this case.

6.3.4.5 Multilevel Models

A simple multilevel model can also be fit to these data by considering transects as a random effect in the regression model. In doing so, we effectively "block" with transects, treating points within transects has having potential spatial dependence (Keitt et al. 2002). Because this structure is not spatially explicit, we effectively assume that dependence is constant within transects (e.g., neighboring points have the same dependence as points located along the ends of the transects). These models can be fit using the `lme4` package. Prior to the model fitting, we need to make sure that transect is considered a factor. Then we can fit the model with `glmer` function.

```
> library(lme4)

#random effects should be a factor
> str(point.data)
> point.data$TRANSECT <- as.factor(point.data$TRANSECT)

#glmm using lme4
> VATH.glmm <- glmer(VATH ~ elev + I(elev^2) + (1|TRANSECT), family =
  "binomial", data = point.data)

> summary(VATH.glmm)

##
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: VATH ~ elev + I(elev^2) + (1 | TRANSECT)
Data: point.data

AIC BIC logLik deviance df.resid
498.4 517.2 -245.2 490.4 801

Scaled residuals:
 Min 1Q Median 3Q Max
-1.3520 -0.1755 -0.1541 -0.1129 5.7688

Random effects:
 Groups Name Variance Std.Dev.
 TRANSECT (Intercept) 4.456 2.111
Number of obs: 805, groups: TRANSECT, 167

Fixed effects:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.470 3.262 -2.596 0.00942 **
elev 9.459 5.155 1.835 0.06653 .
I(elev^2) -4.043 1.979 -2.043 0.04106 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

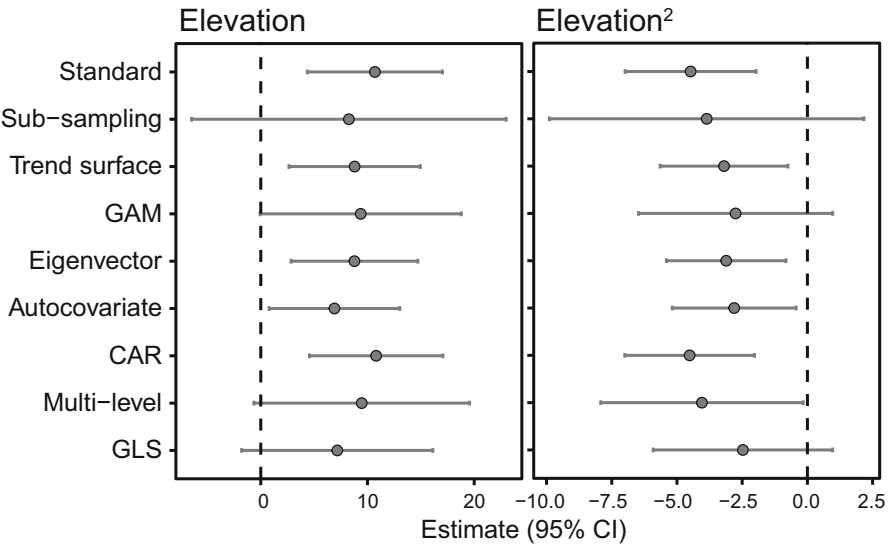


Fig. 6.10 Estimates of elevation relationships based on the spatial models considered

```
Correlation of Fixed Effects:
 (Intr) elev
elev -0.981
I(elev^2) 0.946 -0.988
```

When fitting random effects, we specify (1|TRANSECT), which signals that our transect is being considered as a random intercept. We will see more uses of random effects and their specification in Chap. 11. In this case, we find that by adding a random transect effect to the model structure, positive spatial dependence in the residuals vanishes (Fig. 6.9), although now there is some slight negative autocorrelation in the residuals at short distances. Also, note that the SEs increase and that the elevation effect is only weakly significant (Fig. 6.10).

6.3.4.6 GLS and Mixed Models

GLS and spatially explicit mixed models are difficult to implement for non-normal response data. For normal response data, the `nlme` package can accommodate spatial correlation structures in the model residuals (sometimes referred to as "R-side" correlation structures) or in the random effects (sometimes referred to as "G-side" correlation structures) (Littell et al. 2006).

Given the hierarchical structure of the data, we can fit spatial mixed models where spatial correlation is calculated within transects or across the entire region. The `glmmPQL` function in `MASS` package can be used for GLS and spatial mixed models.

However, this approach uses penalized "quasi-likelihood" and has been shown to have poor properties (Rousset and Ferdy 2014). Because maximum likelihood is not used, we cannot use model selection approaches with this function. Nonetheless, we can still estimate environmental relationships that account for spatial dependence. Here, we fit an exponential correlation function within transects, by identifying transect as a random effect and using the `corExp` command:

```
> library(MASS)
> library(nlme)
> VATH.pql <- glmmPQL(VATH ~ elev + I(elev^2), random = ~1|TRANSECT,
  correlation = corExp(form = ~ EASTING +
  NORTHING), family = "binomial", data = point.data)
```

A similar model can be fit that considers spatial dependence throughout the region (not just within transects). This model takes considerable time to run, but we illustrate it as an example. To do so, we create an observation-level factor. This factor is then fit into the model as a random effect (Dormann et al. 2007).

```
> GROUP <- factor(rep("obs", nrow(point.data)))

> VATH.gls <- glmmPQL(VATH ~ elev + I(elev^2), random = ~1|GROUP,
  correlation = corExp(form = ~ EASTING + NORTHING),
  family = "binomial", data = point.data)
```

Penalized quasi-likelihood has some limitations, including potential bias in estimating random and fixed effects, as well as an inability to use model selection. A recent development that fits similar models without use of penalized quasi-likelihood may overcome some of these limitations (Rousset and Ferdy 2014). The `spaMM` package uses maximum likelihood to estimate a spatial GLMM with Laplace approximation. A similar formulation to that above can be fit with this package using the `corrHLfit` function:

```
> library(spaMM)
> VATH.spamm.ml <- corrHLfit(VATH ~ elev + I(elev^2) +
  Matern(1|EASTING+NORTHING),
  HLmethod = "ML", data = point.data, family = binomial(), ranFix =
  list(nu=0.5)))
```

In this function, we specify a general Matérn spatial correlation structure. The negative exponential function used above is a specific form of a Matérn correlation structure, which in this case is called by setting `nu = 0.5` in the `randFix` statement (see Chap. 5). Overall, this model provides similar estimates and results to the gls model in this situation. For both, spatial autocorrelation is not removed in the residuals when fitting the spatial correlation function across the entire region (Fig. 6.9). However, when only fitting the function within transects with the pql model, we find that spatial autocorrelation is removed in the residuals.

6.4 Next Steps and Advanced Issues

6.4.1 General Bayesian Models for Spatial Dependence

Proper accounting for spatial dependence in non-normal data can be difficult. In this chapter, we have focused on approaches that aim to address this issue in a variety of ways, but each of these approaches has some limitations. Bayesian models that capture spatial dependence provide a flexible means to accommodate spatial dependence. The INLA package provides one straightforward approach to do so, but INLA is limited to only certain types of regression models. More flexibility can be achieved through modeling spatial dependence using the bugs language (via Winbugs or Jags) (Kery and Royle 2016). In these approaches, either CAR/SAR types of models can be fit or GLS and mixed model-like formulations can be fit. This is often accomplished through the hierarchical formulation of spatial dependence coming from a multivariate normal distribution. These types of models are often thought to be useful for accounting for spatial dependence, although they can be challenging to fit (Beale et al. 2010).

6.4.2 Detection Errors and Spatial Dependence

Throughout this chapter, we have ignored the problem of sampling error, such as imperfect detection of species, to focus more simply on the issue of accounting for spatial autocorrelation. However, observation errors are common in data sets and these errors frequently need to be accounted for (MacKenzie et al. 2002). Several models exist for accounting for imperfect detection, both in terms of false positive and false negative errors (Miller et al. 2011). False negative errors are more common, where a species or individual occurs in an area but we fail to detect it. Occupancy, *N*-mixture, and distance sampling models are common approaches to account for these issues (Kery and Royle 2016). False positive errors occur when we misidentify species: we record that a species occurs in an area when in fact it does not. False positive errors are more difficult to account for, but some models exist that do so (Miller et al. 2011). We do not focus on these models in this book, largely because there have been several excellent books that illustrate these models, including their implementation in R (Royle and Dorazio 2008; Kery and Royle 2016).

There has been recent interest in extending these models to account for spatial autocorrelation (Hines et al. 2010; Johnson et al. 2013). Initial attempts used autocovariates like those shown here to account for spatial dependence (Royle and Dorazio 2008). More recently, geostatistical models have been developed as well (Johnson et al. 2013; Broms et al. 2014). Most of these models require customized code and implementation with Winbugs or Jags interfaced through the use of R (Carroll and Johnson 2008; Rota et al. 2011). However, some specialized R

packages can also accommodate spatial dependence in this context. The `hSDM` and `stocc` packages provide occupancy implementations that can accommodate spatial dependence (Johnson et al. 2013).

6.5 Conclusions

Tobler's first law of geography emphasizes that spatial dependence is common in nature. Given that ignoring this fact can lead to spurious inferences (Bivand 1980; Legendre 1993), accounting for spatial dependence in ecological data is often needed. Doing so, however, can be challenging. Here, we illustrate a variety of approaches to accounting for spatial dependence, contrasting their utility when using binary response data. In this case, trend surface and related environmental filtering (GAMs, eigenvector mapping) did not remove the spatial dependence in the residuals. CAR models also did not remove the spatial dependence, presumably because of the small neighborhood considered. Autocovariate and multilevel models did remove spatial dependence in the residuals by appropriately capturing the spatial scale of dependence in the data. Similar to Beale et al. (2010) and Dormann et al. (2007), we find that autocovariate models tended to shrink the effects of the environment relative to other approaches. In general, we recommend the use of mixed models and CAR models that can account for local spatial dependence and adjust uncertainty (SEs/CIs) of environmental relationships. This example illustrates that appropriately capturing the scale of spatial dependence in model structure is important for well-specified spatial models.

Throughout this discussion, we have used geographic coordinates and distances to make inferences about spatial dependence and adjust for this issue in understanding environmental relationships. Yet in many situations, effective distances that capture the complexity of the environment (e.g., shopping malls as barriers to organism movement and resource acquisition) may be more relevant. Spatial weights matrices can capture such complexities when warranted (Dray et al. 2006). Ver Hoef et al. (2018) also emphasized how spatial neighborhoods used in autoregressive models capture similar ideas to the use of network modeling in connectivity assessments (see Chap. 9). This is an interesting and important linkage that we expect will receive more attention in the coming years.

Care should be taken when applying spatial models, particularly for non-normal response variables. There is ongoing debate regarding the utility of different modeling approaches to account for spatial dependence (Dormann et al. 2007; Betts et al. 2009; Dormann 2009). In addition, while several lines of evidence suggest that spatial autocorrelation is problematic for conventional regression modeling, counter examples have also been emphasized (Diniz et al. 2003; Hawkins et al. 2007). Further advances in this area will no doubt provide a useful set of tools for spatial ecologists and conservation biologists alike.

References

Algar AC, Kharouba HM, Young ER, Kerr JT (2009) Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. Ecography 32(1):22–33. https://doi.org/10.1111/j.1600-0587.2009.05832.x

Augustin NH, Mugglestone MA, Buckland ST (1996) An autologistic model for the spatial distribution of wildlife. J Appl Ecol 33(2):339–347

Bardos DC, Guillera-Arroita G, Wintle BA (2015) Valid auto-models for spatially autocorrelated occupancy and abundance data. Methods Ecol Evol 6(10):1137–1149. https://doi.org/10.1111/2041-210x.12402

Bates D, Machler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67(1):1–48

Bauman D, Drouet T, Dray S, Vleminckx J (2018) Disentangling good from bad practices in the selection of spatial or phylogenetic eigenvectors. Ecography 41:1–12

Beale CM, Lennon JJ, Yearsley JM, Brewer MJ, Elston DA (2010) Regression analysis of spatial data. Ecol Lett 13(2):246–264. https://doi.org/10.1111/j.1461-0248.2009.01422.x

Beguin J, Martino S, Rue H, Cumming SG (2012) Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation. Methods Ecol Evol 3 (5):921–929. https://doi.org/10.1111/j.2041-210X.2012.00211.x

Betts MG, Diamond AW, Forbes GJ, Villard MA, Gunn JS (2006) The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. Ecol Model 191 (2):197–224

Betts MG, Ganio LM, Huso MMP, Som NA, Huettmann F, Bowman J, Wintle BA (2009) Comment on "Methods to account for spatial autocorrelation in the analysis of species distributional data: a review". Ecography 32(3):374–378. https://doi.org/10.1111/j.1600-0587.2008.05562.x

Betts MG, Phalan B, Frey SJK, Rousseau JS, Yang ZQ (2018) Old-growth forests buffer climate-sensitive bird populations from warming. Divers Distrib 24(4):439–447. https://doi.org/10.1111/ddi.12688

Bini LM, Diniz JAF, Rangel T, Akre TSB, Albaladejo RG, Albuquerque FS, Aparicio A, Araujo MB, Baselga A, Beck J, Bellocq MI, Bohning-Gaese K, Borges PAV, Castro-Parga I, Chey VK, Chown SL, de Marco P, Dobkin DS, Ferrer-Castan D, Field R, Filloy J, Fleishman E, Gomez JF, Hortal J, Iverson JB, Kerr JT, Kissling WD, Kitching IJ, Leon-Cortes JL, Lobo JM, Montoya D, Morales-Castilla I, Moreno JC, Oberdorff T, Olalla-Tarraga MA, Pausas JG, Qian H, Rahbek C, Rodriguez MA, Rueda M, Ruggiero A, Sackmann P, Sanders NJ, Terribile LC, Vetaas OR, Hawkins BA (2009) Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography 32(2):193–204. https://doi.org/10.1111/j.1600-0587.2009.05717.x

Bivand R (1980) A Monte Carlo study of correlation coefficient estimation with spatially autocorrelated observations. Quaestionies Geographicae 6:5–10

Bivand R, Piras G (2015) Comparing implementations of estimation methods for spatial econometrics. J Stat Softw 63(18):1–36

Bivand RS, Pebesma EJ, Gomez-Rubio V (2013) Applied spatial data analysis with R. Use R! 2nd edn. Springer, New York

Bjørnstad ON, Falck W (2001) Nonparametric spatial covariance functions: estimation and testing. Environ Ecol Stat 8(1):53–70. https://doi.org/10.1023/a:1009601932481

Blangiardo M, Cameletti M (2015) Spatial and spatio-temporal Bayesian models with R-INLA. Wiley, Chichester

Bolker B (2008) Ecological models and data in R. Princeton University Press, Princeton, NJ

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24(3):127–135. https://doi.org/10.1016/j.tree.2008.10.008

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecol Model 157(2–3):281–300. https://doi.org/10.1016/s0304-3800(02)00200-4

Brand LA, George TL (2001) Response of passerine birds to forest edge in coast redwood forest fragments. Auk 118(3):678–686. https://doi.org/10.1642/0004-8038(2001)118[0678:Ropbtf]2.0.Co;2

Broms KM, Johnson DS, Altwegg R, Conquest LL (2014) Spatial occupancy models applied to atlas data show Southern Ground Hornbills strongly depend on protected areas. Ecol Appl 24 (2):363–374. https://doi.org/10.1890/12-2151.1

Carl G, Kuhn I (2010) A wavelet-based extension of generalized linear models to remove the effect of spatial autocorrelation. Geogr Anal 42(3):323–337

Carroll C, Johnson DS (2008) The importance of being spatial (and reserved): assessing Northern Spotted Owl habitat relationships with hierarchical Bayesian Models. Conserv Biol 22 (4):1026–1036. https://doi.org/10.1111/j.1523-1739.2008.00931.x

Carroll C, Johnson DS, Dunk JR, Zielinski WJ (2010) Hierarchical Bayesian spatial models for multispecies conservation planning and monitoring. Conserv Biol 24(6):1538–1548. https://doi.org/10.1111/j.1523-1739.2010.01528.x

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74(368):829–836. https://doi.org/10.2307/2286407

Crase B, Liedloff AC, Wintle BA (2012) A new method for dealing with residual spatial autocorrelation in species distribution models. Ecography 35(10):879–888. https://doi.org/10.1111/j.1600-0587.2011.07138.x

Crase B, Liedloff A, Vesk PA, Fukuda Y, Wintle BA (2014) Incorporating spatial autocorrelation into species distribution models alters forecasts of climate-mediated range shifts. Glob Chang Biol 20(8):2566–2579. https://doi.org/10.1111/gcb.12598

Cressie NAC (1993) Statistics for spatial data. Wiley, Chichester

Dale MRT, Fortin MJ (2014) Spatial analysis: a guide for ecologists, 2nd edn. Cambridge University Press, Cambridge

Diniz JAF, Bini LM, Hawkins BA (2003) Spatial autocorrelation and red herrings in geographical ecology. Glob Ecol Biogeogr 12(1):53–64. https://doi.org/10.1046/j.1466-822X.2003.00322.x

Diniz JAF, Nabout JC, Telles MPD, Soares TN, Rangel T (2009) A review of techniques for spatial modeling in geographical, conservation and landscape genetics. Genet Mol Biol 32(2):203–211. https://doi.org/10.1590/s1415-47572009000200001

Dormann CF (2009) Response to Comment on "Methods to account for spatial autocorrelation in the analysis of species distributional data: a review". Ecography 32(3):379–381. https://doi.org/10.1111/j.1600-0587.2009.05907.x

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kuehn I, Ohlemueller R, Peres-Neto PR, Reineking B, Schroeder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30(5):609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol Model 196(3–4):483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

Dray S, Pelissier R, Couteron P, Fortin MJ, Legendre P, Peres-Neto PR, Bellier E, Bivand R, Blanchet FG, De Caceres M, Dufour AB, Heegaard E, Jombart T, Munoz F, Oksanen J, Thioulouse J, Wagner HH (2012) Community ecology in the age of multivariate multiscale spatial analysis. Ecol Monogr 82(3):257–275. https://doi.org/10.1890/11-1183.1

Emerson S, Wickham C, Ruzicka KJ (2015) Rethinking the linear regression model for spatial ecological data: comment. Ecology 96(7):2021–2025. https://doi.org/10.1890/14-0879.1

Finley AO, Banerjee S, Gelfand AE (2015) spBayes for large univariate and multivariate point-referenced spatio-temporal data models. J Stat Softw 63(13):1–28

Fletcher RJ Jr (2005) Multiple edge effects and their implications in fragmented landscapes. J Anim Ecol 74(2):342–352

Fortin MJ, James PMA, MacKenzie A, Melles SJ, Rayfield B (2012) Spatial statistics, spatial regression, and graph theory in ecology. Spatial Stat 1:100–109. https://doi.org/10.1016/j.spasta.2012.02.004

Garland T, Harvey PH, Ives AR (1992) Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst Biol 41(1):18–32. https://doi.org/10.2307/2992503

Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York

Getis A, Griffith DA (2002) Comparative spatial filtering in regression analysis. Geogr Anal 34 (2):130–140. https://doi.org/10.1111/j.1538-4632.2002.tb01080.x

Griffith DA, Peres-Neto PR (2006) Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. Ecology 87(10):2603–2613. https://doi.org/10.1890/0012-9658(2006)87[2603:smietf]2.0.co;2

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135(2–3):147–186

Haining R (2003) Spatial data analysis: theory and practice. Cambridge University Press, Cambridge

Hastie T, Tibshirani R (1986) Generalized additive models. Stat Sci 1:297–310

Hawkins BA, Diniz JAF, Bini LM, De Marco P, Blackburn TM (2007) Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. Ecography 30 (3):375–384. https://doi.org/10.1111/j.2007.0906-7590.05117.x

Hines JE, Nichols JD, Royle JA, MacKenzie DI, Gopalaswamy AM, Kumar NS, Karanth KU (2010) Tigers on trails: occupancy modeling for cluster sampling. Ecol Appl 20(5):1456–1466. https://doi.org/10.1890/09-0321.1

Hooten MB, Wikle CK, Dorazio RM, Royle JA (2007) Hierarchical spatiotemporal matrix models for characterizing invasions. Biometrics 63(2):558–567. https://doi.org/10.1111/j.1541-0420.2006.00725.x

Horn BKP (1981) Hill shading and the reflectance map. Proc IEEE 69:14–47

Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. Ecol Monogr 54(2):187–211. https://doi.org/10.2307/1942661

Hutto RL, Young JS (2002) Regional landbird monitoring: perspectives from the Northern Rocky Mountains. Wildl Soc Bull 30(3):738–750

Johnson DS, Conn PB, Hooten MB, Ray JC, Pond BA (2013) Spatial occupancy models for large data sets. Ecology 94(4):801–808

Keitt TH, Bjørnstad ON, Dixon PM, Citron-Pousty S (2002) Accounting for spatial pattern when modeling organism-environment interactions. Ecography 25(5):616–625

Kery M, Royle JA (2016) Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS. Academic, San Diego

Kissling WD, Carl G (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models. Glob Ecol Biogeogr 17(1):59–71. https://doi.org/10.1111/j.1466-8238.2007.00334.x

Koenig WD (1999) Spatial autocorrelation of ecological phenomena. Trends Ecol Evol 14 (1):22–26. https://doi.org/10.1016/s0169-5347(98)01533-x

Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? Ecology 74(6):1659–1673

Legendre P, Legendre L (1998) Numerical ecology. Elsevier, Amsterdam

Lennon JJ (2000) Red-shifts and red herrings in geographical ecology. Ecography 23(1):101–113. https://doi.org/10.1034/j.1600-0587.2000.230111.x

Lichstein JW, Simons TR, Shriner SA, Franzreb KE (2002) Spatial autocorrelation and autoregressive models in ecology. Ecol Monogr 72(3):445–463

Littell RC, Millien GA, Stroup RD, Schabenberger O (2006) SAS for mixed models. SAS Institute Inc., Cary, NC

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. Ecology 83(8):2248–2255

McCarthy KP, Fletcher RJ, Rota CT, Hutto RL (2012) Predicting species distributions from samples collected along roadsides. Conserv Biol 26(1):68–77. https://doi.org/10.1111/j.1523-1739.2011.01754.x

McCullagh P, Nelder JA (1989) Generalized linear models. Chapman and Hall, London

Melles SJ, Fortin MJ, Lindsay K, Badzinski D (2011) Expanding northward: influence of climate change, forest connectivity, and population processes on a threatened species' range shift. Glob Chang Biol 17(1):17–31. https://doi.org/10.1111/j.1365-2486.2010.02214.x

Miller JA (2012) Species distribution models: spatial autocorrelation and non-stationarity. Prog Phys Geogr 36(5):681–692. https://doi.org/10.1177/0309133312442522

Miller J, Franklin J, Aspinall R (2007) Incorporating spatial dependence in predictive vegetation models. Ecol Model 202(3–4):225–242. https://doi.org/10.1016/j.ecolmodel.2006.12.012

Miller DA, Nichols JD, McClintock BT, Grant EHC, Bailey LL, Weir LA (2011) Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. Ecology 92(7):1422–1428. https://doi.org/10.1890/10-1396.1

Nelder JA, Wedderburn RW (1972) Generalized linear models. J R Stat Soc Series A General 135 (3):370. https://doi.org/10.2307/2344614

Oksanen J, Guillaume B, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens HH, Szoecs E, Wagner H (2018) Vegan: community ecology package. R version 2.4-6

Rota CT, Fletcher RJ Jr, Evans JM, Hutto RL (2011) Does accounting for detectability improve species distribution models. Ecography 34:659–670

Rousset F, Ferdy JB (2014) Testing environmental and genetic effects in the presence of spatial autocorrelation. Ecography 37(8):781–790. https://doi.org/10.1111/ecog.00566

Royle JA, Dorazio RM (2008) Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations, and communities. Academic, San Diego

Sauer JR, Niven DK, Hines JE, Ziolkowski DJ, Pardieck KL, Fallon JE, Link WA (2017) The North American Breeding Bird Survey, results and analysis 1966–2015, Version 2.07.2017. USGS Patuxent Wildlife Research Center, Laurel, MD

Segurado P, Araújo MB, Kunin WE (2006) Consequences of spatial autocorrelation for niche-based models. J Appl Ecol 43(3):433–444. https://doi.org/10.1111/j.1365-2664.2006.01162.x

Sokal RR, Oden NL (1978) Spatial autocorrelation in biology II: Some biological implications and four applications of evolutionary and ecological interest. Biol J Linn Soc 10(2):229–249. https://doi.org/10.1111/j.1095-8312.1978.tb00014.x

Stephens PA, Buskirk SW, del Rio CM (2007) Inference in ecology and evolution. Trends Ecol Evol 22(4):192–197. https://doi.org/10.1016/j.tree.2006.12.003

Swihart RK, Slade NA (1985) Testing for independence of observations in animal movements. Ecology 66(4):1176–1184. https://doi.org/10.2307/1939170

Teng SN, Xu C, Sandel B, Svenning JC (2018) Effects of intrinsic sources of spatial autocorrelation on spatial regression modelling. Methods Ecol Evol 9(2):363–372

Thorson JT, Minto C (2015) Mixed effects: a unifying framework for statistical modelling in fisheries biology. ICES J Mar Sci 72(5):1245–1256. https://doi.org/10.1093/icesjms/fsu213

Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York

Ver Hoef JM, Peterson EE, Hooten MB, Hanks EM, Fortin MJ (2018) Spatial autoregressive models for statistical inference from ecological data. Ecol Monogr 88(1):36–59. https://doi.org/10.1002/ecm.1283

Wagner HH (2003) Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. Ecology 84(4):1045–1057. https://doi.org/10.1890/0012-9658(2003)084[1045:scipci]2.0.co;2

Wagner HH (2004) Direct multi-scale ordination with canonical correspondence analysis. Ecology 85(2):342–351. https://doi.org/10.1890/02-0738

Wagner HH, Fortin MJ (2005) Spatial analysis of landscapes: concepts and statistics. Ecology 86 (8):1975–1987. https://doi.org/10.1890/04-0914

Wall MM (2004) A close look at the spatial structure implied by the CAR and SAR models. J Stat Plan Inference 121(2):311–324. https://doi.org/10.1016/s0378-3758(03)00111-3

Wilson MFJ, O'Connell B, Brown C, Guinan JC, Grehan AJ (2007) Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continential slope. Mar Geod 30:3–35

Wintle BA, Bardos DC (2006) Modeling species-habitat relationships with spatially autocorrelated observation data. Ecol Appl 16(5):1945–1958. https://doi.org/10.1890/1051-0761(2006)016[1945:msrwsa]2.0.co;2

Wood SN (2006) Generalized additive models: an introduction with R. Chapman and Hall and CRC, Boca Raton, FL

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) Mixed effects models and extensions in ecology with R. Springer, New York

Part II
Ecological Responses to Spatial Pattern and Conservation

Chapter 7
Species Distributions



7.1 Introduction

Understanding and predicting species distributions lies at the heart of ecology. Predictive models of species distributions are increasingly used in both basic and applied ecology to predict the effects of future climate change (Thomas et al. 2004), land-use change (Feeley and Silman 2010; Martin et al. 2013), species invasion (Peterson 2003; Elith et al. 2010; Jimenez-Valverde et al. 2011), agricultural suitability (Evans et al. 2010; Plath et al. 2016), best places for species reintroduction (Hirzel et al. 2004; Martinez-Meyer et al. 2006), identify new protected areas (Wilson et al. 2005), and to refine biodiversity inventories (Raxworthy et al. 2003).

Over the past two decades, there has been an explosion in the advancement and application of predictive distribution models (Guisan and Zimmermann 2000; Elith and Leathwick 2009; Renner et al. 2015). Species distribution models (SDMs), ecological niche models (ENMs), climate envelope models, and habitat suitability models (HSMs) all describe models that relate species distribution (occurrence or abundance) to the environment through the quantification of response surfaces (i.e., relationships of species distribution with environmental variables; Guisan and Zimmermann 2000; Guisan et al. 2017) (Fig. 7.1). Other related models include resource selection functions, occupancy models, and GAP models (Scott et al. 1993; Manly et al. 2002; Rodrigues et al. 2004; MacKenzie et al. 2006). These models are used for both inference on environmental relationships as well as prediction and projection, where the estimated functions are used to map distributions over space and time. These types of models have been developed in different sub-disciplines and each has a unique focus on the types of questions addressed, the scales at which questions are typically asked, and the specific types of data that are used. However, they all emphasize the relationship of species distribution with the environment.

Here, we describe the key concepts relevant to predicting species distributions, the types of data typically used, some common modeling algorithms, and illustrate how models are frequently evaluated. Our general goal is to illustrate how concepts,

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_7

Libros de Estadística en PDF | Statistics Books in PDF

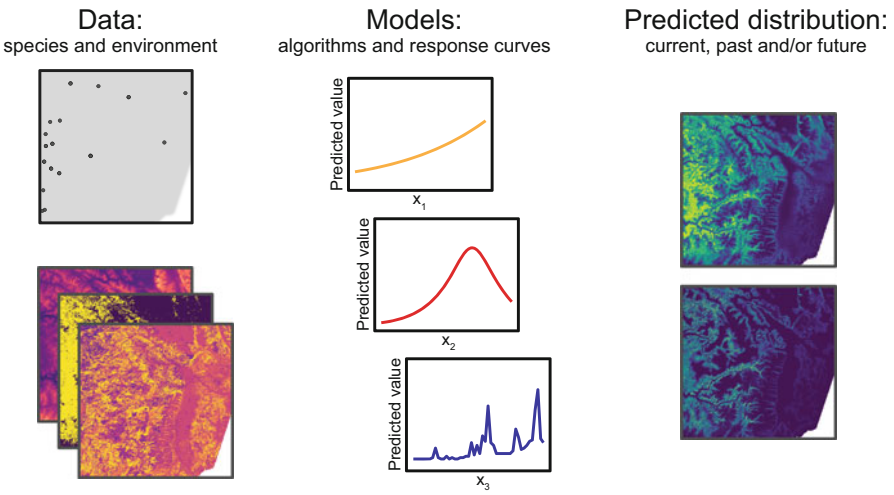


Fig. 7.1 A general framework of modeling species distributions. Data on species location are linked to spatial data on the environment with quantitative models. These models vary considerably in their assumptions about species responses to environmental gradients. With estimated response curves, species distributions are mapped in space and/or time. Modified from Guisan et al. (2017)

data and models are used to create maps of species distributions for addressing ecological questions and conservation problems. For more information regarding general species distribution concepts, see the excellent books by Franklin (2009), Peterson et al. (2011), and Guisan et al. (2017).

7.2 Key Concepts and Approaches

7.2.1 The Niche Concept

> No concept in ecology has been more variously defined or more universally confused than 'niche'. Nonetheless, the concept has become symbolic of the whole field of ecology. Real and Brown (1991)

Species distribution modeling generally relies on niche concepts for developing models (Austin 2002, 2007; Hirzel and Le Lay 2008). Most of the applications of niche theory is heuristic—that is to say that scientists tend to use general ideas that emerge from niche theory. Understanding the relevance of the niche concept is essential for building, interpreting, and applying distribution models to ecological, evolutionary, and conservation problems. Other theoretical developments have also been used in the context of predicting species distributions, such as habitat selection (Fretwell and Lucas 1970) and metapopulation theory (Pulliam 1988; Hanski and Ovaskainen 2003), but here we focus on developments related to niche theory.

7.2.1.1 A Brief History of the Niche Concept and a Plethora of Niches

The term *niche* was originally coined by Joseph Grinnell in the early 1900s. Grinnell was interested in the biogeography of birds and what limited their geographic range. He specifically considered the problem of spatial overlap in congener thrasher species in California (Grinnell 1917). His interpretation of the niche was largely based on the idea of species–environment relationships, emphasizing the role of habitat and behavioral adaptations as key components of a species' niche. For example, the California thrasher (*Toxostoma redivivum*) that he studied is well adapted to its environment, foraging in shrubs and having adaptive behaviors to reduce predation risk (e.g., camouflage).

In the 1920s–1930s, Elton advanced the niche concept, taking a different perspective. He emphasized the functional role of the species in its environment in relation to food and enemies, in which species could impact the environment through trophic interactions (Elton 1927). Elton focused on what a species does rather than where a species occurs, such that he focused on both the species response to, as well as the effect on, the environment. This perspective was quite distinct from Grinnell's perspective.

In the 1950s, Hutchinson took a quantitative perspective on the niche, considering it a "*N*-dimensional hypervolume where a species could persist" (Hutchinson 1957). In this way, an *N*-dimensional hypervolume reflects the idea that there are *N* environmental variables that are required for species persistence, each of which can be viewed as a different dimension in environmental space, and it is the intersection of suitable values of all *N* variables, or the hypervolume, that identifies the niche (Blonder et al. 2014). This work catalyzed the application of niche concepts—including niche breadth, niche overlap, and niche partitioning—by emphasizing measurable properties or dimensions of the niche. Hutchinson also distinguished between fundamental versus realized niches, where the *fundamental niche* was the environmental hypervolume in which a species could potentially persist (sometimes referred to as the physiological or potential niche), while the *realized niche* was a subset of this space where species actually occurred (Fig. 7.2). He assumed that the realized niche was smaller than the fundamental niche due to species interactions, particularly competition. This distinction is often made in the development of species distribution models (e.g., Guisan and Thuiller 2005). Hutchison (1957) defined the niche as a property of the species, not a property of the environment. As a consequence, for Hutchison, there were no "empty niches" in the world.

Pulliam (2000) and a seminal book by Chase and Leibold (2003) advanced the niche concept. Pulliam (2000) emphasized that dispersal limitation could result in many places that have environmental conditions that fall within a species' niche yet remain unoccupied. He also emphasized that the realized niche could in fact be larger than the fundamental niche in situations where species occurred in sink habitats (see Chap. 10). Chase and Liebold (2003) sought to unify niche concepts, integrating Eltonian and Hutchinsonian views in a common framework. They defined the niche

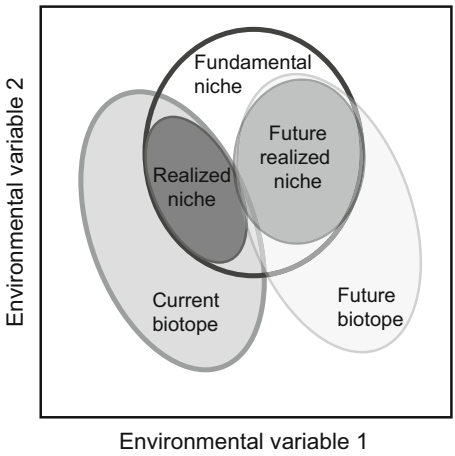


Fig. 7.2 Environmental gradients and the niche. Shown are two environmental variables relevant for the fundamental and realized niche of a species. The current biotope constrains the observed niche of a species, where the current biotope does not include all conditions of the fundamental niche. Changes in the biotope from environmental change causes a shift in the realized niche. Modified from Franklin (2009) and Williams and Jackson (2007)

as, "The joint description of the environmental conditions that allow a species to satisfy its minimum requirements so that the birth rate of a local population is equal to or greater than its death rate along with the set of per capita effects of that species on these environmental conditions" (Table. 7.1).

An important aspect of niches defined by Hutchison, and advanced by Pulliam (2000), Chase and Liebold (2003), and others, is that of fitness. From these perspectives, the niche embodies conditions where positive population growth occurs. Models of species distribution, in contrast, typically only use information on species occurrence or abundance (see below). Because species occurrence or abundance may not correlate with resource quality or positive population growth (Van Horne 1983; Schlaepfer et al. 2002; Robertson and Hutto 2006), using information on occurrence or abundance alone may not be sufficient for modeling the niche. That nuance may be fine under some situations, but it makes it clear that predicting species distributions may not be the same as predicting the niche. Indeed, there has been much debate regarding exactly what species distribution models really predict and how it relates to the niche concept (Franklin 2009; Araújo and Peterson 2012; Peterson and Soberón 2012).

7.2.1.2 Geographic Versus Environmental Space

When translating the niche concept to spatial models, a key distinction is geographic versus environmental space. This distinction was emphasized by Hutchinson's perspective on the niche, where he distinguished the niche from the geographic

Table 7.1 Common terms and definitions used in species distribution modeling

Term	Definition
Biotope	The community's environment (independent of a species).
Correlative distribution model	Predictive models that are based on response functions derived from species distribution and environmental factors.
Fundamental niche	The environmental hypervolume in which a species could potentially persist.
Mechanistic distribution model	Predictive models based on either experiments or known relationships of species with critical limiting factors, such as thermal tolerances of species.
Niche	The joint description of the environmental conditions that allow a species to satisfy its minimum requirements so that the birth rate of a local population is equal to or greater than its death rate along with the set of per capita effects of that species on these environmental conditions.
Realized niche	The environmental hypervolume in which a species occurs, or a subset of the fundamental niche where favorable biotic interactions occur.
Sample selection bias	Bias that can arise when samples of species distribution are nonrandom from the underlying distribution. Common in data collected opportunistically.
Species prevalence	How common a species is across the extent under consideration.

space that contains environmental variation, or the *biotope* (Whittaker et al. 1973; Colwell and Rangel 2009) (Fig. 7.2). A consequence of this distinction is that the niche is clearly an attribute of a species or population (one cannot have "empty niches"), which also helps to understand the differences between niches and habitats (see Chap. 8). Often we aim to make inferences on environmental space (e.g., functions that describe species responses to environmental conditions) and then we wish to map these responses in geographic space to make predictions or projections of species distributions (Fig. 7.1).

The current biotope and future biotope may vary considerably, and our understanding of niches is constrained by the biotope where species currently occur (Fig. 7.2). For instance, when using niche concepts to understand and predict the effects of climate change, current data on species niches may be insufficient because portions of the fundamental niche may not be expressed under current conditions (Williams and Jackson 2007). Similar problems arise with predicting the spread of invasive species using information on the environment in the native range of the species (Peterson 2003; Broennimann et al. 2007). Experiments can help partially resolve this issue (Buckley and Kingsolver 2012).

7.2.1.3 Limiting Factors and the Niche

Several factors can limit the dimensions of the niche (Araújo and Guisan 2006). Soberón categorized these factors as being one of three categories: abiotic, biotic, and movement-related limitations (Soberón 2007, 2010; Soberón and Peterson 2005). He visualized and interpreted these limiting factors using Venn diagrams and set theory, what are referred to a "BAM" diagrams (Biotic-Abiotic-Mobility diagrams; Fig. 7.3). Where these three factors intersect defines the current geographic distribution of a species. **A** captures favorable abiotic conditions and non-interactive variables ("scenopoetic variables") where the intrinsic growth rate of a species is positive, what has been termed the Grinnellian fundamental niche (James et al. 1984). **B** emphasizes the area where biotic interactions (sometimes referred to as Eltonian factors) allow for positive population growth. **M** represents the area that is accessible to organisms, that is, the colonizable area (Barve et al. 2011). In this context, the geographic expression of the *realized niche* has been described as the intersection of B and A, where conditions are suitable but movement limitations may or may not preclude species occurrence (Peterson et al. 2011; Soberón and Peterson 2005). Soberón argued that biotic factors are typically only relevant at fine spatial grains and thus can potentially be ignored in predicting broad-scale distributions and ranges of species (Soberón 2010; Busby 1991), termed the Eltonian Noise Hypothesis (Soberón and Nakamura 2009), although this conclusion is often debated (e.g., Wisz et al. 2013).

In general, the relative importance of these limiting factors may vary across species and across spatial scales (Pearson and Dawson 2003; Soberón and Peterson 2005). For example, Lira-Noriega et al. (2013) found that the importance of dispersal-

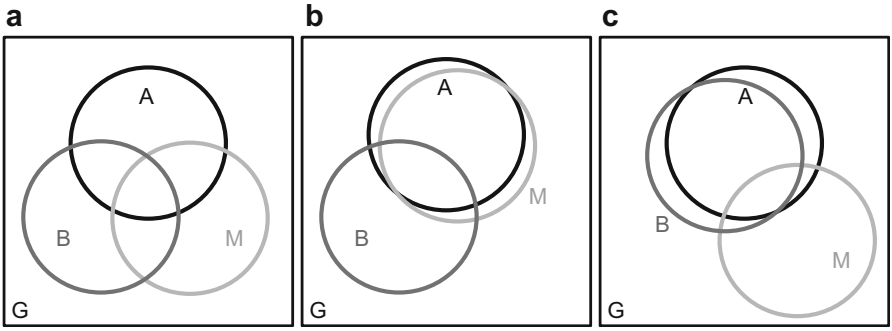


Fig. 7.3 BAM diagrams, illustrating the intersection of abiotic, biotic and movement-related limiting factors of relevance to the niche within a study region, G (or biotope). In this framework, A is considered the Grinnellian fundamental niche, which may or may not be occupied in a region depending on biotic interactions and movement limitations. In (**a**) similar overlap of limiting factors occurs. (**b**) At a fine spatial scale (small grain and extent), the fundamental niche is entirely accessible, but may not be fully occupied due to biotic interactions. (**c**) At a broad spatial scale (coarse grain and large extent), movement limitations may prevent colonization of some portions of the fundamental niche, while biotic interactions at a coarse grain have been hypothesized to have small effects on distribution. Modified from Soberón and Peterson (2005)

related constraints for distributions of mistletoe (*Phoradendron californicum*) varied as a function of spatial resolution of models, where dispersal-related constraints were more important at fine resolutions. It has also been argued that this framework helps illuminate differences in approaches and philosophies for modeling species distributions (Soberón and Nakamura 2009).

7.2.2 Predicting Distributions or Niches?

The focus on predicting and mapping distributions has led to a wide array of terms to describe such models and projections. Often these models are referred to as "ecological niche models," "environmental niche models," "habitat suitability models," or "species distribution models" (Franklin 2009; Peterson et al. 2011; Guisan et al. 2017). In this context, a common question arises: are these efforts actually modeling niches (Peterson and Soberón 2012)? One argument is that if the focus is on environmental space, rather than geographic space, these modeling efforts are more squarely in the vein of modeling the niche (Peterson and Soberón 2012). Yet as all modern concepts regarding the niche emphasize that it is the environmental space is when species can persist (Holt 2009), we suggest refraining from interpreting these models as that of the niche when only distribution information (and no demographic information) is used for model building. While such models can provide hypotheses regarding niches, they are best viewed as modeling distributions.

7.2.3 Mechanistic Versus Correlative Distribution Models

The vast diversity of species distribution models can be organized in several ways. Two useful properties include whether models are correlative (e.g., phenomenological) or mechanistic (i.e., process-based), and the types of response data that are used.

Correlative distribution models take information on species distribution, such as presence records, and relate them to environmental covariates, based on some sort of formal relationship. These models are typically phenomenological—models that describe or explain patterns without regard to underlying mechanisms. In contrast, *mechanistic distribution models* are typically based on either experiments or known relationships of species with critical limiting factors, such as thermal tolerances of species (Buckley 2008; Kearney and Porter 2009). It is often argued that mechanistic models may be more valuable when extrapolating model predictions to new places or times; however, formal comparisons between correlative and mechanistic models have revealed similar model performance in some situations (Buckley et al. 2010). Furthermore, there are strengths and limitations to both approaches and it has been

argued that perhaps this is a false dichotomy in the characterization of distribution models (Dormann et al. 2012).

7.2.4 *Data for Correlative Distribution Models*

Correlative models can also be categorized based on whether presence-only, presence–absence (or detection-non-detection; MacKenzie et al. 2002), or count (abundance) data are used (Brotons et al. 2004; Lutolf et al. 2006; Potts and Elith 2006; Aarts et al. 2012). Presence-only data, or data where only a sample of presence locations are available (and no information is available on absence or abundance), are commonly used in correlative distribution modeling (Elith et al. 2006). There are many sources of such data, including museum and herbaria specimens, information from citizen science programs, and atlas programs (Graham et al. 2004). Presence-only data can be used in isolation or they can be compared to background points, sometimes called "pseudo-absences" to build distribution models. The latter approach has been shown to frequently produce more accurate species distribution models than using presence data alone (Elith et al. 2006). The value of using background points is that it provides information on the biotope and if presence locations reflect a non-random distribution of the underlying environment available to organisms (Fig. 7.4). Two challenges with using background points are determining the number of background points and their spatial distribution (VanDerWal et al. 2009; Barbet-Massin et al. 2012). Some studies have attempted to select background

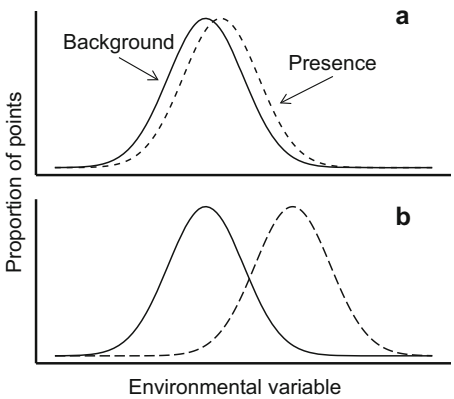


Fig. 7.4 The use of background or pseudo-absence points for presence-only modeling can provide relevant information on the biotope in the region for comparison to presence locations based on the difference between the environment at presence locations relative to the background locations. In (**a**) presence and background points have a similar distribution of environmental values, suggesting random distribution relative to the environmental gradients, whereas in (**b**) presence locations suggest a non-random distribution where the species is more likely to occur at high values of the environmental gradient

points that may be more likely to be considered absences based on using certain rules, such as only creating background points at minimum distances away from presence points; however, it is more common to simply generate randomly distributed background points. Renner et al. (2015) recently argued that many background points should be generated—more than commonly implemented in the literature—based on describing the use of background points in the context of inhomogeneous point process models (see below).

Presence-only data have the benefit of being plentiful across broad geographic areas. Furthermore, it is sometimes argued that such data circumvent the problem of false negatives in presence–absence data (i.e., recording an absence when in fact the species is present) (Guisan et al. 2007). Nonetheless, important limitations of such data include that there is often sample selection bias in opportunistic presence-only data and that the prevalence of the species is unknown. Both of these issues are valid concerns. *Sample selection bias* occurs when samples are a nonrandom sample from the region of interest, which often occurs in presence-only data when observations are more likely to be documented near easily accessible areas, such as near roads or urban areas (Kadmon et al. 2004; Loiselle et al. 2008; Phillips et al. 2009; McCarthy et al. 2012). Such bias can result in the identification of spurious environmental relationships and inaccurate predictions of distributions, in which models may provide predictions of sampling bias rather than underlying distributions. Unknown *species prevalence* arises because presence-only data do not provide information on how common the species is in the extent under consideration, because it is unclear if the presence-only samples reflect a small or large proportion of the underlying distribution. For instance, 30 presence records may be available for a species in a large study region, which could be because the species is rare and this number reflects the prevalence of the rare species, or it could be that the species is common and it was just inadequately sampled. This uncertainty leads to the conclusion that the probability of occurrence cannot be directly estimated with presence-only data (Yackulic et al. 2013) without making strong assumptions (Royle et al. 2012; Hastie and Fithian 2013). Instead, these models predict a relative measure of occurrence that is assumed to be proportional to the true probability of occurrence, similar to interpretations of resource selection functions relative to resource selection probability functions (see Chap. 8).

Presence–absence data, on the other hand, typically come from planned, standardized surveys. These types of data allow for formal modeling of the probability of occurrence of species (and can potentially account for observation errors and imperfect detection in the estimation of occurrence). These types of data are also thought to suffer less from sample selection bias. The rationale is that even when sampling may be biased across space or over time, because models are comparing occurrence observations to absence (or non-detection observations), effects of sample selection bias on estimated environmental relationships should be limited. Some have argued that because absence data may result from observation errors (false-negative errors), that it may be beneficial to only use presence data to help circumvent that problem (Guisan et al. 2007). However, in most situations imperfect

absence data is still useful and can improve model predictions and interpretation (e.g., Brotons et al. 2004; Rota et al. 2011).

Count data are also sometimes used in distribution modeling (Guisan and Harrell 2000; Potts and Elith 2006), and are generally derived from planned survey data. Such data have the potential to provide abundance or density estimates. Count data provide greater information content and resolution in potential species–environment relationships (Cushman and McGarigal 2004); however, count data often require greater sampling intensity. Because distribution models are less frequently built with count data, for the remainder of this chapter we focus primarily on presence-only and presence–absence data.

7.2.5 Common Types of Distribution Modeling Techniques

We provide an overview on common modeling approaches for species distributions. Our summary is not comprehensive; rather we aim to emphasize very different techniques that capture the spectrum of variation in modeling algorithms. We illustrate envelope models (Pearson and Dawson 2003), the use of generalized linear and additive models (Guisan et al. 2002), regression trees and forests (Prasad et al. 2006; Elith et al. 2008), and Maxent (Phillips et al. 2006). We conclude by noting that many of these models can be derived more generally as inhomogeneous point process models (Renner et al. 2015), which may be helpful for better interpreting the relationships among model techniques.

These types of modeling algorithms are sometimes organized into three philosophically different approaches: profile methods, statistical models, and machine-learning algorithms. *Profile methods* are simple approaches that use environmental distances or similarity-based measures to relate environmental variability at presence locations to other locations across the region of interest. Some examples include envelope models (e.g., BIOCLIM), Mahalonobis distance, and DOMAIN (Carpenter et al. 1993; Rotenberry et al. 2006). *Statistical methods* are typically variants of linear models, such as generalized linear and additive models (Guisan and Zimmermann 2000) (see Chap. 6). In these approaches, a model is specified and then fit to the data via maximum likelihood or related techniques (e.g., ordinary least-squares). Statistical methods frequently focus on estimation of parameters and providing measures of uncertainty. *Machine-learning techniques* focus on identifying (and classifying) structure in complex data, often for situations where non-linearities and interactions are expected to occur, with the frequent goal of accurate prediction or classification (Olden et al. 2008). These philosophical distinctions can, however, be unclear, as some algorithms can be described from both a statistical and machine-learning perspective (e.g., Phillips et al. 2006; Elith et al. 2011).

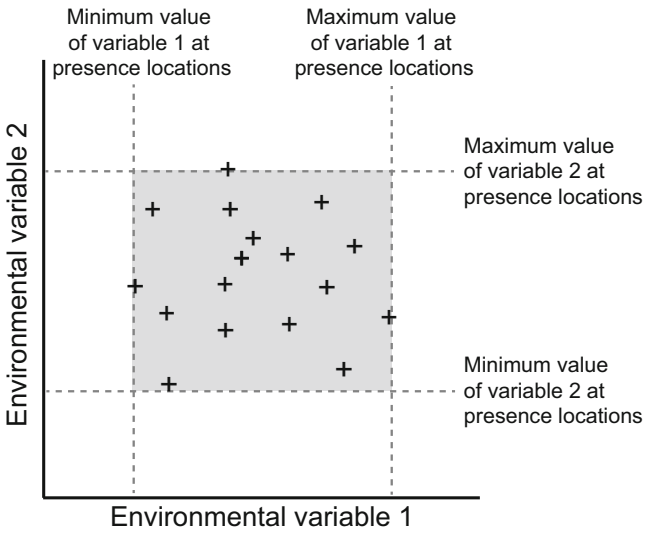


Fig. 7.5 Envelope models use information on either the observed minimum and maximum values of environmental factors at presence locations or quantiles of values (e.g., the 5%, 95% quantile). Pluses denote species occurrences and the grey box denotes the envelope

7.2.5.1 Envelope Models

Envelope models are presence-only models, wherein the distribution of environmental variation at presence locations is used to create an "envelope" of suitability. For example, the upper and lower quantiles of environmental covariates (e.g., 5–95% of elevation values) provide a means to create an envelope, where environmental conditions above or below those quantiles are deemed to be locations outside of the envelope (Fig. 7.5). There are many variations on this theme, but in general these approaches assume that all environmental variables considered are relevant, such that locations must be within the envelope of all variables.

The earliest applications of this approach focused on climatic variables and large-scale geographic range modeling. Busby (1991) developed software for this problem, BIOCLIM, which used climatic variables in a GIS to determine envelopes. More recent developments have attempted to gain more information out of envelope approaches by considering multivariate relationships among variables and through the use of similarity or kernel density measures to obtain relative measures of suitability.

7.2.5.2 GLMs and GAMs

In prior chapters, we introduced the use of generalized linear models (GLMs) and generalized additive models (GAMs). Both of these approaches are frequently applied to the problem of modeling species distributions. For distribution modeling, logistic models are typically used based on binary response data; however, these

models are flexible and can also accommodate abundance response variables (Potts and Elith 2006). For presence-only data, presence points are typically contrasted to background points (Elith et al. 2006). While initial applications of this approach were somewhat ad hoc, this form of logistic regression can approximate more theoretically motivated inhomogeneous point process models (see below). See Chap. 6 for more detailed discussion of these methods.

Although GLMs have been widely used, a primary concern for their implementation in distribution modeling is the fact that may not adequately capture non-linear response functions, which are often emphasized in niche theory (Austin 2007). Because GAMs can accommodate non-linearity through the use of splines, they are frequently used as a logical extension of GLMs. Nonetheless, the types of non-linearity captured by GAMs (see Chap. 6) are less general than some other methods, such as Maxent and regression trees.

7.2.5.3 Regression Trees and Forests

An alternative to the generalized linear (and additive) modeling framework is the use of classification and regression trees (CART), also known as classification tree analysis (CTA), or recursive partitioning (RP). Classification trees work with data whose response variables are discrete, while regression trees work with continuous response variables. Like GAMs, they do not rely on a priori hypotheses about the relationship between independent and dependent variables. This method consists of recursive partitions of the values of predictors into groups that are as homogeneous as possible in terms of the response (Fig. 7.6). The tree is built by repeatedly splitting

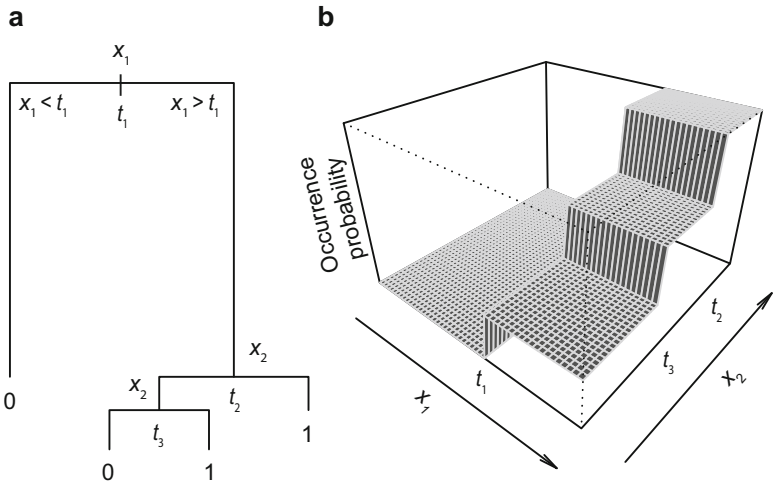


Fig. 7.6 A classification tree. (**a**) The splits of the tree and (**b**) how these splits result in species responses to environmental gradients. Modified from Elith et al. (2008)

the data, defined by a simple rule based on a single explanatory variable. At each split, the data are partitioned into two exclusive groups, each of which is as homogeneous as possible. A common approach is to grow a large tree and then prune it (i.e., reduce its size/complexity) by collapsing the weakest links identified through cross-validation and various indices (e.g., the "Gini" index; Breiman et al. 1984). The result can be thought of as a dichotomous tree that helps to classify locations of species occurrence. When the trees are short, they can be intuitive and visually appealing in terms of describing factors explaining distribution. As tree size grows, their interpretation can become more difficult.

Some advantages of this approach include the ability to easily handle non-linear relationships and interactions, outcomes are unaffected by monotonic transformations, trees are insensitive to outliers, and trees can accommodate missing data in predictor variables by using surrogates (Breiman et al. 1984; De'ath and Fabricius 2000). Nonetheless, CTA often performs poorly compared to GLMs, GAMs, and other models for species distribution modeling (Elith et al. 2008), in part because it has difficulty in modeling smooth functions and that CTA can be sensitive to small changes in the training (model building) data (Hastie et al. 2009; Guisan et al. 2017). However, two extensions of CTA—Boosted Regression Trees and Random Forests—are quickly being adopted because of their high predictive performance. We will focus on these methods, rather than CTA.

Random Forests and Boosted Regression Trees have gained popularity primarily because they typically provide high predictive accuracy relative to CTA and some other SDM algorithms (Elith et al. 2006; Prasad et al. 2006; Cutler et al. 2007). Rather than producing a single classification tree, these approaches are ensemble techniques that compile information from several models, using either "bagging" or "boosting." *Bagging* is a type of a bootstrap procedure, where several models are created through bootstrap sampling of the data (i.e., sampling with replacement) and predictions from models are combined in some way. *Boosting* uses sequential model development (a forward, stage-wise procedure), where with each iteration (sequence) there is an increasing emphasis on the training observations that are difficult to classify.

In Boosted Regression Trees, small, parsimonious trees are fit to the training data, with small trees sequentially added to the existing regression tree (Friedman 2002). The approach is stage-wise (rather than step-wise), meaning that with each iteration where new trees are added, the existing tree is left unchanged. The final model is a linear combination of many trees, analogous to a multiple regression model where each term is a parsimonious tree (Elith et al. 2008). There are two key parameters of interest when fitting a Boosted Regression Tree: the learning rate (or shrinkage parameter), which quantifies the contribution of each individual tree to the model, and the tree complexity, which controls the types of interactions considered. These parameters in combination will determine the number of trees used for predictions. Boosting has been shown to increase predictive abilities of models (Elith et al. 2006), reduce bias, and reduce variance in estimates, even when complex environmental relationships occur. For more on Boosted Regression Trees, see Elith et al. (2008).

Random Forests is a form of bagging, or bootstrap aggregation, where many trees are grown from bootstrap samples of the data, thereby producing a "forest" (Breiman 2001; Cutler et al. 2007). Predictions are made from each tree in the forest. Each tree gives a classification, such that each the tree "votes" for that class. The forest then chooses the classification having the most votes (across all the trees in the forest). Each tree is grown with the following steps. First, the training data are sampled with replacement (i.e., the data are bootstrapped). This sample is the training set for growing the tree. Second, for each node in the tree, n variables are selected at random out of N total variables (typically $n \ll N$) and are used to split the node in the tree. n is held constant during the forest growing, where each tree is grown to the largest extent possible (Breiman 2001). Accuracy and error rates are computed for each sample using the "out-of-bag" samples (those not used in the bootstrap sample) and are then averaged over all predictions. Some benefits of Random Forests include the following: (1) it can run efficiently on large datasets; (2) it can handle many explanatory variables and potential interactions; (3) it is argued to not over-fit; and (4) it can be used in several different types of problems (e.g., classification, survival analysis, clustering, missing value imputation) (Cutler et al. 2007).

7.2.5.4 Maximum Entropy

Maxent is a widely used approach for species distribution modeling, which uses the concept of maximum entropy (Phillips et al. 2006). Elith et al. (2006) provided a comprehensive analysis of the utility of different modeling algorithms for presence-only data and concluded that Maxent was one of the most useful algorithms. This result, coupled with available software that is relatively straightforward to implement, has led to widespread use of Maxent. In addition, it is frequently noted that Maxent is one of the only common distribution modeling algorithms designed specifically for presence-only data, because Maxent does not assume that background points are locations where the species does not occur (i.e., it is not assuming background points are absences), unlike the standard usage of GLMs, GAMs, and regression trees with presence–background data (but see Ward et al. 2009). As such, it might be particularly well suited for presence-only data.

The Maxent modeling framework can be described from several perspectives (Merow et al. 2013). In general, Maxent can be thought of as a log-linear model (Elith et al. 2011) and some parameterizations can be described more generally as an inhomogeneous point process model (Renner and Warton 2013; Phillips et al. 2017). The concept of maximum entropy states the best approximation of an unknown distribution is the one that is most spread out (or uniform), subject to some types of constraints (Franklin 2009). In this case, the constraints are derived from the expected value of the distribution estimated from the presence-only data. In its original formulation, Phillips et al. (2006) provided a geographic perspective regarding a Maxent probability distribution, where the Maxent distribution is equivalent to maximizing the likelihood of a Gibb's probability distribution, which can be written as:

$$p\left(z(\,s_i)\right) = \frac{\exp\left(z(\,s_i)\lambda\right)}{\sum_i \exp\left(z(\,s_i)\lambda\right)}, \tag{7.1}$$

where z is a vector of J environmental variables at locations s_i and λ is a vector of coefficients (Phillips et al. 2006). The numerator of Eq. (7.1) is a log-linear model, while the denominator is a normalization constant, such that $\Sigma\, p = 1$. Note that this latter aspect of the algorithm results in very small values for predictions at individual locations, but one can rescale p to make it more interpretable relative to other modeling algorithms (Elith et al. 2011; Phillips et al. 2017).

The Maxent package commonly used also includes other aspects of modeling that is not based on the idea of maximum entropy per se but rather general techniques employed in machine-learning modeling, such as model regularization and the use of "basis" functions or "features" to create non-linear response functions (Phillips et al. 2006; Phillips and Dudik 2008). In statistics, *model regularization* is an approach of shrinkage of parameter coefficients towards zero, which reduces potential over-fitting of models (Tibshirani 1996). *Basis functions* or features are similar to the use of splines in GAMs, where features are an expanded set of transformations of the original covariates (Elith et al. 2011; Hefley et al. 2017). The practical difference between features used in Maxent and that of splines in GAMs is that Maxent can consider some functions that are not polynomial smoothers (e.g., cubic splines; see below). Maxent considers six types of features: linear, quadratic, product, threshold, hinge, and categorical (Fig. 7.7).

Because of these components to the Maxent program, there has been some confusion regarding *why* Maxent may be useful: is it because of the concept of

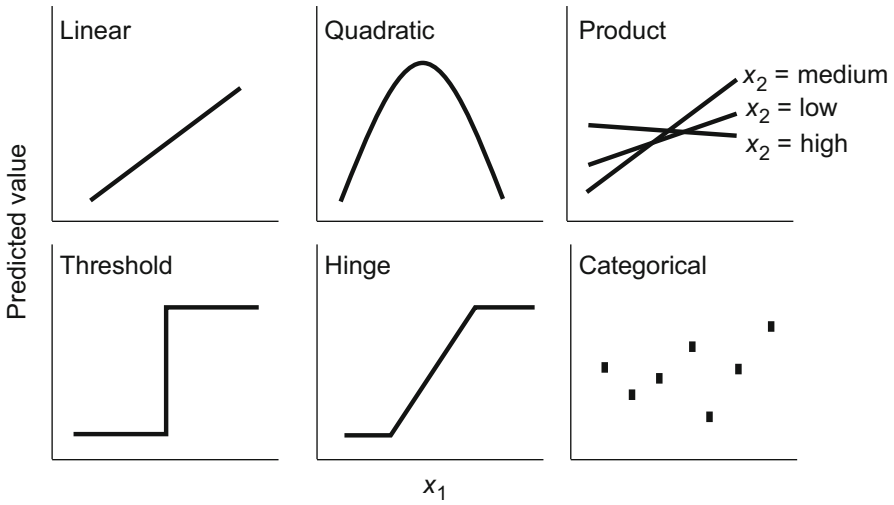


Fig. 7.7 Features considered by Maxent

maximum entropy or is it due to some of these other aspects? For instance, Gaston and Garcia-Vinas (2011) found that logistic regression that used a similar model regularization technique as Maxent (i.e., the "lasso"—least absolute shrinkage and selection operator) performed as well as Maxent, while logistic regression without regularization performed poorly. An important note: while Maxent can be run using presence–absence rather than presence–background (available) data, the underlying theory of the Maxent algorithm is based on presence-only data. Consequently, Maxent should not be used for presence–absence analysis of species distributions (Guillera-Arroita et al. 2014).

7.2.5.5 Point Process Models

It has been recently shown that several of the above modeling frameworks that focus on presence-only data can be derived more generally as spatial point process models (PPMs), including the use of Maxent, GLMs, GAMs, and Boosted Regression Trees (Aarts et al. 2012; Fitian and Hastie 2013; Renner and Warton 2013; Renner et al. 2015; Phillips et al. 2017). Previously, we have discussed point process models in the context of understanding spatial point patterns (Chap. 4). Here, the idea is that presence-only data can be viewed as point locations across a bounded region of interest, such that inhomogeneous point process models can describe the intensity (~density) of species, λ, in the region. The realization that many of the above SDM algorithms can be viewed as inhomogeneous point process models provides a unification of different modeling frameworks and it helps provide guidance for some recurring problems in distribution modeling (Warton and Shepherd 2010; Renner and Warton 2013; Phillips et al. 2017).

A point process is inhomogeneous when intensity varies across a region. Variation in intensity is captured by spatially explicit covariates by modeling intensity based on a log-linear relationship:

$$\log\lambda(s) = \alpha + \beta z(s), \tag{7.2}$$

where s is the species location. Consequently, PPMs are similar to Poisson regression (one type of generalized linear model; see Chap. 6), but the focus is on spatial locations of point occurrences rather than focus being on the point occurrences themselves (Fithian and Hastie 2013). In the likelihood of a point process model, there is a component that focuses specifically on estimating the background environmental conditions. This component can be approximated with background points (Berman and Turner 1992), referred to as "quadrature points" (because these points approximate the function that describes the background environment). Fithian and Hastie (2013) showed that by providing large weights to background points, logistic regression can approximate the inhomogeneous point process model and retrieve reliable parameter estimates of environmental relationships.

In a related way, Renner and colleagues (Renner and Warton 2013; Renner et al. 2015) showed how Maxent and other models can be derived as point process models

in this framework. Why is this useful? By showing a common derivation, it illustrates the relationships among these techniques and better isolates exactly how they are different and implicit ways in which they are similar (e.g., some assumptions thought to not be relevant to Maxent but are relevant to GLMs may need to be reconsidered). In addition, this derivation provides important insight in some key aspects of species distribution modeling. For example, Warton and Shepherd (2010) and Renner et al. (2015) provided interesting discussions on how PPMs help clarify the role of background points and the number background points that should be included in analyses of presence-only data. Renner et al. (2015) emphasized that more background points should be used to estimate point processes than what is typically done in species distribution modeling. Also, the key for background points is that they should adequately capture the environment, such that they suggest that regular grids of points may be helpful, rather than random point generation.

The application of the PPM framework for species distribution modeling frequently only requires minor changes in model development. Renner et al. (2015) and Fithian and Hastie (2013) provided several examples of how these models can be implemented. In general, point process models can be fit in ways similar to other models, but typically more background points are used and presence and background points may be weighted differently (Fithian and Hastie 2013; Renner et al. 2015).

7.2.6 Combining Models: Ensembles

Because of the major differences in assumptions among modeling algorithms and their variable utility under different situations, ecologists have increasing used an "ensemble" approach to modeling (Araújo and New 2007). In a nutshell, ensemble models are typically (weighted) averages or related summaries of different model predictions (Fig. 7.8). For instance, in hurricane forecasting, ensemble predictions of hurricane paths are frequently used to get "consensus" predictions.

To make ensembles, we might take the median probability from a suite of models or take a weighted average, where the weights come from a measure of predictive accuracy (e.g., AUC or TSS). It is often argued that ensemble predictions can be more accurate than predictions from single models (Marmion et al. 2009). Nonetheless, care should be taken when using and interpreting ensembles, because some modeling algorithms are fundamentally predicting different currencies than others (e.g., envelope methods, GLM). For instance, profile methods typically predict environmental similarity while GLM-like models predicts (relative) probabilities of occurrence.

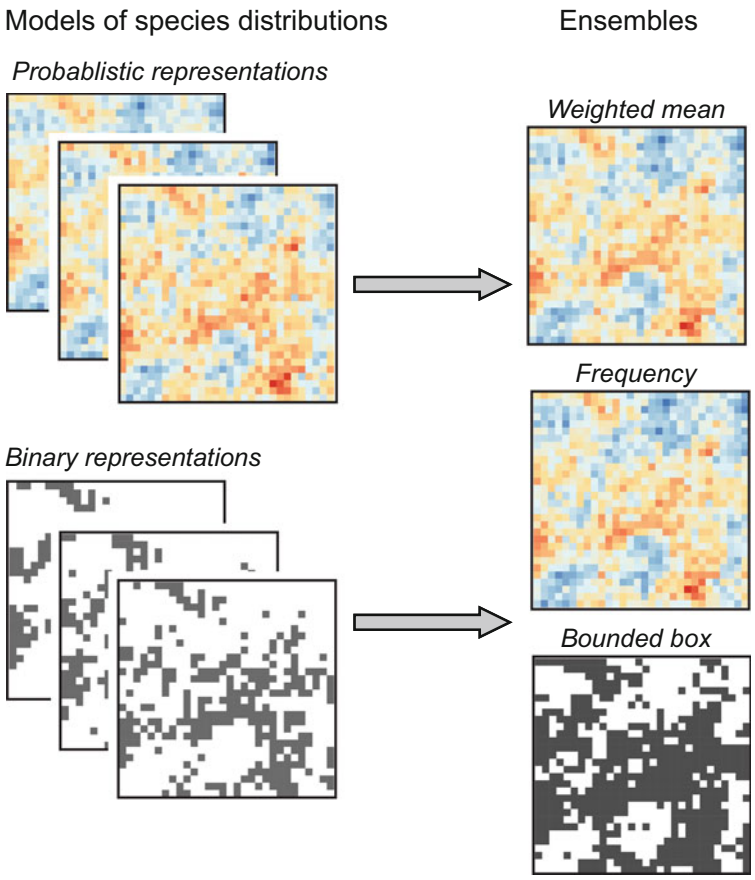


Fig. 7.8 Ensemble modeling integrates predictions from several models to make predictions of species distribution (e.g., occurrence). Ensemble predictions can be based on a variety of approaches. Shown are (weighted) averages of model predictions, frequencies of predicted occurrence based on binary summaries of model predictions, and the use of a bounded box, where at least one model predicts occurrence

7.2.7 Model Evaluation

Models can be evaluated in several ways. In the wildlife literature, there is a strong focus on model selection (e.g., AIC) (Burnham and Anderson 1998). Model selection can be very useful for contrasting the fit of models to data to compare hypotheses. However, model selection alone does not provide explicit information on the predictive performance of models, that is, the ability of models to predict to new locations in space or time, which is often of primary interest in distribution modeling (Hijmans 2012). In this way, predictions from models are frequently used for spatial interpolation (e.g., mapping species distributions), projections (e.g., evaluating

alternative scenarios of expected climate change), and forecasting (e.g., making probabilistic predictions of species distribution in a new time or place).

To evaluate models, the primary approach is to build models with a portion of the data (sometimes called the "training" data) and then use the model to predict observations not used in the model building stage (sometimes called "test data" or "validation data"). This approach is commonly referred to as *external validation* (or *cross validation*) to distinguish it from *internal validation* (or *resubstitution*; see Fielding and Bell 1997), where a model is assessed based on predictions used to build the model. In general, external validation is thought to provide a much more honest assessment of the model performance than resubstitution. There are several approaches for partitioning data to be used for model building and testing. Prospective (independent) sampling, where new data are collected at different locations or time periods than for the data used in model training, is perhaps the most reliable approach but requires greater effort (Fielding and Bell 1997). In the absence of prospective sampling, K-fold partitioning is frequently used (Boyce et al. 2002). In K-fold partitioning, the data are split into K groups, or folds, and $K-1$ folds are used to predict the remaining fold that is left out of modeling training, such that K models are built and evaluated using each data point once as test data. This approach is an efficient way to split data for model evaluation. Folds are often created based on taking a random sample (without replacement) from the data for each fold, although this can result in test and training data being spatially interspersed (with the potential for spatial dependence in responses; see Chap. 6). Other approaches to creating folds include making spatial blocks of data ("K-fold block validation") (Wenger and Olden 2012) or stratifying random samples to ensure the same number of presence locations occur in each fold, such that the spatial distribution between training and testing data share similar characteristics (Hijmans 2012).

Once models are built and predictions are made onto new data, summary metrics are typically used to assess the predictive performance of models. The types of summary metrics and their utility depend on the type of response variable and evaluation data used in assessing models. We briefly summarize some common approaches with presence–absence models, presence-only models, and abundance models.

7.2.7.1 Evaluation with Presence–Absence Data

To evaluate predictions from presence–absence (or detection–non-detection data), we can either consider model discrimination or model calibration (Pearce and Ferrier 2000). *Model discrimination* assesses how well a model can tell the presences from absences (or background points) in the testing data set. In contrast, *model calibration* attempts to measure the agreement between predicted probabilities of occurrence and observed proportions of locations occupied in the testing data set.

Table 7.2 The confusion matrix

	Observed	
Predicted	Present	Absent
Present	a	b
Absent	c	d

Table 7.3 Common metrics derived from the confusion matrix (see Table 7.2 for constants used)

Metric	Equation
False positive rate (errors of commission)	$b/(b + d)$
False negative rate (errors of omission)	$c/(a + c)$
Sensitivity (True positive rate)	$a/(a + c)$
Specificity (True negative rate)	$d/(b + d)$
Correct classification rate	$(a + d)/N$
Prevalence	$(a + c)/N$
Kappa	$[(a + d) - (((a + c)(a + b)\ 1\ (b + d)(c + d))/N)]/$ $[N - (((a + c)(a + b) + (b + d)(c + d))/N)]$
True Skill Statistic	$a/(a + c) + d/(b + d) - 1$

Model Discrimination. For interpreting model discrimination, often the focus is on metrics that can be derived from the *confusion matrix*, or a summary table of predictions of presence–absence relative to observed presence–absence (Table 7.2). Typically, this matrix is obtained by truncating probabilistic predictions to 0/1 data, by selecting a threshold for truncating predictions. However, we note that Lawson et al. (2014) recently showed that the use of the confusion matrix need not require truncating predictions. There are several metrics that can be derived from the confusion matrix (Fielding and Bell 1997), including metrics that focus on certain types of errors in predictions (e.g., false positive or false negative errors), or overall model predictive accuracy (e.g., the correct classification rate). Here we focus on two metrics commonly used in distribution modeling: Kappa and the True Skill Statistic.

Kappa is a commonly used metric that expresses the agreement not obtained randomly between two qualitative variables. Kappa is a popular metric because it takes into account both omission and commission errors (Table 7.3). It is also less problematic than some simpler metrics taken from the confusion matrix, such as the correct classification rate (CCR) (Table 7.3), which can give a misleading interpretation of model performance because high CCR can occur when models predict all presences or all absences for common or rare species, respectively.

The True Skill Statistic (TSS), sometimes called the Hanssen–Kuipers Skill Score, has been traditionally used for assessing the accuracy of weather forecasts. TSS is typically defined as: sensitivity + specificity − 1. Like Kappa, TSS takes into account both omission and commission errors, as well as successes as a result of random guessing. It ranges from −1 to +1, where +1 indicates perfect agreement and values of zero or less indicate a performance no better than random. However, in contrast to Kappa, TSS is less affected by species prevalence (see Alouche et al. 2006). TSS is also thought to not be affected by the size of the validation set. TSS is a

special case of Kappa when the proportions of presences and absences in the validation set are equal.

A common question pertains to how thresholds should be set for defining the confusion matrix. There are several approaches that have been used. Thresholding can be based on a general cutoff (e.g., predicted probability = 0.5), the prevalence of species in the training data, or more complex approaches, such as searching for the threshold that maximizes kappa or some other evaluation metric (Liu et al. 2005, 2013). Simple measures, such as using the prevalence (i.e., the proportion of sites occupied in the training data) can be useful (Liu et al. 2005). Liu et al. (2013) recommended searching for the value that maximizes the sum of specificity and sensitivity. In some cases, the type of error might matter (e.g., false positive or false negative rates may be more problematic in applications) and can be considered in this decision-making process (Fielding and Bell 1997).

Another popular metric for model discrimination is the Area under the Receiver Operating Characteristic (ROC) Curve (AUC), a curve representing the relationship between the false positive fraction (1 − specificity) and the sensitivity (true positive rate) for a range of thresholds. Good model performance is characterized by a curve that maximizes sensitivity for low values of (1 − specificity), that is, when the curve passes close to the upper left corner of the plot. The area under this curve (AUC) measures model discrimination. An AUC value of 0.5 can be interpreted as the model performing no better than a random prediction, with scores approaching 1 indicating progressively better performance. A value of 0.8 for the AUC means that for 80% of the time, a random selection from presence locations will have a prediction greater than a random selection from the absence locations (Fielding and Bell 1997). Thus, it is a rank-based discrimination metric and has a formal relationship to a Wilcoxon sign test. This metric is popular in part because is not dependent on using a threshold. AUC is widely used, but it is not without criticism (Lobo et al. 2008; Peterson et al. 2008). Some known issues with AUC is that it can vary depending on the spatial extent considered, where a larger extent tends to increase AUC. Because of this sensitivity, AUC can be misleading when compared in absolute terms across studies (although within an investigation it may be comparable among model algorithms). This criticism is relevant to other performance metrics as well. It is also frequently argued that the entire range considered by AUC is not biologically meaningful (Lobo et al. 2008). Finally, AUC was developed for presence–absence types of data. Its application to presence-only data should be used with caution.

Model Calibration. Model calibration is an important way to evaluate presence–absence models, where predicted probabilities are contrasted to observed proportion presences (or observed probabilities) in testing data. For example, a model could have good discrimination and yet consistently under (or over) predict the probability of occurrence. Such bias could be problematic when applying models to conservation problems.

Model calibration can be accomplished through two general approaches. First, a common way to interpret how well as model is calibrated is through the use of

calibration plots. In this approach, predicted probabilities of occurrence and observed proportions of sites occupied are contrasted. To do so, often validation data are pooled based on predicted probabilities. By pooling observations from validation data, the proportion of locations occupied can be calculated (rather than relying solely on binary data). This is akin to some types of goodness-of-fit tests in statistics. These plots can be compared qualitatively or more quantitatively, such as comparing regression lines (e.g., intercepts, slopes) fit through different calibration plots (Guisan et al. 2017). Second, in addition to calibration plots, some metrics can be used, such as metrics that focus on the variation explained, error, and likelihoods (Lawson et al. 2014). In particular, the cross-validated log-likelihood and/or deviance ($-2 \times$ log-likelihood) can be calculated as a measure of model calibration (Lawson et al. 2014; Fithian et al. 2015), which have a strong foundation in statistical theory. In this context, the cross-validated log-likelihood (LL_{cv}) is defined as:

$$LL_{cv} = \sum_i \log(p_i y_i + (1 - p_i)(1 - y_i)), \tag{7.3}$$

where p_i is the predicted probability for observation i and y_i is the observed presence or absence of the species in the test data.

7.2.7.2 Evaluation with Presence-Only Data

Evaluating presence-only models can be challenging. When test data are presence–absence, the approaches mentioned above are frequently employed (Elith et al. 2006; Hijmans 2012), although care should be taken because of the subtle differences in model training and testing data. However, when test data are also presence-only, the approaches for presence–absence data should not be used. In such situations, evaluations should be based only on presence locations (and not the background or pseudo-absence locations) (Hirzel et al. 2006). One popular index is the Boyce Index (Boyce et al. 2002). The rationale of this index is to compare the predicted frequency of suitability values at evaluation points for a b classes (where b are bins of suitability; e.g., 0.0–0.2, 0.21–0.4, etc.) to the expected frequency of points based on a random distribution of points across the study area. This approach has been extended to reduce the sensitivity of bin classes on observed outcomes (Hirzel et al. 2006). Phillips and Elith (2010) also provided an extension of calibration plots for presence-only data.

7.2.7.3 Evaluation of Abundance (Count) Responses

Evaluating non-binary responses (e.g., abundance data) is, in many ways, more straightforward that evaluating models based on binary data. In these cases, no transformation of predictions is needed, unlike models based on binary responses.

These approaches typically focus on how well models are calibrated rather than discrimination. Common statistics include the root mean squared error, the coefficient of determination (R^2), and correlation coefficient (Potts and Elith 2006). Root mean squared error, RSME, is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2}, \tag{7.4}$$

where p_i is the prediction for observation i and y_i is the observed value. In addition, statistics such as the deviance or cross-validated log likelihood can be used.

7.3 Examples in R

We illustrate the process of fitting species distribution models to presence-only data. Our goals are to contrast common modeling techniques and illustrate how models can be interpreted and evaluated. We also illustrate how different types of model evaluation can alter the conclusions regarding the utility of species distribution models.

7.3.1 Packages in R

In R, there are several libraries that can be used for species distribution modeling. Four common "wrapper" packages include `dismo` (Hijmans et al. 2017), `sdm` (Naimi and Araújo 2016), `ecospat` (Di Cola et al. 2017), and `biomod2` (Thuiller et al. 2016). These packages call other packages to perform a variety of species distribution models, including all those mentioned above and several others. Each of the models considered in `dismo` and `biomod2` could be implemented with other packages in R. For the purposes of illustration, here we will largely use individual packages because this provides greater flexibility and transparency in model development. We will also use `dismo` for implementation of some models not available in other packages. We use the `PresenceAbsence` package for model evaluation, which has a comprehensive set of evaluation metrics (Freeman and Moisen 2008), but several other packages can also evaluate models.

7.3.2 The Data

We will return to the data used in Chap. 6 on spatial regression for illustrating species distribution modeling techniques: the Northern Region Landbird Monitoring

Program (Hutto and Young 2002). In this monitoring program, sampling locations consisted of point counts (100-m radius) along a transect (10 points/transect; transects were approximately 3 km long), with transects randomly selected within USFS Forest Regions across Montana and Idaho. These points were also resampled over time (temporal repeated measures), although we will not consider these temporal repeated measures here. We will subset data to consider presence-only observations to illustrate presence-only modeling, but we will use presence–absence data for model evaluation, similar to prior syntheses on presence-only modeling techniques (e.g., Elith et al. 2006).

We again focus on the varied thrush. McCarty et al. (2012) modeled the occurrence of several species in this region, including the varied thrush. The varied thrush is a species of conservation interest, in part because it has declined in the region over the past 30 years (see Chap. 6), and it is considered an "interior" and "old-growth" species (Brand and George 2001; Betts et al. 2018). McCarty et al. (2012) considered the following covariates: canopy cover, the presence of mesic forest, elevation, and mean annual precipitation (see also George 2000). We consider each of these factors. Original GIS layers for canopy cover and mesic forest were 15-m resolution digital land-cover maps developed by the United States Forest Service Northern Region Vegetation Mapping Program (USFS R1-VMP), using Landsat TM imagery and aerial photography (Brewer et al. 2004). McCarty et al. (2012) used a Principal Components Analysis (PCA) to reduce the number of canopy cover variables from three to two. One principal component reflected a linear gradient of canopy cover, which we use here, whereas the other component reflected a non-linear gradient (high factor loadings on intermediate categories of canopy cover). We consider the proportion of mesic forest within a 1-km buffer. The 1-km landscape scale was chosen on the basis of other investigations in this region that showed strong correlations of avian distribution at this scale (Tewksbury et al. 2006; Fletcher and Hutto 2008), although other scales could be considered to best determine the scale of effect (see Chap. 2). Elevation was derived from a 30-m resolution Digital Elevation Model. Prior to analysis, all GIS layers were aggregated to a 200-m resolution, reflecting the grain of the sampling unit (100-m-radius point counts). Mean annual precipitation data come from the PRISM Climate Group at Oregon State University (http://www.prismclimate.org).

7.3.3 Prepping the Data for Modeling

There are several steps to prepping data for distribution modeling, depending on the data sources. In particular, working with opportunistic data often requires vetting observation and collating information to create relevant data frames for modeling purposes. See Di Cola et al. (2017) for more on these issues.

We first load the response data and subset the data based on presence–absence, as well as the *x*-*y* coordinates for presence–absence locations, to allow for simple extraction for modeling. There are two sources of data we consider. The first

comes from the entire region collected in 2004 (`vath.data`). The second we consider as independent (prospective sampling) validation data (`vath.val`) collected in the region in 2007–2008 at a subset of points considered in 2004.

```
> vath.data <- read.csv(file = "vath_ 2004.csv", header = T)
> vath.val <- read.csv(file = "vath_VALIDATION.csv", header = T)

#subset to presence-only / absence-only
> vath.pres <- vath.data[vath.data$VATH == 1,]
> vath.abs <- vath.data[vath.data$VATH == 0,]
> vath.pres.xy <- as.matrix(vath.pres[,cbind("x", "y")])
> vath.abs.xy <- as.matrix(vath.abs[,cbind("x","y")])

#validation
> vath.val.pres <-
  as.matrix(vath.val[vath.val$VATH == 1, cbind("x", "y")])
> vath.val.abs <-
  as.matrix(vath.val[vath.val$VATH == 0, cbind("x","y")])
> vath.val.xy <- as.matrix(vath.val[,cbind("x","y")])
```

Next, we will load raster grids that contain relevant spatial information on the covariates we will consider (Fig. 7.9).

```
> library(raster)
> elev <- raster("elev.gri") #elevation layer (km)
> canopy <- raster("cc2.gri") #linear gradient from PCA
> mesic <- raster("mesic.gri") #presence of mesic forest
> precip <- raster("precip.gri") #mean precip (cm)

#check maps
> compareRaster(elev, canopy)

##
[1] TRUE

> compareRaster(elev, mesic)

##
Error in compareRaster(elev, precip) : different extent
```

In this situation, these maps do not align because they are of slightly different resolutions and extent, where the mesic forest is a resolution of 210 $\times$ 210 m while the others are 200 $\times$ 200 m. The elevation and canopy layers have the same extent, but the others are slightly different. As a consequence, we cannot create a raster stack (or brick) of these data. To rectify this problem, we resample the precipitation and mesic forest layers to be consistent with the elevation and canopy layers. Note that we use the `"ngb"` method for `mesic`, a categorical (binary) variable, and `"bilinear"` method for `precip`, a continuous variable.

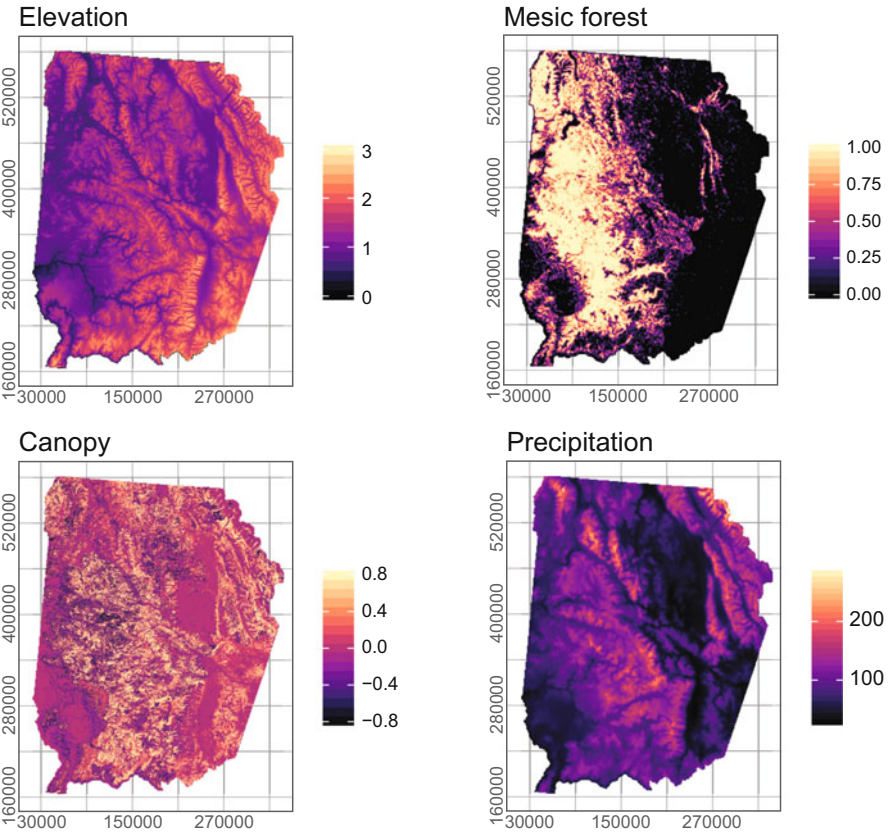


Fig. 7.9 Explanatory variables considered in model building, including elevation (in km), canopy cover (a derived metric taken from a Principal Components Analysis), the percent of mesic forest cover within 1 km of each location, and mean precipitation (cm)

```
#resample to align layers
> mesic <- resample(x = mesic, y = elev, "ngb")
> precip <- resample(x = precip, y = elev, "bilinear")

#crop to same extent
> mesic <- mask(mesic, elev)
> precip <- mask(precip, elev)

> compareRaster(elev, precip, canopy, mesic)

##
[1] TRUE
```

This resampling and masking aligns the raster data. Before we create a raster stack, we also add a larger scale covariate for mesic forest: the proportion of mesic forest in the surround 1 km.

```
#make 1 km wet forest
> fw.1km <- focalWeight(mesic, 1000, 'circle')
> mesic1km <- focal(mesic, w = fw.1km, fun = "sum", na.rm = T)
```

We can now create a raster stack of the environmental covariates (Fig. 7.9).

```
> layers <- stack(canopy, elev, mesic, mesic1km, precip)
> names(layers) <- c("canopy", "elev", "mesic", "mesic1km", "precip")

> plot(layers)
> pairs(layers, maxpixels = 1000)
```

Because mesic and mesic1km are highly correlated, we only consider mesic1km in further modeling. We can use the `dropLayer` function to remove that layer from the raster stack:

```
> layers <- dropLayer(layers, 3)
```

We can generate background points in several ways. The `dismo` package includes the `randomPoints` function for generating random points without replacement. For distribution modeling, we may want to generate points without replacement, because sampling with replacement would potentially create duplicate records (but see Renner et al. 2015). In addition, the `raster` package has the `sampleRandom` and `sampleRegular` functions, which can also generate availability points (without replacement). The `randomPoints` and `sampleRandom` functions are similar, but there is one key difference in the context of distribution modeling: the `randomPoints` function allows the user to also provide the presence points and, if so, it will not generate available points that those locations. We will illustrate the use of this package, generating 2000 background points. We choose this number for computational purposes only. In practice, we may want to increase this number substantially (Renner et al. 2015), but 2000 should be sufficient for illustration here.

```
> library(dismo)
> back.xy <- randomPoints(layers, p = vath.pres.xy, n = 2000)
> colnames(back.xy) <- c("x", "y")
```

With these locations and the points we identified above for presence and validation data, we extract covariate values at each point with the `extract` function, remove potential NAs (where random points were generated but not all environmental data occur), and link them into a single data frame:

```
> pres.cov <- extract(layers, vath.pres.xy)
> back.cov <- extract(layers, back.xy)
> val.cov <- extract(layers, vath.val.xy)

#link data
> pres.cov <- data.frame(vath.pres.xy, pres.cov, pres = 1)
> back.cov <- data.frame(back.xy, back.cov, pres = 0)
> val.cov <- data.frame(vath.val, val.cov)

#remove any potential NAs
> pres.cov <- pres.cov[complete.cases(pres.cov),]
> back.cov <- back.cov[complete.cases(back.cov),]
> val.cov <- val.cov[complete.cases(val.cov),]

> all.cov <- rbind(pres.cov, back.cov) #combine data
```

These data can now be used with a variety of modeling techniques.

7.3.4 Contrasting Models

7.3.4.1 Envelopes

Envelope models can be readily fit in the `dismo` package. In these models, we only use the presence locations. To create the envelope, the `bioclim` function in `dismo` calculates the percentiles of observed environmental covariates at presence locations and the values of covariates at each location on the map are compared to these percentiles. The closer the value of the location to the median value of a covariate at presence locations, the more suitable that location is deemed to be. Then, the minimum similarity value across covariates is used (analogous to Liebig's Law of the Minimum; Austin 2007). In our dataset, the envelope can be calculated as:

```
> bioclim.vath <- bioclim(layers, vath.pres.xy)
```

Here, the model will consider all covariates in the layer stack. We can plot the environmental variation at (~envelopes) the presence locations and produce a predictive map from this model:

```
#envelope plots
> plot(bioclim.vath, a = 1, b = 2, p = 0.85) #canopy-elev 85%
> plot(bioclim.vath, a = 1, b = 2, p = 0.95) #canopy-elev 95%
> plot(bioclim.vath, a = 1, b = 3, p = 0.85) #canopy-mesic 85%
```

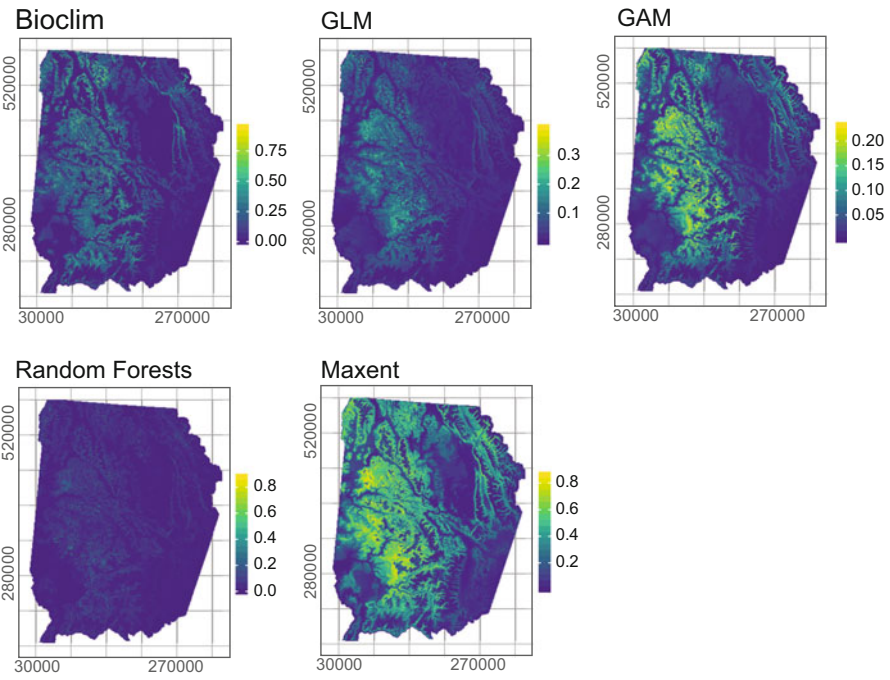


Fig. 7.10 Predictive maps taken from several distribution modeling techniques

```
#map it
> bioclim.map <- predict(layers, bioclim.vath)
> plot(bioclim.map, axes = F, box = F, main = "bioclim")
```

This map (Fig. 7.10) reflects the similarity of locations to environmental covariates at presence locations. It is scaled such that a value of 1 would be locations that have the median value of all covariates considered, while a value of zero would reflect locations where at least one covariate is outside the range of environmental covariates at presence locations.

While this model is simple in form, it illustrates the extent to which locations fall within the environmental variation of observed locations. Note that in doing so, it may often over-predict distributions.

7.3.4.2 GLMs and GAMs

Generalized linear models (GLMs) and generalized additive models (GAMs) are frequently used in distribution modeling. In these cases, presence–absence or

presence–background data are used (Fithian and Hastie 2013). We fit a simple GLM (see Chap. 6 for more on GLMs and spatial regression models):

```
> glm.vath <- glm(pres ~ canopy + elev + I(elev^2) + mesic1km +
  precip, family = binomial(link = logit), data = all.cov)

> summary(glm.vath)

##
Call:
glm(formula = pres ~ canopy + elev + I(elev^2) + mesic1km + precip,
 family = binomial(link = logit), data = all.cov)

Deviance Residuals:
 Min 1Q Median 3Q Max
−0.8053 −0.3377 −0.2130 −0.1274 3.5746

Coefficients:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) −12.186128 2.001925 −6.087 1.15e−09 ***
canopy 0.655128 0.282635 2.318 0.02045 *
elev 13.207998 3.251465 4.062 4.86e−05 ***
I(elev^2) −5.477279 1.293859 −4.233 2.30e−05 ***
mesic1km 1.127415 0.376421 2.995 0.00274 **
precip 0.011051 0.004529 2.440 0.01468 *
- - -
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

 Null deviance: 773.28 on 2093 degrees of freedom
Residual deviance: 667.91 on 2088 degrees of freedom
AIC: 679.91

Number of Fisher Scoring iterations: 8
```

In this model, we consider linear relationships for all covariates, except elevation, which we allow to be a non-linear (quadratic) relationship (`I(elev^2)`) because we expect that thrushes may be most likely to occur at moderate elevations (see Chap. 6). We do not consider model selection here, but model selection could be performed manually with a variety of packages, such as `MuMIn` (Barton 2018). This model suggests that there is a strong non-linear effect of elevation and a linear, positive effect of mesic forest within the surrounding 1 km. The other two covariates show weaker, positive linear relationships.

We can make a predicted map with:

```
> glm.map <- predict(layers, glm.vath, type = "response")
```

In this case, we specify `type ='response'` to make predictions on the probability scale. Otherwise, predictions would be on the link scale (here, the logit scale).

Generalized additive models can be fit with a few packages; here we illustrate the use of the `mgcv` package (Wood 2006). The default approach in `mgcv` is to optimally determine the number of knots via generalized cross-validation and to use thin-plate splines as a smoother. In this syntax, the `s()` function specifies that a spline will be applied to a covariate.

```
> library(mgcv)
> gam.vath <- gam(pres ~ s(canopy) + s(elev) + s(mesic1km) +
  s(precip), family = binomial(link = logit), data = all.cov,
  method = "ML")

> summary(gam.vath)

##
Family: binomial
Link function: logit

Formula:
pres ~ s(canopy) + s(elev) + s(mesic1km) + s(precip)

Parametric coefficients:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) −4.068 0.252 −16.14 <2e−16 ***
- - -
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
 edf Ref.df Chi.sq p-value
s(canopy) 1.000 1.000 4.373 0.03651 *
s(elev) 3.157 3.997 23.796 9.28e−05 ***
s(mesic1km) 1.000 1.000 1.550 0.21316
s(precip) 4.403 5.226 19.671 0.00158 **
- - -
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.0709 Deviance explained = 17.3%
-ML = 335.55 Scale est. = 1 n = 2094
```

Results from this default GAM are generally similar to the GLM. We can tune the GAM by manually setting the number of knots (see Chap. 6, Fig. 6.5), requesting a different type of smoother function, or by allowing for potential interactions between predictor variables. We illustrate examples of each of these types of tuning. Also, note that we could include linear, rather than smoother terms, to the model by

removing the 's' command around covariates. First, we specify the number of knots used manually, for example:

```
> gam.vath.knot3 <- gam(pres ~ s(canopy, k = 3) + s(elev, k = 3)
  + s(mesic1km, k = 3) + s(precip, k = 3), method = "ML", family =
  binomial(link = logit), data = all.cov)

> gam.vath.knot6 <- gam(pres ~ s(canopy, k = 6) + s(elev, k = 6)
  + s(mesic1km, k = 6) + s(precip, k = 6), method = "ML", family =
  binomial(link = logit), data = all.cov)
```

As the number of knots increase, the complexity of the smoother increases. Note that we also ask for model fitting with maximum likelihood (`method="ML"`), which allows us to make formal comparisons among models using model selection criteria. We can incorporate the potential for interactions between smoothers using a "tensor" product term. Tensor product smoothers address the potential for capturing interactions among variables that can be on different units of measurement (Wood 2006). They can be incorporated as:

```
> gam.vath.tensor <- gam(pres ~ te(canopy, elev, precip,
  mesic1km), family = binomial(link = logit), method = "ML", data
  = all.cov)
```

Finally, we can contrast thin-plate spline smoothers (the default in `mgcv`) to other smoother functions, such as a cubic spline ('`cr`'):

```
> gam.vath.cr <- gam(pres ~ s(canopy, bs = "cr") + s(elev, bs =
  "cr") + s(mesic1km, bs = "cr") + s(precip, bs = "cr"), family =
  binomial(link = logit), method = "ML", data = all.cov)
```

The relationships modeled when altering the number of knots and the smoother do not change much at all in this example (see Fig. 7.11 for some example plots). Overall, this model refines our understanding relative to the GLM, suggesting that varied thrush respond non-linearly to precipitation gradients in addition to elevation. This model tuning can be formally evaluated with model selection criteria, such as AIC:

```
> AIC(gam.vath, gam.vath.knot3, gam.vath.knot6, gam.vath.tensor, gam.
  vath.cr)

##
 df AIC
gam.vath 12.2 663.9
gam.vath.knot3 7.7 670.3
gam.vath.knot6 10.4 662.0
gam.vath.tensor 25.8 658.1
gam.vath.cr 12.9 663.9
```

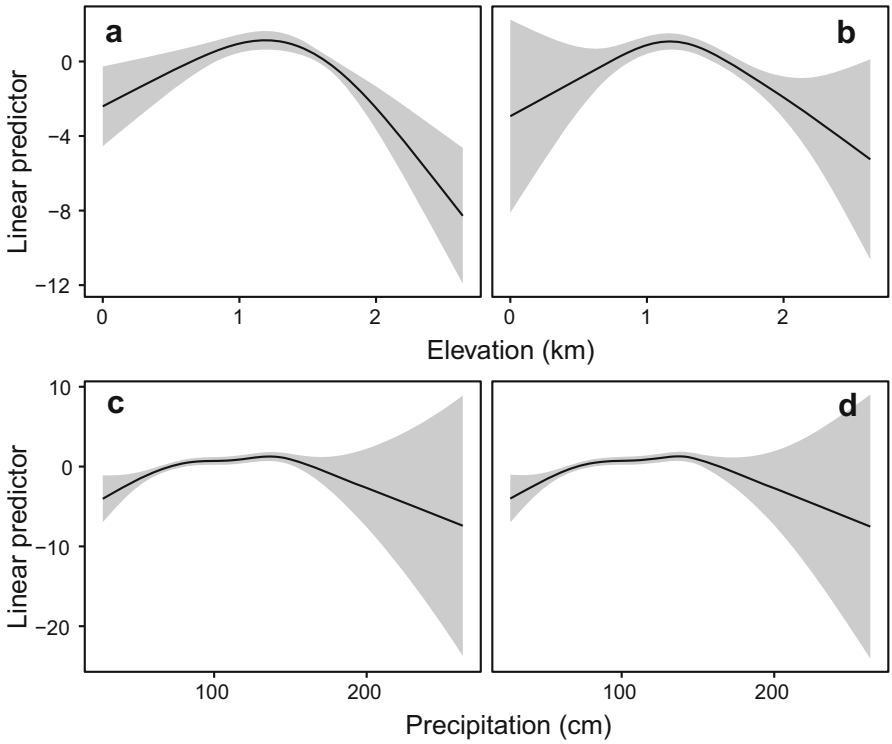


Fig. 7.11 Tuning a generalized additive model. Top panel contrasts (**a**, **b**) the number of knots used for modeling relationships of thrush occurrence with elevation (3 versus 6 knots). Bottom panel contrasts (**c**) Thin-plate spline versus (**d**) cubic spline (for automated knot selection) for precipitation

In this case, we find that there is some support for only having six knots in the smoother and the use of the tensor product. We can then map the model in a way similar to above:

```
> gam.map <- predict(layers, gam.vath.knot6, type = "response")
```

7.3.4.3 Regression Trees and Forests

Here, we focus on the application of Random Forest models using the `randomForest` package (Liaw and Wiener 2002). For Boosted Regression models, see the `gbm` package and the tutorial in Elith et al. (2008). The `randomForest` package can model both categorical (classification) and

continuous (regression) response variables. We will implement a classification model. The default model function can be implemented as:

```
> library(randomForest)
> rf.vath <- randomForest(as.factor(pres) ~ canopy + elev +
  mesic1km + precip, data = all.cov, na.action = na.omit)
```

There are two parameters that are frequently adjusted for model tuning: `mtry` and `ntree`. `mtry` is the number of explanatory variables that are sampled for each tree, while `ntree` is the number of trees that are grown to produce the forest. We use the `tuneRF` function to determine the optimal values for `mtry`:

```
> rf.vath.tune <- tuneRF(y = as.factor(all.cov$pres), x =
  all.cov[,c(3:6)], stepFactor = 0.5, ntreeTry = 500)
```

Here we specify `ntreeTry` = 500, which is the default in the function. In general, it is thought that predictions are less sensitive to `ntree` than `mtry`. The `tuneRF` function adjusts `mtry` at different intervals (`stepFactor`), determining which value minimizes the predictive error (out-of-bag error). With this tuning, we update the model with `mtry=1` based on the out-of-bag error:

```
> rf.vath <- randomForest(as.factor(pres) ~ canopy + elev +
  mesic1km + precip, data = all.cov, mtry = 1, ntree = 500,
  na.action = na.omit)
```

We can then map the Random Forest prediction, similar to other models (Fig. 7.10).

```
> rf.map <- predict(layers, rf.vath, type = "prob", index = 2)
> plot(rf.map)
```

The primary difference here is that we specify `'index = 2'` because the `predict` function will make predictions for each class (there can be $\geq$2). In this case, it provides predictions for 0 and 1, with 1 being the second column from the predict object (thus, we ask for `index = 2` to plot predictions).

7.3.4.4 Maximum Entropy

The use of maximum entropy for species distribution modeling relies on the Maxent program, which is a stand-alone Java software that is freely downloaded (http://biodiversityinformatics.amnh.org/open_source/maxent/). We can call this package in R via `dismo`. Note Phillips et al. (2017) also recently released the `maxnet` package for fitting Maxent models in R based on its relationship to the

inhomogeneous point process model (see Sect. 7.2.5.5 for more). This package may be preferred in many cases because it does not require linking to the stand-alone Maxent program. We first focus on the use of `dismo` to call the Maxent software, because of its widespread use and useful interface for comparing across models, but briefly illustrate the use of `maxnet` in Sect. 7.2.5.5.

To call the stand-alone Maxent software from R, Java (https://www.oracle.com/java/index.html) must be installed on your computer. Note that if you run R on a 64 bit platform, you will need to make sure to install Java for 64 bit. Also, `rJava` (Urbanek 2017) will need to be installed and loaded in R. Once Maxent is downloaded, the maxent.jar file must be placed in the java folder of the `dismo` package. The location of this file can be found with the following:

```
> system.file("java", package = "dismo")
```

The `maxent` function calls Maxent to fit the model:

```
#default
> max.vath <- maxent(layers, p = vath.pres.xy)
```

The default `maxent` function takes presence-only points (`vath.pres.xy`) and generates 10,000 background points for comparison, extracting environmental data from these points and the presence points. We can manually provide background points instead, which can be useful to control the precise number and location of points used when comparing modeling techniques.

```
#default, but provide background points
> max.vath <- maxent(layers, p = vath.pres.xy, a = back.xy)
```

We can tune the Maxent model in several ways (Merow et al. 2013). Two common approaches are to: (1) adjust the regularization of the model; and (2) adjust the types of features considered (Fig. 7.12). We illustrate examples of both of these adjustments.

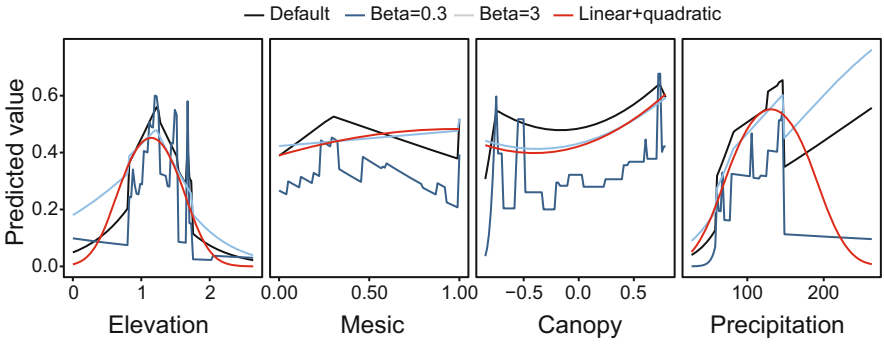


Fig. 7.12 Tuning a Maxent model. Shown are the default response curves, setting the regularization multiplier (β) to 0.3 and 3, and only considering linear and quadratic features

First, the regularization parameter can be changed manually. In this context, Maxent uses the lasso technique for regularization, such that coefficients that do not explain variation in presence locations are penalized and shrink toward zero. In this way, the default value for regulation is proportional to the number of presence locations and the variability in the environmental covariate at presence locations (Elith et al. 2011). The parameter, β, is a constant that is multiplied by the default regularization value. As β increases, a greater penalty is imposed. We can check this by adjusting beta and plotting changes in response curves (Fig. 7.12).

```
> maxent.beta3 <- maxent(layers, p = vath.pres.xy, a = back.xy,
  args = c("betamultiplier=3"))
```

In the above model, we specify a beta multiplier of 3 (the default setting is 1). Typically, this multiplier is altered to be > 1 because of concerns regarding potential overfitting of environmental relationships, but in Fig. 7.12 we illustrate setting the multiplier to be < 1 as well. We can also alter model complexity in terms of the features considered:

```
> maxent.features <- maxent(layers, p = vath.pres.xy, a =
  back.xy, args = c("noproduct", "nohinge", "nothreshold"))
```

In the above model, we tell Maxent to not use product (interactions), hinge, or threshold features. This reduces the model complexity to only consider linear and quadratic features, similar to a simple GLM. We can interpret the impacts of this tuning on partial relationships with the `dismo` package (see below for customizing partial plots) (Fig. 7.12):

```
> response(max.vath, expand = 0)
> response(maxent.beta3, expand = 0)
> response(maxent.features, expand = 0)
```

In the above response functions, we specify `expand = 0` to constrain the response plots only to the range of data considered. We can also evaluate the models with the `evaluate` function from the `dismo` package to get AUC statistics for each model. This function requires passing validation presence and absence points. Here, we use the validation samples (output not shown).

```
> evaluate(p = vath.val.pres, a = vath.val.abs, max.vath, layers)
> evaluate(p = vath.val.pres, a = vath.val.abs, maxent.beta3, layers)
> evaluate(p = vath.val.pres, a = vath.val.abs, maxent.features,
  layers)
```

This comparison suggests that each of these models are similar, in terms of AUC. See Sect. 7.2.7 *Model Evaluation* for a more detailed evaluation assessment. Finally, we can map the model (Fig. 7.10), similar to above :

```
> max.map <- predict(layers, max.vath)
```

Note that the prediction values for this Maxent model tend to be much higher than the GLM, GAM, and Random Forest model (Fig. 7.10). Why is that? Maxent provides different ways to plot and interpret the predictions. The default approach in this function is the "logistic" output, whereas the underlying Maxent model output is termed "raw" output. In the raw output, probabilities across the region sum to 1, such that the probability in any given location is very low and is essentially a probability density, sometimes referred to as relative occurrence rate (ROR; Merow et al. 2013). This can be requested in the `predict` function as:

```
> max.raw.map <- predict(layers, max.vath, args = "outputformat
  = raw")
```

The logistic output is a transformation of the raw output, aimed at providing probabilities that are more akin to probabilities of occurrence (Elith et al. 2011). In doing so, the average prediction for a location where a presence point occurrence with the logistic output approaches 0.5. Another alternative to the logistic and raw outputs is the cumulative log-log (cloglog) output (Fithian et al. 2015), which is better rooted in probability theory and is now the default output in the stand-alone Maxent software (Phillips et al. 2017). These different response outputs should not change the rank suitabilities from models, but they will change the absolute values such that care should be taken when implementing and interpreting output.

7.3.4.5 Point Process Models

Finally, we note that most of the above models can be recast formally as inhomogeneous point process (IPP) models. There are several benefits for doing so, because this perspective provides a means to better understand the number of background points needed, understand the role of spatial dependence, and interpret goodness-of-fit and related model diagnostics (Fithian and Hastie 2013; Phillips et al. 2017).

To implement the above models as point process models, Renner et al. (2015) suggested that many more background points should be considered because they are interpreted as "quadrature" points used for approximating an integral in the point process function that describes the background environment. Warton and Shepherd (2010) argued that it is natural to do so by creating a grid of background points (rather than random point generation), which could be created with the `sampleRegular` function in the `raster` package. With these points, point process models can be fit with a variety of packages. A simple updating of the above GLMs and GAMs with a point process formulation would be (Renner et al. 2015):

```
> glm.ppm <- glm(pres ~ canopy + elev + I(elev^2) + mesic1km +
  precip, family = binomial(link=logit), weights = 1000^(1-pres),
  data = all.cov)
```

```
> gam.ppm <- gam(pres ~ s(canopy) + s(elev) + s(mesic1km) +
  s(precip), family = binomial(link = logit), weights = 1000^(1-
  pres), data = all.cov)
```

We use weighted regressions in the above models to approximate the inhomogeneous point process, where we provide arbitrarily large weights to the background points. Also note that when implementing this model, we should include a larger number of background points than what is shown here, potentially sampled in a regular grid. The number of background points can be formally determined in this context by altering the number of background points until the likelihood of the model stabilizes (Renner and Warton 2013; Renner et al. 2015).

We can also fit a Maxent model with a point process formulation using the `maxent` function in the `dismo` package:

```
> maxent.ppm <- maxent(layers, p = vath.pres.xy, a = back.xy,
  args = c("noremoveduplicates"))
```

The key difference in the above model is that in the `maxent` function we specify to not remove duplicate records (multiple presence locations within a cell on the map). If we did not pass our own background points, we would also need to add `"noaddsamplestobackground"` and increase the number of background points generated (e.g., `"maximumbackground = 50000"` for 50,000 points).

Finally, the `maxnet` package can be used as well, which uses the `glmnet` package (Friedman et al. 2010) to fit a Maxent-formulated IPP (based on the idea of "infinitely weighted logistic regression") (Fithian and Hastie 2013) that uses the same regularization and features that the stand-alone Maxent package provides. In this case, `maxnet` requires a different data format than the `maxent` function in `dismo`, where we provide a vector of presence and background locations and a data frame or matrix of covariates at those locations:

```
> library(maxnet)
> library(glmnet)
> max.cov <- all.cov[,c("canopy", "elev", "mesic1km", "precip")]
> maxnet.ppm <- maxnet(all.cov$pres, max.cov)
```

In this function, features can be requested and the regularization constant adjusted in the following way:

```
> maxnet.beta3.linquad <- maxnet(all.cov$pres, max.cov, regmult
  = 3, maxnet.formula(all.cov$pres, max.cov, classes = "lq"))
```

The `classes` statement provides a means to select features for the model, with all features being '`lqhpt`' (linear, quadratic, hinge, product, threshold). We do not focus on these model IPP formulations below, but the interested reader should see Renner et al. (2015).

7.3.5 *Interpreting Environmental Relationships*

Each of these distribution modeling algorithms uses some sort of function regarding species distribution and environmental factors. A challenge is to interpret these functions in a meaningful way across models.

A common approach to do so is through use of partial response plots (or "partial plots"). In these plots, we vary one environmental covariate across the range of observed variation while setting all other environmental covariates to a constant, typically their mean or median. We then make predictions on this new data set to interpret how the models are relating species occurrence to environmental factors. Note that this approach will not adequately illuminate potential interactions between variables if they are considered in models (e.g., through the use of tensor products in GAMs or in Random Forest models). However, it can still be useful for interpreting patterns that underlie the predictions for each algorithm. Elith et al. (2005) generalized this idea to model algorithms that only make predictions on raster grids with what they term the "evaluation strip," or the addition of data to a raster grid that serves a similar purpose as making predictions to new data with partial response plots.

There are some packages that provide functions for calculating partial plots (e.g., the `response` function used above in the `dismo` package), and some of the wrapper packages, such as `biomod2` provide general functions in this way. Here, we illustrate how users can manually accomplish this task, which provides a means to alter graphics or subtle aspects of predictions (e.g., adding uncertainty in predictions). The following code focuses on creating partial plots for the elevation covariate, but see Fig. 7.13 for plots of all covariates. We first generate a new data set (`elev.partial.data`) for predictions:

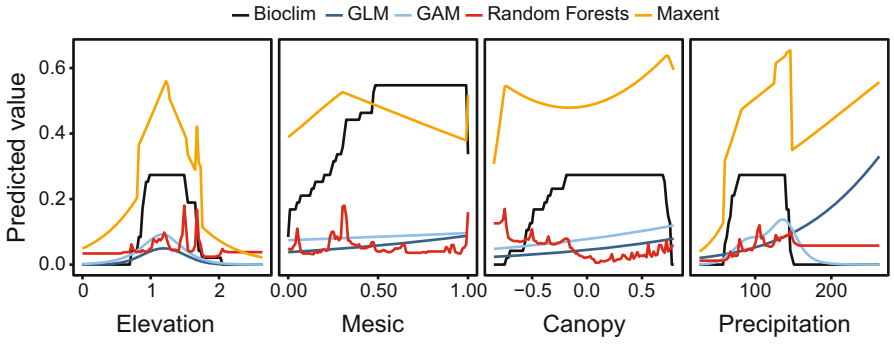


Fig. 7.13 Partial plots from models for elevation (km), proportion of mesic forest within 1km, canopy cover (relative units based on a PCA), and precipitation (cm). For each covariate, all other covariates were set to their median value

```
> canopy.median <- median(back.cov$canopy)
> precip.median <- median(back.cov$precip)
> mesic1km.median <- median(back.cov$ mesic1km)

> elev.range <- seq(min(back.cov$elev),
  max(back.cov$Elev), length = 100)
```

We put the covariates into a data frame and use the `expand.grid` function to expand the data for all possible combinations:

```
> elev.partial.data <- data.frame(expand.grid(Elev = elev.range,
  Canopy = canopy.median, precip = precip.median, mesic1km =
  mesic1km.median))
```

We then make predictions from each model onto this new data set:

```
> bio.pred.elev <- predict(bioclim.vath, elev.partial.data)
> glm.pred.elev <- predict(glm.vath, elev.partial.data, type =
     "response")
> gam.pred.elev <- predict(gam.vath, elev.partial.data, type =
     "response")
> rf.pred.elev <- predict(rf.vath, elev.partial.data, type =
  "prob")
> rf.pred.elev <- rf.pred.elev[,2]
> max.pred.elev <- predict(max.vath, elev.partial.data)
```

Finally, we can use the `plot` function or `ggplot2` (Wickham 2009) to create a partial prediction plot. Here we show the use of `plot` to illustrate plotting the Bioclim, GLM, and Random Forest predictions.

```
#create data frame
> part.elev.df <- data.frame(elevation = elev.range, bioclim =
  bio.pred.elev, glm = glm.pred.elev, gam = gam.pred.elev, rf =
  rf.pred.elev, max = max.pred.elev)

#plot
> plot(part.elev.df$elevation, part.elev.df $bioclim, type = 'l')
> lines(part.elev.df$elevation, part.elev.df$glm, type = 'l',
  col = "red")
> lines(part.elev.df$elevation, part.elev.df$rf, type = 'l', col
  = "blue")
```

These partial plots illustrate the widely divergent environmental functions identified across algorithms in modeling thrush occurrence (Fig. 7.13). Overall, the partial responses for Random Forests are highly complex and non-linear, while the partial responses for the other algorithms are smoother and less complex. Note that the absolute predictions vary as well. This pattern occurs between the modeling algorithms because of the different currencies that they are modeling. Bioclim is

modeling similarity and the predictions from Maxent are based on the logistic output, which tends to make the average prediction for presence locations approximately 0.5 (Elith et al. 2011). In contrast, the GLM, GAM and Random Forests are discriminating presence points versus background points, such that as we increase the number of background points, the probabilities on the *y*-axis will decrease (because increasing the background points decreases the intercept term value in the model). For instance, if we generated the same number of background points as presence points, the intercept on these latter models would generate predictions on these partial plots with means close to 0.5.

7.3.6 Model Evaluation

The above models can be evaluated in a variety of ways and there several packages for model evaluation. The `dismo` package includes the `evaluation` function, but here we use the `PresenceAbsence` package (Freeman and Moisen 2008), which includes a more comprehensive set of evaluation metrics. To use the `PresenceAbsence` package, we create a data frame that includes (in the following order): (1) site IDs for the validation (evaluation) data; (2) the observed responses in the validation data; and (3) model predictions for those locations. This data frame can have predictions from *N* models, where columns for predictions are 3 to *N*+3. We first illustrate model evaluation based on the prospective sampling dataset from 3 to 4 years later in time, and then illustrate how model evaluation can be accomplished with *K*-fold validation (Boyce et al. 2002), a common approach to model evaluation.

For the prospective sampling validation data set, we simply take each of the above models and make predictions for the new locations:

```
> val.cov.pred <- val.cov[,cbind("canopy", "elev", "mesic1km",
  "precip")]
> bio.val <- predict(bioclim.vath, val.cov.pred)
> glm.val <- predict(glm.vath, val.cov.pred, type = "response")
> gam.val <- predict(gam.vath, val.cov.pred, type = "response")
> rf.val <- predict(rf.vath, val.cov.pred, type = "prob")
> rf.val <- rf.val[,2]
> max.val <- predict(max.vath, val.cov.pred)
```

With these predictions, we then create a data frame that is formatted for the `PresenceAbsence` package and we will create a data frame for storing the model evaluation results.

```
> val.data <- data.frame(siteID = 1:nrow(vath.val), obs =
  vath.val$VATH, bio = bio.val, glm = glm.val, gam = gam.val, rf =
  rf.val, max = max.val)
```

```
> summary.eval <- data.frame(matrix(nrow = 0,ncol = 9))
> names(summary.eval)<-c("model", "auc", "corr", "ll",
  "threshold", "sens", "spec", "tss", "kappa")
```

For model evaluation, we will calculate three continuous metrics: AUC, the biserial correlation coefficient, and the cross-validated log-likelihood (Lawson et al. 2014). We will also calculate four binary metrics taken from the confusion matrix: sensitivity, specificity, kappa, and the true skill statistic. The `PresenceAbsence` package can determine thresholds based on a variety of criteria, such as prevalence in the test or training data, maximizing kappa or maximizing the sum of specificity and sensitivity (see `?optimal.thresholds`). Here, we focus on using a threshold that maximizes the sum of specificity and sensitivity (`opt.methods=3` in the `optimal.thresholds` function), which was recommended by Liu et al. (2013). In the following `for` loop, we calculate each of these metrics for each model and populate our summary data frame with the output. We first load the `PresenceAbsence` package and detach `glmnet`, because the latter package also includes a function for calculating AUC.

```
> library(PresenceAbsence)
> detach("package:glmnet")
> nmodels <- ncol(val.data)−2
> for(i in 1:nmodels){
  auc.i <- auc(val.data, which.model = i)
  kappa.opt <- optimal.thresholds(val.data, which.model = i,
    opt.methods = 3)
  sens.i <- sensitivity(cmx(val.data, which.model = i, threshold
= kappa.opt[[2]]))
  spec.i <- specificity(cmx(val.data, which.model = i, threshold
= kappa.opt[[2]]))
  tss.i <- sens.i$sensitivity + spec.i$specificity − 1
  kappa.i <- Kappa(cmx(val.data, which.model = i, threshold =
    kappa.opt[[2]]))
  corr.i <- cor.test(val.data[,2], val.data[,i + 2])$estimate
  ll.i <- sum(log(val.data[,i + 2] * val.data[,2] + (1 −
  val.data[,i + 2]) * (1 − val.data[,2])))
  ll.i <- ifelse(ll.i == "−Inf", sum(log(val.data[,i + 2] +
  0.01) * val.data[,2] + log((1 − val.data[,i + 2])) * (1 −
  val.data[,2])), ll.i)
  summary.i <- c(i, auc.i$AUC, corr.i, ll.i, kappa.opt[[2]],
    sens.i$sensitivity, spec.i$specificity, tss.i, kappa.i[[1]])
  summary.eval <- rbind(summary.eval, summary.i)
}
```

Note that in the above code, we add a small constant to the log-likelihood calculation because the log(0) is undefined (e.g., when the predicted value is 0, as can be the case in the Bioclim model). Based on these summary statistics, it is clear that none of these models appear to predict well to the prospective sampling data set (Table 7.4), despite the fact that these models had clear environmental relationships (see, e.g., `summary(glm.vath)` and `summary(gam.vath)`). This result

Table 7.4 Evaluation of modeling algorithms based on external validation (presence–absence data collected 3–4 years later)

Model	AUC	LL_{cv}	TSS	Kappa
Bioclim	0.586	−685	0.136	0.027
GLM	0.673	−519	0.287	0.106
GAM	0.651	−528	0.237	0.092
Random Forests	0.625	−607	0.182	0.039
Maxent	0.669	−971	0.259	0.164

illustrates the potential challenge of generating accurate species distribution models that can predict accurately over time (Eskildsen et al. 2013; Vallecillo et al. 2009). In this case, the Bioclim and Random Forests models tended to predict the worst of the models considered based on model discrimination, whereas the Maxent model predicts poorly using the logistic output based on the cross-validated log-likelihood, a model calibration metric. We include cross-validated log-likelihoods because they are useful calibration metrics (Lawson et al. 2014); however, they are most properly applied to models trained with presence–absence data rather than presence-only data.

We can also evaluate models with calibration plots. Calibration plots can be easily generated with the `PresenceAbsence` package. For the above models, we use the `calibration.plot` function. An example for the Maxent model is:

```
> calibration.plot(val.data, which.model = 5, N.bins = 5, xlab =
  "Predicted", ylab = "Observed", main = "maxent")
```

Note that this function requires the user to define the number of bins that will be used to pool binary data.

A more common approach is to use *K*-fold validation (Boyce et al. 2002). In that approach we subset the training data into subsets, or folds. We then fit models holding out one fold while using $K-1$ folds for model training. This is then repeated K times. The above evaluation code can be readily applied each fold and then we summarize across folds. We consider fivefolds and apply this approach to the presence–background data used for model training. The `dismo` package has a function `kfold` that will create a vector of k groups based on random allocation to groups, with the constraint that each group is of equal size.

In this *K*-fold case, we are using presence–background data for model evaluation. In general, using such data for model evaluation is limited because no absence data are available for evaluation. In such situations, it is often recommended to use evaluation metrics that only make use of information on presence locations (Guisan et al. 2017). Here, we use the Boyce index, a common metric for evaluating presence-only predictions that does not rely on absence data (Boyce et al. 2002), which can be calculated with the `ecospat` package (Broennimann et al. 2018). We also calculate the same metrics as above for illustrative purposes. In practice, the Boyce index and other metrics aimed specifically for evaluation with presence-only

data should be emphasized (Engler et al. 2004; Hirzel et al. 2004). See Guisan et al. (2017) for more information.

```
#number of k-folds considered
> folds <- 5

#create k-folds
> kfold_pres <- kfold(pres.cov, k = folds)
> kfold_back <- kfold(back.cov, k = folds)
```

Above we apply the `kfold` function separately to the presence and background data. This ensures that each fold will contain the same number of presence points. Then we can apply a `for` loop or something similar to go through each fold. We do not provide the entire for loop here but illustrate how data can be subset for each fold, *k*.

```
#partition data based on each k-fold
 > kfolds <- 1
 > val.pres.k <- pres.cov[kfold_pres == kfolds, ]
 > val.back.k <- back.cov[kfold_back == kfolds, ]
 > val.k <- rbind(val.pres.k, val.back.k)

 > train.pres.k <- pres.cov[kfold_pres != kfolds, ]
 > train.back.k <- back.cov[kfold_back != kfolds, ]
 > train.k <- rbind(train.pres.k, train.back.k)
```

We apply each of these new data sets (either `train.k`, or each component of the training data set, depending on the model algorithm) to the model algorithms of interest described above and make predictions onto the validation data (`val.k`). With this data format, the Boyce index can be calculated for a given model `i` within the `for` loop mentioned above regarding external validation as:

```
> library(ecospat)
> boyce.i <- ecospat.boyce(fit = val.data[,i + 2], obs =
  val.data[1:nrow(val.pres.k),i + 2], res = 100, PEplot = F)
```

Note that `Biomod2` and `sdm` have built-in cross-validation functions; however, here we illustrate how to accomplish *K*-fold validation manually, which allows users to customize how *K*-fold validation is accomplished (see also `ecospat` for functions regarding cross-validation). For example, there has been recent criticism regarding how folds are delineated (Hijmans 2012). In the code above, we randomly select points to folds, yet these points are not likely spatially independent. An alternative is to use "block" *K*-fold validation, where spatial blocks are randomly selected, rather than sample points (Wenger and Olden 2012). In this case, we might randomly select transects or watersheds as blocks for validation purposes. This would be straightforward to accomplish above by sampling transects in lieu of points.

Table 7.5 Evaluating modeling algorithms based on *K*-fold validation

Model	Boyce	AUC	TSS	Kappa
Bioclim	0.525	0.737	0.440	0.080
GLM	0.737	0.781	0.473	0.156
GAM	0.798	0.802	0.462	0.135
Random Forests	0.791	0.839	0.572	0.211
Maxent	0.851	0.803	0.500	0.154

Based on *K*-fold validation, we get a different perspective on the utility of these models (Table 7.5), where summary statistics tend to be higher than with prospective sampling. We also find that the more complex models tend to be favored more, with the Random Forest model tending to predict relatively better when using *K*-fold validation.

7.3.7 Combining Models: Ensembles

With predicted maps, it is straightforward to create model ensembles. A common approach is to make a weighted average of predictions from models based on AUC for each model or some other model evaluation metric (Marmion et al. 2009). We emphasize, however, that because different algorithms model different currencies, we suggest that averaging of predictions should only be made for models that are modeling the same currency. Here, we show how we can create an ensemble based on the GLM and GAM model (Fig. 7.14), which are predicting the same response quantity (and thus on similar currencies; unlike Bioclim and Maxent), but differ in their environmental functions being considered.

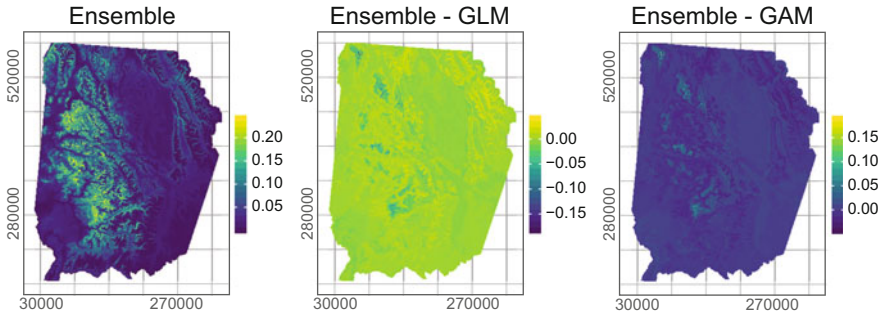


Fig. 7.14 (**a**) An ensemble from the GLM and GAM using a weighted mean based on AUC scores taken from *K*-fold validation, and the difference in ensemble predictions and predictions from (**b**) the GLM, and (**c**) the GAM

```
> models <- stack(glm.map, gam.map)
> names(models) <- c("glm", "gam")

#weighted average based on AUC from prospective sampling
> AUC.glm <- summary.eval[summary.eval$model == "glm", "auc"]
> AUC.gam <- summary.eval[summary.eval$model == "gam", "auc"]
> auc.weight <- c(AUC.glm, AUC.gam)

> ensemble.auc <- weighted.mean(models, auc.weight)
> plot(ensemble.auc)
```

Other approaches to ensemble modeling can include truncating predictions to binary information of predicted presence/absence and then summarizing this information in a variety of ways. Such truncation might be preferred when combining very different modeling techniques. For instance, with that information, models could be integrated by: (1) quantifying a bounded box of predicted occurrence, or the region where at least one algorithm predicts occurrence; or (2) mapping the frequency of predicted occurrence from different model algorithms (Araújo and New 2007) (Fig. 7.8).

7.4 Next Steps and Advanced Issues

7.4.1 Incorporating Dispersal

A common criticism for distribution models is that they typically ignore dispersal-related limitations (Barve et al. 2011). Some approaches simply apply constraints to the mapping process (Cardador et al. 2014), some model colonization processes with time series data (see Chap. 10) (Bled et al. 2013; Yackulic et al. 2015), while others link distribution models with simulations of the dispersal process (Smolik et al. 2010). In general, there is a great need to incorporate movement into the prediction of species distributions and this is an active area of development (Miller and Holloway 2015; Boulangeat et al. 2012). The `MigClim` package provides some functionality for incorporating dispersal constraints into distribution modeling (Engler et al. 2012).

7.4.2 Integrating Multiple Data Sources

Given the limitations of presence-only data and yet the wide availability of such data, it is tempting to integrate presence-only data with other data that suffer fewer biases. Recent modeling advances aim to unite opportunistic presence-only data with presence–absence, occupancy or abundance data to make more reliable predictions by using multiple sources of data simultaneously in model development, termed

integrated species distribution models (Fithian et al. 2015; Koshkina et al. 2017; Pacifici et al. 2017). Such integration can help to minimize bias as well as providing a means to incorporate species prevalence, which is helpful for making predictions of the probability of occurrence. These modeling efforts have been shown to reduce potential bias and increase predictive accuracy of models (Dorazio 2014; Fithian et al. 2015; Fletcher et al. 2016).

7.4.3 Dynamic Models

When time series of location data occur, we may model dynamics of distributions. In this approach, often the focus is on understanding local extinction-colonization dynamics (MacKenzie et al. 2003; Yackulic et al. 2015). In this framework, species distribution (e.g., occupancy) over time is a derived parameter from estimated colonization-extinction dynamics. Benefits of modeling dynamics include that it provides a means to better understand the role of different limiting factors on species distribution (e.g., dispersal limitation) (Broms et al. 2016), whether occurrence at locations reflect underlying habitat quality (Pulliam 2000), it can help identify if species distributions tend to be at equilibrium (a prominent assumption when projecting distributions) (Yackulic et al. 2015), and it can allow predictions of range dynamics (Guillera-Arroita 2017). We will address these dynamical models in Chap. 10.

7.4.4 Multi-species Models

There is also increasing interest in modeling the distribution of multiple species simultaneously (Ferrier and Guisan 2006). This can be done in a variety of ways, and it typically focuses on species co-occurrence (Dorazio et al. 2006; Ferrier et al. 2007; Ovaskainen et al. 2010; Araújo et al. 2011). Simultaneously modeling multiple species can be advantageous for several reasons. First, it can provide insight into potential species interactions and how those change over space or time. Second, one species might be a good predictor of another species, not necessarily due to interactions but because it is an indirect indicator of environmental conditions. We will address some of these techniques in Chap. 11.

7.4.5 Sampling Error and Distribution Models

Throughout this chapter, we have ignored the problem of sampling error, such as imperfect detection of species, to focus more simply on the issues variation in model algorithms and model evaluation. However, observation errors are common in data

sets and these errors frequently need to be accounted for to obtain reliable estimates of environmental relationships. Several models exist for accounting for imperfect detection, both in terms of false positive and false negative errors (Miller et al. 2011; Guillera-Arroita et al. 2017). False negative errors are more common, where a species or individual occurs in an area but we fail to detect it. Several investigations suggest that accounting for false negative errors can improve the predictive performance of distribution models (Rota et al. 2011; Lahoz-Monfort et al. 2014; Guillera-Arroita 2017). One major challenge in the interpretation, however, is that these models predict occupancy across a geographic region, and yet evaluation data are often detection-non detection (typically true occupancy data are not available for evaluating models). Distribution models that account for imperfect detection can be fit with a variety of R packages, including the `unmarked` (Fiske and Chandler 2011), `hSDM` (Vielledent et al. 2014), and `stocc` (Johnson 2015) packages.

7.5 Conclusions

Understanding, predicting, and projecting species distributions provides a means to answer major questions in ecology and can deliver decision support for many conservation problems (Gill et al. 2001; Norris 2004; Wiens et al. 2010; Guisan et al. 2013). The use of species distribution models in ecology, evolution, and conservation has a long tradition (Rotenberry and Wiens 1980; Austin 1987; Donovan et al. 1987), and yet it has exploded over the past 15 years with new advances in modeling algorithms and newly available data sources regarding species locations and geo-spatial data of environmental factors (Graham et al. 2004; Dickinson et al. 2010; Fick and Hijmans 2017).

Many of the species distribution modeling techniques currently being used can be described as inhomogeneous point process models. This realization has several consequences for the implementation and interpretation of species distribution models (Renner and Warton 2013; Renner et al. 2015). We recommend that this framework be generally used to guide correlative species distribution modeling.

Our example illustrates that reliably applying and evaluating species distribution models can be challenging. Extrapolating predictions beyond the environmental conditions used for model building, an issue that commonly occurs when projecting the effects of climate change (Thomas et al. 2004), can be difficult because little information exists on such relationships. Evaluating models with commonly used techniques, such as *K*-fold validation, can sometimes provide a false sense of model performance (Wenger and Olden 2012) and suggest that more complex models are valuable when in fact simpler models may be sufficient for reliable predictions in space and time (cf. Tables 7.4 and 7.5).

Despite this increased use distribution models, these models still have limitations and there use and application should be done with care. Greater focus on mechanistic

modeling and leveraging information on why species distribution varies over space and time may further advance our understanding of species distribution and our ability to predict changes in distribution with ongoing environmental change.

References

Aarts G, Fieberg J, Matthiopoulos J (2012) Comparative interpretation of count, presence-absence and point methods for species distribution models. Methods Ecol Evol 3(1):177–187. https://doi.org/10.1111/j.2041-210X.2011.00141.x

Alouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J Appl Ecol 43(6):1223–1232

Araujo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. J Biogeogr 33(10):1677–1688

Araújo MB, New M (2007) Ensemble forecasting of species distributions. Trends Ecol Evol 22 (1):42–47. https://doi.org/10.1016/j.tree.2006.09.010

Araújo MB, Peterson AT (2012) Uses and misuses of bioclimatic envelope modeling. Ecology 93 (7):1527–1539

Araújo MB, Rozenfeld A, Rahbek C, Marquet PA (2011) Using species co-occurrence networks to assess the impacts of climate change. Ecography 34(6):897–908. https://doi.org/10.1111/j.1600-0587.2011.06919.x

Austin MP (1987) Models for the analysis of species response to environmental gradients. Vegetatio 69(1–3):35–45. https://doi.org/10.1007/bf00038685

Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol Model 157(2–3):101–118

Austin M (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol Model 200(1–2):1–19

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol 3(2):327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Barton K (2018) MuMIn: multi-model inference. R package version 1.40.4

Barve N, Barve V, Jimenez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberon J, Villalobos F (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol Model 222(11):1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Berman M, Turner TR (1992) Approximating point process likelihoods with GLIM. J R Stat Soc C Appl Stat 41(1):31–38

Betts MG, Phalan B, Frey SJK, Rousseau JS, Yang ZQ (2018) Old-growth forests buffer climate-sensitive bird populations from warming. Divers Distrib 24(4):439–447. https://doi.org/10.1111/ddi.12688

Bled F, Nichols JD, Altwegg R (2013) Dynamic occupancy models for analyzing species' range dynamics across large geographic scales. Ecol Evol 3(15):4896–4909. https://doi.org/10.1002/ece3.858

Blonder B, Lamanna C, Violle C, Enquist BJ (2014) The n-dimensional hypervolume. Glob Ecol Biogeogr 23(5):595–609. https://doi.org/10.1111/geb.12146

Boulangeat I, Gravel D, Thuiller W (2012) Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. Ecol Lett 15(6):584–593. https://doi.org/10.1111/j.1461-0248.2012.01772.x

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecol Model 157(2–3):281–300. https://doi.org/10.1016/s0304-3800(02)00200-4

Brand LA, George TL (2001) Response of passerine birds to forest edge in coast redwood forest fragments. Auk 118(3):678–686. https://doi.org/10.1642/0004-8038(2001)118[0678:Ropbtf]2.0.Co;2

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/a:1010933404324

Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. Chapman and Hall/CRC, Boca Raton, FL

Brewer CK, Berglund D, Barber JA, Bush R (2004) Northern region vegetation mapping project summary report and spatial datasets, version 42. Northern Region USFS

Broennimann O, Treier UA, Muller-Scharer H, Thuiller W, Peterson AT, Guisan A (2007) Evidence of climatic niche shift during biological invasion. Ecol Lett 10(8):701–709. https://doi.org/10.1111/j.1461-0248.2007.01060.x

Broennimann O, Di Cola V, Guisan A (2018) ecospat: spatial ecology miscellaneous methods. R package verision 3.0

Broms KM, Hooten MB, Johnson DS, Altwegg R, Conquest LL (2016) Dynamic occupancy models for explicit colonization processes. Ecology 97(1):194–204. https://doi.org/10.1890/15-0416.1

Brotons L, Thuiller W, Araujo MB, Hirzel AH (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27(4):437–448. https://doi.org/10.1111/j.0906-7590.2004.03764.x

Buckley LB (2008) Linking traits to energetics and population dynamics to predict lizard ranges in changing environments. Am Nat 171(1):E1–E19. https://doi.org/10.1086/523949

Buckley LB, Kingsolver JG (2012) Functional and phylogenetic approaches to forecasting species' responses to climate change. Annu Rev Ecol Evol Syst 43:205–226. https://doi.org/10.1146/annurev-ecolsys-110411-160516

Buckley LB, Urban MC, Angilletta MJ, Crozier LG, Rissler LJ, Sears MW (2010) Can mechanism inform species' distribution models? Ecol Lett 13(8):1041–1054. https://doi.org/10.1111/j.1461-0248.2010.01479.x

Burnham KP, Anderson DR (1998) Model selection and inference: a practical information-theoretic approach. Springer, New York

Busby JR (1991) BIOCLIM: a bioclimate analysis and prediction system. In: Margules CR, Austin MP (eds) Nature conservation: cost effective biological surveys and data analysis. CSIRO, Canberra, Australia, pp 64–68

Cardador L, Sarda-Palomera F, Carrete M, Manosa S (2014) Incorporating spatial constraints in different periods of the annual cycle improves species distribution model performance for a highly mobile bird species. Divers Distrib 20(5):515–528. https://doi.org/10.1111/ddi.12156

Carpenter G, Gillison AN, Winter J (1993) DOMAIN—a flexible modeling procedure for mapping potential distributions of plants and animals. Biodivers Conserv 2(6):667–680. https://doi.org/10.1007/bf00051966

Chase JM, Leibold MA (2003) Ecological niches: linking classical and contemporary approaches. University of Chicago Press

Colwell RK, Rangel TF (2009) Hutchinson's duality: the once and future niche. Proc Natl Acad Sci U S A 106:19651–19658. https://doi.org/10.1073/pnas.0901650106

Cushman SA, McGarigal K (2004) Patterns in the species-environment relationship depend on both scale and choice of response variables. Oikos 105(1):117–124

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT (2007) Random forests for classification in ecology. Ecology 88(11):2783–2792. https://doi.org/10.1890/07-0539.1

De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81(11):3178–3192. https://doi.org/10.1890/0012-9658(2000)081[3178:Cartap]2.0.Co;2

Di Cola V, Broennimann O, Petitpierre B, Breiner FT, D'Amen M, Randin C, Engler R, Pottier J, Pio D, Dubuis A, Pellissier L, Mateo RG, Hordijk W, Salamin N, Guisan A (2017) ecospat: an R package to support spatial analyses and modeling of species niches and distributions. Ecography 40(6):774–787. https://doi.org/10.1111/ecog.02671

Dickinson JL, Zuckerberg B, Bonter DN (2010) Citizen science as an ecological research tool: challenges and benefits. Annu Rev Ecol Evol Syst 41:149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636

Donovan ML, Rabe DL, Olson CE (1987) Use of geographic information-systems to develop habitat suitability models. Wildl Soc Bull 15(4):574–579

Dorazio RM (2014) Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. Glob Ecol Biogeogr 23(12):1472–1484. https://doi.org/10.1111/geb.12216

Dorazio RM, Royle JA, Soderstrom B, Glimskar A (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. Ecology 87(4):842–854. https://doi.org/10.1890/0012-9658(2006)87[842:esraab]2.0.co;2

Dormann CF, Schymanski SJ, Cabral J, Chuine I, Graham C, Hartig F, Kearney M, Morin X, Roemermann C, Schroeder B, Singer A (2012) Correlation and process in species distribution models: bridging a dichotomy. J Biogeogr 39(12):2119–2131. https://doi.org/10.1111/j.1365-2699.2011.02659.x

Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. Annu Rev Ecol Evol Syst 40:677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Elith J, Ferrier S, Huettmann F, Leathwick J (2005) The evaluation strip: a new and robust method for plotting predicted responses from species distribution models. Ecol Model 186(3):280–289. https://doi.org/10.1016/j.ecolmodel.2004.12.007

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29 (2):129–151

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77 (4):802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. Methods Ecol Evol 1(4):330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.x

Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. Divers Distrib 17(1):43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Elton C (1927) Animal ecology. Sedgwick and Jackson, London

Engler R, Guisan A, Rechsteiner L (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J Appl Ecol 41 (2):263–274

Engler R, Hordijk W, Guisan A (2012) The MIGCLIM R package—seamless integration of dispersal constraints into projections of species distribution models. Ecography 35(10):872–878. https://doi.org/10.1111/j.1600-0587.2012.07608.x

Eskildsen A, le Roux PC, Heikkinen RK, Hoye TT, Kissling WD, Poyry J, Wisz MS, Luoto M (2013) Testing species distribution models across space and time: high latitude butterflies and recent warming. Glob Ecol Biogeogr 22(12):1293–1303. https://doi.org/10.1111/geb.12078

Evans JM, Fletcher RJ Jr, Alavalapati J (2010) Using species distribution models to identify suitable areas for biofuel feedstock production. Glob Change Biol Bioenergy 2(2):63–78. https://doi.org/10.1111/j.1757-1707.2010.01040.x

Feeley KJ, Silman MR (2010) Land-use and climate change effects on population size and extinction risk of Andean plants. Glob Chang Biol 16(12):3215–3222. https://doi.org/10.1111/j.1365-2486.2010.02197.x

Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. J Appl Ecol 43 (3):393–404. https://doi.org/10.1111/j.1365-2664.2006.01149.x

Ferrier S, Manion G, Elith J, Richardson K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Divers Distrib 13(3):252–264. https://doi.org/10.1111/j.1472-4642.2007.00341.x

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol 37(12):4302–4315. https://doi.org/10.1002/joc.5086
Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24(1):38–49
Fiske IJ, Chandler RB (2011) Unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. J Stat Softw 43(10):1–23
Fithian W, Hastie T (2013) Finite-sample equivalence in statistical models for presence-only data. Ann Appl Stat 7(4):1917–1939. https://doi.org/10.1214/13-aoas667
Fithian W, Elith J, Hastie T, Keith DA (2015) Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods Ecol Evol 6(4):424–438. https://doi.org/10.1111/2041-210x.12242
Fletcher RJ Jr, Hutto RL (2008) Partitioning the multi-scale effects of human activity on the occurrence of riparian forest birds. Landsc Ecol 23:727–739
Fletcher RJ, McCleery RA, Greene DU, Tye CA (2016) Integrated models that unite local and regional data reveal larger-scale environmental relationships and improve predictions of species distributions. Landsc Ecol 31(6):1369–1382. https://doi.org/10.1007/s10980-015-0327-9
Franklin J (2009) Mapping species distributions: spatial inference and prediction. Cambridge University Press, Cambridge, UK
Freeman EA, Moisen G (2008) PresenceAbsence: an R package for presence absence analysis. J Stat Softw 23(11):1–31
Fretwell SD, Lucas HL Jr (1970) On territorial behavior and other factors influencing habitat distribution in birds. I. Theoretical development. Acta Biotheor 19:16–36
Friedman JH (2002) Stochastic gradient boosting. Comput Stat Data Anal 38(4):367–378. https://doi.org/10.1016/s0167-9473(01)00065-2
Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33(1):1–22
Gaston A, Garcia-Vinas JI (2011) Modelling species distributions with penalised logistic regressions: a comparison with maximum entropy models. Ecol Model 222(13):2037–2041. https://doi.org/10.1016/j.ecolmodel.2011.04.015
George TS (2000) Varied thrush (Ixoreus naevius). In: Poole A (ed) The birds of North America Online. Cornell University, Ithaca, NY
Gill JA, Norris K, Potts PM, Gunnarsson TG, Atkinson PW, Sutherland WJ (2001) The buffer effect and large-scale population regulation in migratory birds. Nature 412(6845):436–438
Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT (2004) New developments in museum-based informatics and applications in biodiversity analysis. Trends Ecol Evol 19(9):497–503. https://doi.org/10.1016/j.tree.2004.07.006
Grinnell J (1917) The niche-relationships of the California Thrasher. Auk 34:427–433
Guillera-Arroita G (2017) Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. Ecography 40(2). https://doi.org/10.1111/ecog.02445
Guillera-Arroita G, Lahoz-Monfort JJ, Elith J (2014) Maxent is not a presence-absence method: a comment on Thibaud et al. Methods Ecol Evol 5(11):1192–1197. https://doi.org/10.1111/2041-210x.12252
Guillera-Arroita G, Lahoz-Monfort JJ, van Rooyen AR, Weeks AR, Tingley R (2017) Dealing with false-positive and false-negative errors about species occurrence at multiple levels. Methods Ecol Evol 8(9):1081–1091. https://doi.org/10.1111/2041-210x.12743
Guisan A, Harrell FE (2000) Ordinal response regression models in ecology. J Veg Sci 11 (5):617–626
Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8(9):993–1009
Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135(2–3):147–186

Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol Model 157(2–3):89–100

Guisan A, Zimmermann NE, Elith J, Graham CH, Phillips S, Peterson AT (2007) What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? Ecol Monogr 77(4):615–630

Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, Martin TG, Rhodes JR, Maggini R, Setterfield SA, Elith J, Schwartz MW, Wintle BA, Broennimann O, Austin M, Ferrier S, Kearney MR, Possingham HP, Buckley YM (2013) Predicting species distributions for conservation decisions. Ecol Lett 16(12):1424–1435. https://doi.org/10.1111/ele.12189

Guisan A, Thuiller W, Zimmermann NE (2017) Habitat suitability and distribution models: applications with R. Cambridge University Press, Cambridge, UK

Hanski K, Ovaskainen O (2003) Metapopulation theory for fragmented landscapes. Theor Popul Biol 64(1):119–127. https://doi.org/10.1016/s0040-5809(03)00022-4

Hastie T, Fithian W (2013) Inference from presence-only data; the ongoing controversy. Ecography 36(8):864–867. https://doi.org/10.1111/j.1600-0587.2013.00321.x

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, New York

Hefley TJ, Broms KM, Brost BM, Buderman FE, Kay SL, Scharf HR, Tipton JR, Williams PJ, Hooten MB (2017) The basis function approach for modeling autocorrelation in ecological data. Ecology 98(3):632–646. https://doi.org/10.1002/ecy.1674

Hijmans RJ (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93(3):679–688

Hijmans RJ, Phillips S, Leathwick J, Elith J (2017) dismo: species distribution modeling. R package version 1.1.-4

Hirzel AH, Le Lay G (2008) Habitat suitability modelling and niche theory. J Appl Ecol 45 (5):1372–1381. https://doi.org/10.1111/j.1365-2664.2008.01524.x

Hirzel AH, Posse B, Oggier PA, Crettenand Y, Glenz C, Arlettaz R (2004) Ecological requirements of reintroduced species and the implications for release policy: the case of the bearded vulture. J Appl Ecol 41(6):1103–1116. https://doi.org/10.1111/j.0021-8901.2004.00980.x

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecol Model 199(2):142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

Holt RD (2009) Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. Proc Natl Acad Sci U S A 106:19659–19665. https://doi.org/10.1073/pnas.0905137106

Hutchinson GE (1957) Concluding remarks. Population studies: animal ecology and demography. Cold Spring Harb Symp Quant Biol 22:415–427

Hutto RL, Young JS (2002) Regional landbird monitoring: perspectives from the Northern Rocky Mountains. Wildl Soc Bull 30(3):738–750

James FC, Johnston RF, Wamer NO, Niemi GJ, Boecklen WJ (1984) The grinnellian niche of the wood thrush. Am Nat 124(1):17–30. https://doi.org/10.1086/284250

Jimenez-Valverde A, Peterson AT, Soberon J, Overton JM, Aragon P, Lobo JM (2011) Use of niche models in invasive species risk assessments. Biol Invasions 13(12):2785–2797. https://doi.org/10.1007/s10530-011-9963-4

Johnson DS (2015) stocc: fit a spatial occupancy model via Gibbs sampling. R package version 1.30

Kadmon R, Farber O, Danin A (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecol Appl 14(2):401–413

Kearney M, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecol Lett 12(4):334–350. https://doi.org/10.1111/j.1461-0248.2008.01277.x

Koshkina V, Wang Y, Gordon A, Dorazio RM, White M, Stone L (2017) Integrated species distribution models: combining presence-background data and site-occupany data with imperfect detection. Methods Ecol Evol 8(4):420–430. https://doi.org/10.1111/2041-210x.12738

Lahoz-Monfort JJ, Guillera-Arroita G, Wintle BA (2014) Imperfect detection impacts the performance of species distribution models. Glob Ecol Biogeogr 23(4):504–515. https://doi.org/10.1111/geb.12138

Lawson CR, Hodgson JA, Wilson RJ, Richards SA (2014) Prevalence, thresholds and the performance of presence-absence models. Methods Ecol Evol 5(1):54–64. https://doi.org/10.1111/2041-210x.12123

Liaw A, Wiener M (2002) Classification and regression by randomforest. R News 2(3):18–22

Lira-Noriega A, Soberon J, Miller CP (2013) Process-based and correlative modeling of desert mistletoe distribution: a multiscalar approach. Ecosphere 4(8):99. https://doi.org/10.1890/es13-00155.1

Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28(3):385–393

Liu C, White M, Newell G (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. J Biogeogr 40(4):778–789. https://doi.org/10.1111/jbi.12058

Lobo JM, Jimenez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 17(2):145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x

Loiselle BA, Jorgensen PM, Consiglio T, Jimenez I, Blake JG, Lohmann LG, Montiel OM (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? J Biogeogr 35(1):105–116. https://doi.org/10.1111/j.1365-2699.2007.01779.x

Lutolf M, Kienast F, Guisan A (2006) The ghost of past species occurrence: improving species distribution models for presence-only data. J Appl Ecol 43(4):802–815. https://doi.org/10.1111/j.1365-2664.2006.01191.x

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. Ecology 83(8):2248–2255

MacKenzie DI, Nichols JD, Hines JE, Knutson MG, Franklin AB (2003) Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84(8):2200–2207

MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE (2006) Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier, Amsterdam

Manly BFJ, McDonald LL, Thomas DL, McDonald TL, Erickson WP (2002) Resource selection by animals: statistical design and analysis for field studies. Kluwer Academic Publishers, Dordrecht, the Netherlands

Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W (2009) Evaluation of consensus methods in predictive species distribution modelling. Divers Distrib 15(1):59–69. https://doi.org/10.1111/j.1472-4642.2008.00491.x

Martin Y, Van Dyck H, Dendoncker N, Titeux N (2013) Testing instead of assuming the importance of land use change scenarios to model species distributions under climate change. Glob Ecol Biogeogr 22(11):1204–1216. https://doi.org/10.1111/geb.12087

Martinez-Meyer E, Peterson AT, Servin JI, Kiff LF (2006) Ecological niche modelling and prioritizing areas for species reintroductions. Oryx 40(4):411–418. https://doi.org/10.1017/s0030605306001360

McCarthy KP, Fletcher RJ, Rota CT, Hutto RL (2012) Predicting species distributions from samples collected along roadsides. Conserv Biol 26(1):68–77. https://doi.org/10.1111/j.1523-1739.2011.01754.x

Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36(10):1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Miller JA, Holloway P (2015) Incorporating movement in species distribution models. Prog Phys Geogr 39(6):837–849. https://doi.org/10.1177/0309133315580890
Miller DA, Nichols JD, McClintock BT, Grant EHC, Bailey LL, Weir LA (2011) Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. Ecology 92(7):1422–1428. https://doi.org/10.1890/10-1396.1
Naimi B, Araújo MB (2016) sdm: a reproducible and extensible R platform for species distribution modelling. Ecography 39(4):368–375. https://doi.org/10.1111/ecog.01881
Norris K (2004) Managing threatened species: the ecological toolbox, evolutionary theory and declining-population paradigm. J Appl Ecol 41(3):413–426
Olden JD, Lawler JJ, Poff NL (2008) Machine learning methods without tears: a primer for ecologists. Q Rev Biol 83(2):171–193. https://doi.org/10.1086/587826
Ovaskainen O, Hottola J, Siitonen J (2010) Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. Ecology 91(9):2514–2521. https://doi.org/10.1890/10-0173.1
Pacifici K, Reich BJ, Miller DAW, Gardner B, Stauffer G, Singh S, McKerrow A, Collazo JA (2017) Integrating multiple data sources in species distribution modeling: a framework for data fusion. Ecology 98(3):840–850. https://doi.org/10.1002/ecy.1710
Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. Ecol Model 133(3):225–245
Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Glob Ecol Biogeogr 12(5):361–371. https://doi.org/10.1046/j.1466-822X.2003.00042.x
Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. Q Rev Biol 78(4):419–433. https://doi.org/10.1086/378926
Peterson AT, Soberon J (2012) Species distribution modeling and ecological niche modeling: getting the concepts right. Natureza & Conservacao 10(2):102–107. https://doi.org/10.4322/natcon.2012.019
Peterson AT, Papes M, Soberon J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol Model 213(1):63–72. https://doi.org/10.1016/j.ecolmodel.2007.11.008
Peterson AT, Soberon J, Pearson RG, Anderson RP, Martinez-Mery E, Nakamura M, Araújo MB (2011) Ecological niches and geographic distributions. Princeton University Press, Princeton, NJ
Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31(2):161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x
Phillips SJ, Elith J (2010) POC plots: calibrating species distribution models with presence-only data. Ecology 91(8):2476–2484. https://doi.org/10.1890/09-0760.1
Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190(3–4):231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026
Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl 19(1):181–197
Phillips SJ, Anderson RP, Dudik M, Schapire RE, Blair ME (2017) Opening the black box: an open-source release of Maxent. Ecography 40(7):887–893. https://doi.org/10.1111/ecog.03049
Plath M, Moser C, Bailis R, Brandt P, Hirsch H, Klein AM, Walmsley D, von Wehrden H (2016) A novel bioenergy feedstock in Latin America? Cultivation potential of Acrocomia aculeata under current and future climate conditions. Biomass Bioenergy 91:186–195. https://doi.org/10.1016/j.biombioe.2016.04.009
Potts JM, Elith J (2006) Comparing species abundance models. Ecol Model 199(2):153–163. https://doi.org/10.1016/j.ecolmodel.2006.05.025

Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9(2):181–199. https://doi.org/10.1007/s10021-005-0054-1

Pulliam HR (1988) Sources, sinks, and population regulation. Am Nat 132(5):652–661

Pulliam HR (2000) On the relationship between niche and distribution. Ecol Lett 3(4):349–361

Raxworthy CJ, Martinez-Meyer E, Horning N, Nussbaum RA, Schneider GE, Ortega-Huerta MA, Peterson AT (2003) Predicting distributions of known and unknown reptile species in Madagascar. Nature 426(6968):837–841. https://doi.org/10.1038/nature02205

Real LA, Brown JH (eds) (1991) Foundations of ecology: classic papers with commentaries. University of Chicago Press, Chicago

Renner IW, Warton DI (2013) Equivalence of MAXENT and poisson point process models for species distribution modeling in ecology. Biometrics 69(1):274–281. https://doi.org/10.1111/j.1541-0420.2012.01824.x

Renner IW, Elith J, Baddeley A, Fithian W, Hastie T, Phillips SJ, Popovic G, Warton DI (2015) Point process models for presence-only analysis. Methods Ecol Evol 6(4):366–379. https://doi.org/10.1111/2041-210x.12352

Robertson BA, Hutto RL (2006) A framework for understanding ecological traps and an evaluation of existing evidence. Ecology 87(5):1075–1085

Rodrigues ASL, Akcakaya HR, Andelman SJ, Bakarr MI, Boitani L, Brooks TM, Chanson JS, Fishpool LDC, Da Fonseca GAB, Gaston KJ, Hoffmann M, Marquet PA, Pilgrim JD, Pressey RL, Schipper J, Sechrest W, Stuart SN, Underhill LG, Waller RW, Watts MEJ, Yan X (2004) Global gap analysis: priority regions for expanding the global protected-area network. Bioscience 54(12):1092–1100. https://doi.org/10.1641/0006-3568(2004)054[1092:ggaprf]2.0.co;2

Rota CT, Fletcher RJ Jr, Evans JM, Hutto RL (2011) Does accounting for detectability improve species distribution models. Ecography 34:659–670

Rotenberry JT, Wiens JA (1980) Habitat structure, patchiness, and avian communities in North-American steppe vegetation: a multivariate-analysis. Ecology 61(5):1228–1250. https://doi.org/10.2307/1936840

Rotenberry JT, Preston KL, Knick ST (2006) Gis-based niche modeling for mapping species' habitat. Ecology 87(6):1458–1464

Royle JA, Chandler RB, Yackulic C, Nichols JD (2012) Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. Methods Ecol Evol 3(3):545–554. https://doi.org/10.1111/j.2041-210X.2011.00182.x

Schlaepfer MA, Runge MC, Sherman PW (2002) Ecological and evolutionary traps. Trends Ecol Evol 17(10):474–480

Scott JM, Davis F, Csuti B, Noss R, Butterfield B, Groves C, Anderson H, Caicco S, Derchia F, Edwards TC, Ulliman J, Wright RG (1993) GAP analysis: a geographic approach to protection of biological diversity. Wildl Monogr (123):1–41

Smolik MG, Dullinger S, Essl F, Kleinbauer I, Leitner M, Peterseil J, Stadler LM, Vogl G (2010) Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. J Biogeogr 37(3):411–422. https://doi.org/10.1111/j.1365-2699.2009.02227.x

Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. Ecol Lett 10(12):1115–1123. https://doi.org/10.1111/j.1461-0248.2007.01107.x

Soberón JM (2010) Niche and area of distribution modeling: a population ecology perspective. Ecography 33(1):159–167. https://doi.org/10.1111/j.1600-0587.2009.06074.x

Soberón J, Nakamura M (2009) Niches and distributional areas: concepts, methods, and assumptions. Proc Natl Acad Sci U S A 106:19644–19650. https://doi.org/10.1073/pnas.0901637106

Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. Biodivers Inform 2:1–10

Tewksbury JJ, Garner L, Garner S, Lloyd JD, Saab V, Martin TE (2006) Tests of landscape influence: nest predation and brood parasitism in fragmented ecosystems. Ecology 87(3):759–768

Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, de Siqueira MF, Grainger A, Hannah L, Hughes L, Huntley B, van Jaarsveld AS, Midgley GF, Miles L, Ortega-Huerta MA, Peterson AT, Phillips OL, Williams SE (2004) Extinction risk from climate change. Nature 427(6970):145–148. https://doi.org/10.1038/nature02121

Thuiller W, Georges D, Engler R, Breiner F (2016) biomod2: ensemble platform for species distribution modeling. R package version 3.3.-7

Tibshirani R (1996) Regression shrinkage and selection via the Lasso. J R Stat Soc Series B Methodol 58(1):267–288

Urbanek S (2017) rJava: low-level R to Java interface. R package 0.9-9

Vallecillo S, Brotons L, Thuiller W (2009) Dangers of predicting bird species distributions in response to land-cover changes. Ecol Appl 19(2):538–549. https://doi.org/10.1890/08-0348.1

Van Horne B (1983) Density as a misleading indicator of habitat quality. J Wildl Manag 47:893–901

VanDerWal J, Shoo LP, Graham C, William SE (2009) Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecol Model 220 (4):589–594. https://doi.org/10.1016/j.ecolmodel.2008.11.010

Vielledent G, Merow C, Guelat J, Latimer AM, Kery M, Gelfand AE, Wilson AM, F. Mortier, Silander Jr JA (2014) hSDM: hierachical Bayesian species distribution models. R package version 1.4

Ward G, Hastie T, Barry S, Elith J, Leathwick JR (2009) Presence-only data and the EM algorithm. Biometrics 65(2):554–563. https://doi.org/10.1111/j.1541-0420.2008.01116.x

Warton DI, Shepherd LC (2010) Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. Ann Appl Stat 4(3):1383–1402. https://doi.org/10.1214/10-aoas331

Wenger SJ, Olden JD (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods Ecol Evol 3(2):260–267. https://doi.org/10.1111/j.2041-210X.2011.00170.x

Whittaker RH, Levin SA, Root RB (1973) Niche, habitat, and ecotope. Am Nat 107(955):321–338. https://doi.org/10.1086/282837

Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York

Wiens JJ, Ackerly DD, Allen AP, Anacker BL, Buckley LB, Cornell HV, Damschen EI, Davies TJ, Grytnes JA, Harrison SP, Hawkins BA, Holt RD, McCain CM, Stephens PR (2010) Niche conservatism as an emerging principle in ecology and conservation biology. Ecol Lett 13 (10):1310–1324. https://doi.org/10.1111/j.1461-0248.2010.01515.x

Williams JW, Jackson ST (2007) Novel climates, no-analog communities, and ecological surprises. Front Ecol Environ 5(9):475–482. https://doi.org/10.1890/070037

Wilson KA, Westphal MI, Possingham HP, Elith J (2005) Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biol Conserv 122(1):99–112. https://doi.org/10.1016/j.biocon.2004.07.004

Wisz MS, Pottier J, Kissling WD, Pellissier L, Lenoir J, Damgaard CF, Dormann CF, Forchhammer MC, Grytnes JA, Guisan A, Heikkinen RK, Hoye TT, Kuhn I, Luoto M, Maiorano L, Nilsson MC, Normand S, Ockinger E, Schmidt NM, Termansen M, Timmermann A, Wardle DA, Aastrup P, Svenning JC (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biol Rev 88 (1):15–30. https://doi.org/10.1111/j.1469-185X.2012.00235.x

Wood SN (2006) Generalized additive models: an introduction with R. Chapman and Hall and CRC, Boca Raton, FL

Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Grant EHC, Veran S (2013) Presence-only modelling using MAXENT: when can we trust the inferences? Methods Ecol Evol 4 (3):236–243. https://doi.org/10.1111/2041-210x.12004

Yackulic CB, Nichols JD, Reid J, Der R (2015) To predict the niche, model colonization and extinction. Ecology 96(1):16–23. https://doi.org/10.1890/14-1361.1

Chapter 8
Space Use and Resource Selection



8.1 Introduction

Understanding habitat and resource selection is at the foundation of much research and applications in wildlife ecology (Morrison et al. 2006), and it has been emphasized in ecological theory (Rosenzweig 1981; Pulliam and Danielson 1991; Morris 2003). It is also highly relevant for spatial ecology and conservation (Battin 2004; Resetarits 2005; Fagan and Lutscher 2006), because habitat selection is frequently viewed to operate at different spatial and temporal scales (Johnson 1980; Orians and Wittenberger 1991; Rettie and Messier 2000). Identifying critical habitat is essential for conservation planning and recovery criteria for imperiled species (Thompson and McGarigal 2002; Turner et al. 2004; Taylor et al. 2005).

Habitat selection, resource selection, and space use are related concepts for spatial ecology and conservation. Here, we first distinguish these concepts, briefly describe relevant theory for space use and habitat selection, and finally provide an overview of common ways to quantify space use and resource selection with radiotelemetry data. We focus on radiotelemetry data in this chapter because such data provide a means to illustrate all of these concepts and problems and radiotelemetry is a common technique for addressing applied problems. However, resource and habitat selection can also be understood through the use of species distribution data that are not based on trajectories derived from individuals (e.g., radiotelemetry). For those approaches, see Chap. 7. Throughout, we emphasize how these concepts play out over space and the role of spatial scale in understanding habitat and resource selection.

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_8

8.2 Key Concepts and Approaches

8.2.1 Distinguishing Among the Diversity of Habitat-Related Concepts and Terms

There have been a wide variety of concepts and terms to describe and understand habitat and *resource selection* over the years and often these terms are used loosely. The term *habitat* has two distinct uses. First, it can be used in a species-specific way, wherein habitat refers to a collection of environmental features (e.g., shelter) and resources (e.g., food) necessary for the occupancy of an organism. This definition shares many similarities with the concept of the niche (Chap. 7). Second, habitat is sometimes referred to as a specific set of environmental features that may represent a type of vegetation, plant community, or cover type etc. This latter term suggests that habitat is independent of a specific organism or species. Hall et al. (1997) argued that only the first definition of habitat is correct, yet there has been debate on whether that narrow definition of habitat should only be used and how well that definition cross-walks with related concepts of *habitat use*, *selection*, and *preference* (Table 8.1; Lele

Table 8.1 Common terms and definitions used in resource selection and space use investigations

Term	Definition
Ecological trap	When poor-quality habitats are preferred over better alternatives. Can lead to an "attractive sink."
Habitat	A collection of environmental features and resources necessary for the occupancy of an organism, or a specific set of environmental features that may represent a type of vegetation, plant community, or cover type.
Habitat use/usage	The quantity of a component (food/resource) utilized by a consumer over a fixed period of time.
Habitat availability	The accessibility and procurability of a component to a consumer.
Habitat selection	The process in which an animal chooses the component.
Habitat preference	A particular habitat component being chosen more often if offered on an equal basis with others.
Habitat quality	The extent to which a habitat to promotes positive population growth, through high survival and/or reproduction.
Resource	A feature of the environment required for growth, maintenance, and/or reproduction. Resources may (or may not) be consumed.
Resource selection	The disproportionate use of resources relative to their availability.
Resource selection probability function	A function that estimates the probability of use for resource units of different types.
Resource selection function	A resource selection probability function multiplied by an arbitrary constant, such that it is proportional to resource selection probability functions.
Home range	Area traversed by the individuals in its normal activities of food gathering, mating, and caring for young.
Utilization distribution	The distribution of an animal's position in the plane.
Territory	Area defended by individuals.

et al. 2013). We do not debate these terms here, but rather emphasize that it is necessary to be clear in the use of habitat when considering these topics.

The terms "habitat" and "resources" are sometimes used interchangeably. Yet *resource* is a more general term than habitat, which captures features of the environment that individuals need for growth, maintenance, or reproduction, such that it could include habitat, food, and mates. In contrast, "habitat" does not typically include mates and it may or may not include food (e.g., prey are often not considered habitat for predators; Crowder and Cooper 1982; Underwood et al. 2004; Keim et al. 2011).

8.2.2 Habitat Selection Theory

Habitat selection and resource selection theory were largely developed in the 1960s–1980s, where it primarily advanced from behavioral ecology. Prior to this time, there were many developments in understanding the proximate and ultimate causes of habitat selection (Lack 1933; Svärdson 1949), yet a general theory for habitat selection had not emerged. In the 1960s and 1970s, there was a great emphasis on optimality and borrowing ideas from economics to approach questions and concepts in behavioral ecology. Indeed, some of the key insights from early theory on habitat selection were directly analogous to related developments in behavioral ecology that also focused on optimality, such as the polygyny threshold model (Orians 1969). Models derived from habitat selection theory are often consistent with distribution patterns for some species (Krivan et al. 2008; Hache et al. 2013), even though assumptions are often violated (Kennedy and Gray 1997; Hugie and Grand 1998).

8.2.2.1 Ideal Free Distribution and Its Extensions

A fundamental development in habitat selection theory occurred when Fretwell and Lucas (1970) described the Ideal Free Distribution (IFD). This theory aims at understanding variation in habitat selection and the resulting fitness consequences in situations where habitats vary in quality and population density. In this theory, individuals are assumed to behave ideally, be free to select those habitats where fitness (e.g., reproduction) is highest, and fitness is assumed to decline with increasing population density (Fig. 8.1; Fretwell and Lucas 1970). These assumptions are key components of the IFD model of habitat selection and its subsequent extensions (Fretwell and Lucas 1970; Morris 2003). Extensions included relaxing these primary assumptions to include non-ideal behaviors (Shochat et al. 2005), despotic and preemptive behaviors preventing "free" settlement (Fretwell and Lucas 1970; Pulliam 1988; Rodenhouse et al. 1997), and to account for Allee effects (Fretwell and Lucas 1970), or positive density dependence as low population sizes (Stephens and Sutherland 1999). This general theory has also been extended to include species interactions and predation risk (Rosenzweig 1981; Moody et al. 1996), perceptual

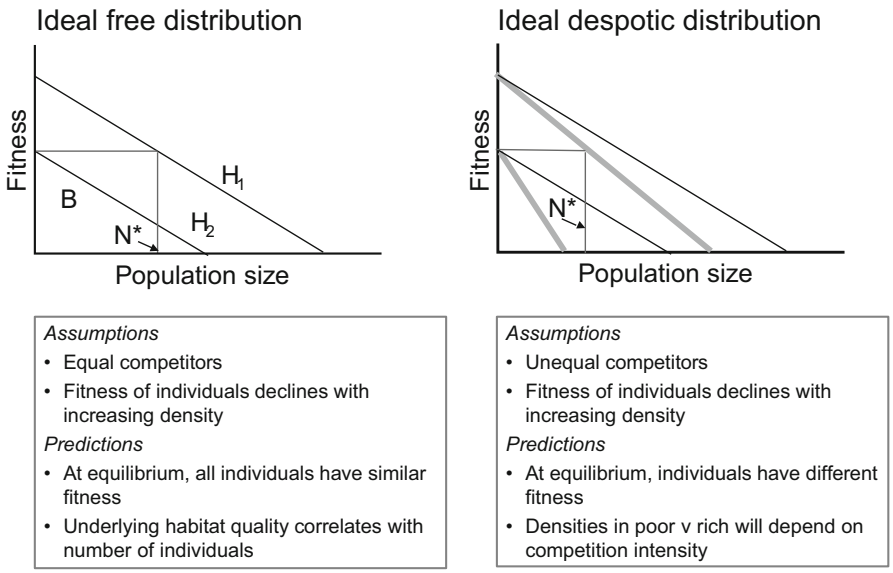


Fig. 8.1 (**a**) Ideal free and (**b**) ideal despotic distributions for two habitats that vary in quality (H_1 and H_2). In (**a**), fitness declines with increasing population size and habitat 2 (H_2) is only used once population size is greater than N^*. In (**b**), dominant competitors occur, shifting the fitness–population size relationship and reducing N^* (shown in bold)

constraints (Abrahams 1986), social behavior (Beauchamp et al. 1997; Nocera et al. 2009), and stochasticity (Morris 2003).

There are several general principles that have arisen from this theoretical development of relevance to interpreting space use and resource selection across landscapes. Theoretical development suggests that high-quality habitats should be preferred, used more consistently over space and time, and may be selected more rapidly and by dominant individuals (e.g., older or larger individuals) (Robertson and Hutto 2006; McLoughlin et al. 2010). Even when individuals are not ideal, if individuals can gain reliable information via sampling the environment, these expectations should generally hold (Pulliam and Danielson 1991). Habitat selection theory also suggests that spatial patterns in distribution may arise, such as spatial autocorrelation in distributional patterns arising from spillover into nearby habitats (McLoughlin et al. 2010). Some developments, such as the idea of site-dependent population regulation in habitat selection (Rodenhouse et al. 1997), provide mechanisms for density-dependence and changes in population growth.

More recently, theoretical developments have focused on understanding how environmental change may impact habitat selection. This work emphasizes that density or abundance can be a poor indicator of *habitat quality* (Van Horne 1983). Such situations frequently capture the idea of an *ecological trap*: where poor-quality habitat is preferred over better alternatives (Schlaepfer et al. 2002), leading to non-ideal habitat selection (Arlt and Part 2007). The concept of ecological traps

has received a great deal of interest, in part because of their potential for detrimental effects on populations (Donovan and Thompson 2001; Kokko and Sutherland 2001; Fletcher et al. 2012; Hale et al. 2015). Though in practice, ecological traps might be relatively rare (Sergio and Newton 2003; Bock and Jones 2004; Robertson and Hutto 2006).

8.2.2.2 Home Range Concepts and Space Use

The concept of the home range dates back to the early twentieth century. Burt (1943) is often credited with first formalizing the concept of *home range* (Fig. 8.2), defining it as the, "area traversed by the individuals in its normal activities of food gathering, mating, and caring for young." He contrasted the home range from the *territory*, or an area defended by individual from other conspecific or heterospecific individuals (Nice 1941). He emphasized that a home range did not need to cover the same area during the entire life of an individual; that is, home ranges could change over time and could be spatially separated, such as breeding and non-breeding home ranges. Typically, home ranges will be larger in size than territories and can vary in size based on both environmental conditions (e.g., resource abundance), individual traits (e.g., age, sex), and population-level issues (e.g., population density).

Most theoretical developments regarding home ranges have focused on understanding why and how mechanistically home ranges may emerge (Borger et al. 2008; Moorcroft and Barnett 2008; Nabe-Nielsen et al. 2013; Potts and Lewis 2014). For central-place foragers (Orians and Pearson 1979), such as nesting birds, it is clear that reproduction places constraints on movement, which can lead to home range behavior. In the absence of such constraints, it can be unclear as to why stabile home ranges emerge, given that simple models of movement, such as diffusion and

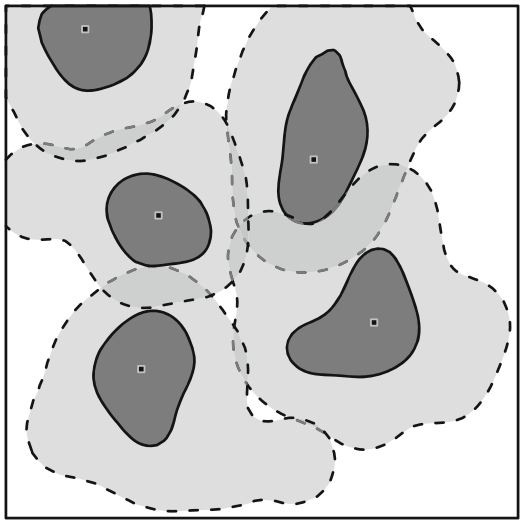


Fig. 8.2 Home range concepts, adapted from Burt (1943). Shown is an example of the relationships between home ranges (ligth grey; dashed lines) and territories (dark grey; solid lines) with key sites (e.g., nesting sites) also overlaid (squares)

random walks (see below) cannot generate stable home ranges (Borger et al. 2008). Several factors can potentially generate stable home ranges, however, such as spatial memory, multiscale resource selection, benefits of motor learning in familiar environments, trade-offs in travel time and resource acquisition, and/or focal point attraction (Stamps 1995; Mitchell and Powell 2004; Gautestad and Mysterud 2005; Gautestad 2011; Merkle et al. 2014; Riotte-Lambert et al. 2015).

Understanding variation in space use within and between home ranges lies at the heart of many problems involved in understanding critical habitats at different scales (Johnson 1980). Interpreting the intensity of space use within home ranges often involves quantifying the utilization distribution of an individual over space (and possibly time). The *utilization distribution* is the two-dimensional relative frequency distribution of an animal over a given period of time (Van Winkle 1975). The utilization distribution can be thought of as a probabilistic representation of use in a home range and is related to some applications of the ecological neighborhood discussed in Chap. 2. Consequently, it provides a spatially explicit representation of the intensity of space use, which can be linked to critical environmental features that explain the intensity of use (Marzluff et al. 2004; Hooten et al. 2013).

8.2.2.3 Movement Concepts and Theory

Resource selection and space use arise from movement of individuals. It is also critical for connectivity (see Chap. 9). In recent years, there has been a convergence of movement-related principles and that of resource selection (Mueller and Fagan 2008; Schick et al. 2008; Zheng et al. 2009; Morales et al. 2010; Zeller et al. 2016; Hooten et al. 2017). Consequently, we briefly discuss relevant ideas and advancements of movement ecology here.

Movement has long been envisioned as a random walk or diffusive process in ecology. Skellam (1951) pioneered this work by incorporating diffusive movement into population dynamics models. Diffusion assumes that movement is random and it provides a means to quantify the probability that a particle moving randomly across the environment will be at a specific location at a given point in time. The expected probability value can be derived from an underlying continuous space, continuous time process (Kareiva 1982). Diffusion models can be extended to account for advection: where certain environments increase movement rates (Skalski and Gilliam 2003; Reeve et al. 2008). In contrast, random walks are typically derived as discrete space, discrete time processes (Fig. 8.3). The general idea is the same: individuals move randomly, but the formulation is different because of the discrete nature of the model. Random walks have been extended in a variety of ways (Codling et al. 2008), the most common of which is the use of correlated random walks (CRWs), wherein there is directionality to movement (Kareiva and Shigesada 1983). In this formulation, it is assumed that the direction traveled at time t will depend on the direction traveled at $t-1$. In general, this extension results in movements that are much more directional and superficially similar to observed movements in nature (Fig. 8.3). Random walks have also been extended to account

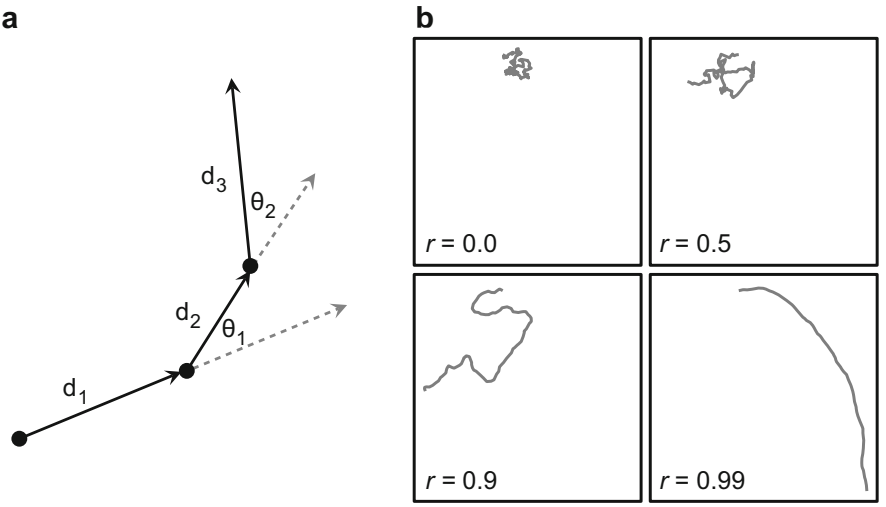


Fig. 8.3 (**a**) Movement trajectories of individuals as described by step lengths, d, (**b**) and relative angles (i.e., turning angles, θ). (**b**) Random walk movement patterns, where both simple ($r = 0$) and correlated random walks ($r > 0$) are illustrated, with the degree of correlation described by r

for memory, social interactions, and other issues (Codling et al. 2008; Gautestad and Mysterud 2010; Smouse et al. 2010; Delgado et al. 2014).

Often in movement ecology a distinction is made regarding whether the focus is on individuals (Lagrangian perspectives) or on the population (Eulerian perspectives) (Nathan et al. 2008). In general, diffusion models are often applied at the population level while random walk models are typically applied at the individual level (Borger et al. 2008). With increasing availability of GPS telemetry data (Tomkiewicz et al. 2010), a Lagrangian perspective is often used to interpret space use and resource selection (Horne et al. 2007; Avgar et al. 2016).

Nathan et al. (2008) outlined a framework aimed at unifying disparate approaches to understanding movement ecology. They emphasized that four general factors drive movement: the intrinsic state of an individual, the motion capacity, the navigation capacity, and the external environment. Intrinsic state captures how variation among individuals, such as their physiological state and personality traits (Zera and Denno 1997; Harrison et al. 2015), may impact the likelihood and direction of movement. Motion capacity focuses on the biomechanics of movement (Damschen et al. 2008; Turlure et al. 2016). Navigation capacity focuses on how and the scale at which organisms acquire information to influence movement decisions, such as individual perceptual constraints and the perceptual range (Zollner and Lima 1997; Fletcher et al. 2013). Finally, the external environment provides a broad category of extrinsic issues that may impact movement, such as land use (Doherty and Driscoll 2018). Feedbacks can occur with each of these four factors. For example, the external environment can impact the intrinsic state of individuals and their navigation capacity.

8.2.3 General Types of Habitat Use and Selection Data

Habitat use and resource selection have been considered over the decades in a variety of ways. This field envisions the problem of selection at different spatial and temporal scales, and a wide variety of data is used to infer selection.

Habitat and resource selection is frequently viewed as being hierarchical (Johnson 1980; Orians and Wittenberger 1991; Rettie and Messier 2000). In a classic article, Johnson (1980) classified resource selection into four hierarchical levels. He defined *first-order selection* as that of the selection of the geographical range of a species. *Second-order selection* occurs at the home range of an individual. *Third-order selection* relates to the use of habitat components within a home range or territory, while *fourth-order selection* relates to the actual procurement of food items. This hierarchy clearly operates at different spatiotemporal scales.

This hierarchy is very relevant to the ways in which resource and space use are evaluated. While there are several methods for evaluating resource selection and preference, such as choice experiments and the timing (order) of use (Robertson and Hutto 2006), most investigations focus on contrasting habitat or resource use to some measure of availability (Beyer et al. 2010; Aarts et al. 2013). Depending on the hierarchy being considered, different sets of resource use and availability data are used.

Thomas and Taylor (1990, 2006) organized investigations regarding the ways in which scientists interpret resource use and availability. They identified four general study designs (see also Manly et al. 2002). In Design I, data are collected at the population level and individuals are not identified, such as focusing on information regarding variation in density, abundance, or occupancy across a resource gradient. We do not focus on Design I here, because it largely overlaps with applications regarding species distribution modeling (see Chap. 7). In Design II, use by individuals is quantified and compared to availability information measured at the population level (not specific to each individual). In Design III, both use and availability data are quantified for each individual. In Design IV, use and availability are measured repeatedly over time in a paired manner for each individual, thereby allowing for variation individual resource selection over time and space (Arthur et al. 1996; Thomas and Taylor 2006).

8.2.4 Home Range and Space Use Approaches

The analysis of home ranges has a long history in animal ecology. The general goal of home-range analysis is typically to estimate (and map) the utilization distribution of individuals and understand variation in utilization distributions. The utilization distribution (UD) provides a quantitative perspective on space use, which can then be related to several problems regarding environmental gradients, critical habitats, and conservation.

In recent years, there has been an explosion of methods that take advantage of more detailed GPS telemetry data (Kie et al. 2010; Tomkiewicz et al. 2010; Kays et al. 2015). Here, we provide an overview of some of the more common methods, noting that each of these methods have been extended in a variety of ways, and other methods exist. For more detailed reviews, see Moorcroft and Lewis (2006) and Worton (1987).

The simplest home range estimator is the *minimum convex polygon* (MCP; Mohr 1947), which is a type of a convex hull. This approach places a convex box, or hull, around the outer most location points (Fig. 8.4). One can consider all points or only use a proportion of points (95%, 90%, etc.) when creating the hull to remove effects

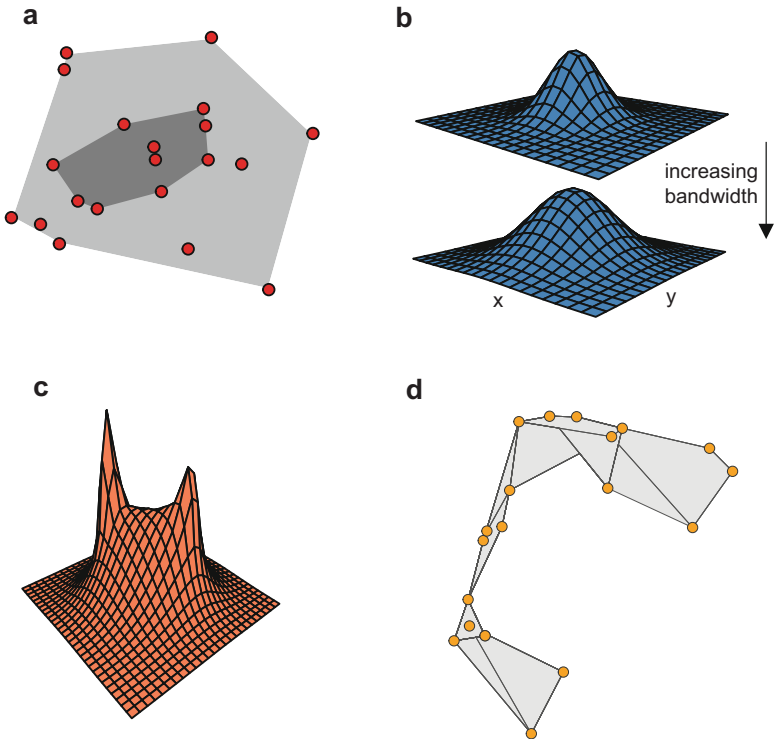


Fig. 8.4 Home range and utilization distribution estimators. (**a**) Minimum convex polygons use convex hulls to delineate the outward boundaries of use. Shown are 100% and 50% minimum convex polygons. (**b**) Kernel density estimators use probability distributions to smooth the expected use around known use locations. Shown is an example with a bivariate normal distribution, illustrating changing bandwidth of the kernel for a single point location. (**c**) Brownian bridge estimators assume random motion between points along a trajectory. Shown is the expected use between two point locations (reflected in the modes of the distribution). Note that the expectation is higher for locations between the two points than locations not between points. (**d**) Local convex hulls take points and create a series of minimum convex polygons using a subset of points that are neighbors, which are stitched together to estimate a home range. Shown is an example where the 4-nearest neighbors are used

of outlier points on the hull. While this approach has been used for decades, it is known to be limited in its ability to quantify utilization distributions for several reasons. First, MCP does not provide a formal estimate of UDs, because it does not estimate a relative intensity of use. Second, as a bounding box, it typically overestimates space use by including areas unsuitable to individuals.

Kernel density estimation (KDE) also has a long history in space use and home range estimation (Worton 1989; Seaman and Powell 1996; Benhamou and Cornelis 2010). Kernel home range estimators use different types of probability density functions (similar to what we used in Chaps. 2–4) for interpolating space use. This general approach is potentially useful because it can capture intensity of use across space. For KDE, we essentially interpolate around each point location based on a kernel function (Fig. 8.4), where the height and spread of the interpolation is based the bandwidth (or "smoothing parameter" or "window width"; Seaman and Powell 1996) of the kernel distribution. Fixed and adaptive kernel estimators have been developed for applying bandwidths (Worton 1989). In a fixed bandwidth, all points use the same bandwidth, while in an adaptive bandwidth, bandwidths can vary across points. Adaptive bandwidths are thought to result in biased estimates of home ranges, where estimates tend to be larger than true home ranges, while fixed bandwidths tend to perform better (Seaman and Powell 1996). For KDEs, point locations are assumed to be independent (Worton 1989), which can be problematic due to temporal autocorrelation (Swihart and Slade 1985), particularly when point locations come from a fine temporal resolution, such as with GPS telemetry data. Violation of this independence assumption can cause problems for bandwidth estimation (Gitzen et al. 2006).

The *Local convex hull* (LoCoH) is a generalization of the MCP and is thought of as a non-parametric kernel method (Getz and Wilmers 2004; Getz et al. 2007). In this approach, convex hulls for successive locations are created using nearest neighbors (defined with different types of criteria; see below and Getz et al. 2007), which are merged together to quantify the home range (Fig. 8.4). This nonparametric approach is potentially useful because it honors hard boundaries and locations that animals may not use (and may otherwise be included in estimates from MCP or KDE). Because LoCoH uses successive locations, it implicitly captures time-related issues; it has been extended to explicitly account for time as well (T-LoCoH; Lyons et al. 2013). See Benhamou and Cornelis (2010) for an alternative way to deal with boundary issues in KDE.

Brownian bridge movement models (Horne et al. 2007) provide a different view on home range analysis by explicitly incorporating movement trajectories into the interpretation of the home range and thus providing more of a mechanistic perspective on the role of movement in home range estimation (in contrast to kernel methods). This approach is based on the idea of a conditional random walk between successive locations (Fig. 8.4), or a random walk conditioned on the time and distance between locations, and the Brownian motion variance ("Brownian" is a term borrowed from physics to describe simple diffusion processes). The model has been used in a variety of ways, but for home range analysis it is relevant because it estimates the probability that an individual occurred in an area of a given period of

time (Horne et al. 2007). This model is more likely to result in estimating connected home ranges than kernel-based approaches, since it estimates use between successive locations via a random-walk process. Importantly, this model does not assume that locations are independent, unlike kernel methods.

Finally, we note that there have been efforts to use mechanistic home range models (Moorcroft et al. 1999; Mitchell and Powell 2004; Moorcroft and Barnett 2008). This general approach focuses on processes that may generate home range patterns in contrast to the above approaches, which tend to focus simply on quantifying home range patterns themselves (without regard to why these patterns emerge). These models are often based on the extension of correlated random walk (CRW) models to capture other key mechanisms that can generate home range patterns, such as response to conspecific scent marks, movement bias toward the center of the home range, avoidance of steep terrain, or memory (Moorcroft et al. 1999, 2006; Hooten et al. 2017). These models typically rely on differential equations and require an advanced understanding of calculus; as a consequence, these approaches will not be considered in more detail here (see Moorcroft and Lewis 2006 for more details).

8.2.5 Resource Selection Functions at Different Scales

Resource selection functions (RSFs) are a group of statistical models that quantify variation in resource use by animals, and are defined as any function that is proportional to the probability of use of a resource unit (Boyce and McDonald 1999; Manly et al. 2002; Lele et al. 2013). They come in many flavors (Manly et al. 2002), depending in part on the type of data being considered as well as the scale in which resource selection is being considered. Resource selection functions also share many features with species distribution models (see Chap. 7) and some RSFs can be considered one type of species distribution model (Franklin 2009; Aarts et al. 2012).

Often a distinction is made between resource selection functions (RSFs) and *resource selection probability functions* (RSPFs) (Boyce and McDonald 1999). The former typically contrasts use to availability of resources such that these functions tend to model a relative measure of use, one that is proportional to the probability of use. In contrast, RSPFs model the probability of use, either from contrasting use to non-use, or through making distributional assumptions regarding the relationships of relative use to the probability of use by capitalizing on the theory of weighted distributions (see Lele and Keim 2006; Lele et al. 2013). Weighted distributions have been shown with simulations to reliably estimate all parameters of RSPFs when sample size is relatively high ($>$500 use points; Rota et al. 2013). The contrast between RSFs and RSPFs is analogous to species distribution models (Chap. 7) that model the probability of occurrence versus a relative probability through the use of presence-only data.

Here, we categorize RSFs based on the type of information being used, such as a collection of use points or a trajectory of use by an individual. We categorize RSFs in this way because of their implementation for spatial ecology concepts and applications. For instance, in the spatial ecology literature, "point selection" is distinguished from "step selection" and "path selection" (Zeller et al. 2012), because these different types of RSFs capture resource selection during movement or dispersal in different ways (see Chap. 9). We describe each briefly below, but for a thorough overview, see the excellent books by Manly et al. (2002) and Hooten et al. (2017).

8.2.5.1 Point Selection

Point selection takes use locations and typically compares use to information about availability, without reference to the trajectory of movement or use that occurred (Fig. 8.5). This approach has a long history in the modeling of resource selection and space use. A simple point selection function for categorical data, termed a "selection ratio", can be described as:

$$w_i = o_i/\pi_i, \tag{8.1}$$

where w_i is the proportion of resource units from category i related to the proportion available, o_i is the proportion of used units in category i, and $\pi_i = a_i/a$, or the proportion of units of i that are available. Here, a_i is the amount or area of category i, and a is the total amount (or area) of all categories. As such, w_i is frequently considered a measure of preference for resource category i. This type of selection ratio has been termed the "Manly selectivity measure", but this general construct dates back to early studies of resource use and availability (Savage 1931). There are several variants on this idea (Manly et al. 2002).

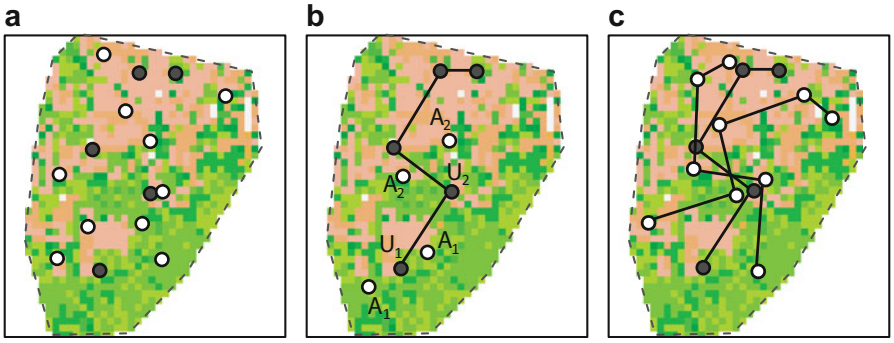


Fig. 8.5 (**a**) Point, (**b**) step, and (**c**) path selection approaches to resource selection. Shown is a minimum convex polygon for a radio-tracked individual, with filled points illustrating a sample of use locations and white points are a sample of available locations. For (**b**), only local availability (A) for the first two usage (U) points is shown

An alternative to this approach for categorical data was the development of compositional analysis (Aebischer et al. 1993). This approach became a popular alternative in the 1990s and early 2000s because addresses some fundamental limitations of selection ratios, such as a lack of independence among resource use samples of an individual (it treats individuals as sample units) and the unit-sum constraint, where avoidance of one type of resource unit may lead to concluding selection for another (Johnson 1980). Compositional analysis is a form of multivariate analysis of variance (MANOVA), applicable to resource selection Designs II and III (Manly et al. 2002). While it became very popular, it has been criticized because it assumes multivariate normality, assumes data from different individuals are independent, and in some cases arbitrary constants must be added to the data (because it is based on log-ratios).

A more general form of point selection is the logistic regression RSF, which can be applied to different types of covariates (e.g., continuous and categorical) and extended to account for spatiotemporal dependence and repeated measures (Manly et al. 2002; Aarts et al. 2008). Here, it is formalized as:

$$w_x = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_k x_k)}, \tag{8.2}$$

where there are K covariates of interest, and β_k is the log odds ratio associated with a 1-unit change in resource k and as such provides a measure of resource use for resource k, and x is the location. Similar RSFs can be constructed based on the amount of use (i.e., log-linear RSFs), or temporally explicit use (e.g., proportional hazards models). The numerator of the RSF in Eq. (8.2) is directly related to Eq. (8.1). See below and Manly et al. (2002) for more details.

This type of RSF shares many similarities of the use of GLMs for predicting species distributions (Chap. 7). Note that similar to presence-only models described in Chap. 6, the intercept, β_0 is not identifiable when applied to use-availability data (i.e., because it will be contingent on the number of available points used). Also, logistic regression RSFs can potentially provide biased estimates in some situations (Keating and Cherry 2004). In particular, concerns have been raised that when using logistic regression with use-availability data, "contamination" can occur, that is, when 0's used in the logistic regression are locations where use actually occurs. Yet there has been some debate regarding if contaminated availability data are really problematic with real-world data (Johnson et al. 2006; Beyer et al. 2010).

Lately, the use of inhomogeneous point process models have been implemented to interpret resource selection (Aarts et al. 2012; Johnson et al. 2013; Northrup et al. 2013; Hooten et al. 2017), similar to their use for modeling species distributions (see Chap. 7). Aarts et al. (2012) illustrated how a variety of RSFs approximate the inhomogeneous point process model. To illustrate this point, they highlight how RSFs can be quantified in environmental space (e.g., Eq. 8.1) or geographic space. For instance, RSFs based on geographic space would consider each location in space as a distinct resource or habitat, such that use is based on where species occur in

geographic space rather than focusing on the frequency of use based on environmental conditions. They also highlight that some models discretize the environment or geographic space (both in terms of use and resource data) while others treat it as continuous. Counts and presence–absence data, for example, are based on a discrete area that is sampled or summarized, whereas point location data may represent an infinitely small area (i.e., it can be treated as continuous). Inhomogeneous point process models for resource selection focus on continuous, geographic space, which can be approximated using discrete data taken from resources summarized in environmental or geographic space. Aarts et al. (2012) showed how logistic regression and Poisson regression can approximate the inhomogeneous point process model. In this way, weights are typically used in the regression model, which may reflect the area or volume sampled with availability points (or "quadrature" points; see Chap. 7). The benefit for using a point process framework is that, similar to species distribution modeling, it clarifies the role of data (e.g., availability points) and it highlights the similarities among models that have been considered more distinct in the past.

Similarly, Johnson et al. (2013) advanced the use of space-time point process models, where events (points where use occurs) are modeled explicitly over time. The benefit of considering space-time point process models is that this framework can readily account for environmental conditions changing over time, impacting availability of resources and use. This framework has been advanced as a more general and flexible framework for resource selection functions (Hooten et al. 2017). Note that for many point process models, standard software can be used (as shown in Chap. 7).

8.2.5.2 Step Selection

With the increased availability of high-resolution GPS telemetry, step-selection functions are frequently used (Fortin et al. 2005; Thurfjell et al. 2014). In a step-selection function, use at time t is contrasted to availability at time t (Fig. 8.5). Availability can be derived at different spatial scales, such that it varies over time. Typically, movement trajectories are used to quantify step lengths (i.e., the distance moved between time $t-1$ and t) and turning angles between successive locations. This information is then used in some way to interpret availability at time t given an individual location at $t-1$ (Fig. 8.5). This data format results in matched data, or a "choice set," where a use location is paired with one or more available locations. Such matched data are more broadly known as a matched case-control design. Statistical models such as the discrete-choice and related conditional logit models are frequently used to interpret resource selection in this context because these types of models can honor the matched sample design (Cooper and Millspaugh 1999; Fortin et al. 2005; Duchesne et al. 2010). While these types of models have largely been assumed to provide synonymous inference to logistic RSFs, the choice probability modeled is subtly different (Lele et al. 2013), where it models the probability of choosing one unit of resource i when offered relative to the choice set at time t.

These models may be useful for linking resource selection and movement across space (Forester et al. 2009; Duchesne et al. 2015), such that they have been widely used in spatial ecology (Thurfjell et al. 2014; Zeller et al. 2016).

Although step selection functions are increasingly used, there are several decisions that are necessary for their development and little consensus currently occurs regarding best practices for their development (Thurfjell et al. 2014). First, most applications of step selection functions focus on the resource used at the end point of each step (and available step), yet some investigators consider resources across the entire step/line segment (Zeller et al. 2012). Second, when generating available points or steps, some have simply buffered around the use location at time t and selected random points within the buffer without respect to distance from the point for comparison with use at time $t + 1$. This latter approach is known to generate bias in selection coefficients and is not recommended (Forester et al. 2009). Rather, available points should follow either the observed step distribution (sampling step lengths from the observed distribution) or through a parametric distribution fit to the observed step lengths. Third, for a step selection function, repeated time steps will typically have a temporal dependence that needs to be accounted for when estimating resource selection. Some applications use an approach based on generalized estimating equations (Fortin et al. 2005; Prima et al. 2017) while others have applied mixed models to accommodate spatiotemporal dependence (Duchesne et al. 2010).

8.2.5.3 Path Selection

An alternative to step selection is to consider entire paths or trajectories of individuals moving across landscapes (Fig. 8.5). The utility of considering entire paths or relevant components of paths (e.g., a daily path, seasonal path, etc.; sometimes referred to collectively as "bursts") is that may provide more useful information on selection during the movement process and it may help reduce problems of temporal autocorrelation in analysis and interpretation of data (Cushman et al. 2011). Path-selection functions have been used in a variety of contexts, such as estimating the resistance of the matrix to movement (Zeller et al. 2016) and interpreting individual variation in dispersal biology (Elliot et al. 2014).

In a path-selection function, a used path is defined and contrasted to a random path. Random paths can be generated in a variety of ways. For instance, the use path can simply be rotated from the origin location. This type of random path preserves the topology of the observed path. In contrast, observed paths can be rotated and shifted in the region of interest. Finally, random paths could be generated from the estimated components of the used path (e.g., the distribution of step lengths and turning angles), similar to a step-selection function but inferred at the level of the entire path. This latter approach has seen less use in resource selection modeling but is commonly used for simulations in spatial ecology (Beyer et al. 2013).

8.3 Examples in R

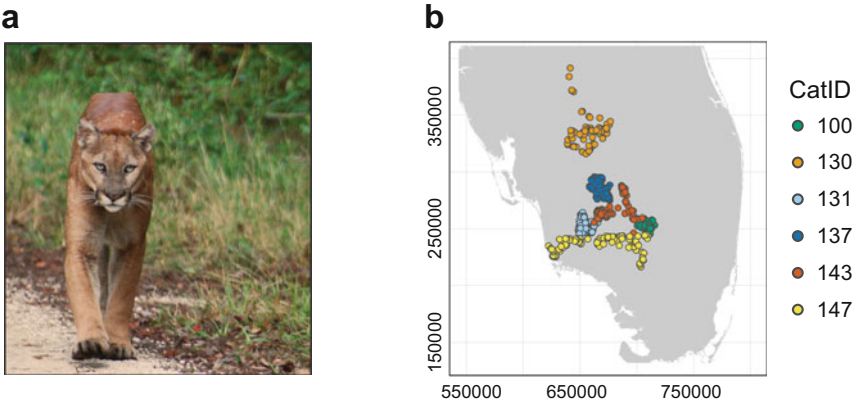
8.3.1 Packages in R

In R, there are several libraries that can be used for resource and habitat selection modeling. We will focus on the use of `adehabitat` (Calenge 2006), which was broken into three packages we will consider: `adehabitatHR`, `adehabitatHS`, and `adehabitatLT`. `adehabitatHR` focuses on home range analysis, `adehabitatHS` focuses on habitat selection analysis, and `adehabitatLT` provides a platform for interpreting and analyzing movement trajectories.

8.3.2 The Data

For the purposes of understanding resource selection and habitat use, we use radiotelemetry data from Florida panthers (*Puma concolor coryi*) in south Florida. The Florida panther is a critically endangered mammal confined to southern Florida. Over the past 30 years, there have been thorough efforts to understand space use, home ranges, and resource selection for the Florida panther using radiotelemetry (Maehr and Cox 1995; Cox et al. 2006; Land et al. 2008; Onorato et al. 2011; Frakes et al. 2015; McCarthy and Fletcher 2015). Across much of this time, panthers were typically located every 1–3 days using fixed wing aircraft and locations had an estimated accuracy of 489 m (Cox et al. 2006). The overwhelming majority (97%) of telemetry locations were collected from 0700 to 1559 and most (88%) from 0700 to 1159 (Cox et al. 2006), limiting inference primarily to the early day when panthers are typically resting. However, Land et al. (2008) contrasted inferences from VHF and GPS telemetry data where points were taken throughout the day and night (when panthers are more active), finding that resource selection was similar with these two data sources.

Here, we illustrate resource selection and space use topics from a small portion of VHF data of radiotelemetered panthers. Our data come from six panthers, three subadults and three adults, each of which has a variable number of relocations (Fig. 8.6). Our goals for this example are twofold. First, we illustrate how such data can be used to interpret space use through quantifying utilization distributions with various methods, each of which provides complementary insights. Second, we contrast approaches to understanding resource selection, which differ in the way they capture issues of spatial and temporal scale during the process of resource selection.



c

CatID	Sex	Age	# Locations	# Months monitored
100	Male	Adult	127	11
131	Male	Adult	118	11
137	Male	Adult	131	11
130	Male	Sub-adult	85	9
143	Male	Sub-adult	129	11
147	Male	Sub-adult	123	9

Fig. 8.6 Telemetry data for the Florida panther in south-central Florida. (**a**) A Florida panther. (**b**) Map of use locations for each panther across the study region. (**c**) The number of panthers and the number of use locations for each panther

8.3.3 Prepping the Data for Modeling

First, we load a raster layer of land-cover types (`panther_landcover.grd`) for the region and a shape file of panther telemetry locations. We consider a raster-based, land-cover map, aggregated to a 500 m resolution (reflecting the approximate grain of telemetry error). This map was generated by the Florida Fish and Wildlife Commission in 2003, with the original map containing 43 land-cover types and a map resolution of 15 m. While there are more recent land-cover mapping efforts, this map is the best mapping effort for the time period of the panther data (2006). We also import an .shp file of panther locations (`panthers.shp`) and overlay the two data sources.

```
> library(raster)
> library(rgdal)

#land cover
> land <- raster("panther_landcover.grd")
> projection(land)
```

```
#Add panther data
> panthers <- readOGR("panthers.shp")
> projection(panthers)
```

These two maps have the same CRS, so we will not need to modify either map for visualizing and analyzing the data (we do store the CRS as `crs.land` for later use). We briefly explore the data and plot it to understand variation in the data (Fig. 8.6).

```
#explore panther data
> summary(panthers)
> unique(panthers$CatID) #the unique cat IDs
> panthers$CatID <- as.factor(panthers$CatID)

#plot
> plot(land)
> points(panthers, col = panthers$CatID)
```

To simplify this analysis, we reclassify the land-cover data into fewer categories with the reclassify function from the `raster` package:

```
#reclassify map into fewer landcover categories
> classification <- read.table("landcover reclass.txt", header = T)
> head(classification)

##
Landcover Description ChangeTo Description2
1 1 CoastalStrand 0 coastalwetland
2 2 Sand 0 coastalwetland
3 23 saltmarsh 0 coastalwetland
4 24 mangroveswamp 0 coastalwetland
5 25 scrubmangrove 0 coastalwetland
6 26 tidalflat 0 coastalwetland
```

This table describes the original classification scheme and the scheme that we are reclassifying. The new categories can be seen with:

```
> levels(classification$Description2)

##
[1] "barrenland" "coastalwetland" "cropland" "cypressswamp"
"dryprairie"
[6] "exotics" "freshwatermarsh" "hardwoodswamp" "openwater"
[12] "pasture/grassland" "pinelands" "scrub/shrub" "uplandforest"
"urban"
```

We can then use the `reclassify` function, but need to convert the data frame into matrix format to do so.

```
> class <- as.matrix(classification[,c(1,3)])
> land_sub <- reclassify(land, rcl = class)
```

We also create some new layers that represent continuous covariates of key land-cover types for panthers by using a moving-window analysis (see Chap. 3) of two forest types, specifically upland forests and forested wetlands. Both of these forest types have previously been shown to be important for resource selection in Florida panthers (Kautz et al. 2006; McCarthy and Fletcher 2015). For this moving window, we calculate the portion of each forest in a 5 km radius, roughly reflecting the median distance moved between successive locations in this data set:

```
#forested wetlands
> wetforest <- land_sub
> values(wetforest) <- 0
> wetforest[land_sub == 9 | land_sub == 11] <- 1

#forested uplands
> dryforest <- land_sub
> values(dryforest) <- 0
> dryforest[land_sub == 10 | land_sub == 12] <- 1

#moving window to get neighborhood proportion
> fw <- focalWeight(land_sub, 5000, 'circle')
> dry.focal <- focal(dryforest,w = fw, fun = "sum", na.rm = T)
> wet.focal <- focal(wetforest,w = fw, fun = "sum", na.rm = T)

#merge raster data
> layers <- stack(land_sub, wet.focal, dry.focal)
> names(layers) <- c("landcover", "wetforest", "dryforest")
> plot(layers)
```

Plotting these maps highlights that wet forest habitats are largely in the southwest portion of Florida, while dry forest (upland) forest is scattered throughout the study area (Fig. 8.7). With these data, we can turn to estimating space use through home range analysis.

8.3.4 Home Range Analysis

There are a variety of approaches to estimating home ranges of animals. Here, we will illustrate several approaches. For `adhabitatHR`, we take `panthers` and select the column of the data frame that includes the animal identification number (`CatID`) to calculate Minimum Convex Polygons with the `mcp` function:

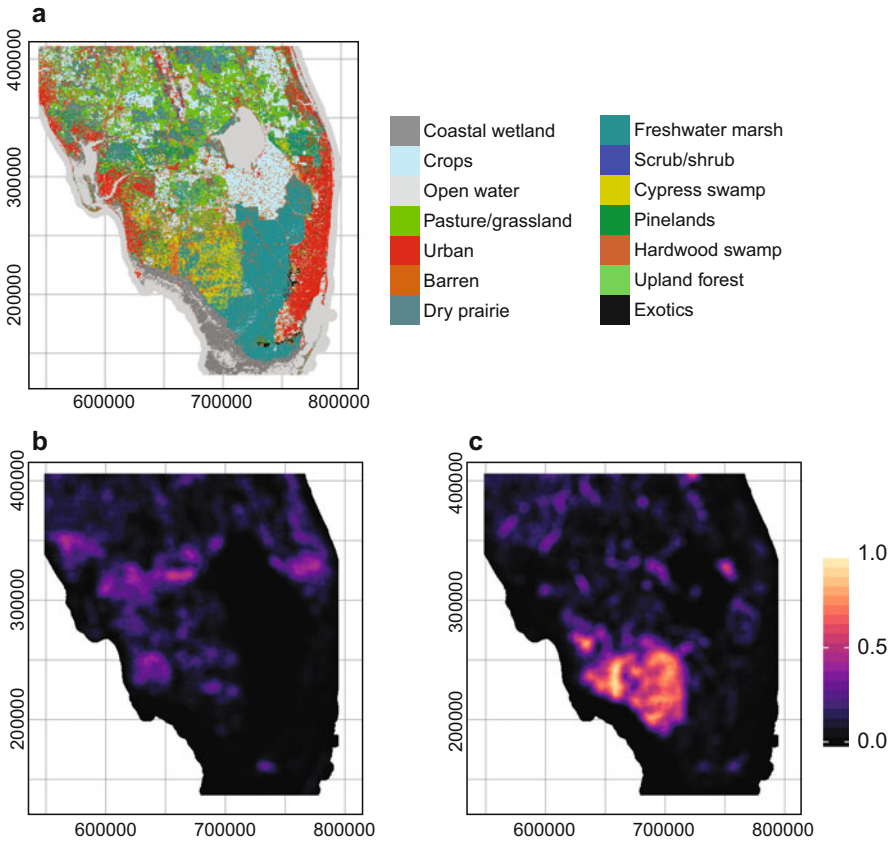


Fig. 8.7 Covariates considered in resource selection analyses. (**a**) A reclassified land-cover layer, (**b**) the proportion of wetland forest within 5 km, and (**c**) the proportion of upland forest within 5 km

```
> library(adhabitatHR)
> mcp95 <- mcp(panthers[,"CatID"], percent = 95)
> mcp50 <- mcp(panthers[,"CatID"], percent = 50)

> plot(land_sub)
> plot(panthers, col = panthers$CatID, add = T)
```

Above we plot the panther locations and overlay the 100%, 95%, and the 50% MCP (Fig. 8.8), the latter of which is sometimes referred to as the "core area" used (Seaman and Powell 1996). The objects created by `mcp` are spatial polygons. We can extract more information from this analysis, such as the area of MCPs (see below).

Another common approach to home range analysis is through the use of kernel methods. We considered kernels briefly when considering spatial point patterns in Chap. 4, and for scale issues in Chaps. 2 and 3. Here, we use a similar approach.

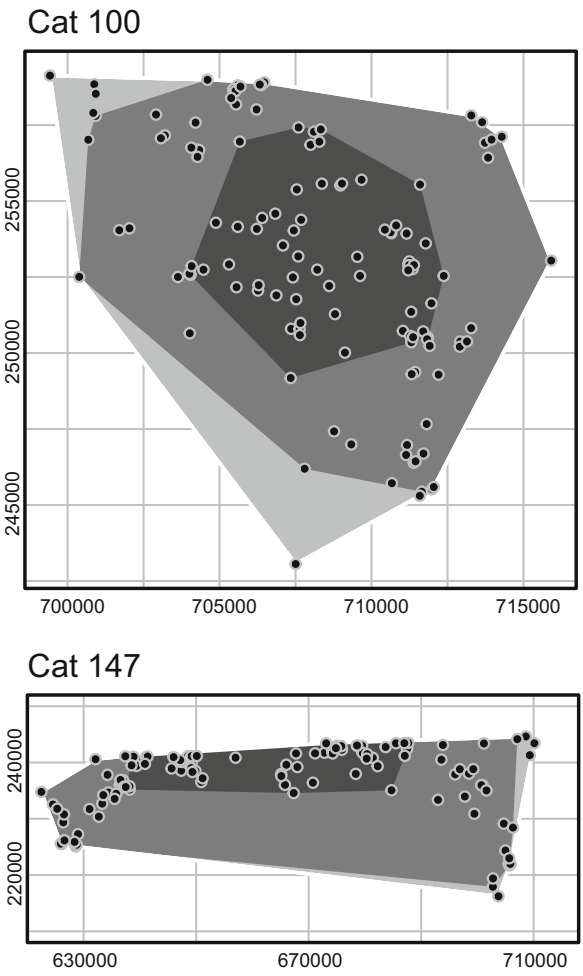


Fig. 8.8 Home ranges based on minimum convex polygons for two panthers, one adult (Cat 100) and one sub-adult (Cat 147). Shown are the raw locations (points) and the 50%, 95%, and 100% minimum convex polygons for each panther

For KDEs, there are two key issues to consider. First, there are a few different types of kernels that can be applied. A classic kernel relies on the bivariate normal distribution (aka, a Gaussian kernel); it is bivariate because the normal distribution is considered in both the x and y dimension. One common non-normal approach includes the Epanechnikov kernel (Epanechnikov 1969). Second, there are a variety of ways to identify appropriate bandwidths, which alter the degree of smoothing in the kernel and the extent of interpolation (Worton 1995). The default technique for estimating a smoothing bandwidth is an ad hoc approach based simply on the number of locations and the variance of the x–y coordinates, termed h_{ref}. Below we calculate both the bivariate normal kernel and the Epanechnikov kernel (Fig. 8.9) using h_{ref}:

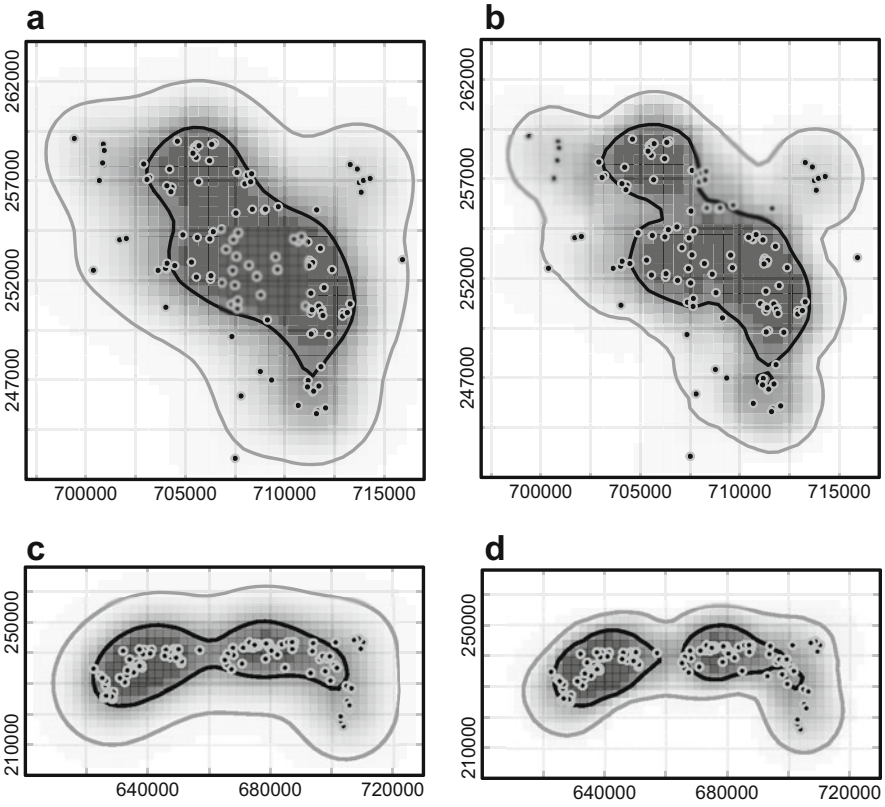


Fig. 8.9 Home ranges estimated using kernel methods for Cat 100 (**a**, **b**) and Cat 147 (**c**, **d**). Shown are the raw locations (points), utilization distributions quantified with bivariate normal kernels (**a**, **c**) and the Epanechnikov kernel (**b**, **d**), and the 50% and 95% contours

```
> kernel.href.bivar <- kernelUD(panthers[,"CatID"], h = "href", kern =
  "bivnorm")
> kernel.href.epa <- kernelUD(panthers[,"CatID"], h = "href", kern =
  "epa")

> image(kernel.href.bivar)
> image(kernel.href.epa)
```

The object created is an `estUD` object, which essentially stores information for each individual in list form. We can extract information regarding the utilization distribution based on the bivariate normal kernel for the first individual as:

```
> kernel.href.bivar[[1]]@data
```

Note that this information is in vector format, where each value reflects a pixel on the map used; this vector does not include values for all pixels on our map (i.e., the

length is not equal to `ncell(land_sub)`). We can also extract what values of h were used for each animal (here shown is animal 2):

```
> kernel.href.bivar[[2]]@h
```

h_{ref} can sometimes lead to over-smoothing of the utilization distribution. We can manually change h, such as setting it to a lower value than that estimated by h_{ref} (which ranged from approximately 1600–9000 in this example). Alternatively, least-squares cross validation (h_{lscv}) is another approach recommended to estimate reliable bandwidths (Worton 1989). Yet, sometimes this method (and related ones) does not converge. We can estimate home-ranges use h_{lscv} in the above function by calling, `h = "LSCV"`. For this example, h_{lscv} does not converge for some of the animals.

We can also estimate home ranges using the local convex hull approach (LoCoH). With this general approach, there are different ways to generate local convex hulls. Getz et al. (2007) argued that *a*-LoCoH, or the "adaptive sphere-of-influence" is a superior approach for LoCoH in contrast to a fixed sphere, *r*-LoCoH, or a fixed number of points, *k*-LoCoH, all of which can be implemented with `adehabitatHR`.

We illustrate both *a*-LoCoH and *k*-LoCoH for panther 147. In each approach, we first need to search for the best parameter regarding neighbors that delineate local hulls. This is frequently done by increasing the parameters that are used for hull construction (k, a, and r) and calculating changes in home range area. In this context, k is the number of nearest neighbors minus one from which hulls are created, and a is a parameter where convex hulls are created such that the sum of their distances from the maximum number of nearest neighbors is $\leq a$. In general, as k, a, or r increases, there will be fewer unused areas (or holes), such that the home range polygon will have a coarser grain (Getz et al. 2007). As these parameters increase, the estimated home range area may level off once all spurious holes within the home range are covered, but home range area should increase again when one or more real holes in use become covered (Getz et al. 2007). Getz et al. (2007) recommended starting k at $N^{0.5}$ (where N is the number of points for an individual), while a can be initiated at the maximum distance between pairs of points. We use these rules of thumb below.

```
#subset to only consider panther 147
> panther147 <- panthers[panthers$CatID == 147,]

#initialize
> k.int <- round(nrow(coordinates(panther147))^0.5, 0)
> a.int <- round(max(dist(coordinates(panther147))), 0)

> k.search <- seq(k.int, 10*k.int, by = 5)#number of points
> a.search <- seq(a.int, 2*a.int, by = 3000)#distance in m
```

```
> LoCoH.a.range <- LoCoH.a.area(SpatialPoints(coordinates
  (panther147)), arange = a.search)
> LoCoH.k.range <- LoCoH.k.area(SpatialPoints(coordinates
  (panther147)), krange = k.search)

> plot(LoCoH.a.range)
> plot(LoCoH.k.range)
```

In the above search, we evaluate the changes in home range area as k and a increase. In this case, we select some parameter estimates from these graphs where home range area tends to asymptote and then calculate the home range based on those parameters (top panels of Fig. 8.10). Note that in these data, there is not a strong asymptote, so we simply pick values where the rate of change slows.

```
> a.search[5]
> k.search[11]

> LoCoH.a.124062 <-
  LoCoH.a(SpatialPoints(coordinates(panther147)), a = a.search[5])
> plot(LoCoH.a.124062)

> LoCoH.k.61 <- LoCoH.k(SpatialPoints(coordinates(panther147)),
  k = k.search[11])
> plot(LoCoH.k.61)
```

By altering the parameters based on the `LoCoH.a.area` and `LoCoH.k.area` plots, we see that a-LoCoH appears to be less sensitive to changes in a than k-LoCoH is to changes in k (lower panels of Fig. 8.10). Overall, the LoCoH approach appears to adequately capture the highly irregular space use by panther 147, unlike the minimum convex polygon (and to a lesser extent, the kernel approach).

Brownian bridge models can also be estimated with `adebhabitatHR`. The general approach of Horne et al. (2007) can be implemented with the `kernelbb` function. These models require a different format for the data that captures the trajectory of each individual. Our location data can be converted to trajectory data with the `adehabitatLT` package. For applications to Brownian bridges, we must ensure that the trajectory has time-related information (e.g., date; `typeII = TRUE` below). To make this trajectory information, we format the date information as a `POSIXct` object in R. This is a useful format for doing calculations based on dates and is needed for the creation of a trajectory-based object.

```
#reformat Juldate to a POSIXct object for date
substrRight <- function(x, n){
 substr(x, nchar(x)-n+1, nchar(x))
}

panthers$Juldate <- as.character(panthers$Juldate)
panther.date <- as.numeric(substrRight(panthers$Juldate, 3))
panthers$Date <- as.Date(panther.date, origin=as.Date("2006-01-01"))
```

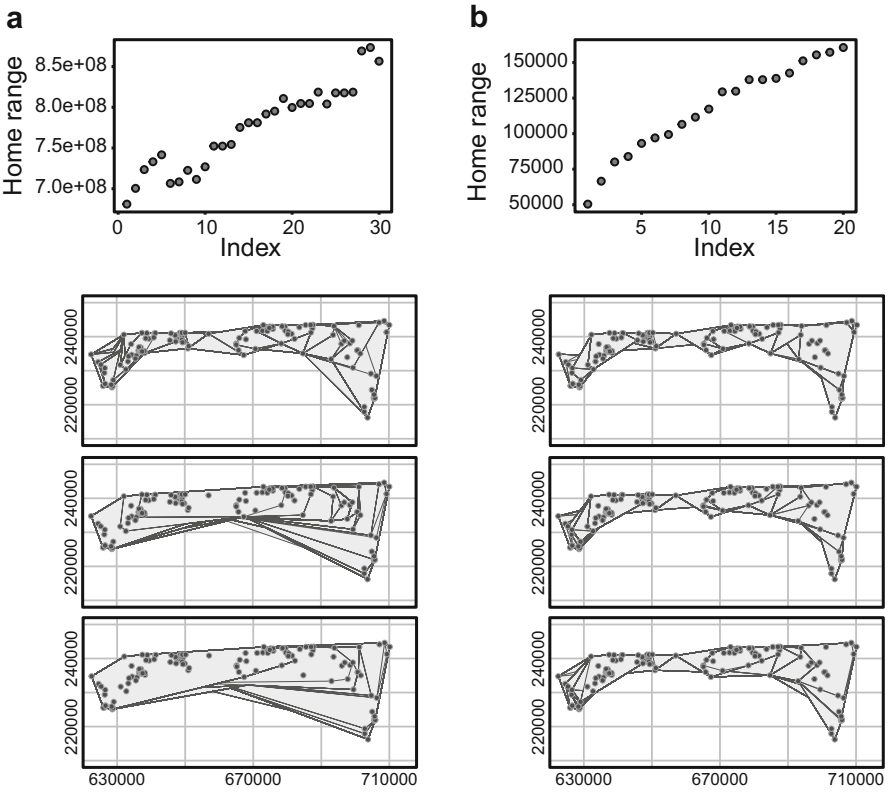


Fig. 8.10 The use of local convex hulls for home range estimation. (**a**) *k*-LoCoH, which uses the *k*-nearest neighbors in hull delineation. (**b**) *a*-LoCoH, which uses an adaptive radius for delineating neighbors. For each, top panels show searches for identifying the best values for *k* and *a*. Maps show changes in home range delineations with (**a**) increases in *k*, and (**b**) increases in *a* (Values increase from upper to lower panels)

```
#create POSIXct object
panthers$Date <- as.POSIXct(panthers$Date,"%Y-%m-%d")

> panther.ltraj <- as.ltraj(xy = coordinates(panthers), date = panthers
  $Date, id = panthers$CatID, typeII = T)

> plot(panther.ltraj)
```

Storing the data in this way provides a useful means to visualize trajectories (Fig. 8.11) from radiotelemetry data, as well as to quantify aspects of the trajectory (see Step Selection Section 8.3.5.2 below). With this trajectory, we must specify two parameters to fit a Brownian bridge. First, the parameter `sig1` must be estimated. This parameter is related to the speed of the individual. As the value of this parameter increases, the assumed tortuosity of the path increases. Second, the parameter `sig2`

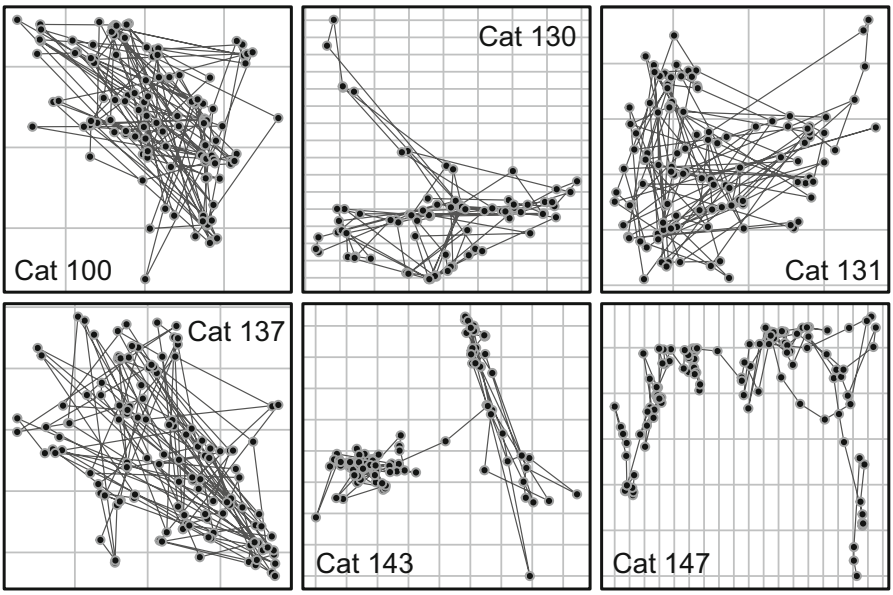


Fig. 8.11 Trajectories of each panther. Background lines reflect 10 × 10 km grids

must be quantified. This parameter is related to *h* in KDE methods and can reflect the error in the location data. As this parameter increases, the smoothing increases.

We estimate the first parameter using maximum likelihood with the `liker` function (see Horne et al. 2007):

```
#telemetry error
> sigma2 <- 450

> sigma1 <- liker(panther.ltraj, sig2 = sigma2, rangesig1 = c(2, 100))

> sigma1

##
Maximization of the log-likelihood for parameter
sig1 of brownian bridge

100 : Sig1 = 13.7718 Sig2 = 450
130 : Sig1 = 18.0881 Sig2 = 450
131 : Sig1 = 13.968 Sig2 = 450
137 : Sig1 = 15.1451 Sig2 = 450
143 : Sig1 = 10.8288 Sig2 = 450
147 : Sig1 = 7.1992 Sig2 = 450
```

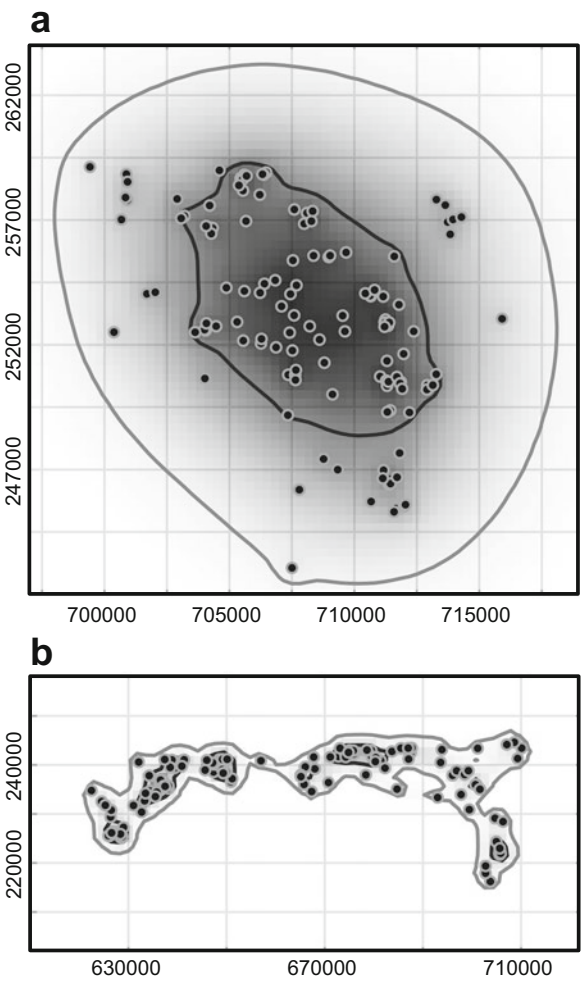


Fig. 8.12 The use of a Brownian bridge for home range estimation for (**a**) Cat 100 and (**b**) Cat 147. Also shown are the 50% and 95% utilization distributions

Given the variation in estimates of `sig1`, we may want to fit the Brownian bridge separately for each cat. We fit a Brownian bridge model to panther 147 as (Fig. 8.12):

```
> bb.147 <- kernelbb(panther.ltraj[6], sig1 = 7.2, sig2 = sigma2, grid =
  100)
```

where `grid` refers to the size of the grid needed to be estimated. The grid sets the grain of the analysis. As grid size increases, so does the computation time. Increasing grid size may be useful for visualizing the utilization distribution, although if interest is solely on estimating home range size, a fine grain may not be needed (although it may increase the precision of the estimate). The object created by `kernelbb` is an `estUDm` object, from which we can extract a variety of information.

Beyond simply mapping, we can summarize the home range analyses in a variety of ways. We can contrast home range area estimates between the MCP, the kernel approach, and the Brownian bridge (output not shown):

```
#home range area estimates based on 95%
> kernel.95 <- getverticeshr(kernel.href.bivar, percent = 95)
> bb.95 <- getverticeshr(bb.147, percent = 95)

#contrast area
> mcp95$area
> kernel.95$area
> bb.95$area
```

While home range estimators have proven valuable for interpreting space use and utilization distributions, most applications do not directly provide inferences on resource selection and habitat preferences. For this, we turn to resource selection functions.

8.3.5 Resource Selection Functions

With the above data, we use resource selection functions to interpret habitat use patterns. This can be done in a variety of ways. We illustrate several general approaches based on point selection, step selection, and path selection functions.

8.3.5.1 Point Selection Functions

The most common approach for estimating resource selection functions is to contrast use points versus available points (a "point selection" analysis). Available points can be generated in a variety of ways, depending on the scale at which resource selection is interpreted.

We first consider the situation where we want to estimate selection ratios for land-cover (or vegetation) types—categorical variables. We then extend the general idea to models that can accommodate categorical and continuous covariates.

For a Design II analysis, we contrast use versus availability across the entire study area extent. First, we reformat the use point data in a way that has the appropriate format for `adehabitatHS`. That package requires that each individual is a row in a data frame and each column is a land-cover type. We extract land-cover types at use points with the `extract` function in the `raster` package, and then reshape the data to a proper format with the `reshape2` package (Wickham 2007) for input into `adehabitatHS`:

```
> library(reshape2)
> use <- extract(layers, panthers)
```

```
> use <- data.frame(use)
> use$CatID <- as.factor(panthers$CatID)

#use reshape2, dcast function:
> useCatID <- dcast(use, CatID ~ landcover, length, value.var = "CatID")
> newclass.names <- unique(classification[,3:4])
> names(useCatID) <- c("CatID", as.character(newclass.names[1:13,2]))
```

We generate 1000 random points (though in practice, more points might be needed; Northrup et al. 2013) and extract land-cover categories at those points to gain information on availability across the extent of interest with the `sampleRandom` function:

```
#use sampleRandom function from raster to create availability
> set.seed(8)
> rand.II <- sampleRandom(landcover, size = 1000)
> rand.II.land <- data.frame(rand.II)

#sum up counts of each landcover type
> table(rand.II.land)
```

With these available points, we reshape them for inclusion with use points.

```
#sum up counts of each landcover type
> avail.II <- tapply(rand.II.land, rand.II.land, length)

> names(avail.II) <- as.character(newclass.names[1:14, 2])
> avail.II

#remove exotics, which was not observed in sample of use
> avail.II <- avail.II[c(-14)]
```

For a Design III resource selection analysis, we derive available points within the home ranges of individuals. A common approach to do so is through the sampling of a minimum convex polygon, although other approaches are possible (e.g., Rota et al. 2014). To illustrate, we take 200 samples from each individual home range, based on a 99% MCP for each individual. To do so, we use a `for` loop, where for each individual, we calculate the home range and then use the `spsample` function in the `sp` package to generate random points within the home range polygon that is created (note `sampleRandom` from the `raster` package used above will not constrain sampling to the MCP polygon):

```
> library(sp)
> cat.unique <- unique(panthers$CatID)
> samples <- 200
> rand.III <- matrix(nrow = 0, ncol = 2)
```

```
#loop for all individuals
> for(i in 1:length(cat.unique)){
  id.i <- cat.unique[i]
  cat.i <- panthers[panthers$CatID == id.i,]
  mcp.i <- mcp(SpatialPoints(coordinates(cat.i)), percent = 99)
  rand.i <- spsample(mcp.i, type = "random", n = samples)
  rand.i.sample <- extract(land_sub, rand.i)

 #make a matrix of CatID and rand samples
 cat.i <- rep(cat.unique[i], length(rand.i))
 rand.cat.i <- cbind(cat.i, rand.i.sample)
 rand.III <- rbind(rand.III, rand.cat.i)
}
```

Now we reshape the data with the `dcast` function in `reshape2` package.

```
> rand.III <- data.frame(rand.III)
> rand.III$cat.i <- as.factor(rand.III$cat.i)
> colnames(rand.III)=c("cat.i","landcover")
> avail.III <- dcast(rand.III, cat.i ~ landcover, length, value.var =
  "cat.i")
> names(avail.III) <- c("CatID", as.character(newclass.names[1:13,2]))
> avail.III
```

With such formatted data for Design II and III resource selection, we use the `adehabitatHS` package to calculate selection ratios (Eq. 8.1). For selection ratios in a Design II (one set of availability data for all individuals; Fig. 8.13):

```
> library(adehabitatHS)
> sel.ratioII <- widesII(u = useCatID[,c(2:ncol(useCatID))],
  a = as.vector(avail.II),
  avknown = F, alpha = 0.05)
> summary(sel.ratioII)
> sel.ratioII
> sel.ratioII$wi #selection ratios
> sel.ratioII$se.wi #selection ratio SEs
> plot(sel.ratioII)
```

In the above approach, we provide use samples and available samples, and we note that availability is assumed to be not known (`avknown = F`) but rather it is estimated through the sample availability points. In this case, we could have taken proportions from GIS data (or potentially sampled an arbitrarily large number of availability points) and set `avknown = T`, but we use a random sample approach to be consistent with other resource selection approaches described below. For selection ratios in a Design III (different availability for each individual):

```
> sel.ratioIII <- widesIII(u = useCatID[,c(2:ncol(useCatID))],
  a = avail.III[,2:14],
  avknown = FALSE, alpha = 0.05)
```

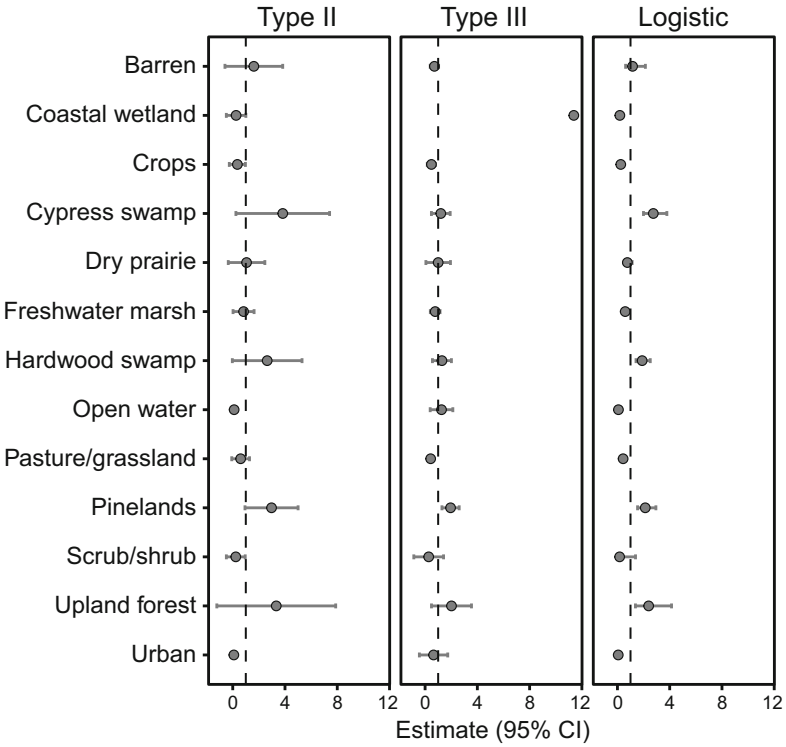


Fig. 8.13 Selection ratios calculated with a (**a**) Type II analysis for the Manly selectivity measure, (**b**) Type III analysis for the Manly selectivity measure, and (**c**) logistic regression (Design II analysis). Note that for the logistic regression, we show odds ratios from the log-linear function, $w = \exp(\beta_1)$

```
> summary(sel.ratioIII)
> sel.ratioIII
> sel.ratioIII$wi #selection ratios
> sel.ratioIII$se.wi #selection ratio SEs
> plot(sel.ratioIII)
```

These selection ratios are useful when covariates of interest are categorical. The tests assumed that individuals are independent samples and that selection (preferences) is the same among individuals (Manly et al. 2002). These assumptions can be potentially overcome with mixed effects models (Thomas et al. 2006).

The above approaches do not accommodate cases when covariates are continuous. When covariates are continuous, a more common approach is the use of regression-based RSF models (and related extensions) to make inference on resource selection (Manly et al. 2002; Keating and Cherry 2004; Johnson et al. 2006). Logistic RSFs can be fit with standard functions in R, such as the `glm` function.

This approach is very similar to fitting a GLM model for presence-only distribution data (Chap. 7). Here, we briefly illustrate this for a Design II analysis.

Logistic RSF type models can be fit in a variety of ways. For a Design II analysis, we specify two RSF models as:

```
#create data frame
> use.cov <- data.frame(use[,1:3], use = 1)
> back.cov <- data.frame(rand.II, use = 0)
> all.cov <- data.frame(rbind(use.cov, back.cov))

#run two models
> rsf.all <- glm(use ~ landcover + wetforest + dryforest, family
  = binomial(link = logit), data = all.cov)

> rsf.forest <- glm(use ~ wetforest + dryforest, family =
  binomial(link = logit), data = all.cov)
```

In the first model, we include categorical land-cover type (at used and available points), as well as continuous metrics of the proportion of wet and dry forest in the surrounding landscape (see above). In the second model, we remove land-cover type, because there are several categories and thus several df (13) are required for this model. We can ask whether the inclusion of land-cover type is warranted either with model selection criteria (e.g., Akaike's Information Criterion) or using a likelihood-ratio test as:

```
> anova(rsf.forest, rsf.all, test = "LRT")
```

In this case, the likelihood ratio test, as well as model selection criteria (with landcover, AIC = 1706.7; removing land-cover, AIC = 1795.4), suggest that land-cover type should be included in the model. Note that this logistic RSF is assuming that all points are independent of each other, which might not be warranted. The GLM framework can be extended to deal with such non-independence (see Chap. 5) (Aarts et al. 2008). We can contrast the results of using a logistic regression framework to Type II selection ratios calculated above by fitting a land cover only model and extracting w by exponentiating the beta coefficients (e.g., selection for upland forest, $w = \exp(\beta_{\text{uplandforest}})$; Fig. 8.13), also known as the odds ratios estimated in the logistic regression (Manly et al. 2002). In this context, the odds ratios from the logistic regression are nearly identical to the Type II selection ratios ($r = 1$); however, there tends to be greater uncertainty in the estimates when using logistic regression (Fig. 8.13).

Finally, we note that an inhomogeneous point process model could be approximated with the logistic model to interpret resource selection (Aarts et al. 2012). An efficient way to approach resource selection from this perspective is to use a regular grid of available points across the region of interest. Weights can then be applied based on the area sampled (e.g., the grain of the grid). Note that Aarts et al. (2012) used weights for a Poisson regression approximation in this way, but not for logistic

regression; however, see Fithian and Hastie (2013) for another example of using weighted logistic regression to approximate the inhomogeneous point process model. The inhomogeneous point process model for resource selection can be approximated with a weighted logistic regression as:

```
> library(sp)

#get polygon boundary for study area
> raster.extent <- land > -Inf
> studyregion <- rasterToPolygons(raster.extent, dissolve=TRUE)

#create regular grid
> rand.grid <- spsample(studyregion, cellsize = 1000, type="regular")
> grid.1km <- SpatialPoints(rand.grid, proj4string = CRS(crs.land))
> grid.area <- 1000 * 1000

#extract covariates
> rand.cov.grid <- extract(layers, grid.1km)
> use.cov.grid <- data.frame(use[,1:3], use = 1, grid.area = 1)
> back.cov.grid <- data.frame(rand.cov.grid, use = 0, grid.area = grid.
  area)
> all.cov.grid <- data.frame(rbind(use.cov.grid, back.cov.grid))

#logistic approximation to ipp
> rsf.ipp.forest <- glm(use ~ wetforest + dryforest, weight = grid.area,
  family = binomial(link = logit),data = all.cov.grid)
```

We use the `sp` package to generate a regular grid of points, spaced 1 km apart. The `spsample` function requires a polygon boundary of the extent of interest for sampling, which we create with the `rasterToPolygons` function. We then extract the covariates and create a data frame that includes the weights. Overall, this model provides similar estimates to the logistic model described above for `dryforest`, but slightly higher estimates for `wetforest`.

8.3.5.2 Step Selection Functions

For resource selection functions that use information on steps and paths of trajectories, we will use the `adehabitatLT` package. This package can accommodate trajectory data from individuals and will provide several helpful summary statistics from such types of data. The `as.ltraj` function that we used to create the `panther.ltraj` object above for Brownian bridges creates a list file for each trajectory (here summarized at the cat level; that is, one trajectory per individual cat). This list has several useful summary statistics, including: (1) the original x–y locations (note that the number of rows in the `ltraj` object is the same as the number of points); (2) the change in x–y coordinates (distance moved in x–y directions from time t to $t + 1$); (3) the distance moved from t to $t + 1$; (4) change

in time (difference in time intervals between successive locations); (5) mean squared displacement; (6) the absolute angle change; and (7) the relative angle change. The absolute angle change is the angle moved relative to the *x*-axis, while the relative angle change is the turning angle between moves (e.g., 0 means that the individual proceeded in the same direction). Note that these angles are provided in radians (0–2π) rather than degrees (0–360°). The object created can be visualized in several ways.

```
#plot trajectories
> plot(panther.ltraj)
> plot(panther.ltraj, id = "147")
```

We then use this information to calculate the average step lengths and turning angles by individuals. For instance, we can plot histograms and rose diagrams to visualize the variation in step lengths and turning angles, respectively (plots not shown):

```
#distance for second CatID
> panther.ltraj[[2]][,6]
> hist(panther.ltraj[[2]][,6], main = "Second CatID")

#plots of relative movement angles for second CatID
#relative angles: change in direction from previous time step
> rose.diag(na.omit(panther.ltraj[[2]][,10]), bins = 12, prop = 1.5)
> circ.plot(panther.ltraj[[2]][,10], pch = 1)
```

Step selection functions can be generated by contrasting the location selected at time *t* to alternative, available locations to the individual at time *t*, based on information regarding step lengths and potentially turning angles. Here, we consider an example where for each location used, we generate three available locations that sample habitat availability. In practice, we might want to generate many more availability points (Northrup et al. 2013).

```
> stepdata <- data.frame(coordinates(panthers))
> stepdata$CatID <- as.factor(panthers$CatID)
> names(stepdata) <- c("X", "Y", "CatID")
> n.use <- dim(stepdata)[[1]]
> n.avail <- n.use * 3
```

Now we generate random samples of step lengths and relative turning angles from our observed distribution:

```
#convert trajectory back to data frame for easy manipulation
> traj.df <- ld(panther.ltraj)

#sample steps/angles with replacement
> avail.dist <- matrix(sample(na.omit(traj.df$dist), size =
```

```
  n.avail, replace = T), ncol = 3)
> avail.angle <- matrix(sample(na.omit(traj.df$rel.angle), size
  = n.avail, replace = T), ncol = 3)

#name columns
> colnames(avail.dist) <- c("a.dist1", "a.dist2", "a.dist3")
> colnames(avail.angle) <- c("a.angle1", "a.angle2", "a.angle3")

#link available distances/angles to observations
> traj.df <- cbind(traj.df, avail.dist, avail.angle)
```

Here, we created available data for step lengths and relative turning angles from our observed data. When considering multiple animals, it is common to draw relative turn angles and step distances from observed data for all individuals, except the focal animal, to reduce problems of potential circularity (Thurfjell et al. 2014).

With these available distances and relative turn angles, we can calculate the x–y coordinates of available locations. To do so, we need to clearly understand how the trajectory data frame stores information in relation to locations at time t and $t + 1$. What we want is to take the location at time t and use our available distances and turn angles to generate available coordinates in $t + 1$ to contrast to use in $t + 1$. In our data frame, x–y coordinates for each use location are based on time t, where the distance and turn angles for that row in the data frame can generate x–y locations in $t + 1$ (the next row of data). For example, we can take the x–y coordinates for row 2 and calculate the x–y coordinates in $t + 1$ (row 3) with the absolute or relative angles using trigonometry (output not shown):

```
#calculate coordinates in t+1 from t using absolute angle:
> traj.df[2, "x"] + traj.df[2, "dist"] * cos(traj.df[2,
  "abs.angle"])
> traj.df[2, "y"] + traj.df[2, "dist"] * sin(traj.df[2,
  "abs.angle"])

#calculate coordinates in t + 1 from t using relative angle:
> traj.df[2, "x"] + traj.df[2, "dist"] *
  cos(traj.df[1,"abs.angle"] + traj.df[2, "rel.angle"])
> traj.df[2, "y"]+ traj.df[2, "dist"] * sin(traj.df[1,
  "abs.angle"]+ traj.df[2, "rel.angle"])

#check
> traj.df[3, c("x", "y")]
```

Note that for the use of relative angles, we need information on the absolute angle in $t-1$. With this structure in mind, we create new values in the data frame, where available x–y coordinates are created and linked to the appropriate use coordinates. We first create a new column in the data frame, which contains the absolute angle for $t-1$. We illustrate calculations using a `for` loop. The reason for this loop is that we do not want the absolute angle for $t-1$ in situations where the previous row of data is a different individual. In that case, absolute angle should be an NA.

```
> traj.df$abs.angle_t_1 <- NA
> for(i in 2:nrow(traj.df)){
  traj.df$abs.angle_t_1[i] <- ifelse(traj.df$id[i] ==
  traj.df$id[i - 1], traj.df$abs.angle[i - 1], NA)
}
```

Alternatively, we could apply the same logic without the use of a `for` loop as:

```
> traj.df$abs.angle_t_1 <- c(NA,
  traj.df$abs.angle[1:nrow(traj.df) − 1])
> traj.df[!duplicated(traj.df$id), "abs.angle_t_1"] <- NA
```

Then, we calculate new x,−y coordinates using trigonometry. For brevity, we show code for calculating only one availability point:

```
#calculate use coords for t+1
> traj.df$x_t1 <-
  traj.df[, "x"] + traj.df[,"dist"] * cos(traj.df[, "abs.angle"])
> traj.df$y_t1 <- traj.df[, "y"] + traj.df[, "dist"] * sin(traj.df[,
  "abs.angle"])

#calculate avail coords for t+1
> traj.df$x_a1 <- traj.df[, "x"] + traj.df[, "a.dist1"] *
  cos(traj.df[, "abs.angle_t_1"] + traj.df[, "a.angle1"])
> traj.df$y_a1 <- traj.df[, "y"] + traj.df[, "a.dist1"] *
  sin(traj.df[, "abs.angle_t_1"] + traj.df[, "a.angle1"])
```

With these new coordinates, we reshape the data for step selection into a long format. Again, we only show one set of availability points, but note that when appending, we append all data together:

```
#reformat data for step selection
> traj.df <- traj.df[complete.cases(traj.df),] #remove NAs

> traj.use <- data.frame(use = rep(1,nrow(traj.df)),
  traj.df[,c("id", "pkey", "date", "x_t1", "y_t1")])
> traj.a1 <- data.frame(use = rep(0,nrow(traj.df)),
  traj.df[,c("id", "pkey", "date", "x_a1", "y_a1")])

> names(traj.use) <- c("use", "id", "pair", "date", "x", "y")
> names(traj.a1) <- c("use", "id", "pair", "date", "x", "y")

#append use and available data together
#note that traj.a2/a3 should be created in same way as traj.a1
> stepdata.final <- rbind(traj.use, traj.a1, traj.a2, traj.a3)
```

This new data frame stores a use and availability choice set based on each `pair` (`pkey`) for each individual cat (`id`). With this information, we then use the

`extract` function in the `raster` package to get information on environmental covariates between paired use and availability locations.

```
#create a spatial points data frame
> step.coords <- SpatialPoints(stepdata.final[,c("x", "y")],
  proj4string = CRS("crs.land"))

#extracts covariates from layers
> cov <- extract(layers, step.coords)

#add covariates to dataframe of use/available
> stepdata.final <- data.frame(cbind(stepdata.final, cov))
```

Finally, we fit two types of conditional logit models and contrast them to a conventional logistic regression RSF. Conditional logit models are related to standard logistic regression but honor the matched data by including strata that identify each choice set. In the first model, we simply have `pair` as our strata of interest. In the second model, we add individual cats (`CatID`) as a cluster (Fortin et al. 2005). This latter approach is sometimes used to account for a lack of independence within individuals, and is akin to the use of generalized estimating equations for accounting for spatial or temporal dependence (Fieberg et al. 2009).

```
> library(survival)
#conditional logistic
> logit.ssf <- clogit(use ~ wetforest + dryforest + strata(pair), data =
  stepdata.final)

#including catID as cluster (~GEE)
> logit.cat.ssf <- clogit(use ~ wetforest + dryforest + strata(pair) +
  cluster(id), method = "approximate", data = stepdata.final)

#logistic ignoring the local pairing structure of the data
> logit.rsf <- glm(use ~ wetforest + dryforest, family = "binomial",
  data = stepdata.final)
```

In this case, the conditional logit RSFs based on step-selection functions provide qualitatively similar conclusions regarding habitat–use relationships (Table 8.2; note that the values might change slightly because of the stochastic nature of selecting a small number of availability points). The primarily qualitative difference is that they identify forested wetlands as being selected, while the standard logistic regression does not identify significant relationships based on 95% CI with this land-cover type (though the point estimates are positive). It is important to keep in mind that the standard RSF and an RSF for matched case-control data are subtly different in terms of the interpretation of coefficients (see Sect. 8.2.5.2; Lele et al. 2013) and it is clear in this case that quantitative values differ between approaches, where coefficients tend to be larger with the conditional logit model (Table 8.2).

Table 8.2 Coefficients from RSF models

	Forested wetland		Upland forest	
Model	β (SE)	95% CI	β (SE)	95% CI
Logistic	0.32 (0.20)	−0.08 to 0.68	1.24 (0.49)	0.27–2.19
Conditional logistic	0.72 (0.34)	0.06 to 1.40	2.18 (0.70)	0.81–3.55
Conditional logistic (GEE)	0.71 (0.27)	0.20 to 1.24	2.18 (1.11)	0.00–4.37

8.3.5.3 Path Selection Functions

The previous use of trajectories to implement step selection functions can be readily generalized to the problem of path selection functions. For a path selection function, we contrast resource use on the observed path to that of available paths. To calculate available paths, we simply need to randomly shift the trajectory in a different direction. Random shifts can be accomplished by using the `NMs.randomShiftRotation` function in `adehabitatLT`. This function allows for shifting a trajectory angle or location, without changing the trajectory shape. We illustrate this process for one trajectory, which can then be extended the problem to generating an RSF based on a path selection function similar to that described above.

We will focus on panther 147. We generate one realization (`nrep = 1`) of each path by specifying:

```
> panther147.traj <- panther.ltraj[6]
> path.model <- NMs.randomShiftRotation(panther147.traj, rshift
  = F, rrot = T, nrep = 1)
```

Note, here we set `rshift = F` and `rrot = T`, which means that the trajectories will be randomly rotated along the barycenter of the path (the center of mass of the path), but the trajectory will not be shifted to a different location. An alternative approach is using the `Rotation` function in `spdep` package (Bivand 2006).

The above call sets up the model. To simulate realizations from the model we call:

```
> path.avail <- testNM(path.model)

#reformat list output for plotting:
> path.avail.df <- data.frame(path.avail[[1]])
> path.avail.ltraj <- as.ltraj(xy = path.avail.df[, c("x","y")], date =
  path.avail.df[, "date"], id =
  rep(147,nrow(path.avail.df)))

#plot to compare
> plot(path.avail.ltraj)
> plot(panther.ltraj, id = "147")
```

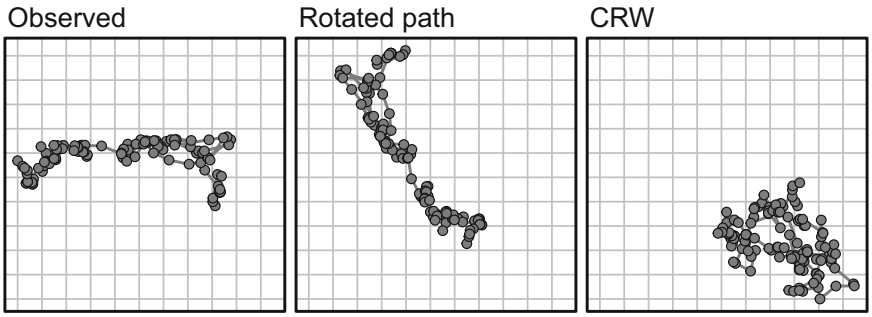


Fig. 8.14 An example showing the approach to path selection. (**a**) The observed trajectory of Cat 147, (**b**) a randomly rotated path to be used as an "available path," and (**c**) a realization of a correlated random walk (CRW) based on the trajectory of Cat 147. Background lines are a 10×10 km grid

We can see from these plots that the trajectory was rotated along approximately the center of the trajectory (Fig. 8.14).

Rather than simply shifting the trajectory, one could also consider correlated random walks (Fig. 8.3) for the entire pathway. In this way, we estimate the parameters of a correlated random walk from observed data and generate trajectories to contrast observed trajectories. We can use the `NMs.randomCRW` function to simulate correlated random walks based on parameters estimated from trajectory data.

```
> CRW.model <- NMs.randomCRW(panther147.traj, rangles = T, rdist
  = T, nrep = 1)
> CRW.avail <- testNM(CRW.model)
```

This object has the same structure as above. We reformat it to plot the CRW trajectory:

```
> CRW.avail.df <- data.frame(CRW.avail[[1]])
> CRW.avail.ltraj <- as.ltraj(xy = CRW.avail.df[,c("x", "y")],
  date = CRW.avail.df[, "date"], id = rep(147, nrow(CRW.avail.df)))

> plot(panther.ltraj, id = "147")
> plot(CRW.avail.ltraj)
```

Comparing plots clearly shows how in this null model, we arrive at a different shaped trajectory, but one that is based on the CRW parameters estimated from the observed trajectory. Note also that in this case, the CRW starts at the origin of the observed trajectory (Fig. 8.14).

Either approach can be repeated for different individuals or "bursts" (i.e., different trajectories that may occur within individuals, such as migratory movements in the spring and fall). With this information, it is straightforward to extract land-cover

data from used and random trajectories and contrast them with conditional logit models, like what we described above for step selection functions (Section 8.3.5.2).

8.4 Next Steps and Advanced Issues

8.4.1 Mechanistic Models and the Identification of Hidden States

There is increasing interest in using "mechanistic" models for interpreting home ranges. Benhamou and Cornelis (2010) extended the Brownian bridge model in a few useful ways, termed *biased random bridge* models. This approach does not assume simple diffusion like the Brownian bridge model but rather incorporates an advection term which causes directionality in the assumed movement process from one location to the next. These models can be implemented with the `BRB` function in the `adehabitatHR` package.

With trajectory data, we may want to classify behavioral states of animals at different times (Gurarie et al. 2016). For many animals, behavioral states can be inferred based on variation in turning angles and step lengths from trajectory data. For instance, foraging states may reflect short step lengths and a wide variance in turning angles, while dispersal or related states are expected to reflect longer step lengths and smaller turning angles. While much of the focus of the identification of hidden states lies in understanding movement trajectories (Patterson et al. 2009), such state identification is also useful for more generally interpreting variation in resource selection and space use.

Several approaches have been developed to interpret hidden states. Gurarie et al. (2016) classified this diversity of approaches into (1) metric-based, (2) classification and segmentation, (3) phenomenological time-series analysis, and (4) mechanistic movement modeling. In metric-based analyses, variation in metrics, such as fractal dimension or first-passage time (Fauchald and Tveraa 2003), are used to qualitatively understand variation in movements. In classification and segmentation, partitioning and/or clustering algorithms are used (Calenge 2006). In phenomenological time-series analysis, changes over time are inferred through temporal autocorrelation functions or through temporal change-point analysis (Gurarie et al. 2009). Finally, mechanistic movement modeling typically uses "hidden Markov" approaches, where random walk models assumed for different latent behavioral states (Jonsen et al. 2005; Morales et al. 2004; Beyer et al. 2013). In general, these types of approaches can be useful if different behavioral states result in very different distributions of step lengths and turning angles (Beyer et al. 2013). However, approaches can be sensitive to misspecification and biases in data collection (Gurarie et al. 2016). Most of these approaches can be implemented in R. For instance, hidden Markov models can be fit with a recent package, `moveHMM` (Michelot et al. 2016) and change point analyses can be fit with `bcpa`.

8.4.2 Biotic Interactions

Resource selection analyses have also been extended in recent years to account for potential biotic interactions, either through conspecific interactions or through heterospecific interactions. Of great interest has been predator–prey resource selection (Hebblewhite et al. 2005). In addition, there is increasing interest on the role of conspecifics in driving movements and resource selection (Fletcher 2006; Campomizzi et al. 2008; McLoughlin et al. 2010). With telemetry data, simultaneous models of potential interactions can be fit when multiple conspecific or heterospecific individuals are in proximity (Delgado et al. 2014; Perez-Barberia et al. 2015). This approach is, however, often limited by available data and in inference regarding impacts of other non-tagged individuals impacting movement and resource selection.

8.4.3 Sampling Error and Resource Selection Models

A common form of error in the use of radiotelemetry data is measurement error (i.e., location error). Location error can be particularly problematic when errors are spatially biased (Frair et al. 2010). For example, location errors can be larger in areas of large elevational relief or in some types of vegetation that may interfere with GPS fixes (Frair et al. 2004).

Location errors can be accounted for formally in a variety of ways. One approach is to use a spatially weighted scheme from which GPS fixes are sampled (Frair et al. 2004). Another common approach is the use state-space models (Patterson et al. 2008; Breed et al. 2012). In state-space models, the observation process and ecological process are defined separately, where the observation process includes information regarding sampling error and/or bias. This information can be based on prior studies or can be estimated directly with the data. State-space models are often fit with Bayesian techniques, but see the `bsam` package in R.

8.5 Conclusions

Habitat and resource selection are fundamental to animal ecology and modeling resource selection is generally relevant to modeling species distributions more broadly (Franklin 2009; Dorazio 2012). Over the past 15 years, there have major advances in resource selection modeling. In general, the incorporation of fine spatiotemporal resolution of animal space use and movement via GPS telemetry and related sensors is allowing new questions in resource selection and space use to be answered (Cagnacci et al. 2010; Kays et al. 2015; Wilmers et al. 2015). These advances have made, and will likely continue to make, important

contributions to the conservation of critical habitats across landscapes and regions (Kautz et al. 2006; Hebblewhite and Haydon 2010; Colchero et al. 2011; Costa et al. 2012; Queiroz et al. 2016).

With these advances, the intersection of movement and resource selection is coming more into focus (Moorcroft and Barnett 2008; Van Moorter et al. 2016). The use of step selection and path selection functions embrace that resources are selected as organisms move through the landscape (or seascape) (Cushman et al. 2011; Thurfjell et al. 2014). We expect the nexus of resource selection and movement will continue to develop, which may help unify these concepts and link them to population dynamics and connectivity (Morales et al. 2010; Vasudev et al. 2015).

References

Aarts G, MacKenzie M, McConnell B, Fedak M, Matthiopoulos J (2008) Estimating space-use and habitat preference from wildlife telemetry data. Ecography 31(1):140–160. https://doi.org/10.1111/j.2007.0906-7590.05236.x

Aarts G, Fieberg J, Matthiopoulos J (2012) Comparative interpretation of count, presence-absence and point methods for species distribution models. Methods Ecol Evol 3(1):177–187. https://doi.org/10.1111/j.2041-210X.2011.00141.x

Aarts G, Fieberg J, Brasseur S, Matthiopoulos J (2013) Quantifying the effect of habitat availability on species distributions. J Anim Ecol 82(6):1135–1145. https://doi.org/10.1111/1365-2656.12061

Abrahams MV (1986) Patch choice under perceptual constraints: a cause for departures from an ideal free distribution. Behav Ecol Sociobiol 19(6):409–415. https://doi.org/10.1007/bf00300543

Aebischer NJ, Robertson PA, Kenward RE (1993) Compositional analysis of habitat use from animal radio-tracking data. Ecol 74(5):1313–1325

Arlt D, Part T (2007) Nonideal breeding habitat selection: a mismatch between preference and fitness. Ecology 88(3):792–801

Arthur SM, Manly BFJ, McDonald LL, Garner GW (1996) Assessing habitat selection when availability changes. Ecology 77(1):215–227

Avgar T, Potts JR, Lewis MA, Boyce MS (2016) Integrated step selection analysis: bridging the gap between resource selection and animal movement. Methods Ecol Evol 7(5):619–630. https://doi.org/10.1111/2041-210x.12528

Battin J (2004) When good animals love bad habitats: ecological traps and the conservation of animal populations. Conserv Biol 18(6):1482–1491

Beauchamp G, Belisle M, Giraldeau LA (1997) Influence of conspecific attraction on the spatial distribution of learning foragers in a patchy habitat. J Anim Ecol 66(5):671–682

Benhamou S, Cornelis D (2010) Incorporating movement behavior and barriers to improve kernel home range space use estimates. J Wildl Manag 74(6):1353–1360. https://doi.org/10.2193/2009-441

Beyer HL, Haydon DT, Morales JM, Frair JL, Hebblewhite M, Mitchell M, Matthiopoulos J (2010) The interpretation of habitat preference metrics under use-availability designs. Phil Trans R Soc B 365(1550):2245–2254. https://doi.org/10.1098/rstb.2010.0083

Beyer HL, Morales JM, Murray D, Fortin MJ (2013) The effectiveness of Bayesian state-space models for estimating behavioural states from movement paths. Methods Ecol Evol 4(5):433–441. https://doi.org/10.1111/2041-210x.12026

Bivand R (2006) Implementing spatial data analysis software tools in R. Geogr Anal 38(1):23–40. https://doi.org/10.1111/j.0016-7363.2005.00672.x

Bock CE, Jones ZF (2004) Avian habitat evaluation: should counting birds count? Front Ecol Environ 2(8):403–410

Borger L, Dalziel BD, Fryxell JM (2008) Are there general mechanisms of animal home range behaviour? A review and prospects for future research. Ecol Lett 11(6):637–650. https://doi.org/10.1111/j.1461-0248.2008.01182.x

Boyce MS, McDonald LL (1999) Relating populations to habitats using resource selection functions. Trends Ecol Evol 14(7):268–272

Breed GA, Costa DP, Jonsen ID, Robinson PW, Mills-Flemming J (2012) State-space methods for more completely capturing behavioral dynamics from animal tracks. Ecol Model 235:49–58. https://doi.org/10.1016/j.ecolmodel.2012.03.021

Burt WH (1943) Territoriality and home range concepts as applied to mammals. J Mammal 24 (3):346–352

Cagnacci F, Boitani L, Powell RA, Boyce MS (2010) Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. Phil Trans R Soc B 365 (1550):2157–2162. https://doi.org/10.1098/rstb.2010.0107

Calenge C (2006) The package "adehabitat" for the R software: a tool for the analysis of space and habitat use by animals. Ecol Model 197(3–4):516–519. https://doi.org/10.1016/j.ecolmodel.2006.03.017

Campomizzi AJ, Butcher JA, Farrell SL, Snelgrove AG, Collier BA, Gutzwiller KJ, Morrison ML, Wilkins RN (2008) Conspecific attraction is a missing component in wildlife habitat modeling. J Wildl Manag 72(1):331–336. https://doi.org/10.2193/2007-204

Codling EA, Plank MJ, Benhamou S (2008) Random walk models in biology. J R Soc Interface 5 (25):813–834. https://doi.org/10.1098/rsif.2008.0014

Colchero F, Conde DA, Manterola C, Chavez C, Rivera A, Ceballos G (2011) Jaguars on the move: modeling movement to mitigate fragmentation from road expansion in the Mayan Forest. Anim Conserv 14(2):158–166. https://doi.org/10.1111/j.1469-1795.2010.00406.x

Cooper AB, Millspaugh JJ (1999) The application of discrete choice models to wildlife resource selection studies. Ecology 80(2):566–575

Costa DP, Breed GA, Robinson PW (2012) New insights into pelagic migrations: implications for ecology and conservation. Annu Rev Ecol Evol Syst 43:73–96. https://doi.org/10.1146/annurev-ecolsys-102710-145045

Cox JJ, Maehr DS, Larkin JL (2006) Florida panther habitat use: new approach to an old problem. J Wildl Manag 70(6):1778–1785. https://doi.org/10.2193/0022-541x(2006)70[1778:fphuna]2.0.co;2

Crowder LB, Cooper WE (1982) Habitat structural complexity and the interaction between bluegills and their prey. Ecology 63(6):1802–1813. https://doi.org/10.2307/1940122

Cushman SA, Raphael MG, Ruggiero LF, Shirk AS, Wasserman TN, O'Doherty EC (2011) Limiting factors and landscape connectivity: the American marten in the Rocky Mountains. Landsc Ecol 26(8):1137–1149. https://doi.org/10.1007/s10980-011-9645-8

Damschen EI, Brudvig LA, Haddad NM, Levey DJ, Orrock JL, Tewksbury JJ (2008) The movement ecology and dynamics of plant communities in fragmented landscapes. Proc Natl Acad Sci U S A 105(49):19078–19083. https://doi.org/10.1073/pnas.0802037105

Delgado MD, Penteriani V, Morales JM, Gurarie E, Ovaskainen O (2014) A statistical framework for inferring the influence of conspecifics on movement behaviour. Methods Ecol Evol 5 (2):183–189

Doherty TS, Driscoll DA (2018) Coupling movement and landscape ecology for animal conservation in production landscapes. Proc R Soc B 285(1870). https://doi.org/10.1098/rspb.2017.2272

Donovan TM, Thompson FR (2001) Modeling the ecological trap hypothesis: a habitat and demographic analysis for migrant songbirds. Ecol Appl 11(3):871–882

Dorazio RM (2012) Predicting the geographic distribution of a species from presence-only data subject to detection errors. Biometrics 68(4):1303–1312. https://doi.org/10.1111/j.1541-0420.2012.01779.x

Duchesne T, Fortin D, Courbin N (2010) Mixed conditional logistic regression for habitat selection studies. J Anim Ecol 79(3):548–555. https://doi.org/10.1111/j.1365-2656.2010.01670.x

Duchesne T, Fortin D, Rivest LP (2015) Equivalence between step selection functions and biased correlated random walks for statistical inference on animal movement. PLoS One 10(4). https://doi.org/10.1371/journal.pone.0122947

Elliot NB, Cushman SA, Macdonald DW, Loveridge AJ (2014) The devil is in the dispersers: predictions of landscape connectivity change with demography. J Appl Ecol 51(5):1169–1178. https://doi.org/10.1111/1365-2664.12282

Epanechnikov VA (1969) Non-parametric estimation of a multivariate probability density. Theory Probab Appl 14(1):153–158

Fagan WF, Lutscher F (2006) Average dispersal success: linking home range, dispersal, and metapopulation dynamics to reserve design. Ecol Appl 16(2):820–828. https://doi.org/10.1890/1051-0761(2006)016[0820:adslhr]2.0.co;2

Fauchald P, Tveraa T (2003) Using first-passage time in the analysis of area-restricted search and habitat selection. Ecology 84(2):282–288. https://doi.org/10.1890/0012-9658(2003)084[0282:Ufptit]2.0.Co;2

Fieberg J, Rieger RH, Zicus MC, Schildcrout JS (2009) Regression modelling of correlated data in ecology: subject-specific and population averaged response patterns. J Appl Ecol 46 (5):1018–1025. https://doi.org/10.1111/j.1365-2664.2009.01692.x

Fithian W, Hastie T (2013) Finite-sample equivalence in statistical models for presence-only data. Ann Appl Stat 7(4):1917–1939. https://doi.org/10.1214/13-aoas667

Fletcher RJ Jr (2006) Emergent properties of conspecific attraction in fragmented landscapes. Am Nat 168(2):207–219

Fletcher RJ Jr, Orrock JL, Robertson BA (2012) How the type of anthropogenic change alters the consequences of ecological traps. Proc R Soc B 279(1738):2546–2552. https://doi.org/10.1098/rspb.2012.0139

Fletcher RJ Jr, Maxwell CW Jr, Andrews JE, Helmey-Hartman WL (2013) Signal detection theory clarifies the concept of perceptual range and its relevance to landscape connectivity. Landsc Ecol 28(1):57–67. https://doi.org/10.1007/s10980-012-9812-6

Forester JD, Im HK, Rathouz PJ (2009) Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. Ecology 90(12):3554–3565. https://doi.org/10.1890/08-0874.1

Fortin D, Beyer HL, Boyce MS, Smith DW, Duchesne T, Mao JS (2005) Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park. Ecology 86 (5):1320–1330. https://doi.org/10.1890/04-0953

Frair JL, Nielsen SE, Merrill EH, Lele SR, Boyce MS, Munro RHM, Stenhouse GB, Beyer HL (2004) Removing GPS collar bias in habitat selection studies. J Appl Ecol 41(2):201–212. https://doi.org/10.1111/j.0021-8901.2004.00902.x

Frair JL, Fieberg J, Hebblewhite M, Cagnacci F, DeCesare NJ, Pedrotti L (2010) Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. Phil Trans R Soc B 365(1550):2187–2200. https://doi.org/10.1098/rstb.2010.0084

Frakes RA, Belden RC, Wood BE, James FE (2015) Landscape analysis of adult Florida panther habitat. PLoS One 10(7). https://doi.org/10.1371/journal.pone.0133044

Franklin J (2009) Mapping species distributions: spatial inference and prediction. Cambridge University Press, Cambridge, UK

Fretwell SD, Lucas HL Jr (1970) On territorial behavior and other factors influencing habitat distribution in birds. I. Theoretical development. Acta Biotheor 19:16–36

Gautestad AO (2011) Memory matters: influence from a cognitive map on animal space use. J Theor Biol 287:26–36. https://doi.org/10.1016/j.jtbi.2011.07.010

Gautestad AO, Mysterud I (2005) Intrinsic scaling complexity in animal dispersion and abundance. Am Nat 165(1):44–55. https://doi.org/10.1086/426673

Gautestad AO, Mysterud I (2010) The home range fractal: from random walk to memory-dependent space use. Ecol Complex 7(4):458–470. https://doi.org/10.1016/j.ecocom.2009.11.005

Getz WM, Wilmers CC (2004) A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. Ecography 27(4):489–505. https://doi.org/10.1111/j.0906-7590.2004.03835.x

Getz WM, Fortmann-Roe S, Cross PC, Lyons AJ, Ryan SJ, Wilmers CC (2007) LoCoH: nonparameteric Kernel methods for constructing home ranges and utilization distributions. PLoS One 2(2). https://doi.org/10.1371/journal.pone.0000207

Gitzen RA, Millspaugh JJ, Kernohan BJ (2006) Bandwidth selection for fixed-kernel analysis of animal utilization distributions. J Wildl Manag 70(5):1334–1344. https://doi.org/10.2193/0022-541x(2006)70[1334:bsffao]2.0.co;2

Gurarie E, Andrews RD, Laidre KL (2009) A novel method for identifying behavioural changes in animal movement data. Ecol Lett 12(5):395–408. https://doi.org/10.1111/j.1461-0248.2009.01293.x

Gurarie E, Bracis C, Delgado M, Meckley TD, Kojola I, Wagner CM (2016) What is the animal doing? Tools for exploring behavioural structure in animal movements. J Anim Ecol 85(1):69–84. https://doi.org/10.1111/1365-2656.12379

Hache S, Villard MA, Bayne EM (2013) Experimental evidence for an ideal free distribution in a breeding population of a territorial songbird. Ecology 94(4):861–869. https://doi.org/10.1890/12-1025.1

Hale R, Treml EA, Swearer SE (2015) Evaluating the metapopulation consequences of ecological traps. Proceedings of the Royal Society B: Biological Sciences 282 (1804):20142930–20142930

Hall LS, Krausman PR, Morrison ML (1997) The habitat concept and a plea for standard terminology. Wildl Soc Bull 25(1):173–182

Harrison PM, Gutowsky LFG, Martins EG, Patterson DA, Cooke SJ, Power M (2015) Personality-dependent spatial ecology occurs independently from dispersal in wild burbot (Lota lota). Behav Ecol 26(2):483–492. https://doi.org/10.1093/beheco/aru216

Hebblewhite M, Haydon DT (2010) Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. Phil Trans R Soc B 365(1550):2303–2312. https://doi.org/10.1098/rstb.2010.0087

Hebblewhite M, Merrill EH, McDonald TL (2005) Spatial decomposition of predation risk using resource selection functions: an example in a wolf-elk predator-prey system. Oikos 111(1):101–111. https://doi.org/10.1111/j.0030-1299.2005.13858.x

Hooten MB, Hanks EM, Johnson DS, Alldredge MW (2013) Reconciling resource utilization and resource selection functions. J Anim Ecol 82(6):1146–1154. https://doi.org/10.1111/1365-2656.12080

Hooten MB, Johnson DS, McClintock BT, Morales JM (2017) Animal movement: statistical models for telemetry data. CRC Press, Boca Raton, FL

Horne JS, Garton EO, Krone SM, Lewis JS (2007) Analyzing animal movements using Brownian bridges. Ecology 88(9):2354–2363. https://doi.org/10.1890/06-0957.1

Hugie DM, Grand TC (1998) Movement between patches, unequal competitors and the ideal free distribution. Evol Ecol 12(1):1–19

Johnson DH (1980) The comparison of usage and availability measurements for evaluating resource preference. Ecology 61(1):65–71

Johnson CJ, Nielsen SE, Merrill EH, McDonald TL, Boyce MS (2006) Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. J Wildl Manag 70(2):347–357. https://doi.org/10.2193/0022-541x(2006)70[347:Rsfbou]2.0.Co;2

Johnson DS, Hooten MB, Kuhn CE (2013) Estimating animal resource selection from telemetry data using point process models. J Anim Ecol 82(6):1155–1164. https://doi.org/10.1111/1365-2656.12087

Jonsen ID, Flemming JM, Myers RA (2005) Robust state-space modeling of animal movement data. Ecology 86(11):2874–2880. https://doi.org/10.1890/04-1852

Kareiva P (1982) Experimental and mathematical analyses of herbivore movement: quantifying the influence of plant spacing and quality on foraging discrimination. Ecol Monogr 52(3):261–282. https://doi.org/10.2307/2937331

Kareiva PM, Shigesada N (1983) Analyzing insect movement as a correlated random walk. Oecologia 56(2–3):234–238

Kautz R, Kawula R, Hoctor T, Comiskey J, Jansen D, Jennings D, Kasbohm J, Mazzotti F, McBride R, Richardson L, Root K (2006) How much is enough? Landscape-scale conservation for the Florida panther. Biol Conserv 130(1):118–133. https://doi.org/10.1016/j.biocon.2005.12.007

Kays R, Crofoot MC, Jetz W, Wikelski M (2015) Terrestrial animal tracking as an eye on life and planet. Science 348(6240). https://doi.org/10.1126/science.aaa2478

Keating KA, Cherry S (2004) Use and interpretation of logistic regression in habitat selection studies. J Wildl Manag 68(4):774–789. https://doi.org/10.2193/0022-541x(2004)068[0774:uaiolr]2.0.co;2

Keim JL, DeWitt PD, Lele SR (2011) Predators choose prey over prey habitats: evidence from a lynx-hare system. Ecol Appl 21(4):1011–1016. https://doi.org/10.1890/10-0949.1

Kennedy M, Gray RD (1997) Habitat choice, habitat matching and the effect of travel distance. Behaviour 134:905–920. https://doi.org/10.1163/156853997x00223

Kie JG, Matthiopoulos J, Fieberg J, Powell RA, Cagnacci F, Mitchell MS, Gaillard JM, Moorcroft PR (2010) The home-range concept: are traditional estimators still relevant with modern telemetry technology? Phil Trans R Soc B 365(1550):2221–2231. https://doi.org/10.1098/rstb.2010.0093

Kokko H, Sutherland WJ (2001) Ecological traps in changing environments: ecological and evolutionary consequences of a behaviourally mediated Allee effect. Evol Ecol Res 3(5):537–551

Krivan V, Cressman R, Schneider C (2008) The ideal free distribution: a review and synthesis of the game-theoretic perspective. Theor Popul Biol 73(3):403–425. https://doi.org/10.1016/j.tpb.2007.12.009

Lack D (1933) Habitat selection in birds - with special reference to the effects of afforestation on the Breckland avifauna. J Anim Ecol 2:239–262. https://doi.org/10.2307/961

Land ED, Shindle DB, Kawula RJ, Benson JF, Lotz MA, Onorato DP (2008) Florida panther habitat selection analysis of concurrent GPS and VHF telemetry data. J Wildl Manag 72(3):633–639. https://doi.org/10.3193/2007-136

Lele SR, Keim JL (2006) Weighted distributions and estimation of resource selection probability functions. Ecology 87(12):3021–3028. https://doi.org/10.1890/0012-9658(2006)87[3021:wdaeor]2.0.co;2

Lele SR, Merrill EH, Keim J, Boyce MS (2013) Selection, use, choice and occupancy: clarifying concepts in resource selection studies. J Anim Ecol 82(6):1183–1191. https://doi.org/10.1111/1365-2656.12141

Lyons AJ, Turner WC, Getz WM (2013) Home range plus: a space-time characterization of movement over real landscapes. Movement Ecol 1(2):1–14

Maehr DS, Cox JA (1995) Landscape features and panthers in Florida. Conserv Biol 9(5):1008–1019. https://doi.org/10.1046/j.1523-1739.1995.9051008.x

Manly BFJ, McDonald LL, Thomas DL, McDonald TL, Erickson WP (2002) Resource selection by animals: statistical design and analysis for field studies. Kluwer Academic, Dordrecht, the Netherlands

Marzluff JM, Millspaugh JJ, Hurvitz P, Handcock MS (2004) Relating resources to a probabilistic measure of space use: forest fragments and Steller's Jays. Ecology 85(5):1411–1427. https://doi.org/10.1890/03-0114

McCarthy KP, Fletcher RJ Jr (2015) Does hunting activity for game species have indirect effects on resource selection by the endangered Florida Panther? Anim Conserv 18:138–145

McLoughlin PD, Morris DW, Fortin D, Vander Wal E, Contasti AL (2010) Considering ecological dynamics in resource selection functions. J Anim Ecol 79(1):4–12. https://doi.org/10.1111/j.1365-2656.2009.01613.x

Merkle JA, Fortin D, Morales JM (2014) A memory-based foraging tactic reveals an adaptive mechanism for restricted space use. Ecol Lett 17(8):924–931. https://doi.org/10.1111/ele.12294

Michelot T, Langrock R, Patterson TA (2016) moveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models. Methods Ecol Evol 7 (11):1308–1315. https://doi.org/10.1111/2041-210x.12578

Mitchell MS, Powell RA (2004) A mechanistic home range model for optimal use of spatially distributed resources. Ecol Model 177(1–2):209–232. https://doi.org/10.1016/j.ecolmodel.2004.01.015

Mohr CO (1947) Table of equivalent populations of North American small mammals. Am Midl Nat 37(1):223–249. https://doi.org/10.2307/2421652

Moody AL, Houston AI, McNamara JM (1996) Ideal free distributions under predation risk. Behav Ecol Sociobiol 38(2):131–143. https://doi.org/10.1007/s002650050225

Moorcroft PR, Barnett A (2008) Mechanistic home range models and resource selection analysis: a reconciliation and unification. Ecology 89(4):1112–1119. https://doi.org/10.1890/06-1985.1

Moorcroft PR, Lewis MA, Crabtree RL (1999) Home range analysis using a mechanistic home range model. Ecology 80(5):1656–1665. https://doi.org/10.1890/0012-9658(1999)080[1656:hrauam]2.0.co;2

Moorcroft P, Lewis MA (2006) Mechanistic home range analysis. Princeton Monograph in Population Biology

Moorcroft PR, Lewis MA, Crabtree RL (2006) Mechanistic home range models capture spatial patterns and dynamics of coyote territories in Yellowstone. Proc R Soc B 273 (1594):1651–1659. https://doi.org/10.1098/rspb.2005.3439

Morales JM, Haydon DT, Frair J, Holsiner KE, Fryxell JM (2004) Extracting more out of relocation data: building movement models as mixtures of random walks. Ecology 85(9):2436–2445. https://doi.org/10.1890/03-0269

Morales JM, Moorcroft PR, Matthiopoulos J, Frair JL, Kie JG, Powell RA, Merrill EH, Haydon DT (2010) Building the bridge between animal movement and population dynamics. Phil Trans R Soc B 365(1550):2289–2301. https://doi.org/10.1098/rstb.2010.0082

Morris DW (2003) Toward an ecological synthesis: a case for habitat selection. Oecologia 136 (1):1–13

Morrison ML, Marcot BG, Mannan RW (2006) Wildlife-habitat relationships: concepts and applications. Island Press, Washington, DC

Mueller T, Fagan WF (2008) Search and navigation in dynamic environments - from individual behaviors to population distributions. Oikos 117(5):654–664. https://doi.org/10.1111/j.0030-1299.2008.16291.x

Nabe-Nielsen J, Tougaard J, Teilmann J, Lucke K, Forchhammer MC (2013) How a simple adaptive foraging strategy can lead to emergent home ranges and increased food intake. Oikos 122(9):1307–1316. https://doi.org/10.1111/j.1600-0706.2013.00069.x

Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE (2008) A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci U S A 105 (49):19052–19059. https://doi.org/10.1073/pnas.0800375105

Nice MM (1941) The role of territory in bird life. Am Midl Nat 26(3):441–487

Nocera JJ, Forbes GJ, Giraldeau L-A (2009) Aggregations from using inadvertent social information: a form of ideal habitat selection. Ecography 32:143–152

Northrup JM, Hooten MB, Anderson CR, Wittemyer G (2013) Practical guidance on characterizing availability in resource selection functions under a use-availability design. Ecology 94 (7):1456–1463. https://doi.org/10.1890/12-1688.1

Onorato DP, Criffield M, Lotz M, Cunningham M, McBride R, Leone EH, Bass OL, Hellgren EC (2011) Habitat selection by critically endangered Florida panthers across the diel period:

implications for land management and conservation. Anim Conserv 14(2):196–205. https://doi.org/10.1111/j.1469-1795.2010.00415.x
Orians GH (1969) On evolution of mating systems in birds and mammals. Am Nat 103(934):589. https://doi.org/10.1086/282628
Orians GH, Pearson NE (1979) On the theory of cental place foraging. In: Horn DJ, Mitchell RD, Stairs GR (eds) Analysis of ecological systems. The Ohio State University Press, Columbus, pp 154–177
Orians GH, Wittenberger JF (1991) Spatial and temporal scales in habitat selection. Am Nat 137: S29–S49. https://doi.org/10.1086/285138
Patterson TA, Thomas L, Wilcox C, Ovaskainen O, Matthiopoulos J (2008) State-space models of individual animal movement. Trends Ecol Evol 23(2):87–94. https://doi.org/10.1016/j.tree.2007.10.009
Patterson TA, Basson M, Bravington MV, Gunn JS (2009) Classifying movement behaviour in relation to environmental conditions using hidden Markov models. J Anim Ecol 78 (6):1113–1123. https://doi.org/10.1111/j.1365-2656.2009.01583.x
Perez-Barberia FJ, Small M, Hooper RJ, Aldezabal A, Soriguer-Escofet R, Bakken GS, Gordon IJ (2015) State-space modelling of the drivers of movement behaviour in sympatric species. PLoS One 10(11). https://doi.org/10.1371/journal.pone.0142707
Potts JR, Lewis MA (2014) How do animal territories form and change? Lessons from 20 years of mechanistic modelling. Proc R Soc B 281(1784). https://doi.org/10.1098/rspb.2014.0231
Prima MC, Duchesne T, Fortin D (2017) Robust inference from conditional logistic regression applied to movement and habitat selection analysis. PLoS One 12(1). https://doi.org/10.1371/journal.pone.0169779
Pulliam HR (1988) Sources, sinks, and population regulation. Am Nat 132(5):652–661
Pulliam HR, Danielson BJ (1991) Sources, sinks, and habitat selection: a landscape perspective on population dynamics. Am Nat 137:S50–S66
Queiroz N, Humphries NE, Mucientes G, Hammerschlag N, Lima FP, Scales KL, Miller PI, Sousa LL, Seabra R, Sims DW (2016) Ocean-wide tracking of pelagic sharks reveals extent of overlap with longline fishing hotspots. Proc Natl Acad Sci U S A 113(6):1582–1587. https://doi.org/10.1073/pnas.1510090113
Reeve JD, Cronin JT, Haynes KJ (2008) Diffusion models for animals in complex landscapes: incorporating heterogeneity among substrates, individuals and edge behaviours. J Anim Ecol 77 (5):898–904. https://doi.org/10.1111/j.1365-2656.2008.01411.x
Resetarits WJ (2005) Habitat selection behaviour links local and regional scales in aquatic systems. Ecol Lett 8(5):480–486
Rettie WJ, Messier F (2000) Hierarchical habitat selection by woodland caribou: its relationship to limiting factors. Ecography 23(4):466–478. https://doi.org/10.1034/j.1600-0587.2000.230409.x
Riotte-Lambert L, Benhamou S, Chamaille-Jammes S (2015) How memory-based movement leads to nonterritorial spatial segregation. Am Nat 185(4):E103–E116. https://doi.org/10.1086/680009
Robertson BA, Hutto RL (2006) A framework for understanding ecological traps and an evaluation of existing evidence. Ecology 87(5):1075–1085
Rodenhouse NL, Sherry TW, Holmes RT (1997) Site-dependent regulation of population size: a new synthesis. Ecology 78(7):2025–2042
Rosenzweig ML (1981) A theory of habitat selection. Ecology 62:327–335
Rota CT, Millspaugh JJ, Kesler DC, Lehman CP, Rumble MA, Jachowski CMB (2013) A re-evaluation of a case-control model with contaminated controls for resource selection studies. J Anim Ecol 82 (6):1165–1173. https://doi.org/10.1111/1365-2656.12092
Rota CT, Rumble MA, Millspaugh JJ, Lehman CP, Kesler DC (2014) Space-use and habitat associations of Black-backed Woodpeckers (Picoides arcticus) occupying recently disturbed forests in the Black Hills, South Dakota. For Ecol Manag 313:161–168. https://doi.org/10.1016/j.foreco.2013.10.048

Savage RE (1931) The relation between the feeding of the herring off the east coast of England and teh plankton of the surrounding waters. Fishery Investigation, Ministry of Agriculture, Food and Fisheries, Series 2, 12:1–88

Schick RS, Loarie SR, Colchero F, Best BD, Boustany A, Conde DA, Halpin PN, Joppa LN, McClellan CM, Clark JS (2008) Understanding movement data and movement processes: current and emerging directions. Ecol Lett 11(12):1338–1350. https://doi.org/10.1111/j.1461-0248.2008.01249.x

Schlaepfer MA, Runge MC, Sherman PW (2002) Ecological and evolutionary traps. Trends Ecol Evol 17(10):474–480

Seaman DE, Powell RA (1996) An evaluation of the accuracy of kernel density estimators for home range analysis. Ecology 77(7):2075–2085. https://doi.org/10.2307/2265701

Sergio F, Newton I (2003) Occupancy as a measure of territory quality. J Anim Ecol 72(5):857–865

Shochat E, Patten MA, Morris DW, Reinking DL, Wolfe DH, Sherrod SK (2005) Ecological traps in isodars: effects of tallgrass prairie management on bird nest success. Oikos 111(1):159–169

Skalski GT, Gilliam JF (2003) A diffusion-based theory of organism dispersal in heterogeneous populations. Am Nat 161(3):441–458. https://doi.org/10.1086/367592

Skellam JG (1951) Random dispersal in theoretical populations. Biometrika 28:196–218

Smouse PE, Focardi S, Moorcroft PR, Kie JG, Forester JD, Morales JM (2010) Stochastic modelling of animal movement. Phil Trans R Soc B 365(1550):2201–2211. https://doi.org/10.1098/rstb.2010.0078

Stamps J (1995) Motor learning and the value of familiar space. Am Nat 146(1):41–58. https://doi.org/10.1086/285786

Stephens PA, Sutherland WJ (1999) Consequences of the Allee effect for behaviour, ecology and conservation. Trends Ecol Evol 14(10):401–405

Svärdson G (1949) Competition and habitat selection in birds. Oikos 1:157–174

Swihart RK, Slade NA (1985) Testing for independence of observations in animal movements. Ecology 66(4):1176–1184. https://doi.org/10.2307/1939170

Taylor MFJ, Suckling KF, Rachlinski JJ (2005) The effectiveness of the endangered species act: a quantitative analysis. Bioscience 55(4):360–367. https://doi.org/10.1641/0006-3568(2005)055[0360:teotes]2.0.co;2

Thomas DL, Taylor EJ (1990) Study designs and tests for comparing resource use and availability. J Wildl Manag 54(2):322–330. https://doi.org/10.2307/3809050

Thomas DL, Taylor EJ (2006) Study designs and tests for comparing resource use and availability II. J Wildl Manag 70(2):324–336. https://doi.org/10.2193/0022-541x(2006)70[324:sdatfc]2.0.co;2

Thomas DL, Johnson D, Griffith B (2006) A Bayesian random effects discrete-choice model for resource selection: population-level selection inference. J Wildl Manag 70(2):404–412

Thompson CM, McGarigal K (2002) The influence of research scale on bald eagle habitat selection along the lower Hudson River, New York (USA). Landsc Ecol 17(6):569–586. https://doi.org/10.1023/a:1021501231182

Thurfjell H, Ciuti S, Boyce MS (2014) Applications of step-selection functions in ecology and conservation. Movement Ecol 2:4

Tomkiewicz SM, Fuller MR, Kie JG, Bates KK (2010) Global positioning system and associated technologies in animal behaviour and ecological research. Phil Trans R Soc B 365 (1550):2163–2176. https://doi.org/10.1098/rstb.2010.0090

Turlure C, Schtickzelle N, Van Dyck H, Seymoure B, Rutowski R (2016) Flight morphology, compound eye structure and dispersal in the bog and the cranberry fritillary butterflies: an inter- and intraspecific comparison. PLoS One 11(6). https://doi.org/10.1371/journal.pone.0158073

Turner JC, Douglas CL, Hallam CR, Krausman PR, Ramey RR (2004) Determination of critical habitat for the endangered Nelson's bighorn sheep in southern California. Wildl Soc Bull 32 (2):427–448. https://doi.org/10.2193/0091-7648(2004)32[427:dochft]2.0.co;2

Underwood AJ, Chapman MG, Crowe TP (2004) Identifying and understanding ecological preferences for habitat or prey. J Exp Mar Biol Ecol 300(1–2):161–187. https://doi.org/10.1016/j.jembe.2003.12.006

Van Horne B (1983) Density as a misleading indicator of habitat quality. J Wildl Manag 47:893–901

Van Moorter B, Rolandsen CM, Basille M, Gaillard JM (2016) Movement is the glue connecting home ranges and habitat selection. J Anim Ecol 85(1):21–31. https://doi.org/10.1111/1365-2656.12394

Van Winkle W (1975) Comparison of several probabilistic home-range models. J Wildl Manag 39:118–123

Vasudev D, Fletcher RJ Jr, Goswami VR, Krishnadas M (2015) From dispersal constraints to landscape connectivity: lessons from species distribution modeling. Ecography 38:967–978. https://doi.org/10.1111/ecog.01306

Wickham H (2007) Reshaping data with the reshape package. J Stat Softw 21(12):1–20

Wilmers CC, Nickel B, Bryce CM, Smith JA, Wheat RE, Yovovich V (2015) The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. Ecol 96 (7):1741–1753. https://doi.org/10.1890/14-1401.1

Worton BJ (1987) A review of models of home range for animal movement. Ecol Model 38 (3–4):277–298. https://doi.org/10.1016/0304-3800(87)90101-3

Worton BJ (1989) Kernel methods for estimating the utilization distribution in home-range studies. Ecology 70(1):164–168. https://doi.org/10.2307/1938423

Worton BJ (1995) Using Monte-Carlo simulation to evaluate kernel-based home-range estimators. J Wildl Manag 59(4):794–800. https://doi.org/10.2307/3801959

Zeller KA, McGarigal K, Whiteley AR (2012) Estimating landscape resistance to movement: a review. Landsc Ecol 27(6):777–797. https://doi.org/10.1007/s10980-012-9737-0

Zeller KA, McGarigal K, Cushman SA, Beier P, Vickers TW, Boyce WM (2016) Using step and path selection functions for estimating resistance to movement: pumas as a case study. Landsc Ecol 31(6):1319–1335. https://doi.org/10.1007/s10980-015-0301-6

Zera AJ, Denno RF (1997) Physiology and ecology of dispersal polymorphism in insects. Annu Rev Entomol 42:207–230. https://doi.org/10.1146/annurev.ento.42.1.207

Zheng C, Pennanen J, Ovaskainen O (2009) Modelling dispersal with diffusion and habitat selection: analytical results for highly fragmented landscapes. Ecol Model 220 (12):1495–1505. https://doi.org/10.1016/j.ecolmodel.2009.02.024

Zollner PA, Lima SL (1997) Landscape-level perceptual abilities in white-footed mice: perceptual range and the detection of forested habitat. Oikos 80(1):51–60

Chapter 9
Connectivity



9.1 Introduction

The importance of space for ecology and conservation relies on the importance of connectivity. It is well known that connectivity can influence populations and communities through a variety of mechanisms, including demographic rescue, inbreeding avoidance, colonization of unoccupied habitat, mass effects, and the spread of disease (Hanski 1998; Chisholm et al. 2011; Rudnick et al. 2012). Consequently, knowledge of connectivity can enhance our understanding of species current and potential distribution patterns, population demography, genetic variability, evolutionary processes, and overall viability of species in heterogeneous landscapes, as well as provide insights into the dynamics of metacommunities (Leibold et al. 2004; Carrara et al. 2012). It is also critical for the maintenance of key ecosystem processes and services (Margosian et al. 2009; Mitchell et al. 2013). Connectivity is increasingly relevant for conservation aimed at ameliorating negative impacts of human-induced environmental change on long-term species persistence and biodiversity (Crooks and Sanjayan 2006; Heller and Zavaleta 2009). As a consequence, there has been an explosion our understanding and quantification of connectivity over the past 20 years.

Here, we provide an overview on the concept of connectivity and its relevance for applied ecology. We first outline the various interpretations regarding connectivity and theoretical developments that emphasize its importance. Connectivity has been conceptualized from a variety of perspectives; here we aim to show the similarities and differences of these perspectives. We then describe three general approaches to quantifying connectivity. We illustrate how these approaches are implemented through two examples on endangered species.

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_9

9.2 Key Concepts and Approaches

9.2.1 The Multiple Meanings of Connectivity

The term "connectivity" has been used loosely in ecology, evolution, and conservation. Despite this loose usage, the concept of connectivity consistently emphasizes two issues: the structure of the landscape (or seascape) and the movement or flow of organisms, matter, or energy. A distinction is frequently made in connectivity research regarding whether connectivity should capture structural or functional concepts. Structural concepts emphasize landscape configuration and contiguity only, implicitly assuming that physical proximity is the key issue that underlies connectivity. Structural concepts do not attempt to capture species- or process-specific variability. Functional concepts for connectivity explicitly attempt to capture the movement or flow process and integrate this process with the structure of the landscape to interpret and quantify connectivity (Fig. 9.1).

The terms for connectivity can be organized in at least three ways: (1) the way in which it is measured (e.g., structural versus functional connectivity), (2) the scale(s) at which connectivity is measured; and (3) the process that is being captured. Calabrese and Fagan (2004) classified connectivity metrics into three general categories based on the way connectivity is measured: structural, potential, and actual connectivity (Fig. 9.1). *Structural connectivity* focuses on simply quantifying

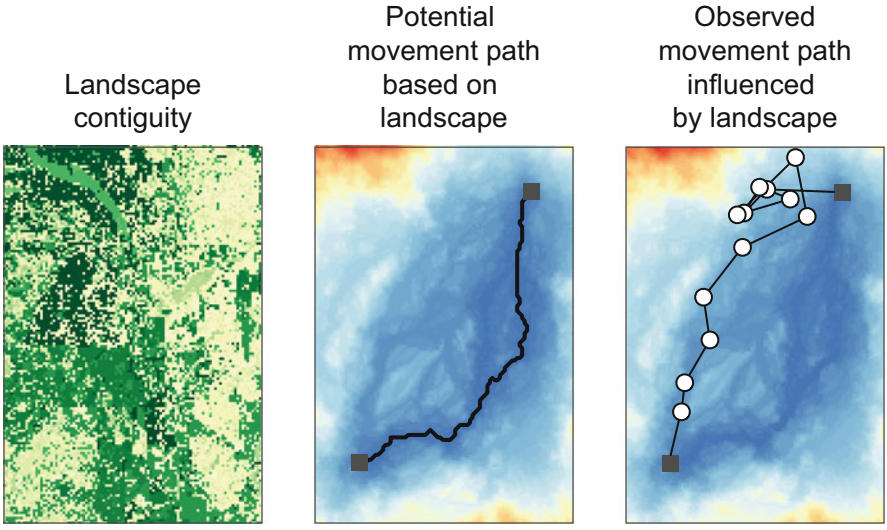


Fig. 9.1 Hierarchical relationships regarding landscape connectivity. Landscape structure can describe structural connectivity, while potential connectivity occurs when landscape structure is linked to movement capacity (e.g., motion capacity, navigation capacity) of species or other processes. Realized, or actual, connectivity quantifies observed movements across landscapes. Such movements can impact a variety of biological patterns and processes (e.g., genetic connectivity, demographic connectivity)

landscape contiguity to interpret connectivity. *Potential connectivity* is a type of functional connectivity that quantifies how individuals could move across landscapes and how the landscape may alter movement. It typically uses indirect, auxiliary information regarding species- or process-specific movements and flow. Actual (i.e., observed) or *realized connectivity* is another type of functional connectivity that measures movement and flow to directly quantify or estimate landscape connectivity (Table 9.1).

Connectivity can also be measured at different scales. Metapopulation biology has long focused on patch isolation (or its inverse, patch connectivity), while landscape ecology has frequently quantified connectivity for entire landscapes (Tischendorf and Fahrig 2000; Moilanen and Hanski 2001). Meso-scales—above the patch and below the entire landscape—have also been considered (Bodin and Noberg 2007; Fletcher et al. 2013b). Patch connectivity is frequently of interest when conservation practices are aimed toward local or site level actions (Acevedo et al. 2015; Rubio et al. 2015). Meso-scale connectivity can be important in cases where movements between nearby patches are relatively common (e.g., gap-crossing; Bélisle and Desrochers 2002; Richard and Armstrong 2010), such

Table 9.1 Different types of connectivity

Term	Definition
Demographic connectivity	The degree to which population growth is affected by dispersal across the landscape. Most sensitive to the relative contribution of dispersal to local recruitment.
Effective connectivity	Connectivity that includes the successful reproduction of individuals post-movement. A landscape-scale extension of effective dispersal.
Functional connectivity	Connectivity that incorporates information on observed or potential movement properties of an organism.
Genetic connectivity	Linkages between local populations based on gene flow. Requires effective dispersal and is one form of population and effective connectivity. Most sensitive to the absolute number of dispersers.
Hydrologic connectivity	Water-mediated transfer of matter, energy, and/or organisms across the landscape.
Landscape connectivity	Most frequently defined as the degree to which the landscape influences movement.
Migratory connectivity	The movement of individuals annually or seasonally between locations (e.g., between breeding and non-breeding areas).
Population connectivity	Linkages between local populations through dispersal across a landscape.
Potential connectivity	A type of functional connectivity that focuses on where individuals could move and how the landscape may alter movement. It may differ from realized or actual connectivity for a variety of behavioral (e.g., natal preferences) and population (e.g., density dependence) processes.
Realized connectivity	A type of functional connectivity that focuses directly on the extent of movement across landscapes and how the landscape alters observed movement. Sometimes referred to as "actual connectivity."
Structural connectivity	Connectivity that only considers habitat contiguity in its interpretation.

that connectivity is more limited at scales beyond the patch level. Landscape-scale connectivity may be most relevant for regional and continental conservation planning (Minor and Lookingbill 2010).

In addition to these aspects of quantifying pattern, connectivity terms also vary based on the process of interest. Demographic connectivity, genetic connectivity, migratory connectivity, and hydrologic connectivity are just a few common examples of this use (Table 9.1) (Pringle 2001; Webster et al. 2002; Lowe and Allendorf 2010). With this usage, the component of movement or flow being considered should be relevant to the overarching process of interest. For example, Lowe and Allendorf (2010) argued that genetics is frequently used to make inferences on *demographic connectivity*, or population connectivity that focuses on the relative contribution of dispersal to population growth occurring on ecological time scales (Kool et al. 2013), but that genetic data have little information to provide regarding the effects of movement on population growth. Movement for genetic connectivity emphasizes that infrequent movements can be sufficient for genetic mixing, even though such movements may play little role in population growth rates on ecological time scales.

9.2.2 The Connectivity Concept

Over the past several decades, there has been considerable theoretical development to understand and predict connectivity. Here, we briefly summarize these developments across different levels of organization (individuals, populations, communities), focusing on two issues. First, how is the space (or the landscape/seascape) considered? Second, what is the role of movement and how is movement predicted to influence biological patterns and processes across space?

At the individual level, developments have primarily occurred in the context of applying foraging theory and related behavioral ecology theory (e.g., information theory) to landscapes (Ims 1995; Bélisle 2005; Fletcher et al. 2013a). This theory often focuses on proximate, short-term movement responses of individuals to landscape structure, such as the scale at which individuals perceive habitat and the role of different types of decision-making on dispersal and/or searching behavior through landscapes, with an emphasis on the subsequent effects of these decisions on individual fitness (Zollner and Lima 1999; Fletcher 2006; Pe'er and Kramer-Schadt 2008). Early aspects of foraging theory were concerned with travel time between patches to explain expected residency time (Charnov 1976), which translated to focus simply on distances between resource patches. Since that time, theoretical development has used spatially explicit, individual-based simulations to capture aggregation of habitat and other structural issues of connectivity to interpret individual dispersal success (e.g., Tyler and Hargrove 1997; Fletcher 2006; Pe'er and Kramer-Schadt 2008). This theoretical development has suggested that several biotic and abiotic factors may alter realized connectivity, such as local habitat quality

leading to little movement in landscapes of high structural connectivity (Bélisle 2005).

At the population level, much theoretical development has occurred in both metapopulation ecology and population genetics. We briefly outline each of these domains (see Chap. 10 for more information). Early metapopulation theory predicted that distance among patches influences colonization rates of unoccupied habitats and can also influence rescue effects (Hanski 1998). More recent metapopulation theory has incorporated other aspects of landscape structure on colonization dynamics, such as patch aggregation (Hiebeler 2000), matrix effects (Moilanen and Hanski 1998), disturbance (Johst and Drechsler 2003; Kallimanis et al. 2005), asymmetric resistance (Vuilleumier et al. 2010), and succession (Verheyen et al. 2004). Results from metapopulation theory suggest that connectivity has benefits on population persistence at intermediate levels: when connectivity is too low, frequent local extinctions occur that are not balanced with colonization, yet when connectivity is too high, population synchrony results, which can increase vulnerability to global (metapopulation) extinction (Heino et al. 1997; Matter 2001). Related source–sink theory incorporates the effects of connectivity through variation in immigration and emigration rates (Pulliam 1988; Thomas and Kunin 1999). In this theory, landscape structure is frequently emphasized simply as the proportion of source and sink habitats on the landscape (Pulliam and Danielson 1991; Runge et al. 2006). In both of these sets of theoretical development, regional population size is emphasized, either indirectly via variation in patch size (assuming larger patches harbor more individuals that serve as propagules for dispersal), or directly through estimates of population abundance. Landscape ecology theory focusing on populations has also emphasized the role of the matrix in terms of disperser mortality and movements near patch boundaries (Fahrig 1998; Bender and Fahrig 2005).

For population genetics, early theory incorporated migration, or the extent to which a local population's alleles are replaced by immigrant alleles, as a probabilistic process between subpopulations. Propagule movement is a critical component of migration, with effective dispersal (i.e., dispersal followed by successful post-dispersal reproduction) being required (Pfluger and Balkenhol 2014). Geographic distance and some aspects of habitat configuration were incorporated early on in this theoretical development, such as stepping-stone models, which predicted increased genetic homogenization at shorter distances (Wright 1943; Kimura and Weiss 1964). The role of migration has been shown to have direct effects on pairwise genetic distance (genetic divergence), where the genetic distance between populations cannot be large unless the migration rate is very low (Larson et al. 1984). This result emphasizes that small genetic distances can be maintained even at infrequent levels of effective dispersal (Lowe and Allendorf 2010). Landscape genetics theory has recently emphasized the role of the landscape matrix, such as isolation-by-resistance relationships (McRae 2006), and landscape heterogeneity, in regard to isolation-by-environment relationships (Sexton et al. 2014; Wang and Bradburd 2014). There is also increasing emphasis on integrating both population-level and individual-level genetic variation with spatial statistics to better capture complex landscape structure

and isolate the roles of movement on genetic structure, despite a general lack of theoretical development (Guillot et al. 2009).

At the community level, much of the theoretical underpinnings on effects of connectivity stems from island biogeography and metacommunity ecology (see also Chap. 11). In island biogeography, MacArthur and Wilson (1967) identified several aspects of the landscape (island configuration) that can influence immigration rates of species. These include distance to mainland, aggregation of islands, and the presence of corridors and stepping-stones, each of which is highly relevant when interpreting landscape connectivity (Saura et al. 2014). Classic island biogeography theory neglected aspects of the landscape matrix on predictions of connectivity, although more recent theory has attempted to capture such issues (Cook et al. 2002). Less theory has focused specifically on the effects of connectivity for species interactions, with most of this theory being developed in the context of specific types of interactions (e.g., Roy et al. 2004).

These theoretical developments that focus on different levels of organization shares common themes, yet connectivity is often interpreted differently in terms of the role of movement and the way the landscape is considered. Movement has been captured in different ways, with some developments requiring that movements result in individuals recruiting into a new breeding population (e.g., population genetics theory), what has been termed *effective connectivity* (Robertson et al. 2018). Furthermore, movement is frequently interpreted at different temporal and spatial scales (Lowe and Allendorf 2010). The effects of landscape structure on this diverse theoretical development is also variable, in terms of the complexity of landscape structure being captured and the scales at which connectivity is interpreted (Tischendorf and Fahrig 2000; Moilanen and Hanski 2001). While recent theory tends to consistently emphasize the role of the matrix in terms of matrix *resistance* (Table 9.2), some matrix-related issues such as the role of hard barriers have not been consistently treated among theoretical developments. For instance, physical and ecological barriers that generate population structure are often emphasized in genetics, yet in population ecology such hard barriers are less frequently emphasized in understanding of connectivity (but see McRae et al. 2012; Fletcher et al. 2013b).

9.2.3 Factors Limiting Connectivity

Traditionally, connectivity assessments have emphasized the role of space as the primary limitation of observed connectivity, such as the focus on isolation-by-distance relationships in population genetics. In recent years, there has been growing awareness that other factors can limit observed connectivity. Much of the work on the role of the matrix impacting movement rates fall into this realm of limiting factors for connectivity.

Vasudev et al. (2015) borrowed from niche theory to organize factors influencing observed connectivity into three general categories of constraints: spatial, environmental, and intrinsic constraints (Fig. 9.2). Spatial constraints focus on the role of

Table 9.2 Common terms and definitions used in connectivity

Term	Description
Conductance	A measure of permeability of a location to movement. Typically quantified as the inverse of resistance (1/resistance).
Corridor	A relatively linear feature of habitat (frequently assumed to be low quality) that structurally connects habitats across a landscape.
Cost layer	Spatial data (e.g., raster data) that describe the difficulty posed (inverse of permeability) by each cell in the landscape to movement.
Circuit theory	Models that apply the concept of random walks to networks (electrical circuits). Used for connectivity mapping.
Dispersal traps	Locations across landscapes that capture situations where there are not formidable spatial or intrinsic constraints to movement, yet environment factors make the locations have high risk of mortality.
Effective distance	A measure of distance that accounts for non-geographic (environmental) factors.
Least-cost analysis	A type of connectivity analysis where potential paths are identified based on the minimum cumulative cost to potential movement.
Graph theory	Models that apply algorithms for relational data to interpret connectivity. Typically applied to patch networks, where nodes are considered patches, and links represent actual or potential movement.
Habitat availability	Considering connectivity based on reachable habitat, where the patch itself is a space where connectivity occurs, in addition to connections between patches.
Linkage	Locations on the landscape that facilitate movements between habitats or patches. Corridors are one type of linkage. Linkages are also sometimes referred to in a population context as describing dispersal or flow between patches (i.e., population linkage).
Network	A collection of patches across a landscape that is linked by movement or flow of materials, energy, or organisms.
Pinch-point	A location where a bottleneck in potential or actual connectivity occurs, such that movement may be funneled through the location.
Resistance	A measure of the relative barrier of a location to movement. Typically considered the inverse of conductance (or permeability).
Undervalued dispersal routes	Locations across the landscape that are could be used for dispersal and promote connectivity based on the spatial location and environmental attributes, yet are rarely used due to intrinsic constraints to dispersal.

physical distance among locations as a limiting factor. Environmental constraints focus on how environmental variation can limit connectivity in a variety of ways, including the role of abiotic and biotic factors extrinsic to the organism that can alter movement. Intrinsic constraints are organismal constraints that can limit connectivity at different levels of biological organization. For example, at the individual level, physiological condition or phenotypic attributes (e.g., sex) may influence the likelihood, distance, and direction of movement (Turlure et al. 2011), thereby impacting realized connectivity (Baguette et al. 2013). At the population level, issues such as density-dependent dispersal or variation in dispersal modes can impact connectivity.

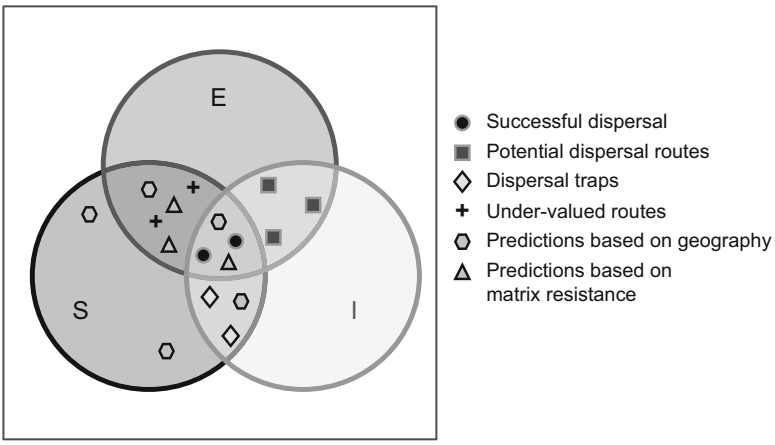


Fig. 9.2 Visualizing limiting factors on successful dispersal and realized connectivity based on spatial (S), environmental (E), and intrinsic (I) constraints. The overlap of these limiting factors in geographic space illustrate dispersal with dispersal traps, undervalued dispersal routes, and potential dispersal routes. Modified from Vasudev et al. (2015)

At the species level, motion capacity and other species-specific traits can impact movement and thus connectivity (Nathan et al. 2008). Some issues, such as the quantification of *effective distance* based on matrix resistance (see below) represent an intersection of these constraints (e.g., space and environment for effective distances). This organization of the constraints of connectivity suggests that connectivity can be limited for a variety of reasons, even for highly mobile species. As a consequence, understanding factors that limit dispersal and connectivity may be necessary for the accurate prediction and conservation of connectivity.

9.2.4 *Three Common Perspectives on Quantifying Connectivity*

There have been three common approaches to quantifying connectivity in ecology. First, connectivity has been, and continues to be, quantified based primarily on structural features (i.e., structural connectivity) of the landscape and the use of programs such as Fragstats to calculate proximity-related measures (McGarigal et al. 2002) (see Chap. 3). Second, the use of spatially explicit measures of connectivity based on the *resistance* of the landscape (i.e., potential functional connectivity) are increasingly used to map connectivity across landscapes and make inferences on issues such as potential conservation *corridors* (Zeller et al. 2012). Third, the use of patch-based network analysis or *graph theory* measures of both structural and potential connectivity are increasingly common (Urban and Keitt 2001; Fall et al. 2007; Urban et al. 2009; Rayfield et al. 2011). Many of these latter measures can be

thought of as extensions of metapopulation concepts that attempt to better capture aspects of landscape configuration.

9.2.4.1 Structural Connectivity of Land Cover

There is a long history in quantifying connectivity of land cover. These metrics typically focus on distances between patches or cells in the landscape, the area of patches or land cover, and/or their juxtaposition. Consequently, these connectivity metrics are generally thought to be measures of landscape configuration or a combination of landscape configuration and composition (see Chap. 3). Distances between patches (e.g., nearest-neighbor distance) are simple measures of proximity, and, while intuitive, tend to not predict movement and distributions very well (Moilanen and Nieminen 2002; Winfree et al. 2005). Metrics that use multiple distances between patches coupled with their areas (e.g., the proximity index; Chap. 3) can capture more realistic complexity (Gustafson and Parker 1994). Patch area is generally thought to be a useful proxy to incorporate into connectivity assessments for two reasons: (1) the target effect (i.e., larger area patches are more likely to have propagules intercept them, thus influencing immigration/colonization) (Lomolino 1990); and (2) population size/propagule pressure, where larger patches tend to have more individuals or propagules that can be a source of colonists into other areas (Hanski 1999). For the target effect, the emphasis in connectivity metrics lies in the size of the focal patch (more precisely, the circumference), while for propagule pressure, the emphasis lies in the size of patches surrounding the focal patch.

Fahrig (2003) argued that habitat area in the landscape is typically inversely correlated with metrics of patch isolation. As the amount of habitat surrounding a fragment declines, that fragment must be more isolated. Consequently, habitat area may be a useful proxy for connectivity surrounding a patch or fragment, although with habitat area comes several other changes in potential processes (e.g., population size), so it is unclear if this structural measure can isolate the role of connectivity in practice.

9.2.4.2 Landscape (Matrix) Resistance

An alternative approach to quantifying connectivity focuses on mapping potential corridors and/or the permeability of the landscape to movement. In this approach, spatially explicit maps (typically raster layers) are used to quantify potential flow across the landscape (Fig. 9.3). Land-cover and related raster maps are transformed into maps of "resistance" or "friction," where values of the raster maps describe the potential for movement (or lack of movement) through the cell. These resistance values are then used to create a transition matrix, which is a sparse matrix based on local connections between adjacent cells (typically using either a four-neighbor or eight-neighbor rule; Fig. 9.4). Resistance can be determined from a variety of

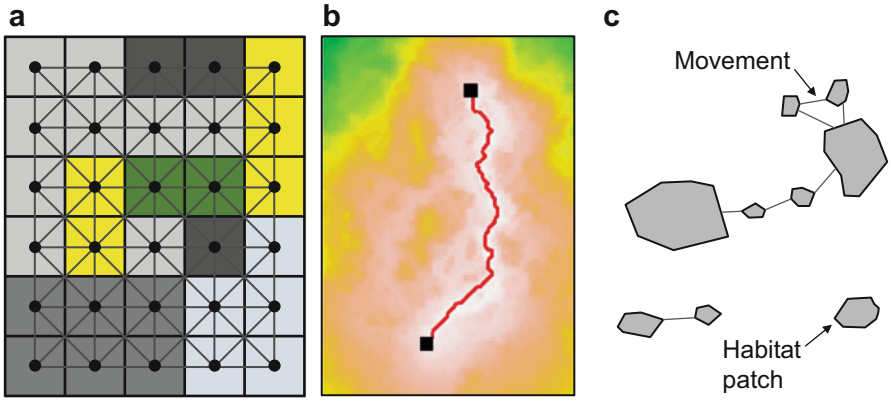


Fig. 9.3 Resistance-based and network-based (graph theory) perspectives for connectivity in ecology. (**a**) Across landscapes, network theory is also used to convert raster maps into sparse networks, where cells (pixels) are linked to neighbors based on "resistance" to movement, such that (**b**) connectivity can then be mapped (red = least-cost path). (**c**) Patch-based networks focus on habitat patches (e.g., forest fragments) and the potential movement of individuals between them

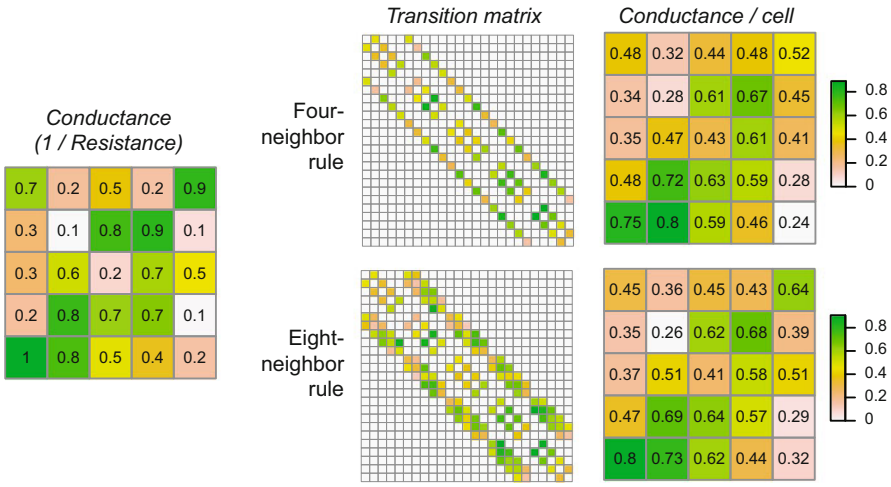


Fig. 9.4 From land-cover permeability to transition matrices. Left panel shows a toy land-cover raster layer where increasing values reflect increasing permeability to movement (conductance = 1 / resistance). Middle panel shows the resulting sparse transition matrices created by taking the mean value between pairs of cells connected with a four-neighbor or eight-neighbor rule (darker values indicate greater conductance values). Right panel illustrates new raster layers that show the average conductance per cell

sources, such as information from expert opinion, data on habitat use, or movement data (Zeller et al. 2012). The determination of resistance values is a critical step in these approaches, as several techniques have been shown to be sensitive to resistance values (Rayfield et al. 2010).

To estimate resistance, information on species distribution, habitat use, or movement can be used. When working with empirical data (rather than expert opinion) on species distribution, the use of point selection and related methods can be used (see Chap. 8). These approaches quantify resistance based on the inverse of metrics of habitat use based on species presence locations. For instance, point selection functions can be used that are essentially selection ratios regarding habitat preference (see Chap. 8). Yet the use of habitat use data for interpreting resistance has been criticized for several reasons. For example, Vasudev et al. (2015) argued that by definition movement through the matrix is outside of a species niche (see Chap. 7), while information on habitat use is typically within the niche of species.

In contrast to information on species distribution and habitat use, movement data can be used to interpret resistance, which are often based on radiotelemetry data or mark–recapture data. For radiotelemetry data, step selection functions, and path selection functions are frequently used (see Chap. 8). Step selection functions focus on contrasting steps from one point to the next (e.g., consecutive radiotelemetry fixes) and how land-cover alters steps. Path selection functions focus on the land-cover used along entire paths and contrasts this to alternative paths of the same length and shape. Both of these approaches can potentially be useful, but ideally should be implemented on individuals and paths where dispersal is occurring rather than focusing on steps and paths of individuals moving within territories or home ranges (Harju et al. 2013). Mark–recapture data (e.g., spatial capture–recapture or multistate capture–recapture; Brownie et al. 1993; Royle et al. 2018) have been used less frequently to estimate resistance, but can be useful in this context by optimizing resistance values based on assumed movement processes (e.g., random walks, least-cost movement) to fit observed movement data between locations (Hanks and Hooten 2013; Royle et al. 2013; Graves et al. 2014; Peterman 2018). In a similar way, Howell et al. (2018) recently showed how resistance values can be estimated from data on colonization–extinction dynamics, which may be more readily available across landscapes than mark–recapture data.

With these resistance maps, potential paths between locations can be identified through several algorithms, such as the use of *least-cost analysis* or the use of *circuit theory* (Bunn et al. 2000; McRae et al. 2008; Etherington 2016). These approaches generally convert resistance maps to transition matrices and use network analysis to interpret potential connectivity.

The application of least-cost paths identifies paths (or corridors) of least resistance between two or more locations. This is frequently accomplished through the use of Dijkstra's algorithm for identifying shortest paths (Dijkstra 1959). This algorithm has been widely used in corridor and connectivity analysis, in part because of the fact that it is computationally quicker to calculate than other algorithms and in part because it can identify "optimal" paths. Nonetheless, the use of least-cost paths has been criticized for several reasons, including the implicit assumptions of goal-oriented search between locations, animals will take the optimal least-cost path regarding potential movement, and its reliance on a single path (Sawyer et al. 2011). The latter concern has been addressed through the use of *least-cost corridor* analysis, an extension of least-cost paths that attempts to identify wider corridors

(rather than paths) of low resistance (Pinto and Keitt 2009). In addition, applications of what has been termed *factorial least-cost path* analysis can address some of these concerns (Rudnick et al. 2012). In that approach, least-cost paths are identified between several points in the landscape (e.g., a grid of point locations) and the least-cost paths are summed for the entire landscape.

The use of *circuit theory* assumes random walks between locations to calculate a resistance distance (McRae 2006). Incorporating variation in resistance into the random walk theory leads to a biased random walk process, such that the average probability of movement is assumed to vary based on resistance. By assuming a random walk, rather than a path of least resistance, flow can be predicted across entire landscapes. This approach has been argued to be useful by accounting for redundancy in paths for connectivity (McRae et al. 2008; Fletcher et al. 2014), and in its ability to identify "pinch-points" or bottlenecks that may arise across landscapes. Some criticisms of this approach include that plants and animals may not move in a way that resembles simple random walks, it is computationally more expensive than least-cost path analysis, and that sometimes results are difficult to interpret across large landscapes because flow tends to be predicted as being diffuse. Note that on very large graphs, it has been argued that the use of random walks, like those in circuit theory, become less helpful because the resistance distance converges on the local properties (i.e., the degree or strength; see below) at a cell level (von Luxburg et al. 2014). This issue has been neglected in the ecological literature but has been repeatedly acknowledged in the network literature.

Saerens et al. (2009) derived theory that formally links the idea of least-cost paths and resistance distances as being along a continuum of movement possibilities. They show that these ideas can be coupled through what they call a *randomized shortest-path* by using a tuning parameter, θ. When $\theta = 0$, the model reduces to a random walk (equivalent to a circuit theory approach). As θ increases, the model approaches a least-cost path. Note that the absolute value of θ will vary depending on the extent of the study region.

9.2.4.3 Patch-Based Graphs

Over the last two decades, graph theory is increasing used for quantifying connectivity (Urban and Keitt 2001; Urban et al. 2009; Galpern et al. 2011; Albert et al. 2017; Drake et al. 2017; Haase et al. 2017; Martensen et al. 2017). Graph theory is a mathematical framework widely used in computer science, operations research and information technology. It deals with the efficient flow of information or connectivity between objects, sometimes referred to relational data. Graph theory and network analysis are related. Traditionally, graph theory has been more of a subject of basic mathematics that focused on proofs regarding relational data, while network analysis has had a more applied focus that emphasizes the use of spectral methods and linear algebra to interpret flow on networks (Strogatz 2001).

The potential benefits of graph theory and related network analysis for spatial ecology is at least fourfold (Proulx et al. 2005; Fall et al. 2007; Urban et al. 2009).

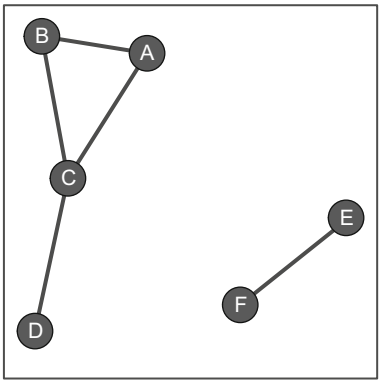
First, these approaches provide a means to formally capture potential indirect linkages that may be important for connectivity, such as that of stepping stones (Saura et al. 2014). Second, these approaches are often useful for quantifying connectivity at different spatial scales. Third, graph theory and network analyses are often helpful for visualizing potential connectivity in space (and time). Finally, such analyses tend to be computationally efficient, allowing for the analysis and visualization of very large numbers of linkages and patches across landscapes.

Conceptually, a graph is simply a collection of points that are interconnected by lines (Dale and Fortin 2010). Points are called *vertices* (or sometimes called *nodes*) and connection between vertices are represented by *edges* (or sometimes called *links* or *arcs*). In the context of connectivity, patches are nodes, links describe information about movement, and the *network* or graph represents the entire collection of patches and the links between them (Fig. 9.2). Consequently, this approach requires patch delineation and focuses on both patch connectivity and summaries of connectivity for the entire landscape based on these patches. Hence, a key step in creating a graph is the quantification of links that reflect movement or flow (see below).

Mathematically, most graphs are represented in matrix form by an *adjacency matrix* **A** (also sometimes referred to as a transition matrix). The indices of this matrix represent the vertices of the graph (Fig. 9.5). For instance, if the graph has seven patches, then **A** will be a 7 × 7 matrix (adjacency matrices are "square" matrices). If i and j represent a row and a column in the adjacency matrix, an element A_{ij} in the matrix **A** will represent the relationship between vertex i and vertex j (Fig. 9.5). Then:

$$A_{ij}\begin{cases} = 1 \text{ if there is a connection between } i \text{ and } j; \\ = 0 \text{ otherwise} \end{cases} \tag{9.1}$$

Note that the diagonal of the adjacency matrix is frequently set to zero; self-links would represent fidelity or philopatry in the context of movement and landscape



$$A = \begin{array}{c} \begin{array}{cccccc} A & B & C & D & E & F \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \end{array}$$

Fig. 9.5 A toy patch-based graph with binary links and the resulting adjacency matrix

connectivity. Element $A_{3,2} = 1$, which means that there is a connection between v_3 and v_2 (vertex C and B in Fig. 9.5, respectively).

In the above example, we focused on one of the simplest types of graphs, which is called an *undirected, unweighted graph*. Undirected means that if there is a connection between v_1 and v_2, there will be necessarily a connection between v_2 and v_1. Unweighted means that links or edges are binary (0,1), while a *weighted graph* can have links that are not binary (integers, continuous values, probabilities). Another way of representing the relationship between vertices is using a *directed graph*. In directed graphs, if there is a connection between say v_1 and v_2 that does not necessarily mean that there is a connection between v_2 and v_1. In these graphs, we typically plot the directional relationship between vertices using arrows. Clearly, weighted, directed graphs have the potential to contain the most information on movement and connectivity. However, there is often interest to simplify adjacency matrices because weighted, directed graphs can be more difficult to visualize, information on weights is sometimes limited, and it can be computationally more demanding than using simpler unweighted graphs. We emphasize, however, that when information on weights is available, typically this information will provide better inference for connectivity problems.

Frequently, we make assumptions regarding the likelihood that patches are potentially connected, rather than having observed data on movements (Fletcher et al. 2016). One common way to define a potential connectivity is to use the maximum known dispersal distance, h, for the species (Urban and Keitt 2001). Then if the distance between patches is less than h then there is a potential connection between the two patches. A second way is to use dispersal kernel information to parameterize links. In this situation, distance (or effective distance, such as least-cost distance) between patches is used to quantify the probability of potential movement. For example, in metapopulation biology, a negative exponential kernel is frequently considered. In this case:

$$A_{ij} = \exp\left(-\alpha d_{ij}\right), \tag{9.2}$$

where α is the inverse of the mean dispersal distance for the species. Finally, if we have observed dispersal events, we can parameterize $\mathbf{A}$ based on the number of observed dispersal events between i and j (Fletcher et al. 2011).

With this general framework, there have been dozens of metrics used to quantify connectivity at different scales, ranging from the patch scale to the entire landscape (Rayfield et al. 2011). Rayfield et al. (2011) classified these metrics along two dimensions: the scale at which the metric captured connectivity (e.g., patch, landscape) and the connectivity property being captured. The connectivity properties they considered were: route-specific flux, route redundancy, route vulnerability, and connected habitat area. Route-specific flux emphasizes the (relative) amount of movement or flow through a patch. Route redundancy aims to capture the degree to which alternate routes for movement occur between patches. Route vulnerability accounts for the degree to which the landscape funnels potential movement, such as

the occurrence of *pinch-points* or bottlenecks in potential movement. Connected habitat area focuses on all movement pathways to quantify the effective colonizable (or reachable) area from the perspective of the species. Depending on the application (e.g., habitat restoration or conservation planning), different properties of connectivity may be more or less relevant.

Here, we do not attempt to show all these metrics but rather focus on a handful of metrics that have been fruitfully employed and capture connectivity at different spatial scales. Some of these approaches share strong similarities with connectivity metrics based on metapopulation theory (see Chap. 10). In fact, in some of the original applications of graph theory to connectivity, much of the rationale and development came directly from ideas in metapopulation ecology (Urban and Keitt 2001). These approaches extend metapopulation ecology to capture connectivity more readily at different scales, as well as capturing indirect linkages that might be masked with metapopulation approaches (Saura and Rubio 2010).

9.3 Examples in R

We illustrate these common approaches to predicting and mapping connectivity by addressing the problem of movement across landscapes for two endangered species. Our goals for these examples are to show how landscape resistance can be incorporated into connectivity modeling, show the similarity and differences in approaches for mapping connectivity and prioritizing patches for connectivity conservation, and highlight how different types of connectivity metrics can capture properties of connectivity operating at different spatial scales.

9.3.1 Packages in R

In R, there are a few libraries that can be used for connectivity analysis. `gdistance` allows for calculations of least-cost paths and related distance-based metrics (van Etten 2012). For patch-based connectivity, `igraph` (Csardi and Nepusz 2006) and `statnet` (Handcock et al. 2008) provide comprehensive network-based and graph-based metrics. For statistical modeling of connectivity (movement) data on patch-based graphs (Snijders 2011), `ergm` and `latentnet` provide useful, but underused, statistical modeling approaches that are part of the `statnet` platform, and several packages provide (partial) Mantel tests that are commonly used for assessing significance of connectivity with genetics data. `SDMTools` provides some Fragstats-like summaries of structural connectivity (VanDerWal et al. 2010), although note that this package does not include some of the common metrics available in Fragstats (see Chap. 3; McGarigal et al. 2002). Finally, `MetaLandSim` (Mestre et al. 2016) calculates some recently derived patch-based network measures for connectivity (Pascual-Hortal and Saura 2006; Saura and

Pascual-Hortal 2007) that are not available in the more general purpose `igraph` and `statnet` libraries. We focus on the use of `gdistance` and `igraph`.

9.3.2 The Data

We use two data sets to illustrate the diverse ways in which functional connectivity is quantified (see Chap. 3 for quantifying structural connectivity). Our goal is to contrast approaches and inferences for connectivity when we focus on connectivity mapping based on matrix resistance, and connectivity prioritization based on patch-based graphs.

In the first example, we return to the endangered Florida panther considered in Chap. 8. For this chapter, we interpret connectivity and potential corridors among protected lands in south Florida. Florida panthers are critically endangered mammals confined to southern Florida. For the recovery of this species, the US Fish and Wildlife Service requires that populations expand north of their current distribution. This issue has led to great interest in identifying potential corridors for panthers and understanding connectivity for this species (Kautz et al. 2006). Prior work has used radiotelemetry data to interpret landscape resistance based on point and home range selection functions (Kautz et al. 2006). We use this information to quantify potential connectivity between protected areas in south Florida.

The second example focuses on patch-based network approaches through the use of occurrence and mark–recapture data on the endangered snail kite (*Rostrhamus sociabilis plumbeus*). In the USA, snail kites are restricted to patchily distributed wetlands in central and south Florida (Reichert et al. 2015). Long-term mark–resight monitoring has been conducted for this species, providing information on dispersal of individuals among wetlands over time. Consequently, a patch-based graph theory approach is a natural framework for interpreting connectivity among wetlands (nodes) based on observed and potential movement from mark–resight data (Fletcher et al. 2011; Reichert et al. 2016). In the example provided, we use data on within-breeding season movements (March–June) of snail kites, from 1997 to 2013, based on standardized resight surveys approximately 18–21 days apart (Reichert et al. 2016). These movements are assumed to reflect snail kites moving between successive potential breeding locations (i.e., within-season breeding dispersal) (Fletcher et al. 2011).

9.3.3 Functional Connectivity Among Protected Areas for Florida Panthers

For interpreting functional connectivity among protected areas, we use two spatial data layers (Fig. 9.6): a vegetation raster layer developed by the Florida Fish and

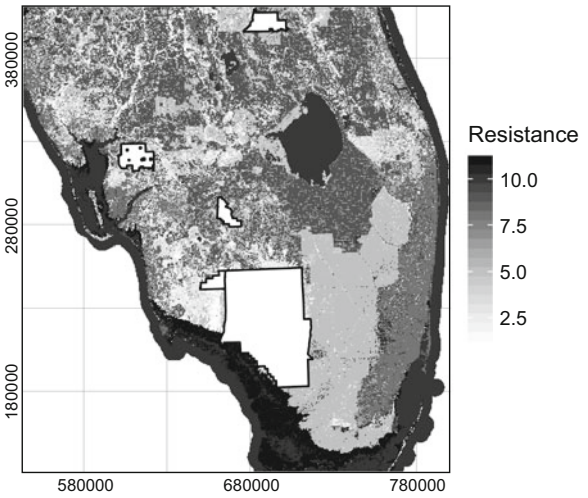


Fig. 9.6 Landscape resistance and the Florida panther. Florida panthers are known to move widely across southern Florida. As part of the recovery plan for this endangered species, dispersal of panther from southern Florida into central and northern Florida is needed. Landscape connectivity analyses have been done based on least-cost paths to address this goal. Map shows a resistance layer for panthers, parameterized from point selection analyses by Kautz et al. (2006). Polygons show five protected areas, where panthers commonly use southern protected areas

Wildlife Conservation Commission, which is largely consistent with the information used in Kautz et al. (2006). This layer was used in Chap. 8, but here we reclassify the layer to interpret land-cover resistance to movement. We also use a vector-based shape file of some protected areas in southern Florida. We begin by importing the data with the `raster` and `rgdal` packages and taking a look at some of the data properties.

```
> library(raster)
> library(rgdal)
> land <- raster("panther_landcover")
> projection(land)

##
[1] "+proj=aea +lat_1=24 +lat_2=31.5 +lat_0=24 +lon_0=-84
+x_0=400000 +y_0=0 +ellps=GRS80 +units=m +no_defs"

> res(land)

##
[1] 500 500
```

```
#public areas in need of connections
> public <- readOGR("panther_publicland.shp")
> proj4string(public)

##
[1] "+proj=aea +lat_1=24 +lat_2=31.5 +lat_0=24 +lon_0=-84
+x_0=400000 +y_0=0 +ellps=GRS80 +units=m +no_defs"
```

In the above, we find that both layers have the same projection and that the raster layer has a 500 m resolution. We can also view the attribute table for the `public` shape file with `public@data`. For connectivity analyses, we focus on identifying connections from the centroids of each protected area. We can calculate centroids with the `gCentroid` function from the `rgeos` package:

```
#get centroids of protected areas
> library(rgeos)
> public_centroids <- gCentroid(public, byid = T)
```

We reclassify the land-cover map to create a cost (or resistance) map. To do so, we import a reclassification table that includes four columns: (1) current land-cover value; (2) current land-cover description; (3) new land-cover value (resistance value); and (4) new land-cover description. This reclassification is taken from Kautz et al. (2006), who used a resource-selection analysis on telemetered panthers to parameterize a *cost layer*. Once imported, we can use the `reclassify` function in the `raster` package to create the cost map.

```
> classification <- read.table("resistance reclass.txt",
  header = T)
> class <- as.matrix(classification[,c(1,3)])
> land_cost <- reclassify(land, rcl = class)
```

9.3.3.1 Effective Distances

With this cost layer, we can calculate distances between protected areas. Several distance metrics have been proposed. We consider four metrics. First, we calculate the Euclidean distance between locations, or the distance "as the crow flies." This distance measure ignores the resistance of the landscape. Second, we calculate the least-cost distance between locations. Least-cost distances are the shortest distances based on the shortest path between locations (i.e., the path that includes the minimum sum of costs). With these paths, either the least-cost distance (cumulative cost) is used or the least-cost path length (i.e., the physical length of the path) is used; Etherington and Holland (2013) argued that least-cost distance is generally more appropriate than least-cost path length. Third, we can calculate an effective distance based on circuit theory. In this approach, random walks are assumed, which acknowledge the potential for alternative paths to alter the effective distance

calculation. Here, we calculate commute distance, which quantifies the expected time for an individual to move from one location to another and back again. This metric is slight different than the more commonly used "resistance distance" (a metric that describes the effective resistance between two locations) in interpretation, but it is tends to be linearly proportional to, and highly correlated with, resistance distance (McRae et al. 2008; Marrotte and Bowman 2017). Finally, we consider a randomized-shortest path algorithm (Saerans et al. 2009; Panzacchi et al. 2016, which can provide results similar to least-cost paths and circuit theory by altering a parameter, θ, in the model. The `gdistance` package can calculate each of these distance metrics. We first need to create a transition layer from which distances will be calculated:

```
> library(gdistance)
#create a 'conductance' transition layer:
> land_cond <- transition(1 / land_cost, transitionFunction = mean, 8)

#make correction for diagonal connections (8-neighbor rule)
> land_cond <- geoCorrection(land_cond, type = "c", multpl = F)
```

What is this transition layer? It converts the nodes (raster cells) into a sparse network of connections (see Fig. 9.4 for a toy example), here based on the means of *conductance* values between pairs of nodes (see `transitionFunction = mean`). Using the mean of conductance values is the most common approach to building transition layers, but such layers could be built with different functions (Etten and Hijmans 2010). For example, variables such as the slope of the environment (Chap. 6) may be important to resistance in some situations, yet this effect likely depends upon the direction an organism is moving (with moving uphill having greater resistance than downhill). Note that we also make a correction for this transition layer to account for diagonal connections; it is recommended that for least-cost analysis, `type = "c"`, whereas for circuit theory analysis, `type = "r"` (Etten and Hijmans 2010). In this example these two types of corrections provide identical results, so we simplify by only correcting with `type = "c"`. Now, we calculate each distance measure and look at how correlated they are:

```
#Euclidean distance
> geo.dist <- pointDistance(public_centroids, lonlat = F)
> geo.dist <- as.dist(geo.dist)

#least-cost distance
> lc.dist <- costDistance(land_cond, public_centroids)

#Commute distance (proportional to resistance distance)
> circuit.dist <- commuteDistance(land_cond, public_centroids)

#randomized shortest path distance
> rSP.dist_t001 <- rSPDistance(land_cond, from =
  public_centroids, to = public_centroids, theta = 0.001)
```

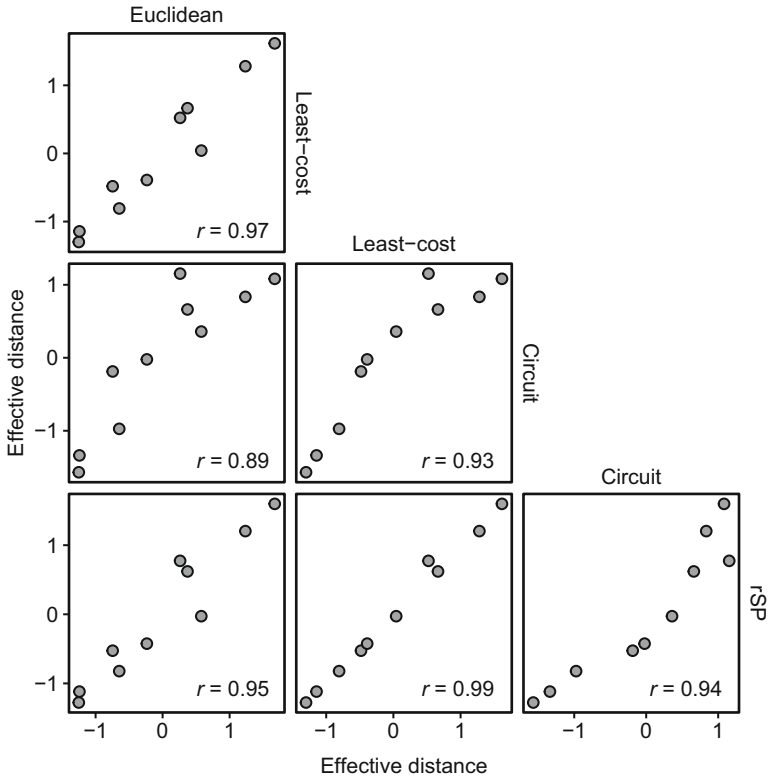


Fig. 9.7 Correlations of effective distances (scaled to mean of 0 variance of 1) between protected areas, calculated based on Euclidean distance, least-cost distance, commute distance (random walk/circuit theory) and randomized shortest paths (with $\theta = 0.001$)

For our calculation of the randomized shortest path, we set $\theta = 0.001$. We arrived at this value after exploring alternative values to identify a model that makes predictions intermediate of least-cost and circuit analysis (see Sect. 9.3.3.4 for more details). In this situation, the effective distance metrics are all highly correlated with Euclidean distance (Fig. 9.7). This correlation is common, but it has been shown that the relationship between these distance metrics can be non-linear (Marrotte and Bowman 2017). Also, note that these distance calculations illustrate the computational demands of using circuit-theoretic approaches: on the authors' computer, the `commuteDistance` and `rSPDistance` functions took *much longer* to complete than did the `costDistance` function (`costDistance` <1 s; `commuteDistance` ~400 s; `rSPDistance` ~747 s).

9.3.3.2 Least-Cost Paths

We now focus in on mapping these potential connectivity metrics between two protected areas: the Florida Panther Wildlife Refuge, a critical area for panther populations, and a nearby protected area, Okaloacoochee Slough State Forest. We can use the above approaches to map potential paths for connectivity for which the effective distances were calculated. First, we look at the least-cost path. To speed computation, we crop our raster layer to focus simply on this pair of protected areas as an illustrative example.

```
#crop maps to speed computation and zoom in for interpretation
> fpwr_ossf_extent <- extent(642000, 683000, 237000, 298000)
> land_sub <- crop(land, fpwr_ossf_extent)
> land_cost_sub <- crop(land_cost, fpwr_ossf_extent)
> land_cond_sub <- transition(1 / land_cost_sub, transitionFunction =
  mean, 8)
> land_cond_sub <- geoCorrection(land_cond_sub, type = "c", multpl = F)

> fpwr_ossf_lcp <- shortestPath(land_cond,
  public_centroids@coords[5,], public_centroids@coords[3,],
  output = "SpatialLines")

> plot(land_cost_sub, axes = F)
> plot(public, add = T)
> points(public_centroids, col = "grey30")
> lines(fpwr_ossf_lcp, col = "red")
```

In this case, the path illustrates that the route for least resistance is not equivalent to the Euclidean distance (Fig. 9.8). Instead, it includes some circuitry due to the resistance of the landscape. Yet this path only identifies a single line segment, the width of the grain of the landscape map. We can remove this potential limitation with the application of least-cost corridors.

9.3.3.3 Least-Cost Corridors

One criticism of least-cost paths is that they focus on a single path, which may not be reliable for connectivity in the long-run. A potential solution is to map a *least-cost corridor*, or the wider area in which there is a low cost between two (or more areas) (Pinto and Keitt 2009). Mapping least-cost corridors is straightforward but requires a few steps. First, we calculate the cumulative costs/resistances from each location, creating two maps (Fig. 9.9a, b).

```
> fpwr.cost <- accCost(land_cond_sub,
  public_centroids@coords[5,])
> ossf.cost <- accCost(land_cond_sub,
  public_centroids@coords[3,])
```

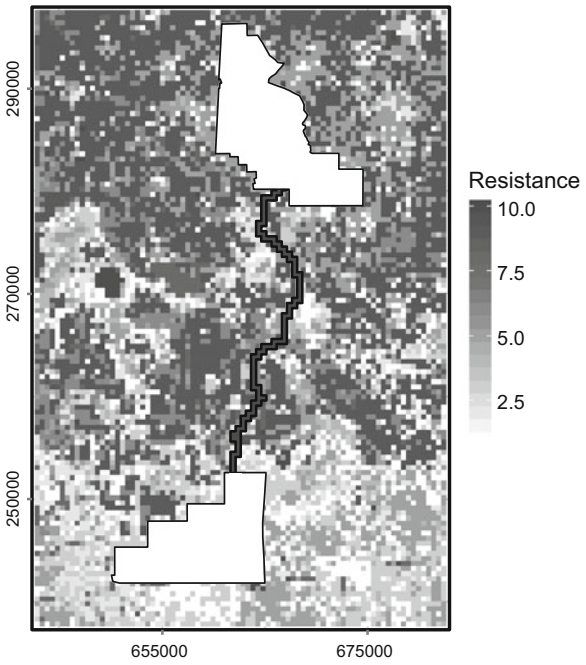


Fig. 9.8 Landscape resistance and the least-cost path

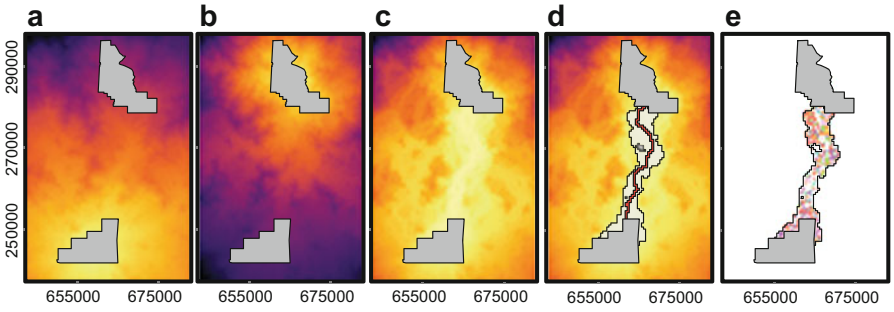


Fig. 9.9 The steps for creating a least-cost corridor. (**a**, **b**) Cumulative resistance from centroids of each protected area. (**c**) The sum of the cumulative resistances. (**d**) The least-cost corridor, based on the lowest 10% quantile of resistance. (**e**) The land cover identified in the least-cost corridor

We then overlay the costs by adding the two layers (Fig. 9.9c) and clip the new layer based on a specific quantile of cost. Here, we clip based on the lowest 10% cost (Fig. 9.9d), but this can be altered to increase or decrease the corridor width (where increasing the quantile will increase the width).

```
> leastcost_corridor <- overlay(fpwr.cost, ossf.cost, fun = function
  (x, y){return (x + y)})

#get lower quantile of accumulated cost
> quantile10 <- quantile(leastcost_corridor, probs = 0.10, na.rm =
  TRUE)

> leastcost_corridor10 <- leastcost_corridor
> values(leastcost_corridor10) <- NA
> leastcost_corridor10[leastcost_corridor < quantile10] <- 1

> plot(leastcost_corridor10,legend = F,axes = F)
> points(public_centroids, col = "grey30")
> lines(fpwr_ossf_lcp, col = "red")
```

Finally, we might be interested in understanding the land-cover (or related attributes, such as ownership) along these predicted paths for conservation planning. We can use the `raster` package to extract this information in a straightforward way from least-cost paths or corridors (Fig. 9.9e).

```
#identify land-cover along the lcp
> lcp.land <- extract(land, fpwr_ossf_lcp)
> table(lcp.land)

##
lcp.land
 7 8 9 10 12 15 17 18 19 20 27 30 31 32 35 41
 1 11 19 6 19 4 16 1 17 14 1 5 5 2 2 1

#identify land-cover along the least-cost corridor
> corridor.land <- crop(land, leastcost_corridor10)
> corridor.land <- mask(corridor.land, leastcost_corridor10)
> plot(corridor.land, axes = F, legend = F)
> table(as.vector(corridor.land))

##
 6 7 8 9 10 12 15 17 18 19 20 27 30 31 32 34 35 37 41 42
 40 4 33 143 16 137 36 220 11 97 76 16 41 72 6 15 26 4 6 1
```

For both cases, we fine that the most common land-cover type identified in the least-cost path and least-cost corridor is freshwater marsh (ID = 12) and cypress swamp (ID = 17; see `classification`), two land-cover types that are conducive to movement for Florida panthers (Kautz et al. 2006).

9.3.3.4 Flow Mapping

We now contrast the least-cost path and corridor to *randomized shortest-path* analysis and alter the tuning parameter, θ, to interpret this model. This model can

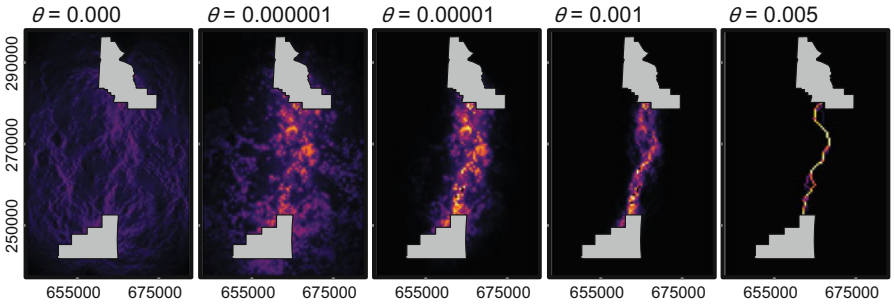


Fig. 9.10 Mapping probabilities of passage based on randomized shortest paths. When $\theta = 0$ the model converges onto a simple random walk (akin to circuit theory). As θ increases, the model converges on a least-cost path

be implemented in `gdistance` and the resulting predicted maps of flow can be visualized.

We use the `passage` function to map the potential flow, similar in spirit to the mapping of current flow in circuit theory (McRae et al. 2008). This function maps the number of potential movements through cells before arriving in a destination location from a source location (van Etten 2012). We focus only on mapping flow between refuges considered in our mapping of least-cost paths and corridors (Fig. 9.10).

```
#random walk (akin to a current map in Circuitscape)
> passage.map_t0 <- passage(land_cond_sub,
  origin = public_centroids@coords[3,],
  goal = public_centroids@coords[5,], theta = 0)

#alter theta to converge on a lcp
> passage.map_t001 <- passage(land_cond_sub,
  origin = public_centroids@coords[3,],
  goal = public_centroids@coords[5,], theta = 0.001)
```

When mapping probabilities of passage as a function of the tuning parameter, θ, we find some interesting conclusions. First, when $\theta = 0$, such that we are mapping a biased random walk, the probability of movement at each cell in the landscape is very low, suggesting that movement is very diffuse and there are no hard boundaries to movement in this portion of the landscape (i.e., there is high route redundancy and little route vulnerability; Rayfield et al. 2011). As θ increases, the probabilities of passage become more well-defined and increase around the least-cost corridor and least-cost path. These comparisons illustrate the relationship between random walks and circuit theory to that of least-cost paths.

9.3.4 *Patch-Based Networks and Graph Theory*

We now turn to the use of patch-based graphs (networks) to interpret connectivity. We contrast some common metrics for quantifying potential connectivity to realized connectivity for the endangered snail kite.

In our example on snail kites, we consider several types of graphs. We construct three networks: a binary network based on the mean observed movement distance, a probability network based on a negative exponential function (using the mean movement distance), and an observed movement network that uses the number of observed movements within breeding seasons for snail kites (Fig. 9.11) (Reichert et al. 2016). Our focus is on within-season breeding dispersal of snail kites, where kites track wetland conditions across their geographic range for breeding.

To create a graph, `igraph` can handle several types of data formats. We will focus on simply passing adjacency matrices into `igraph`. First, we import the observed movement matrix (`kite_movement.csv`) along with some node attributes on coordinates and patch areas. We can then use these data to create new adjacency matrices:

```
> A.obs <- read.csv("kite_movement.csv", header = T)
> nodes <- read.csv("kite_nodes.csv", header = T)
> area <- nodes$area #in km^2

#create distance matrix
> coords <- cbind(nodes$XCoord, nodes$YCoord)
> distmat <- pointDistance(coords, lonlat = F)
> distmat <- distmat / 1000 #in km
```

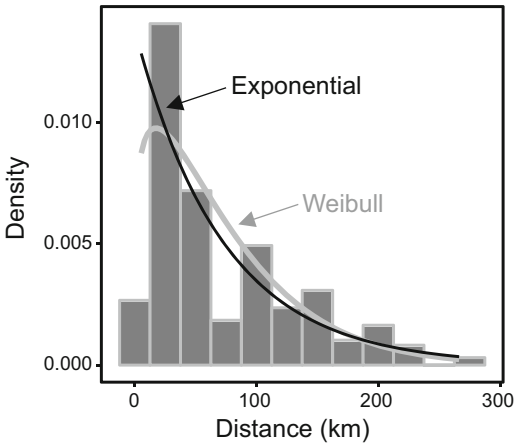


Fig. 9.11 Dispersal kernel for observed within-season movements of the snail kite. Shown are exponential and Weibull kernels and observed within-season movements (histogram) in the snail kite. Note that populating patch-based graphs based on mean or maximum movement distances would imply a segmented (threshold) relationship in dispersal kernels

9.3.4.1 Dispersal Kernels

Before we create graphs that represent movement, we illustrate the construction of dispersal kernels. Dispersal kernels quantify the probability of dispersal as a function of distance from the original location, such as the natal site for natal dispersal, or a breeding site for breeding dispersal (Greenwood 1980; Greenwood and Harvey 1982). In this case, we create dispersal kernels from within-season movements. Dispersal kernels can be quantified in a variety of ways; here we provide a simple illustration. We first reformat the observed movement matrix to represent the number of movements between patches as a function of distance.

```
> link.loc <- which(A.obs > 0, arr.ind = T)
> within_disp <- cbind(distmat[link.loc], A.obs[link.loc])

#repeat distances based on frequency of observations
> within_disp <- rep(within_disp[,1], within_disp[,2])
> names(within_disp) <- "distance"
```

With these reformatted data, we fit a variety of potential dispersal kernel functions to the data. This can be accomplished with the `fitdistrplus` and `fdrtool` packages (Delignette-Muller and Dutang 2015; Klaus and Strimmer 2015). The `fitdistrplus` package can fit many kernel densities to the data. We use the `fdrtool` package to fit a 1Dt distribution, which is a commonly used distribution that captures the potential for "fat-tailed" dispersal kernels, that is, situations where infrequent, but long-distance dispersal occurs (Clark et al. 1999). We contrast the 1Dt distribution with other commonly assumed dispersal kernels, including the log-normal, exponential, and Weibull distributions:

```
> library(fitdistrplus)
> library(fdrtool)
> disp.lnorm <- fitdist(data = within_disp,
  distr = "lnorm", method = "mle")
> disp.exp <- fitdist(data = within_disp,
  distr = "exp", method = "mle")
> disp.weib <- fitdist(data = within_disp,
  distr = "weibull", method = "mle")
> disp.1dt <- fitdist(data = within_disp,
  distr ="halfnorm", start = list(theta = 0.01),
  method = "mle")

> disp.AIC <- gofstat(list(disp.exp, disp.lnorm, disp.weib,
  disp.1dt), fitnames = c("exponential", "lognormal", "Weibull", "1Dt"))
> disp.AIC$aic

##
exponential lognormal Weibull 1Dt
 4110.798 4111.573 4093.400 4103.977
```

Based on model AIC, the Weibull distribution appears to be the best fit to the data (Fig. 9.11). However, below we use the exponential distribution, given its common usage in metapopulation ecology and landscape connectivity (Hanski 1994; Saura and Pascual-Hortal 2007). The exponential distribution is frequently used because it requires only one parameter, α, which describes the inverse of the mean movement (or dispersal) distance for a species. Such information is more frequently available in the literature for use for connectivity modeling (Calabrese and Fagan 2004).

9.3.4.2 Creating a Network or Graph

With the above information on dispersal, we now create graphs based on different transition matrices, **A**.

```
#create adjacency matrix with mean distance
> A.mean.dist <- mean(within_disp)
> A.mean.dist

##
[1] 72.32945

> A.mean <- matrix(0, nrow = nrow(A.obs), ncol = ncol(A.obs))
> A.mean[distmat < A.mean.dist] <- 1
> diag(A.mean) <- 0

#create adjacency matrix with negative exponential
> A.prob <- matrix(0, nrow = nrow(A.obs),ncol = ncol(A.obs))
> alpha <- 1 / A.mean.dist #inverse of mean distance
> A.prob <- exp(-alpha * distmat)
> diag(A.prob) <- 0
```

Above, we create a binary matrix based on the mean dispersal distance (72 km); however, the maximum dispersal distance is also often used in connectivity modeling (Urban and Keitt 2001). With these matrices, we create `igraph` objects for visualization and analysis (Fig. 9.12). Note that each of these matrices have different types of information, with `A.mean` being an undirected, unweighted network, `A.prob` being an undirected, weighted network, and `A.obs` being a directed, weighted network. We create an `igraph` object for each matrix using the `graph.adjacency` function:

```
> graph.Amean <- graph.adjacency(A.mean, mode = "undirected")
> graph.Aprob <- graph.adjacency(A.prob, mode = "undirected",
  weighted = T)
> graph.Aobs <- graph.adjacency(A.obs, mode = "directed", weighted = T)
```

In `igraph`, we can access information about the network structure in a variety of ways. For instance, to access the information on the nodes/vertices, we use the `V` function, whereas for edges/links, we use the `E` function:

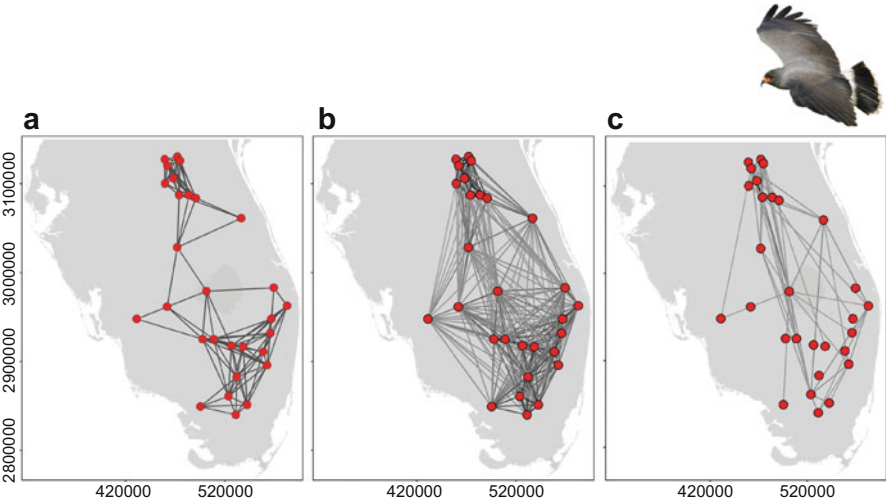


Fig. 9.12 Networks (graphs) based on assumptions about linkages for within breeding season movements of the snail kite. (**a**) An undirected, binary graph based on the mean movement distance. (**b**) An undirected, weighted graph based on a negative exponential dispersal kernel. (**c**) Observed movements taken from mark–resight data (a weighted, directed graph)

```
V(graph.Aobs) #vertex labels

##
+ 29/29 vertices:
 [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

head(E(graph.Aobs)) #edge pairs (edge list)

##
+ 6/95 edges:
[1] 3->5 3->26 3->27 4->9 4->10 4->14

head(E(graph.Aobs)$weight) #edge weights

##
[1] 1 1 1 2 11 2
```

We visualize the patch-based graphs with the standard `plot` function in R. Note that here we use the coordinates of the patches to position the nodes (patches) in space:

```
#visualize
> plot(graph.Amean, layout = coords, vertex.label = NA)
> plot(graph.Aprob, layout = coords, edge.width =
  E(graph.Aprob)$weight, vertex.label = NA)
```

There are several ways to use the `plot` function in `igraph` to make stunning network figures. In particular, we could overlay the patch-based graph on raster or vector GIS layers by first plotting these layers and then making sure that the above function includes `'add=T'` (under the assumption that the geographic layers are all based on the same CRS). In addition, `ggplot2` (Wickham 2009) can be used as well (which was used for Fig. 9.12). We do not focus on those details here, however.

9.3.4.3 Patch-Scale Connectivity

Connectivity metrics frequently describe node-level, or patch-scale, connectivity. These measures are sometimes referred to as *centrality* measures. In the social sciences, Borgatti and Everett (2006) classified these measures based on the type of flow (e.g., random walks) and whether they are "radial" or "medial" measures.

Radial measures focus on quantifying flow that starts or terminates from a patch. In contrast, medial measures focus on the number of paths that flow through a patch. The latter is highly relevant to the concept of stepping stones (Gilpin 1980), whereas the former is relevant to classical measures of patch connectivity (Hanski 1994). Some common radial centrality measures include degree, strength (similar to degree but for weighted graphs), eigenvector centrality, and closeness centrality. Degree simply describes the number of links and strength is the sum of the weights of links to/from patch i to all other directly connected patches. For a directed graph, it is defined as:

$$w_i = \sum_j A_{ij}, \tag{9.3}$$

where w_i is the strength when using a weighted adjacency matrix (when using a binary adjacency matrix, this equation would provide degree). For an undirected graph, only the upper triangle of $\mathbf{A}$ would be summed, such that would be: $w_i = \sum_{j>i} A_{ij}$. Eigenvector centrality attempts to capture neighbor linkages (indirect links), such that if a neighbor (direct link) is connected to many other patches, the focal patch would have higher eigenvector centrality than if the neighbor was unconnected to other patches. It is based on the leading eigenvector of the adjacency matrix (compare this with some metapopulation concepts in Chap. 10). Closeness centrality is the inverse of the average shortest-path distance from patch i to all other connected patches in the network. Closeness centrality has also been extended to random walks, rather than shortest paths. In particular, "Information centrality" has been shown to be a random-walk equivalent of closeness centrality based on shortest paths (note that `igraph` will not calculate information centrality but the package `statnet` can). Consequently, closeness centrality captures patch connectivity but at a larger scale than degree or strength, where indirect linkages (shortest paths) throughout the network are considered. Each of these measures can also be extended to directed graphs. We note that for weighted graphs, both closeness and

betweenness algorithms in `igraph` are based on costs/resistances etc., *not* conductance/movements (e.g., the inverse in linkage values). That is, when using `igraph` above, we focused on links reflecting movements, such that higher values represent greater flow; however, for closeness and betweenness, we use link values that reflect greater cost or lower probabilities of movements.

The most common medial measure is betweenness centrality. Betweenness centrality for patch k quantifies the number of shortest paths between all i and j in the network that go through patch k. This measure has also been extended to random walks, rather than shortest paths (Carroll et al. 2012; Newman 2005). Each of these centrality measures is straightforward to calculate in `igraph`. For example:

```
> Amean.degree <- degree(graph.Amean)
> Amean.eigen <- evcent(graph.Amean)
> Amean.close <- closeness(graph.Amean)
> Amean.between <- betweenness(graph.Amean)
```

When mapping these metrics across the landscape (Fig. 9.13), it is clear that these patch-based metrics capture different aspects of connectivity. To calculate betweenness (or closeness) for a weighted graph, we would alter links as:

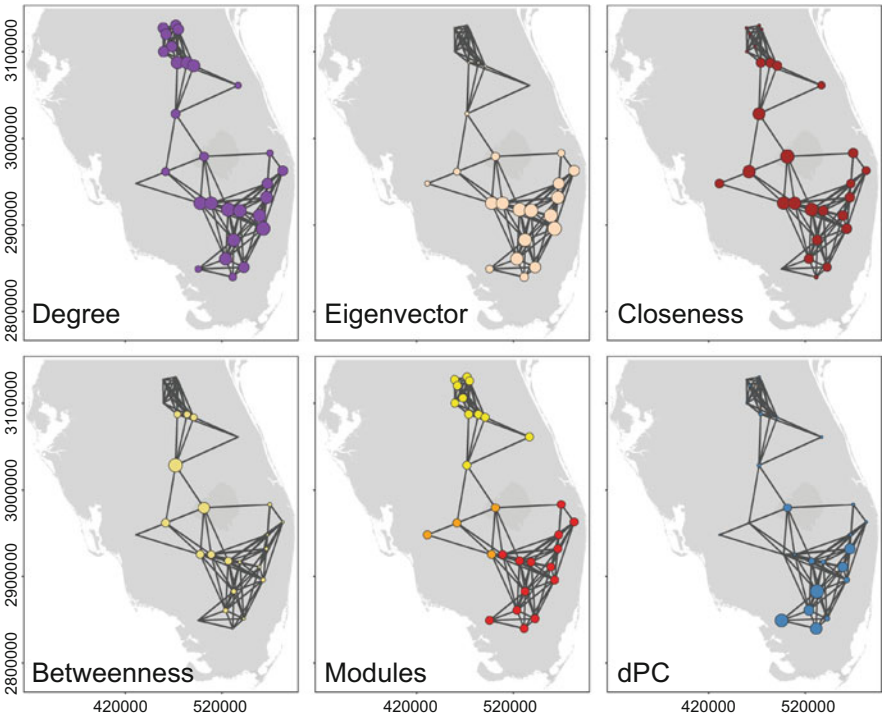


Fig. 9.13 Mapping patch-based and meso-scale measures of connectivity for the snail kite. For degree, eigenvector centrality, closeness centrality, dPC, and betweenness centrality, dot size is proportional to patch connectivity. For modules, dot color represents different modules

```
> Aprob.between <- betweenness(graph.Aprob, weights = 1 / E(graph.
  Aprob)$weight)
```

Area can also be incorporated into patch-based metrics of connectivity. In fact, most metapopulation metrics of connectivity explicitly include area into calculations, as do some related graph metrics (Urban and Keitt 2001). See Chap. 10 for more details regarding metapopulation concepts and metrics.

9.3.4.4 Meso-Scale Connectivity

At the meso-scale, we can calculate components (or clusters) or modularity. Components, c, are subgraphs (or clusters), where any two vertices are connected to each other by paths, but different components are not connected. Modularity, Q, is based on a similar idea, where modules are groups of nodes identified that have more linkages within modules than what would be expected based on chance:

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - P_{ij}\right)\delta\left(I_i, I_j\right), \tag{9.4}$$

where m is the total number of possible links in an undirected network, A_{ij} is the element of the adjacency matrix that describes movement/gene flow between patches i and j, P_{ij} is an expected value, and $\delta(I_i, I_j)$ is an indicator matrix that is equal to 1 if i and j are members of the same module and zero otherwise. The common expected value is based on degree (or strength), $w_i w_j/2m$, where w is the patch degree (strength). This expected value is useful in that both the distribution and the total amount of observed (or potential) movement on the network are conserved ($\sum_{ij}Aij = \sum_{ij}Pij = 2m$); however, other null models can be used. Modules are typically identifying by maximizing Q based on iteratively altering δ (note that this maximization is a computationally hard problem, and there have been dozens of algorithms proposed to maximize Q). This algorithm can accommodate weights and directionality, and it can identify structure within components. It has also been extended to account for spatial issues, such as localized dispersal (Fletcher et al. 2013b). Modularity can be calculated with observed movements (Reichert et al. 2016), data on gene flow (Fortuna et al. 2009), or using networks based on potential connectivity (Foltete and Vuidel 2017; Fletcher et al. 2018). Modularity assumes a "hard partition" in which there is no mixing or overlap among modules, but there have been related concepts that relax that assumption (Valle et al. 2017). `igraph` can calculate components and modularity for simple graphs in a straightforward way:

```
#number of components and node memberships
> Amean.Ncomponents <- clusters(graph.Amean)$no
> Amean.Membcomponents <- clusters(graph.Amean)$membership
```

For components, we find that the entire network is connected—there is only one component. However, there may still be spatial structure in connectivity that can be revealed via modularity.

```
#modularity
> Amean.modularity <- cluster_louvain(graph.Amean)
> modularity(Amean.modularity)

##
[1] 0.4559035

> membership(Amean.modularity)

##
[1] 1 2 1 3 1 2 1 1 3 3 2 3 3 3 3 1 3 2 3 3 1 1 1 3 1 1 1 1 1
```

The `cluster_louvain` function uses the "Louvain" method for maximizing modularity and identifying modules, which is has been shown to work well in terms of its accuracy and computational speed, particularly for very large graphs (Blondel et al. 2008). This function cannot be implemented in `igraph` on a directed graph; however, modularity on directed graphs is best considered by customizing the modularity function to explicitly account for directionality in the null model (Fletcher et al. 2013b). Finally, note that patch-based metrics of connectivity can be quantified that account for modular structure. That is, we can determine the connectivity of patches within modules versus between them (Fletcher et al. 2013b). When we apply the modularity function to the snail kite adjacency matrix, we find that modularity identifies three modules within the one component (Fig. 9.13).

9.3.4.5 Landscape-Scale Connectivity

There are several graph-based metrics for quantifying connectivity across the entire network. Some common measures include connectance, the landscape coincidence probability, the integral index of connectivity, and the probability of connectivity. Connectance is a simple measure for unweighted (binary) graphs, and is defined as the number of observed links over the total number of possible links. The landscape coincidence probability is based on understanding connectivity of components in the landscape and is defined as:

$$\mathrm{LCP} = \sum_{i=1}^{\mathrm{NC}} \left(\frac{c_i}{A_\mathrm{L}} \right)^2, \tag{9.5}$$

where NC is the number of components, c_i is the total area of each component, and A_L is the total area of the landscape considered (area of habitat or nodes plus

non-habitat). This metric could also be applied to modules, rather than components. In R, these metrics can be calculated as:

```
#connectance
> connectance <- graph.density(graph.Amean)

#LCP
> ci <- tapply(area, clusters(graph.Amean)$membership, sum)
> AL <- 63990 #approximate study area in km^2, taken from study area
  polygon (not shown)
> LCP <- sum((ci / AL)^2)
```

The integral index of connectivity (IIC) and probability of connectivity (PC) have been shown to have useful properties for quantifying connectivity at the landscape scale (Pascual-Hortal and Saura 2006; Saura and Pascual-Hortal 2007). IIC is based on a binary network and is described as:

$$\mathrm{IIC} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{a_i a_j}{1 + nl_{ij}}}{A_{\mathrm{L}}^2}, \tag{9.6}$$

where a_i is the area of patch i and nl_{ij} is the number of links in the shortest path from i to j. In R, this is straightforward to calculate. First, we need to create a matrix of shortest paths and then take a vector of patch areas to calculate IIC.

```
> nl.mat <- shortest.paths(graph.Amean)

#Replace inf (in cases with isolated patches) with large values

> nl.mat[is.infinite(nl.mat)] <- 1000

> IICmat <- outer(area, area) / (1 + nl.mat)
> IIC <- sum(IICmat) / AL^2
```

The PC index (Saura and Pascual-Hortal 2007) is currently one of the most frequently used metrics for connectivity at the landscape scale. It is based on a weighted matrix, rather than a binary matrix, and it shares some similarities with metapopulation capacity (Hanski and Ovaskainen 2000; see Chap. 10; Saura and Rubio 2010). It is defined as:

$$\mathrm{PC} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j p_{ij}^*}{A_{\mathrm{L}}^2}, \tag{9.7}$$

where p_{ij}^* is the maximum product probability between patches. p_{ij}^* is calculated from a probabilistic transition matrix, typically with links described as in Eq. (9.2) ($\exp(-\alpha d_{ij})$).This metric is directly related to the neighborhood habitat area index

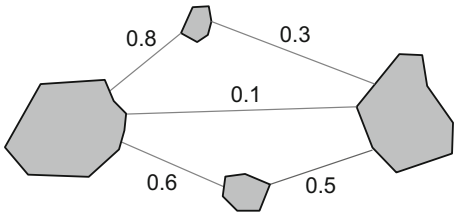


Fig. 9.14 A toy example illustrating the idea of maximum product probability paths for networks (modified from Hock and Mumby 2015). Shown are two patches with three potential paths between them, with the probability of movement for links shown. In this case, the cost distance between the two large patches would be equal going through the upper and lower paths with stepping stones (i.e., the sum of costs is identical), both of which would have a greater likelihood than the direct link probability between patches (0.1). However, the maximum product probability path would go through the lower stepping stone ($0.6 \times 0.5 = 0.3$) rather than the upper stepping stone ($0.8 \times 0.3 = 0.24$). This can be quantified by using least-cost path algorithms based on the negative log of the link probabilities

H_n of Hanski (p 83 of Hanski 1999), where the only difference is the use of p_{ij}^* taken from $\exp(-\alpha d_{ij})$ rather than $\exp(-\alpha d_{ij})$ directly, and the denominator in H_n is the sum of the patch areas rather than the total area (habitat + non-habitat). What is p_{ij}^*, exactly? The maximum product probability comes from theoretical developments on network reliability. It reflects the idea that the shortest path between two locations, i and j, may not reflect direct routes between the two patches but rather may require passing through k. Hock and Mumby (2015) described the rationale for this measurement and its relationship to least-cost paths in detail. If we view links as probabilistic connections (rather than resistance), then we might expect that the probability of movement between two patches will be reflected as the product of probabilities along the path, assuming each transition (step) is independent. The use of least-cost algorithms, however, determine paths based on the sum—rather than the product—of linkages. This can fact can result in potentially identifying inappropriate least-cost paths (or distances; Fig. 9.14). But if we take the negative logarithm of these probabilities, we can use conventional least-cost (or shortest-path) algorithms to quantify p_{ij}^*. We can do this in R with the weighted probabilistic graph we created above:

```
#shortest path based on negative log-transformed weights
> pstar.mat <- shortest.paths(graph.Aprob, weights = log(E(
  graph.Aprob)$weight))
> pstar.mat <- exp(-pstar.mat) #back-transform to probabilities

#numerator of PC
> PCnum <- outer(area, area) * pstar.mat

#probability of connectivity
> PC <- sum(PCnum)/AL^2
```

Note that a patch-based metric, dPC, is related to PC (Saura and Pascual-Hortal 2007). This metric describes the patch-level importance to PC and is calculated based on the change in PC with the removal of a given patch. It is straightforward to calculate based on above, by sequentially removing patches and calculating the change in PC (Saura and Pascual-Hortal 2007). Consequently, dPC describes the proportional contribution of a patch to PC.

dPC has been further decomposed into three fractions (Bodin and Saura 2010; Saura and Rubio 2010). These fractions represent the role that a patch plays in regards to connectivity within the patch (dPC_{intra}), connectivity between patches based on flow to or from the focal patch (dPC_{flux}), and the patch's role in connectivity between other patches and not the focal patch, such as being a stepping stone ($dPC_{connector}$). These fractions are described as:

$$dPC = dPC_{intra} + dPC_{flux} + dPC_{connector}. \tag{9.8}$$

dPC_{intra} acknowledges that the larger the patch is, the more habitat it provides that is reachable for organisms (e.g., within patch dispersal opportunities). It is defined as:

$$dPC_{intra,i} = \frac{a_i^2}{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j p_{ij}^*} \times 100. \tag{9.9}$$

dPC_{flux} is related to other radial connectivity metrics that capture area-weighted flow. This metric is analogous to some other patch-based connectivity metrics used in metapopulation ecology.

$$dPC_{flux,i} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j p_{ij}^* - a_i^2}{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j p_{ij}^*} \times 100. \tag{9.10}$$

$dPC_{connector}$ is a medial-based metric of connectivity that isolates the role of patches in acting as stepping stones or connectors between other patches. It can be most readily quantified based on the remaining difference as:

$$dPC_{connector} = dPC - dPC_{intra} - dPC_{flux}. \tag{9.11}$$

A function for calculating PC, dPC and its related fractions is:

```
> prob.connectivity <-
  function(prob.matrix,area,landarea){
  pc.g <- graph.adjacency(prob.matrix, mode = "undirected",weighted = T)
  pstar.mat <- shortest.paths(pc.g, weights = -
  log(E(pc.g)$weight))
  pstar.mat <- exp(-pstar.mat) #back-transform
  PCmat <- outer(area, area) * pstar.mat
  PC <- sum(PCmat) / landarea^2
```

```
 N <- nrow(prob.matrix)
 dPC <- rep(NA,N)
 for (i in 1:N) {
 prob.mat.i <- prob.matrix[-i, -i]
 area.i <- area[-i]
 pc.g.i <- graph.adjacency(prob.mat.i, mode = "undirected",
 weighted = T)
 pstar.mat.i <- shortest.paths(pc.g.i,
 weights = -log(E(pc.g.i)$weight))
 pstar.mat.i <- exp(-pstar.mat.i)
 PCmat.i <- outer(area.i, area.i) * pstar.mat.i
 PC.i <- sum(PCmat.i) / landarea^2
 dPC[i] <- (PC-PC.i) / PC * 100
}
#fractions
dPCintra <- area^2 / sum(PCmat) * 100
dPCflux <- 2*(rowSums(PCmat) - area^2)/sum(PCmat) * 100
dPCconn <- dPC - dPCintra - dPCflux

patch.metrics <- data.frame(dPC, dPCintra, dPCflux, dPCconn)
pc.list <- list(PC = PC,patch = patch.metrics)
return(pc.list)
}
```

For the snail kite data, we compare relationships among the patch-level connectivity metrics (Fig. 9.15). When using the probability adjacency matrix, it is clear that these metrics capture different elements of connectivity. Patch strength and eigenvector centrality are highly correlated (and closeness to a lesser extent). dPC is highly correlated with patch area, while betweenness appears less correlated with these other patch metrics.

9.3.5 Combining Connectivity Mapping with Graph Theory

Patch-based graph analyses can be applied to raster maps (Carroll et al. 2012). To illustrate this idea, we can revisit the panther example where we calculated least-cost paths. But now, we calculate betweenness centrality for each cell in the landscape. This metric is very relevant in the context of raster maps (not all graph measures are as useful for raster data) because, when applied to resistance or cost surfaces, it can identify locations where the most least-cost paths pass on a landscape. That is, for each cell on the map, it will identify the number of least-cost paths that cross that cell (based on every pair of cells in the landscape).

To illustrate this idea, we convert our raster layer used in the Florida panther example into a transition matrix. Note that for this analysis, we use the cost raster, rather than converting it to a conductance layer as we did previously. Then, we use `igraph` to calculate betweenness for each pixel in the map. Finally, we re-project these betweenness metrics back onto a raster map.

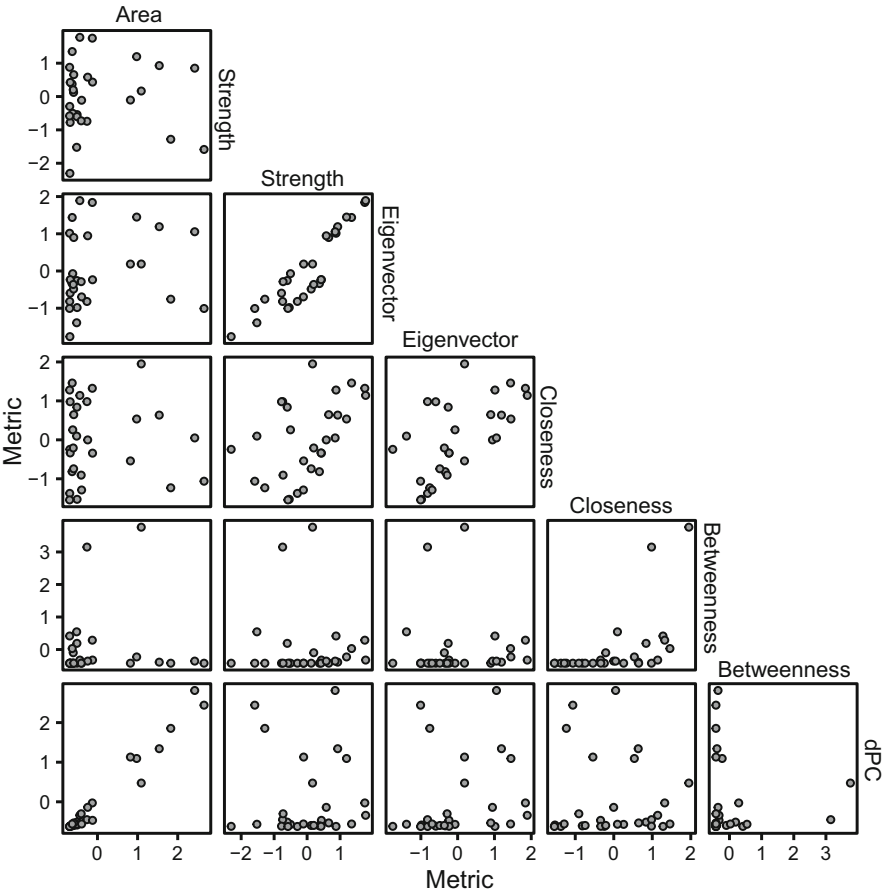


Fig. 9.15 Correlations of patch-based connectivity metrics for the snail kite. Shown are patch values based on the probability adjacency matrix. Also included is patch area (area)

```
#create a transition matrix for use in igraph
> land_cost_subt <- transition(land_cost_sub, transitionFunction
  = mean, 8)
> land_cost_subt <- geoCorrection(land_cost_subt, type = "c")

> land.matrix <- transitionMatrix(land_cost_subt)#sparse matrix


#now take the sparse matrix and use igraph for analysis
> land.graph <- graph.adjacency(land.matrix, mode =
  "undirected", weighted = T)
> land.between <- betweenness(land.graph)

#map betweenness
> land.between.map <- setValues(land_cost_sub, land.between)
> plot(land.between.map)
```

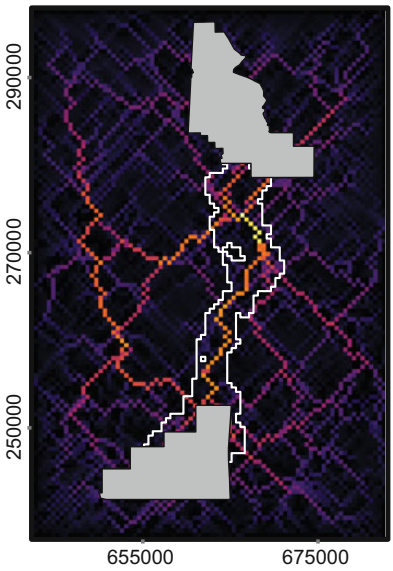


Fig. 9.16 Mapping betweenness centrality to interpret connectivity in a landscape with two protected areas. Brighter values reflect greater betweenness centrality across the landscape. The least-cost corridor is outlined in white for comparison

When we contrast this with the least-cost path analysis, there are some interesting similarities (Fig. 9.16). First, the betweenness mapping reveals the least-cost path identified previously (Fig. 9.8). However, it also reveals another path in the landscape that may be relevant for connectivity.

Note that the mapping of betweenness centrality shares many similarities with the idea of factorial least-cost paths (Rudnick et al. 2012; Elliot et al. 2014). In that approach a grid is placed over the landscape, least-cost paths are calculated between all points on the grid, and these paths are overlaid to generate a connectivity map. This approach could easily be implemented in R using the `sampleRegular` function in the `raster` package. The main difference then between factorial least-cost paths and mapping betweenness centrality is the grain of the analysis: for mapping betweenness centrality, the grain is the grain of the map being used, while for factorial least cost paths, the grain is contingent on the grain of the grid being considered.

9.4 Next Steps and Advanced Issues

9.4.1 Connectivity in Space and Time

While connectivity modeling emphasizes the role of space in potential and realized movement, connectivity can change over time for a variety of reasons (Zeigler and Fagan 2014), such as population trends impacting the number of individuals available for movement, seasonality effects, or effect of changes in climate or land use.

Such variation can alter interpretation of connectivity (Reichert et al. 2016). Importantly, there is always a temporal component to connectivity, although frequently this dimension is implicit rather than explicit in modeling. For instance, when using patch-based graphs, are linkages based on long-term potential movement or short-term (e.g., within-season, annual movements)?

An active area of development in network analysis is that of temporal networks: changes in network structure over time (Holme and Saramaki 2012). These developments have extended standard metrics and models to explicitly model changes in network flow over time. These developments have yet to be well integrated into connectivity modeling (but see Martensen et al. 2017). However, such approaches may prove valuable for interpreting effects of ongoing environmental change.

9.4.2 Individual-Based Models

An alternative approach to the methods presented in this Chapter is the use of individual-based (agent-based) simulation models for predicting and interpreting connectivity. These approaches are very useful for formally capturing individual variation in movement, behavioral issues such as the effect of perceptual ranges on connectivity (Pe'er and Kramer-Schadt 2008), how landscape change may impact movement, and understanding the impacts of connectivity on population viability and community dynamics (Schumaker et al. 2014). Nonetheless, these approaches require very detailed knowledge about the process of interest. These approaches have been used much less frequently than the network approaches described in this chapter (Fletcher et al. 2016), but they can be helpful in many situations.

9.4.3 Diffusion Models

Diffusion models have also been developed for interpreting connectivity (Reeve et al. 2008; Ovaskainen et al. 2008). These models frequently focus on population dynamics and the spread of individuals through a diffusive process, akin to the use of random walks (Ferrari et al. 2014). Extensions include adding advection, or direction, to the diffusion process and allowing for habitat-specific diffusion rates (Reeve et al. 2008). These models are elegant and can concisely track aspects of connectivity; however, they are challenging to fit to data.

9.4.4 Spatial Capture–Recapture

Capture–recapture modeling has been extended to spatially explicit modeling of mark–recapture data across sampling grids (Royle et al. 2014). While the

focus is often on density estimation, these approaches have been extended to estimate aspects of connectivity, such as least-cost paths and estimates of resistance (Royle et al. 2013, 2018; Sutherland et al. 2015). These efforts might be particularly valuable for certain types of data, such as camera-trapping grids.

9.5 Conclusions

Our understanding and quantification of connectivity has been largely catalyzed by increased interest in conserving connectivity with ongoing environmental change (Crooks and Sanjayan 2006; Lawler et al. 2013). Connectivity is now being incorporated into applied problems involving population viability analysis (Stevens and Baguette 2008), invasive species management (Drake et al. 2017; Minor and Gardner 2011), pest control (Margosian et al. 2009), and conservation planning for climate and land-use change (Albert et al. 2017). We now know that connectivity has consistent, positive impacts on populations and communities (Gilbert-Norton et al. 2010; Haddad et al. 2003, 2015, 2014; Fletcher et al. 2016), and it is frequently championed as a means to promote the persistence of biodiversity with ongoing environmental change (Heller and Zavaleta 2009).

Future advancements in our understanding and application of connectivity will require a greater emphasis on estimating connectivity effects for wild populations and communities, particularly in regard to using explicit movement and dispersal data to understand how such movement impact populations and communities. Better incorporation of the problem of scale and how connectivity varies with spatial and temporal scale is also needed (Pascual-Hortal and Saura 2007). Finally, isolating connectivity effects relative to other issues influencing biodiversity will help understand the importance of connectivity relative to other known issues for populations and communities (e.g., habitat amount; Hodgson et al. 2011). By addressing these outstanding issues, we will better understand the relative importance of connectivity for biodiversity and deliver more effective conservation strategies.

References

Acevedo MA, Sefair JA, Smith JC, Reichert B, Fletcher RJ Jr (2015) Conservation under uncertainty: optimal network protection strategies for worst-case disturbance events. J Appl Ecol 52:1588–1597. https://doi.org/10.1111/1365-2664.12532

Albert CH, Rayfield B, Dumitru M, Gonzalez A (2017) Applying network theory to prioritize multispecies habitat networks that are robust to climate and land-use change. Conserv Biol 31 (6):1383–1396. https://doi.org/10.1111/cobi.12943

Baguette M, Blanchet S, Legrand D, Stevens VM, Turlure C (2013) Individual dispersal, landscape connectivity and ecological networks. Biol Rev 88(2):310–326. https://doi.org/10.1111/brv.12000

Bélisle M (2005) Measuring landscape connectivity: the challenge of behavioral landscape ecology. Ecology 86(8):1988–1995
Bélisle M, Desrochers A (2002) Gap-crossing decisions by forest birds: an empirical basis for parameterizing spatially-explicit, individual-based models. Landsc Ecol 17(3):219–231
Bender DJ, Fahrig L (2005) Matrix structure obscures the relationship between interpatch movement and patch size and isolation. Ecology 86(4):1023–1033. https://doi.org/10.1890/03-0769
Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp. https://doi.org/10.1088/1742-5468/2008/10/p10008
Bodin O, Norberg J (2007) A network approach for analyzing spatially structured populations in fragmented landscape. Landsc Ecol 22(1):31–44. https://doi.org/10.1007/s10980-006-9015-0
Bodin O, Saura S (2010) Ranking individual habitat patches as connectivity providers: integrating network analysis and patch removal experiments. Ecol Model 221(19):2393–2405. https://doi.org/10.1016/j.ecolmodel.2010.06.017
Borgatti SP, Everett MG (2006) A graph-theoretic perspective on centrality. Soc Networks 28 (4):466–484. https://doi.org/10.1016/j.socnet.2005.11.005
Brownie C, Hines JE, Nichols JD, Pollock KH, Hestbeck JB (1993) Capture-recapture studies for multiple strata including non-markovian transitions. Biometrics 49(4):1173–1187
Bunn AG, Urban DL, Keitt TH (2000) Landscape connectivity: a conservation application of graph theory. J Environ Manag 59(4):265–278
Calabrese JM, Fagan WF (2004) A comparison-shopper's guide to connectivity metrics. Front Ecol Environ 2(10):529–536
Carrara F, Altermatt F, Rodriguez-Iturbe I, Rinaldo A (2012) Dendritic connectivity controls biodiversity patterns in experimental metacommunities. Proc Natl Acad Sci U S A 109 (15):5761–5766. https://doi.org/10.1073/pnas.1119651109
Carroll C, McRae BH, Brookes A (2012) Use of linkage mapping and centrality analysis across habitat gradients to conserve connectivity of gray wolf populations in western North America. Conserv Biol 26(1):78–87. https://doi.org/10.1111/j.1523-1739.2011.01753.x
Charnov EL (1976) Optimal foraging: marginal value theorem. Theor Popul Biol 9(2):129–136. https://doi.org/10.1016/0040-5809(76)90040-x
Chisholm C, Lindo Z, Gonzalez A (2011) Metacommunity diversity depends on connectivity and patch arrangement in heterogeneous habitat networks. Ecography 34(3):415–424. https://doi.org/10.1111/j.1600-0587.2010.06588.x
Clark JS, Silman M, Kern R, Macklin E, HilleRisLambers J (1999) Seed dispersal near and far: patterns across temperate and tropical forests. Ecology 80(5):1475–1494
Cook WM, Lane KT, Foster BL, Holt RD (2002) Island theory, matrix effects and species richness patterns in habitat fragments. Ecol Lett 5(5):619–623. https://doi.org/10.1046/j.1461-0248.2002.00366.x
Crooks KR, Sanjayan M (eds) (2006) Connectivity conservation. Cambridge University Press, New York
Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJournal Complex Syst 1695(5):1–9
Dale MRT, Fortin MJ (2010) From graphs to spatial graphs. Annu Rev Ecol Evol Syst 41:21–38
Delignette-Muller ML, Dutang C (2015) fitdistrplus: an R package for fitting distributions. J Stat Softw 64(4):1–34
Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1:269–271
Drake JC, Griffis-Kyle KL, McIntyre NE (2017) Graph theory as an invasive species management tool: case study in the Sonoran Desert. Landsc Ecol 32(8):1739–1752. https://doi.org/10.1007/s10980-017-0539-2
Elliot NB, Cushman SA, Macdonald DW, Loveridge AJ (2014) The devil is in the dispersers: predictions of landscape connectivity change with demography. J Appl Ecol 51(5):1169–1178. https://doi.org/10.1111/1365-2664.12282
Etherington TR (2016) Least-cost modelling and landscape ecology: concepts, applications, and opportunities. Curr Landsc Ecol Rep 1:40–53

Etherington TR, Holland EP (2013) Least-cost path length versus accumulated-cost as connectivity measures. Landsc Ecol 28(7):1223–1229. https://doi.org/10.1007/s10980-013-9880-2

Etten J, Hijmans RJ (2010) A geospatial modelling approach integrating archaeobotany and genetics to trace the origin and dispersal of domesticated plants. PLoS One 5:e12060. https://doi.org/10.1371/journal.pone.0012060

Fahrig L (1998) When does fragmentation of breeding habitat affect population survival? Ecol Model 105(2–3):273–292

Fahrig L (2003) Effects of Habitat Fragmentation on Biodiversity. Ann Rev Ecol Evol Syst 34 (1):487–515. https://doi.org/10.1146/annurev.ecolsys.34.011802.132419

Fall A, Fortin MJ, Manseau M, O'Brien D (2007) Spatial graphs: principles and applications for habitat connectivity. Ecosystems 10(3):448–461. https://doi.org/10.1007/s10021-007-9038-7

Ferrari JR, Preisser EL, Fitzpatrick MC (2014) Modeling the spread of invasive species using dynamic network models. Biol Invasions 16(4):949–960. https://doi.org/10.1007/s10530-013-0552-6

Fletcher RJ Jr (2006) Emergent properties of conspecific attraction in fragmented landscapes. Am Nat 168(2):207–219

Fletcher RJ, Reichert BE, Holmes K (2018) The negative effects of habitat fragmentation operate at the scale of dispersal. Ecology 99(10):2176–2186

Fletcher RJ Jr, Acevedo MA, Reichert BE, Pias KE, Kitchens WM (2011) Social network models predict movement and connectivity in ecological landscapes. Proc Natl Acad Sci U S A 108:19282–19287

Fletcher RJ Jr, Maxwell CW Jr, Andrews JE, Helmey-Hartman WL (2013a) Signal detection theory clarifies the concept of perceptual range and its relevance to landscape connectivity. Landsc Ecol 28(1):57–67. https://doi.org/10.1007/s10980-012-9812-6

Fletcher RJ Jr, Revell A, Reichert BE, Kitchens WM, Dixon JD, Austin JD (2013b) Network modularity reveals critical scales for connectivity in ecology and evolution. Nat Commun 4:2572. https://doi.org/10.1038/ncomms3572

Fletcher RJ Jr, Acevedo MA, Robertson EP (2014) The matrix alters the role of path redundancy on patch colonization rates. Ecology 95(6):1444–1450

Fletcher RJ Jr, Burrell N, Reichert BE, Vasudev D (2016) Divergent perspectives on landscape connectivity reveal consistent effects from genes to communities. Curr Landsc Ecol Rep 1 (2):67–79

Foltete JC, Vuidel G (2017) Using landscape graphs to delineate ecologically functional areas. Landsc Ecol 32(2):249–263. https://doi.org/10.1007/s10980-016-0445-z

Fortuna MA, Albaladejo RG, Fernandez L, Aparicio A, Bascompte J (2009) Networks of spatial genetic variation across species. Proc Natl Acad Sci U S A 106(45):19044–19049. https://doi.org/10.1073/pnas.0907704106

Galpern P, Manseau M, Fall A (2011) Patch-based graphs of landscape connectivity: a guide to construction, analysis and application for conservation. Biol Conserv 144(1):44–55. https://doi.org/10.1016/j.biocon.2010.09.002

Gilbert-Norton L, Wilson R, Stevens JR, Beard KH (2010) A meta-analytic review of corridor effectiveness. Conserv Biol 24(3):660–668. https://doi.org/10.1111/j.1523-1739.2010.01450.x

Gilpin ME (1980) The role of stepping-stone islands. Theor Popul Biol 17(2):247–253. https://doi.org/10.1016/0040-5809(80)90009-x

Graves T, Chandler RB, Royle JA, Beier P, Kendall KC (2014) Estimating landscape resistance to dispersal. Landsc Ecol 29(7):1201–1211. https://doi.org/10.1007/s10980-014-0056-5

Greenwood PJ (1980) Mating systems, philopatry and dispersal in birds and mammals. Anim Behav 28(NOV):1140–1162. https://doi.org/10.1016/s0003-3472(80)80103-5

Greenwood PJ, Harvey PH (1982) The natal and breeding dispersal of birds. Annu Rev Ecol Syst 13:1–21. https://doi.org/10.1146/annurev.es.13.110182.000245

Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. Mol Ecol 18(23):4734–4756. https://doi.org/10.1111/j.1365-294X.2009.04410.x

Gustafson EJ, Parker GR (1994) Using an index of habitat patch proximity for landscape design. Landsc Urban Plan 29(2–3):117–130. https://doi.org/10.1016/0169-2046(94)90022-1
Haase CG, Fletcher RJ, Slone DH, Reid JP, Butler SM (2017) Landscape complementation revealed through bipartite networks: an example with the Florida manatee. Landsc Ecol 32 (10):1999–2014. https://doi.org/10.1007/s10980-017-0560-5
Haddad NM, Bowne DR, Cunningham A, Danielson BJ, Levey DJ, Sargent S, Spira T (2003) Corridor use by diverse taxa. Ecology 84(3):609–615. https://doi.org/10.1890/0012-9658(2003)084[0609:cubdt]2.0.co;2
Haddad NM, Brudvig LA, Damschen EI, Evans DM, Johnson BL, Levey DJ, Orrock JL, Resasco J, Sullivan LL, Tewksbury JJ, Wagner SA, Weldon AJ (2014) Potential negative ecological effects of corridors. Conserv Biol 28(5):1178–1187. https://doi.org/10.1111/cobi.12323
Haddad NM, Brudvig LA, Clobert J, Davies KF, Gonzalez A et al (2015) Habitat fragmentation and its lasting impact on Earth. Sci Adv 1:e1500052
Handcock M, Hunter D, Butts C, Goodreau S, Morris M (2008) Statnet: software tools for the representation, visualization, analysis and simulation of network data. J Stat Softw 24(1):1–11
Hanks EM, Hooten MB (2013) Circuit theory and model-based inference for landscape connectivity. J Am Stat Assoc 108(501):22–33. https://doi.org/10.1080/01621459.2012.724647
Hanski I (1994) A practical model of metapopulation dynamics. J Anim Ecol 63(1):151–162
Hanski I (1998) Metapopulation dynamics. Nature 396(6706):41–49
Hanski I (1999) Metapopulation ecology. Oxford University Press, Oxford
Hanski I, Ovaskainen O (2000) The metapopulation capacity of a fragmented landscape. Nature 404 (6779):755–758
Harju SM, Olson CV, Dzialak MR, Mudd JP, Winstead JB (2013) A flexible approach for assessing functional landscape connectivity, with application to greater sage grouse (*Centrocercus urophasianus*). PLoS One 8(12). https://doi.org/10.1371/journal.pone.0082271
Heino M, Kaitala V, Ranta E, Lindstrom J (1997) Synchronous dynamics and rates of extinction in spatially structured populations. Proc R Soc B 264(1381):481–486. https://doi.org/10.1098/rspb.1997.0069
Heller NE, Zavaleta ES (2009) Biodiversity management in the face of climate change: a review of 22 years of recommendations. Biol Conserv 142(1):14–32. https://doi.org/10.1016/j.biocon.2008.10.006
Hiebeler D (2000) Populations on fragmented landscapes with spatially structured heterogeneities: landscape generation and local dispersal. Ecology 81:1629–1641
Hock K, Mumby PJ (2015) Quantifying the reliability of dispersal paths in connectivity networks. J Royal Soc Int 12(105). https://doi.org/10.1098/rsif.2015.0013
Hodgson JA, Moilanen A, Wintle BA, Thomas CD (2011) Habitat area, quality and connectivity: striking the balance for efficient conservation. J Appl Ecol 48(1):148–152. https://doi.org/10.1111/j.1365-2664.2010.01919.x
Holme P, Saramaki J (2012) Temporal networks. Phys Rep Rev Sect Phys Lett 519(3):97–125. https://doi.org/10.1016/j.physrep.2012.03.001
Howell PE, Muths E, Hossack BR, Sigafus BH, Chandler RB (2018) Increasing connectivity between metapopulation ecology and landscape ecology. Ecology 99(5):1119–1128
Ims RA (1995) Movement patterns related to spatial structures. In: Hansson L, Fahrig L, Merriam G (eds) Mosaic landscapes and ecological processes. Chapman & Hall, London, UK, pp 85–109
Johst K, Drechsler M (2003) Are spatially correlated or uncorrelated disturbance regimes better for the survival of species? Oikos 103(3):449–456
Kallimanis AS, Kunin WE, Halley JM, Sgardelis SP (2005) Metapopulation extinction risk under spatially autocorrelated disturbance. Conserv Biol 19(2):534–546
Kautz R, Kawula R, Hoctor T, Comiskey J, Jansen D, Jennings D, Kasbohm J, Mazzotti F, McBride R, Richardson L, Root K (2006) How much is enough? Landscape-scale conservation for the Florida panther. Biol Conserv 130(1):118–133. https://doi.org/10.1016/j.biocon.2005.12.007

Kimura M, Weiss GH (1964) Stepping stone model of population structure and decrease of genetic correlation with distance. Genetics 49(4):561
Klaus B, Strimmer K (2015) fdrtool: estimation of (local) false discovery rates and higher criticism. R package version 1.2.15
Kool JT, Moilanen A, Treml EA (2013) Population connectivity: recent advances and new perspectives. Landsc Ecol 28(2):165–185. https://doi.org/10.1007/s10980-012-9819-z
Larson A, Wake DB, Yanev KP (1984) Measuring gene flow among populations having high levels of genetic fragmentation. Genetics 106(2):293–308
Lawler JJ, Ruesch AS, Olden JD, McRae BH (2013) Projected climate-driven faunal movement routes. Ecol Lett 16(8):1014–1022. https://doi.org/10.1111/ele.12132
Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman D, Loreau M, Gonzalez A (2004) The metacommunity concept: a framework for multi-scale community ecology. Ecol Lett 7(7):601–613. https://doi.org/10.1111/j.1461-0248.2004.00608.x
Lomolino MV (1990) The target area hypothesis: the influence of island area on immigration rates of non-volant mammals. Oikos 57(3):297–300. https://doi.org/10.2307/3565957
Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? Mol Ecol 19(15):3038–3051. https://doi.org/10.1111/j.1365-294X.2010.04688.x
MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press, Princeton, NJ
Margosian ML, Garrett KA, Hutchinson JMS, With KA (2009) Connectivity of the American agricultural landscape: assessing the national risk of crop pest and disease spread. Bioscience 59 (2):141–151. https://doi.org/10.1525/bio.2009.59.2.7
Marrotte RR, Bowman J (2017) The relationship between least-cost and resistance distance. PLoS One 12(3). https://doi.org/10.1371/journal.pone.0174212
Martensen AC, Saura S, Fortin MJ (2017) Spatio-temporal connectivity: assessing the amount of reachable habitat in dynamic landscapes. Methods Ecol Evol 8(10):1253–1264. https://doi.org/10.1111/2041-210x.12799
Matter SF (2001) Synchrony, extinction, and dynamics of spatially segregated, heterogeneous populations. Ecol Model 141(1–3):217–226. https://doi.org/10.1016/s0304-3800(01)00275-7
McGarigal K, Cushman SA, Neel MC, Ene E (2002) FRAGSTATS: Spatial pattern analysis program for categorical maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: http://www.umass.edu/landeco/research/fragstats/fragstats.html
McRae BH (2006) Isolation by resistance. Evolution 60(8):1551–1561
McRae BH, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. Ecology 89(10):2712–2724
McRae BH, Hall SA, Beier P, Theobald DM (2012) Where to restore ecological connectivity? Detecting barriers and quantifying restoration benefits. PLoS One 7(12). https://doi.org/10.1371/journal.pone.0052604
Mestre F, Canovas F, Pita R, Mira A, Beja P (2016) An R package for simulating metapopulation dynamics and range expansion under environmental change. Environ Model Softw 81:40–44. https://doi.org/10.1016/j.envsoft.2016.03.007
Minor ES, Gardner RH (2011) Landscape connectivity and seed dispersal characteristics inform the best management strategy for exotic plants. Ecol Appl 21:739–749. https://doi.org/10.1890/10-0321.1
Minor ES, Lookingbill TR (2010) A multiscale network analysis of protected-area connectivity for mammals in the United States. Conserv Biol 24(6):1549–1558. https://doi.org/10.1111/j.1523-1739.2010.01558.x
Mitchell MGE, Bennett EM, Gonzalez A (2013) Linking landscape connectivity and ecosystem service provision: current knowledge and research gaps. Ecosystems 16(5):894–908. https://doi.org/10.1007/s10021-013-9647-2

Moilanen A, Hanski I (1998) Metapopulation dynamics: effects of habitat quality and landscape structure. Ecology 79(7):2503–2515

Moilanen A, Hanski I (2001) On the use of connectivity measures in spatial ecology. Oikos 95 (1):147–151

Moilanen A, Nieminen M (2002) Simple connectivity measures in spatial ecology. Ecology 83 (4):1131–1145

Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE (2008) A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci U S A 105 (49):19052–19059. https://doi.org/10.1073/pnas.0800375105

Newman MEJ (2005) A measure of betweenness centrality based on random walks. Soc Networks 27(1):39–54. https://doi.org/10.1016/j.socnet.2004.11.009

Ovaskainen O, Luoto M, Ikonen I, Rekola H, Meyke E, Kuussaari M (2008) An empirical test of a diffusion model: predicting clouded apollo movements in a novel environment. Am Nat 171 (5):610–619. https://doi.org/10.1086/587070

Panzacchi M, Van Moorter B, Strand O, Saerens M, Ki IK, St Clair CC, Herfindal I, Boitani L (2016) Predicting the continuum between corridors and barriers to animal movements using Step Selection Functions and Randomized Shortest Paths. J Anim Ecol 85(1):32–42. https://doi.org/10.1111/1365-2656.12386

Pascual-Hortal L, Saura S (2006) Comparison and development of new graph-based landscape connectivity indices: towards the priorization of habitat patches and corridors for conservation. Landsc Ecol 21(7):959–967. https://doi.org/10.1007/s10980-006-0013-z

Pascual-Hortal L, Saura S (2007) Impact of spatial scale on the identification of critical habitat patches for the maintenance of landscape connectivity. Landsc Urban Plan 83(2–3):176–186. https://doi.org/10.1016/j.landurbplan.2007.04.003

Pe'er G, Kramer-Schadt S (2008) Incorporating the perceptual range of animals into connectivity models. Ecol Model 213(1):73–85. https://doi.org/10.1016/j.ecolmodel.2007.11.020

Peterman WE (2018) ResistanceGA: an R package for the optimization of resistance surfaces using genetic algorithms. Methods Ecol Evol. https://doi.org/10.1111/2041-210x.12984

Pfluger FJ, Balkenhol N (2014) A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. Mol Ecol 23 (9):2146–2156. https://doi.org/10.1111/mec.12712

Pinto N, Keitt TH (2009) Beyond the least-cost path: evaluating corridor redundancy using a graph-theoretic approach. Landsc Ecol 24(2):253–266. https://doi.org/10.1007/s10980-008-9303-y

Pringle CM (2001) Hydrologic connectivity and the management of biological reserves: a global perspective. Ecol Appl 11(4):981–998. https://doi.org/10.2307/3061006

Proulx SR, Promislow DEL, Phillips PC (2005) Network thinking in ecology and evolution. Trends Ecol Evol 20(6):345–353

Pulliam HR (1988) Sources, sinks, and population regulation. Am Nat 132(5):652–661

Pulliam HR, Danielson BJ (1991) Sources, sinks, and habitat selection: a landscape perspective on population dynamics. Am Nat 137:S50–S66

Rayfield B, Fortin MJ, Fall A (2010) The sensitivity of least-cost habitat graphs to relative cost surface values. Landsc Ecol 25(4):519–532. https://doi.org/10.1007/s10980-009-9436-7

Rayfield B, Fortin M-J, Fall A (2011) Connectivity for conservation: a framework to classify network measures. Ecology 92(4):847–858. https://doi.org/10.1890/09-2190.1

Reeve JD, Cronin JT, Haynes KJ (2008) Diffusion models for animals in complex landscapes: incorporating heterogeneity among substrates, individuals and edge behaviours. J Anim Ecol 77 (5):898–904. https://doi.org/10.1111/j.1365-2656.2008.01411.x

Reichert BE, Cattau CE, Fletcher RJ Jr, Sykes PW Jr, Rodgers JA Jr, Bennetts RE (2015) Snail kite (*Rostrhamus sociabilis*). In: Poole A (ed) The birds of North America online. Cornell Lab of Ornithology, Ithaca

Reichert BE, Fletcher RJ Jr, Cattau CE, Kitchens WM (2016) Consistent scaling of population structure despite intraspecific variation in movement and connectivity. J Anim Ecol 85:1563–1573

Richard Y, Armstrong DP (2010) Cost distance modelling of landscape connectivity and gap-crossing ability using radio-tracking data. J Appl Ecol 47(3):603–610. https://doi.org/10.1111/j.1365-2664.2010.01806.x

Robertson EP, Fletcher RJ Jr, Cattau CE, Udell BJ, Reichert BE, Austin JD, Valle D (2018) Isolating the roles of movement and reproduction on effective connectivity alters conservation priorities for an endangered bird. Proc Natl Acad Sci U S A 115(34):8591–8596

Roy M, Pascual M, Levin SA (2004) Competitive coexistence in a dynamic landscape. Theor Popul Biol 66(4):341–353. https://doi.org/10.1016/j.tpb.2004.06.012

Royle JA, Chandler RB, Gazenski KD, Graves TA (2013) Spatial capture-recapture models for jointly estimating population density and landscape connectivity. Ecology 94(2):287–294

Royle JA, Chandler RB, Sollmann R, Gardner B (2014) Spatial capture-recapture. Elsevier, Amsterdam

Royle JA, Fuller AK, Sutherland C (2018) Unifying population and landscape ecology with spatial capture-recapture. Ecography 41(3):444–456. https://doi.org/10.1111/ecog.03170

Rubio L, Bodin O, Brotons L, Saura S (2015) Connectivity conservation priorities for individual patches evaluated in the present landscape: how durable and effective are they in the long term? Ecography 38(8):782–791. https://doi.org/10.1111/ecog.00935

Rudnick DA, Ryan SJ, Beier P, Cushman SA, Dieffenbach F, Epps CW, Gerber LR, Hartter J, Jenness JS, Kintsch J, Mernlender AM, Perkl RM, Preziosi DV, Trombulak SC (2012) The role of landscape connectivity in planning and implementing conservation and restoration priorities. Issues Ecol 16:1–20

Runge JP, Runge MC, Nichols JD (2006) The role of local populations within a landscape context: defining and classifying sources and sinks. Am Nat 167(6):925–938. https://doi.org/10.1086/503531

Saerens M, Achbany Y, Fouss F, Yen L (2009) Randomized shortest-path problems: two related models. Neural Comput 21(8):2363–2404. https://doi.org/10.1162/neco.2009.11-07-643

Saura S, Pascual-Hortal L (2007) A new habitat availability index to integrate connectivity in landscape conservation planning: comparison with existing indices and application to a case study. Landsc Urban Plan 83(2–3):91–103. https://doi.org/10.1016/j.landurbplan.2007.03.005

Saura S, Rubio L (2010) A common currency for the different ways in which patches and links can contribute to habitat availability and connectivity in the landscape. Ecography 33(3):523–537. https://doi.org/10.1111/j.1600-0587.2009.05760.x

Saura S, Bodin O, Fortin MJ (2014) Stepping stones are crucial for species' long-distance dispersal and range expansion through habitat networks. J Appl Ecol 51(1):171–182. https://doi.org/10.1111/1365-2664.12179

Sawyer SC, Epps CW, Brashares JS (2011) Placing linkages among fragmented habitats: do least-cost models reflect how animals use landscapes? J Appl Ecol 48(3):668–678. https://doi.org/10.1111/j.1365-2664.2011.01970.x

Schumaker NH, Brookes A, Dunk JR, Woodbridge B, Heinrichs JA, Lawler JJ, Carroll C, LaPlante D (2014) Mapping sources, sinks, and connectivity using a simulation model of northern spotted owls. Landsc Ecol 29(4):579–592. https://doi.org/10.1007/s10980-014-0004-4

Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution 68(1):1–15. https://doi.org/10.1111/evo.12258

Snijders TAB (2011) Statistical models for social networks. Annu Rev Sociol 37:131–153. https://doi.org/10.1146/annurev.soc.012809.102709

Stevens VM, Baguette M (2008) Importance of habitat quality and landscape connectivity for the persistence of endangered natterjack toads. Conserv Biol 22(5):1194–1204. https://doi.org/10.1111/j.1523-1739.2008.00990-x

Strogatz SH (2001) Exploring complex networks. Nature 410(6825):268–276

Sutherland C, Fuller AK, Royle JA (2015) Modelling non-Euclidean movement and landscape connectivity in highly structured ecological networks. Methods Ecol Evol 6(2):169–177. https://doi.org/10.1111/2041-210x.12316

Thomas CD, Kunin WE (1999) The spatial structure of populations. J Anim Ecol 68(4):647–657. https://doi.org/10.1046/j.1365-2656.1999.00330.x
Tischendorf L, Fahrig L (2000) How should we measure landscape connectivity? Landsc Ecol 15 (7):633–641
Turlure C, Baguette M, Stevens VM, Maes D (2011) Species- and sex-specific adjustments of movement behavior to landscape heterogeneity in butterflies. Behav Ecol 22(5):967–975. https://doi.org/10.1093/beheco/arr077
Tyler JA, Hargrove WW (1997) Predicting spatial distribution of foragers over large resource landscapes: a modeling analysis of the Ideal Free Distribution. Oikos 79(2):376–386
Urban D, Keitt T (2001) Landscape connectivity: a graph-theoretic perspective. Ecology 82 (5):1205–1218
Urban DL, Minor ES, Treml EA, Schick RS (2009) Graph models of habitat mosaics. Ecol Lett 12 (3):260–273. https://doi.org/10.1111/j.1461-0248.2008.01271.x
Valle D, Cvetojevic S, Robertson EP, Reichert BE, Hochmair HH, Fletcher RJ (2017) Individual movement strategies revealed through novel clustering of emergent movement patterns. Sci Rep 7. https://doi.org/10.1038/srep44052
van Etten J (2012) gdistance: distances and routes on geographical grids. R package version 1.1-4. http://CRAN.R-project.org/package=gdistance
VanDerWal J, Shoo L, Januchowski S (2010) SDMTools: species distribution modelling tools: tools for processing data associated with species distribution modelling exercises. R package version 1.1
Vasudev D, Fletcher RJ Jr, Goswami VR, Krishnadas M (2015) From dispersal constraints to landscape connectivity: lessons from species distribution modeling. Ecography 38:967–978. https://doi.org/10.1111/ecog.01306
Verheyen K, Vellend M, Van Calster H, Peterken G, Hermy M (2004) Metapopulation dynamics in changing landscapes: a new spatially realistic model for forest plants. Ecology 85 (12):3302–3312
von Luxburg U, Radl A, Hein M (2014) Hitting and commute times in large random neighborhood graphs. J Mach Learn Res 15:1751–1798
Vuilleumier S, Bolker BM, Leveque O (2010) Effects of colonization asymmetries on metapopulation persistence. Theor Popul Biol 78(3):225–238. https://doi.org/10.1016/j.tpb.2010.06.007
Wang IJ, Bradburd GS (2014) Isolation by environment. Mol Ecol 23(23):5649–5662. https://doi.org/10.1111/mec.12938
Webster MS, Marra PP, Haig SM, Bensch S, Holmes RT (2002) Links between worlds: unraveling migratory connectivity. Trends Ecol Evol 17(2):76–83. https://doi.org/10.1016/s0169-5347(01)02380-1
Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York
Winfree R, Dushoff J, Crone EE, Schultz CB, Budny RV, Williams NM, Kremen C (2005) Testing simple indices of habitat proximity. Am Nat 165(6):707–717
Wright S (1943) Isolation by distance. Genetics 28(2):114–138
Zeigler SL, Fagan WF (2014) Transient windows for connectivity in a changing world. Movement Ecol 2(1):1–1. https://doi.org/10.1186/2051-3933-2-1
Zeller KA, McGarigal K, Whiteley AR (2012) Estimating landscape resistance to movement: a review. Landsc Ecol 27(6):777–797. https://doi.org/10.1007/s10980-012-9737-0
Zollner PA, Lima SL (1999) Search strategies for landscape-level interpatch movements. Ecology 80(3):1019–1030

Chapter 10
Population Dynamics in Space



10.1 Introduction

Populations and their dynamics have intrigued scientists for centuries (Malthus 1798). Population biology was initially primarily interested in changes in population numbers at single locations and why populations may persist or go extinct (Gause 1932; Lotka 1927). The relevance of life-history variation among populations and species was also heavily emphasized in being relevant for interpreting population dynamics and evolutionary change (Cole 1954; Pianka 1970).

As population biology continued to develop, experiments and theoretical development in the 1950s–1970s showed that spatial variation can be an important mechanism for population persistence and change (Huffaker 1958; Roff 1974). This work largely emphasized the role of movement and dispersal in altering population dynamics (Skellam 1951; Levin 1976). In the 1980s, an emphasis on spatial heterogeneity in resources and its impacts on birth and death rates flourished (Keddy 1981), leading to the concept of source–sink dynamics (Keddy 1981; Holt 1985; Pulliam 1988). Since that time, there has been a consistent interest in spatial population biology and its relevance for understanding the effects of environmental change.

Population biology also has a prominent place in conservation and management (Ehrlich and Daily 1993; Caughley 1994; Reed et al. 2002; Carroll et al. 2003; Traill et al. 2010). Preventing population extinction lies at the core of policy throughout the world (Mace et al. 2008; Schwartz 2008). Restoration strategies frequently emphasize restoring populations, and management aimed at reducing impacts of pest and invasive species focus on limiting population growth and spread (Sakai et al. 2001; McKay et al. 2005; Schrott et al. 2005). Over the past several decades, the role of space in population dynamics and trends has been illuminated (Hanski et al. 1996; Channell and Lomolino 2000; Wilson et al. 2004), and spatially structured population dynamics have emerged as a key component for conservation and management strategies (Lundberg et al. 2000; Chan et al. 2006).

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_10

Our general objective for this chapter is to provide an overview of the role of space on populations, with an eye toward conservation-relevant issues. To do so, we first discuss some common frameworks for understanding and conserving spatially structured populations. We use the term *spatially structured population* as a broad umbrella term for a population that includes some amount of spatial heterogeneity (Table 10.1). This term captures many spatially focused concepts in population biology. We then illustrate some of these concepts with data on spatial variation in abundance and colonization–extinction dynamics.

Table 10.1 Common population terms using in spatial population modeling

Term	Definition
Area-isolation paradigm	A common framework for metapopulation biology, where patch area is assumed to explain local extinction rates and patch isolation is assumed to explain local colonization rates.
Colonization	The establishment of a species in a formerly unoccupied habitat via dispersal.
Closed population	When a population does not change over a specified period of time (i.e., no movements in or out of the population and no births or deaths).
Extinction	When a species becomes absent from a formerly occupied patch or region. Can be "local," "regional," or for the entire species.
Extinction debt	The number of species committed to extinction, but are not yet extinct, following a disturbance event and/or land clearing.
Demography	The study of the size, structure, and distribution of populations, with a focus on the vital rates that alter these patterns (e.g., births, deaths).
Immigration credit	The number of species committed to eventual immigration following a disturbance event and/or land clearing.
Landscape demography	The demographic properties of populations and their drivers at multiple scales.
Local population	A group of individuals that are geographically distinct from other individuals. Similar to a "population," but term typically used to contrast with a metapopulation.
Metapopulation	A set of local populations connected through dispersal.
Open population	A population where over a specified period of time, births, deaths, immigration or emigration may occur. Contrast with closed population.
Population	A group of individuals of the same species that are genetically or geographically distinct from other groups of individuals.
Population projection model	A model that predicts the state of a population in a future time. Often used for population viability analyses.
Population viability analysis	A general approach to assess the risk of extinction of a population or species over a given time frame. Involves projecting changes in population structure over time, often under different environmental scenarios.
Pseudo-sink	A local habitat that appears to have characteristics of a sink but is above carrying capacity, such that as the local population size decreases below carrying capacity, it may no longer resemble a sink.
Rescue effect	When local extinction is reduced due to higher connectivity causing a greater number of immigrants into patches, which elevates local population size.

(continued)

Table 10.1 (continued)

Term	Definition
Sieve	A local habitat or population where births exceed deaths and emigration $\gg$ immigration, such that it may be unlikely to persist due to high rates of emigration.
Sink	A local habitat or population where deaths exceed births and immigration exceeds emigration, such that the population may not persist (or may persist at a lower population size) in the absence of immigration.
Source	A local habitat or population where births exceed deaths and emigration exceeds immigration, such that the population provides individuals to other populations.
Spatial synchrony	The correlated temporal fluctuations among populations, where typically the magnitude of correlation varies with geographic distance.
Spatially structured population	A population that has some form of spatial heterogeneity in terms of individual distributions and/or variation in demographic rates.
Stochastic patch occupancy model	A spatial modeling approach that focuses on modeling the presence and absence of species on discrete habitat patches as a Markov chain.

10.2 Key Concepts and Approaches

10.2.1 Foundational Population Concepts

Understanding spatial population biology requires understanding fundamental concepts for populations. It is beyond the scope of this book to go into details regarding population biology in general. Interested readers should consult Gotelli (2008), Begon et al. (2009), or Rockwood (2009). Here, we introduce only key concepts and issues that will be relevant for interpreting spatially structured populations.

Populations have been defined in a variety of ways, using both ecological and evolutionary criteria (Waples and Gaggiotti 2006). All definitions of populations generally assume that individuals within a population interact equally (in genetics literature, this assumption is generally referred to as panmixia). Population definitions can also vary as a function of spatial scale, an issue that has been emphasized in metapopulation biology (see Sect. 10.2.2.1).

Populations are often described as being either "open" or "closed". An *open population* is one in which dynamics may occur over a specified period of time, such as movement into (immigration) or out of (emigration) the population, or changes in numbers through birth and death processes. In contrast, a *closed population* is one in which populations numbers do not change over a given time period. Clearly, as the time period of interest increases in length, a formerly closed population may become an open one. Consequently, temporal scale is also important for interpreting populations and their dynamics. Open versus closed populations is an important distinction in spatial ecology, because open populations can be impacted through movement-related processes, a focus of spatially structured population biology.

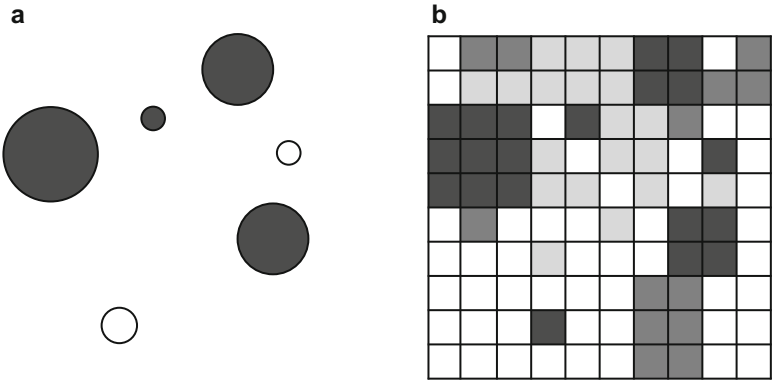


Fig. 10.1 Spatial population dynamics viewed through different spatial lenses. (**a**) A metapopulation approach that focuses on patch dynamics, where open circles represent unoccupied patches and filled circles represent occupied patches; and (**b**) spatially explicit approaches that focus on spatial demography and/or track individuals. Shown is a cellular automata that acknowledges gradients in land cover

10.2.2 Spatial Population Concepts

Spatial population concepts have a long history in ecology and conservation. In general, these ideas can be organized based on two different paradigms. First, there is a large branch of theory and empirical work that focuses on patch occupancy, *extinction*, and *colonization* dynamics (Fig. 10.1). This work ignores demographic vital rates, such as fecundity or survival, and tends to be emphasized when working across broad spatial and temporal scales (many locations or samples over time). This work has been the focus of classic metapopulation biology (Levins 1969; Hanski 1999). In contrast, *spatial demography*, or *landscape demography* (Gurevitch et al. 2016), focuses on variation in demographic rates over space and time. As a consequence of the challenging nature of estimating demographic rates at multiple locations, this focus has largely occurred at relatively small spatial and/or temporal scales (Norris 2004), where demographic rates are measured and interpreted at a small collection of patches or locations (Table 10.2).

10.2.2.1 Metapopulations

Metapopulation Theory. Metapopulation concepts and theory have been essential to the development of spatial population biology. Metapopulation concepts tend to distinguish between local dynamics occurring at the patch level versus regional (or metapopulation) dynamics. For instance, patches can suffer high probabilities of local extinction even though the likelihood of regional extinction is low.

Simple probabilistic processes can illustrate why metapopulations are so important (Gotelli and Kelley 1993). For instance, if we assume that the probability of

Table 10.2 Different paradigms for spatially structured populations

Characteristic	Metapopulation	Spatial demography
State variable(s)[a]	Occupancy, colonization, extinction	Demographic rates, population growth
Number of patches or locations	Many	Few
Ability to capture individual variation	No	Yes
Dispersal component	Colonization, rescue effects	Immigration, emigration
Consideration of spatial heterogeneity in habitat quality and the matrix	Limited to moderate	Highly variable
Types of models	Stochastic patch occupancy, Spatially structured Levins' model	Matrix models, Cellular automata, Individual-based models

[a]Major elements of the model whose rates of change are given by differential or difference equations

local extinction in one year is p_e, and that local extinction probabilities over time are independent, then the probability of persistence over 2 years is:

$$(1 - p_e)(1 - p_e). \tag{10.1}$$

For t years, the probability of persistence is then:

$$(1 - p_e)^t. \tag{10.2}$$

If we consider a situation where there are N patches with p_e for each patch, the probability of metapopulation persistence in one year is:

$$1 - (p_e)^N. \tag{10.3}$$

From this simple equation, the importance of multiple patches in a metapopulation becomes clear (Fig. 10.2). The above equation arrives at this conclusion by assuming that each patch operates independently, such that there is no *spatial synchrony*. It is essentially calculating the probability that at least one patch does not go locally extinct (hence, 1 minus the product of each going extinct). This probability of persistence can be calculated for different numbers of patches, illustrating the importance of "spreading the risk" for population viability (den Boer 1968).

Levins (1969) first developed the ideas of metapopulation dynamics in his classic metapopulation model. In this model, Levins distinguished a *local population* from the *metapopulation*. A local population is similar to the population concept mentioned above, while a metapopulation is a "population of populations." In this model, Levins assumes that local patches are occupied as a balance of extinction, E (proportion of sites going extinct/time), and immigration rates I (proportion of sites

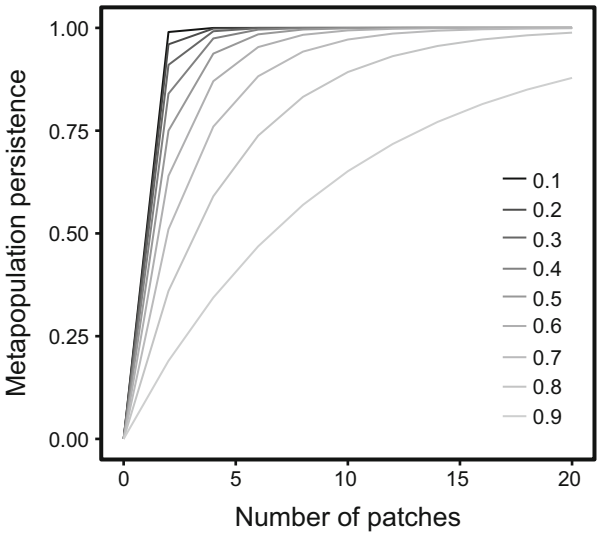


Fig. 10.2 Metapopulation persistence over 1 year, as a function of the number of patches in the metapopulation. Each line represents a local patch extinction probability

colonized/time) of local populations. We can formalize the change in the fraction of patches occupied, f, per unit time with a differential equation as:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = I - E. \tag{10.4}$$

As a consequence, when metapopulations are at equilibrium (i.e., no change in f), we set the left-hand side of Eq. (10.4) to zero, such that $E = I$.

Levins took this idea and then assumed that immigration rates would be a function of the number of sites occupied (providing propagules) and the proportion of sites unoccupied that could become colonized $(1 - f)$:

$$I = cf(1 - f), \tag{10.5}$$

where c is the rate of individuals emigrating from occupied patches. Thus, I is a product of c and f, or the total rate of individuals emigrating from all occupied patches (i.e., internal colonization is permitted, unlike the basic model of island biogeography; see Chap. 11), times $(1 - f)$, or the proportion of sites that could be colonized. Levins then assumed that:

$$E = ef, \tag{10.6}$$

where e is the probability of local extinction, such that the extinction rate for the entire metapopulation is the product of e times the sites occupied (because only occupied sites can become extinct). Putting this together, we have the Levins model:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = cf(1-f) - ef. \tag{10.7}$$

With this model, we can determine the equilibrium fraction of sites occupied, f^*, by setting $\mathrm{d}f/\mathrm{d}t = 0$ and solving the equation, which becomes:

$$f^* = 1 - \frac{e}{c}. \tag{10.8}$$

Consequently, this model predicts that the metapopulation will only persist ($f^* > 0$) when $c > e$.

There are several assumptions of this classic metapopulation model. The model assumes that patches are homogeneous in their characteristics (e.g., area, quality, isolation), there is no spatial structure, patches are asynchronous in their dynamics (i.e., they have independent colonization and extinction dynamics), c and e are constant, there are no time lags, and there is a large number of patches (Gotelli 2008). There are several variants of this model, such as altering the formulation to allow for *rescue effects*, or the effects of recurrent immigration lowering extinction rates, and a focus on external propagule pressure only (similar to island biogeography; Gotelli 1991).

One assumption that has received a great deal of interest is the assumption that local populations are not synchronized in their dynamics (Ranta et al. 1997; Koenig 1998; Bjørnstad et al. 1999; Liebhold et al. 2004; Koenig and Liebhold 2016; Walter et al. 2017). Understanding spatial synchrony (or lack thereof) is of interest for a variety of reasons. First, high spatial synchrony may lead to greater vulnerability of species to environmental stochasticity and it can increase the likelihood of local and metapopulation extinction (Heino et al. 1997; Matter 2001). Second, there is interest in why synchrony occurs (Bjørnstad et al. 1999): does it occur due to localized dispersal, spatial dependence in trophic interactions, or through abiotic (e.g., weather) variation limiting population dynamics (which is known as the Moran effect)?

The metapopulation framework has also been recasted at the individual level. For instance, Tilman and Lehman (1997) considered the problem where "local populations" are considered to be discrete locations on a landscape capable of only housing a single individual, similar to a territory of an individual bird or occupancy of a sessile individual, where priority effects and preemption prevents other individuals from also occupying that location.

Hanski (1994) relaxed some of these assumptions and developed a framework of applying these metapopulation ideas to real data by developing what he termed the incidence function model, one type of *stochastic patch occupancy model* (Caswell and Etter 1993; Moilanen 1999). In the incidence function model (IFM), colonization and extinction rates are assumed to vary across patches as a function of patch size and isolation, thereby relaxing assumptions regarding homogenous patches, no spatial structure and constant c and e. This general set of assumed relationships in

colonization–extinction dynamics is often referred to as the *area-isolation paradigm* (Hanski 1998; Fleishman et al. 2002).

Hanski assumed that local extinction rates were dependent on population size. He specified the functional form of extinction rate for patch i with patch area A as:

$$E_i = \frac{z}{A_i^x}, \tag{10.9}$$

where z and x are constants. Based on this equation, as patch area increases, the likelihood of extinction in patch i declines, but the functional form of the decline varies with z and x. When x is small (<1) even large patches are expected to suffer extinction; z can be considered a parameter that captures the magnitude of environmental stochasticity on patch extinction (Fig. 10.3). In general, extinction–area relationships are commonly assumed in population biology. There are at least three reasons for this relationship: (1) larger areas contain greater numbers of individuals (i.e., larger population size); (2) larger areas have greater spatial heterogeneity in resources, such that as conditions change, there may still be resources available somewhere within patches; and (3) larger areas allow for dynamics akin to metapopulation dynamics that operate within patches (Connor and McCoy 1970; Holt 1992; Hanski 1999). A common extension of this framework is to adjust patch area with a measure of "effective area" if patches vary in habitat quality and/or edge effects (Moilanen and Hanski 1998).

Colonization rates for patch i, C_i, in the incidence function model were based on spatial isolation and were quantified as a function of the number of potential immigrants, M_i:

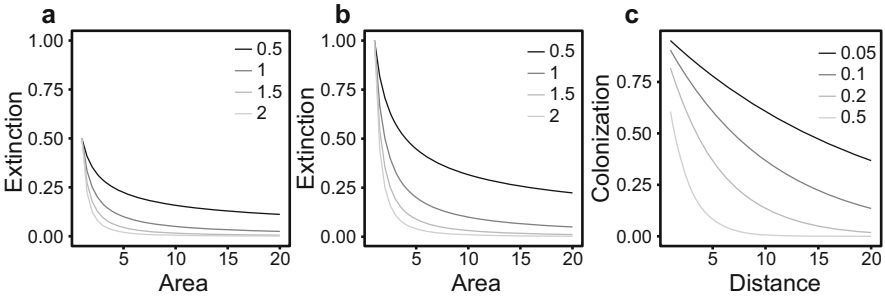


Fig. 10.3 Functional relationships assumed in the incidence function model and related stochastic patch occupancy models. Extinction relationships are often formalized as $E_i = \frac{z}{A_i^x}$, where z and x are constants that modify the shape of the extinction relationship E with patch area, A. (**a**) Extinction relationships when z is varied, (**b**) extinction relationships were x is varied. For variation in colonization, a negative exponential kernel, $\exp(-\alpha d_{ij})$, is frequently used, where d_{ij} is the distance between patch i and j and α is the inverse of the mean dispersal distance. (**c**) Changes in the negative exponential kernel under different values of α

$$C_i = \frac{M_i^2}{M_i^2 + y^2}, \tag{10.10}$$

where y is a constant that determines how fast the colonization rate increases with the number of immigrants. M was quantified as:

$$M_i = \beta S_i \tag{10.11}$$

$$S_i = \sum_j p_j \exp(-\alpha d_{ij}) A_j, \tag{10.12}$$

where p_j is an indicator variable regarding whether patch j is occupied, α is the inverse of the mean dispersal distance for the species, and β is a constant to be estimated. As α increases, the effect of distance increases (Fig. 10.3c). S_i essentially sums the areas of occupied patches, weighting them by their distance to the focal patch j (where the weight is described by dispersal distance). Note that S_i has frequently been used as a potential connectivity metric (Moilanen and Nieminen 2002) and d_{ij} has sometimes been replaced with effective distance measures, such as least-cost distances that capture matrix resistance (see Chap. 9).

The above formulation emphasizes a negative exponential dispersal kernel for colonization (Fig. 10.3). Hanski (1999) argued that the exact for dispersal kernel is often not that important in metapopulation dynamics, because such dynamics are frequently driven by relatively short dispersal events. However, for some problems, such as range expansion and the spread of invasive species, this function may not be warranted, particularly when dispersal has a "fat tail": when there are occasional very long-distance colonization events (see Chap. 9). A variety of fat-tail relationships have been suggested (Clark et al. 1999); an approach that is a minor extension to the negative exponential function above is (Shaw 1995; Moilanen 2002):

$$\frac{1}{1 + \alpha d_{ij}^{\gamma}}, \tag{10.13}$$

where γ is a constant.

With these descriptions of extinction and colonization rates, Hanski then specified an equilibrium measure of incidence or occurrence, J_i, as:

$$J_i = \frac{C_i}{C_i + E_i}, \tag{10.14}$$

(see also Diamond 1972 for a similar approach in the context of island biogeography). This last point was provocative, in the sense that this relationship could be used to potentially take a snapshot of occurrence data to then estimate C_i and E_i. However, doing so makes the assumption that the metapopulation is in long-term, quasi-equilibrium, such that the metapopulation is persistent (Moilanen 2000). Removing

this assumption means that any observed trend in metapopulation occupancy is real (i.e., not simply a stochastic realization of a population at quasi-equilibrium). There are consequences for either assumption. This general idea has been applied in a variety of conservation-focused problems, such as interpreting long-term effects of deforestation (Ferraz et al. 2007) and occupancy dynamics of Golden Eagles (*Aquila chrysaetos*; Martin et al. 2009).

This metapopulation framework has been applied to a variety of species and systems. For example, Hokit et al. (2001) applied this approach to the Florida scrub lizard (*Sceloporus woodi*), an endangered lizard found in patchily distributed uplands in Florida. They contrasted predictions from the incidence function model to a *population projection model* (see Sect. 10.2.4.3), finding that both types of models provided generally similar results of the predicted distribution of this species.

A related topic from metapopulation theory is the concept of metapopulation capacity (Adler and Nuernberger 1994; Hanski and Ovaskainen 2000). Metapopulation capacity was derived from extensions of the Levins' model to generalize to interpreting metapopulation persistence in irregular landscapes. Hanski and Ovaskainen (2000) popularized this idea by showing how this concept could be used to evaluate metapopulation viability based on what they termed an appropriate landscape matrix. They defined the landscape connectivity matrix $\mathbf{M}$ as:

$$\mathbf{M} = A_i A_j \Big(\exp(-\alpha d_{ij}) \Big), \tag{10.15}$$

and $\mathbf{M}_{ii} = 0$. The metapopulation capacity is defined as the leading eigenvalue of $\mathbf{M}$, $\lambda_{\mathbf{M}}$. Patch importance to metapopulation capacity, P_i, can be approximated using the leading eigenvectors of $\mathbf{M}$, x:

$$P = x^2 \lambda_{\mathbf{M}}. \tag{10.16}$$

Note the similarity of using this measure of patch importance to that described in Chap. 9 (using eigenvector centrality of a connectivity matrix). In addition, we highlight that metapopulation capacity shares a similar functional form to the probability of connectivity described in Chap. 9 (Saura and Pascual-Hortal 2007), as well as the neighborhood habitat area metric of Hanski (1999). See Saura and Rubio (2010) for a discussion on some of the similarities and differences between these metrics. Metapopulation capacity can be applied to any landscape where patch locations and sizes are known and an estimate of α exists. It can be used to evaluate issues of both habitat loss and fragmentation (Fig. 10.4), as well as potential habitat restoration (Hanski and Ovaskainen 2000; Ovaskainen et al. 2002).

Another approach for interpreting metapopulation viability is to quantify mean metapopulation lifetime (Frank and Wissel 2002). This metric has a similar goal to the metapopulation capacity, but it is an absolute metric that assumes stochastic, rather than deterministic, extinction. The metapopulation mean lifetime is derived from a stochastic metapopulation model that aims to approximate the viability of metapopulations. This model was initially developed by Frank and Wissel (2002)

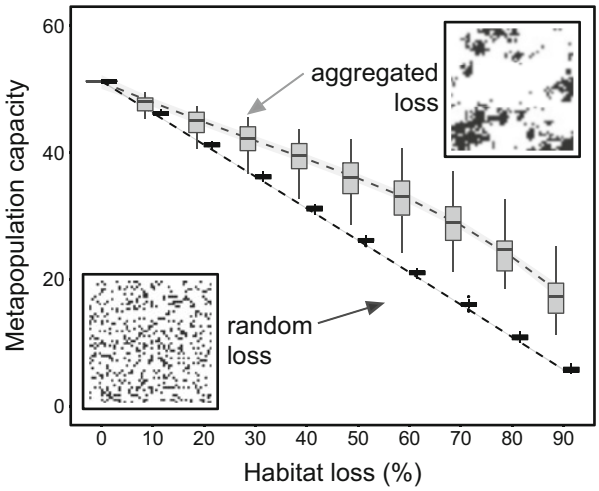


Fig. 10.4 An illustration of lessons from metapopulation theory for habitat loss and fragmentation, where random habitat loss is predicted to have negative effects on metapopulation capacity relative to aggregated loss. Shown is an example where random and aggregated loss occur. In this example, we use Gaussian Random Fields (see Chap. 5) to generate landscapes. Metapopulation capacity is calculated under the two scenarios of habitat loss. Shown are 20 realizations (with boxplots and lowess ribbons illustrating variation across realizations). In this situation, dispersal limitation is imposed by assuming a small mean dispersal distance relative to the landscape size and cells (rather than patches) are used in quantifying metapopulation capacity (such that patch area does not vary). If dispersal is not limited, predictions for aggregated and random loss converge

and later extended to metapopulations on complex networks (Drechsler 2009; Drechsler and Johst 2010; Kininmonth et al. 2010; Johst et al. 2011). The model makes similar assumptions to metapopulation capacity regarding area-isolation relationships on extinction and colonization probabilities, yet contains more parameters to describe extinction–colonization dynamics and effects of environmental variability. It has been applied to a variety of conservation problems (Drechsler and Johst 2010; Johst et al. 2011).

Conservation Lessons from Metapopulation Theory. Metapopulation theory emphasizes that currently unoccupied patches can be critical for long-term metapopulation persistence (Hanski 1998). The colonization–extinction dynamics captured by metapopulation theory highlight that unoccupied patches may be colonized in the future. This issue is important in interpreting species distribution (Barve et al. 2011), habitat quality (Mortelliti et al. 2010), and conservation strategies (Bulman et al. 2007).

Metapopulation theory also suggests that minimum viability metapopulation sizes may occur (Hanski et al. 1996), where a minimum number of patches may be required to ensure long-term metapopulation persistence. This general idea extends arguments in conservation biology regarding minimum viable populations required for long-term persistence (Flather et al. 2011; Traill et al. 2007).

Several spatial aspects of metapopulation networks have been suggested to be important for metapopulation persistence and how metapopulations should be managed. For example, Frank and Wissel (1998) emphasized the importance of the correlation length of local extinction processes (or the spatial correlation in extinction) relative to the dispersal range of species in metapopulation dynamics. From their model development, if the dispersal range is shorter than the correlation length of extinction processes, then metapopulation viability over the long term may be low.

A variety of metapopulation models have focused on patterns of habitat destruction, suggesting that critical thresholds of loss occur for metapopulation persistence (Bascompte and Sole 1996). In addition, many models suggest that the pattern of destruction matters, where random loss is often expected to be more detrimental to metapopulation persistence than non-random (often, aggregated) loss (Fig. 10.4) (Dytham 1995; Moilanen and Hanski 1995). Some metapopulation modeling has also considered the roles of patch loss relative to the reduction in patch size on metapopulation persistence. Patch loss ultimately impacts colonization rates, while reductions in patch size are expected in increase local extinction rates and reduce colonization rates (Hanski 1999), such that reductions in patch size may have greater effects than patch removal for a given amount of loss (Hanski and Ovaskainen 2000). Habitat disturbance (rather than destruction) has also been considered extensively, where the effects of disturbance vary depending on the spatiotemporal patterns that disturbance creates (Johst and Drechsler 2003; Kallimanis et al. 2005). In these applications, it is often emphasized that *extinction debts* may occur: when species are committed to extinction (but have not yet gone extinct) from disturbance or habitat destruction (Tilman et al. 1994; Hanski and Ovaskainen 2002; Kuussaari et al. 2009; Hylander and Ehrlen 2013). *Immigration credits* may also occur and are relevant for conservation, where species may eventually immigrate (or colonize) an area after a disturbance or restoration event (Jackson and Sax 2010; Talluto et al. 2017).

Limitations. While metapopulation theory and application have seen widespread use, there are recurrent questions regarding its relevance to many systems (Harrison 1991; Baguette 2004). Hanski (1999, 2004) argued that the metapopulation approach is potentially useful when three criteria exist. First, space is discrete (e.g., patchy habitats). Second, processes operate on (at least) two scales: the local population and metapopulation scales. Third, habitat is large and permanent enough to allow persistence of breeding local populations for at least a few generations.

10.2.2.2 Spatial Demography

Metapopulation ideas regarding colonization–extinction dynamics can be contrasted with interpreting spatial variation in demography and related vital rates, such as births (B), deaths (D), emigration (E), and immigration (I). This focus was at the core of ideas regarding source–sink dynamics (Holt 1985), balanced dispersal models

(Diffendorfer 1998), and what is termed "structured metapopulation models" (Diekmann 1993).

Source–Sink Dynamics. The concept of source–sink dynamics was popularized in a seminal article by Ron Pulliam (1988), but other important early work comes from Holt (1985) and Paul Keddy (1981, 1982). Pulliam focused on how vital rates can explain population size N at time t as:

$$N_t = N_{t-1}(B + I - D - E), \tag{10.17}$$

such that:

$$\lambda_t = N_t / N_{t-1} = B + I - D - E, \tag{10.18}$$

where λ_t is the finite rate of increase (or loosely, the population growth rate). Pulliam defined a *source* where $\lambda_t > 1$ and a *sink* where $\lambda_t < 1$, such that $B > D$ and $E > I$ for sources, whereas for sinks $B < D$ and $E < I$. However, after defining sources and sinks in this way, he then focused solely on variation in birth and death processes, ignoring the roles of emigration and immigration. That shift in focus led to much confusion in spatial demography (Runge et al. 2006). Sources and sinks have also sometimes been coined as "net exporters" and "net importers" (focusing on the E and I components), which has further muddled the concept. Two related concepts are *pseudo-sinks*, which are locations that appear to be sinks but are essentially above carrying capacity, and *sieves* (Cronin 2007), where emigration rates are very high, such that a population with high birth rates may go locally extinct due to high emigration.

Thomas and Kunin (1999) generalized these ideas to spatially structured populations by noting that most populations can be described along two dimensions, a mobility axis and a compensation axis. The mobility axis relates to how a population is involved in regional (immigration and emigration) relative to local (births, deaths) population dynamics. The compensation axis describes whether populations are exporting individuals (sources, sieves) or importing them (e.g., sinks, pseudo-sinks). This framework emphasizes a continuum in the dynamics and relationships of vital rates.

Overall, the source–sink concept has provided a great deal of insights to population biology and conservation. Pulliam (1988) emphasized several novel insights. First, the source–sink concept acknowledges that when individuals reside in sinks, those environments are outside of the fundamental niche of a species (Pulliam 2000). Second, source–sink dynamics can complicate the interpretation of potential species interactions. For instance, if a habitat is a source for one species but a sink for another, one might inappropriately conclude that asymmetric interspecific competition is occurring between species. Third, when sources occur, classic metapopulation dynamics may be less likely, because local extinction of sources may be rare. Those situations may be more akin to "island–mainland" metapopulations (Gotelli 1991). Finally, this concept has been very important in regard to habitat prioritization for

conservation purposes, where practitioners may be less willing to invest in sink habitats for prioritization (Liu et al. 2011).

Balanced Dispersal. The balanced dispersal model for spatial demography shares some similarities with source–sink models but it makes different assumptions (Diffendorfer 1998). In the balanced dispersal model, sites vary in their carrying capacities but sinks do not occur (McPeek and Holt 1992). Dispersal rates (emigration from patches) are inversely related to local carrying capacities, such that dispersal probability is high from sites with low carrying capacities and dispersal probability is low from sites with high carrying capacities. Consequently, the number of individuals dispersing between patches becomes "balanced" (Diffendorfer 1998). This model is essentially an extension of the Ideal Free Distribution model (see Chap. 8). This model has received some empirical support over the years (Doncaster et al. 1997).

Landscape Demography. The ideas regarding source–sink dynamics and related issues have recently coalesced into the concept of *landscape demography* (Gurevitch et al. 2016). Landscape demography is "the study of demographic properties of populations and their drivers at multiple scales, and of how the relationships among populations and their drivers at any one scale influence demographic outcomes at other scales" (Gurevitch et al. 2016). Issues of scaling in population demography have long been acknowledged (Thomas and Kunin 1999), although empirical work on these issues has been slow to accumulate (but see Cavanaugh et al. 2014).

10.2.3 Population Viability Analysis

In a conservation context, *population viability analysis* (PVA) is often used to interpret the likelihood of (quasi) extinction in populations (Morris and Doak 2002; Reed et al. 2002). Several types of PVAs have been formulated over the years, many of which use different types of data and information for modeling. The utility of PVA has been debated; however, PVA can be useful particularly when applied in a relative way by comparing alternative scenarios of future change rather than for absolute predictions or forecasting of viability (Beissinger and Westphal 1998; Reed et al. 2002). In addition, PVA has been shown to accurately predict population viability in many systems (Brook et al. 2000).

For spatially structured populations, four general approaches tend to be used (Morris and Doak 2002). First, time-series data of changes in population abundance across locations can be used. Second, stochastic patch occupancy models are often used. Third, site-specific demographic data can be used, such as site-specific variation in birth rates or death rates. Finally, spatially explicit, individual-based models can be used.

Time-series data of changes in population abundance across locations are often available from monitoring programs. These data can be harnessed to understand and predict changes in abundance over time. If locations can be considered independent,

then estimates of population change can be quantified for each population using standard, non-spatial approaches (Morris and Doak 2002). With such estimates, the likelihood of extinction across all sites can be calculated with probability theory. If population dynamics across locations are correlated, then this correlation needs to be accounted for because it can have strong effects on population viability. For instance, if there are strong correlations among locations, then strings of "bad years" will occur spatially across sites, increasing the likelihood of extinction. In contrast, temporally autocorrelated environments can sometimes lead to population growth based on strings of "good years," termed inflationary effects (Gonzalez and Holt 2002; Roy et al. 2005; Matthews and Gonzalez 2007). We would expect correlations to arise across sites due to dispersal processes or spatial dependence in abiotic and biotic conditions across sites.

Stochastic patch occupancy models (SPOMs) share some similarities with the abundance approach, but work with sparser information in terms of changes in occurrence and these approaches tend to emphasize metapopulation processes, particularly local colonization and extinction dynamics (Moilanen 2004). The focus is on using time-series of occurrence or occupancy to estimate colonization and extinction dynamics. These estimates can then be used to project dynamics forward in time to interpret population viability. The incidence function model (Hanski 1994) described above is one example of a SPOM. The value of this approach is its ability to work with relatively limited data, which can potentially be collected across many locations in a region.

The third approach is to use demographic population viability analyses. These approaches typically involve the use of population matrix models (Caswell 2001). Population matrix models are widely used in population ecology to estimate the finite rate of increase (~population growth rate), and to understand which demographic parameters (e.g., juvenile survival) limit population growth via various methods, such as sensitivity and elasticity analysis and life table response experiments (de Kroon et al. 2000). With these matrices, population dynamics can be projected forward in time to interpret population viability. In addition, both asymptotic (i.e., long-term) and transient dynamics can be interpreted (Ezard et al. 2010). These types of approaches can be used to infer the roles of local demography relative to movement (immigration and emigration) on population dynamics at different scales (Runge et al. 2006).

Finally, spatially explicit, individual-based models are often used in population viability analyses (Dunning et al. 1995; DeAngelis and Mooij 2005). These models typically couple spatially explicit maps, such as raster grids, with individual-based (or agent-based) algorithms regarding dispersal, reproduction and survival (Grimm et al. 2005). Individual-based approaches allow for straightforward inclusion of individual variation in population processes (e.g., individual variation in dispersal or reproduction), as well as spatial variation in composition, configuration, and quality of habitats (Fahrig 1997, 2001; Wiegand et al. 2005). Yet these approaches are often very complex, require detailed data, and are problem-specific, making it difficult to extend results to interpret generalizable patterns and processes (Norris 2004).

10.2.4 Common Types of Spatial Population Models

10.2.4.1 Stochastic Patch Occupancy Models

Stochastic patch occupancy models (SPOMs) have been developed in a variety of ways (Moilanen 2004), with the Incidence Function Model being among the first (Hanski 1994). The spatially realistic Levins' model is another example (Hanski 1999), as is the dynamic occupancy model (MacKenzie et al. 2003).

SPOMs are discrete time, first-order Markov chain models. They are discrete time in the sense that they are typically applied with temporal data that occurs across years or seasons. First-order Markov chain models are those in which the state at time t depends only on the state in time $t - 1$ (e.g., a second-order Markov chain would depend on $t - 1$ and $t - 2$). The Markov chain has 2^N possible states, where N is the number of patches. It is 2^N because in each patch there are two possible states (occupied or unoccupied), yield a total combination of 2^N. There has historically been a focus on the area-isolation paradigm when modeling states; however, other factors have long been considered as well (Moilanen and Hanski 1998). Once the parameters of SPOMs have been estimated, simulations can be used to project potential metapopulation dynamics over time.

There have been three major limitations of these models (Moilanen 2002). First, they require that all sites are sampled. If not all sites are sampled, estimates for colonization can be biased, where colonization probability tends to be over-estimated. Second, they assume that area is properly measured. If there are errors in estimates of patch size, the scaling of occupancy and area is expected to decrease. Third, they assume that sites are accurately surveyed, with no observation error (i.e., sampling error). When false negative error rates occur, where sites that are occupied are estimated as being unoccupied, this issue can cause large errors in SPOMs (Moilanen 2002). Dynamic occupancy models address this later issue.

10.2.4.2 Dynamic Occupancy Models

The stochastic patch occupancy framework can be generalized to issues of sampling error using state-space models. State-space models distinguish ecological processes (states) with observations (space), which are related to the state but may have error and/or bias associated with them. Because the states, here z_i, the occupancy of a species at site i, are generally only partially observed, we treat them as latent variables: unmeasured (or indirectly measured) variables for which we are interested in estimating. More broadly, $z_i \sim \text{Bernoulli}(\psi_i)$, where ψ_i is the probability of occurrence at site i (a Bernoulli distribution is one form of a binomial distribution where only one trial occurs). This general framework in the context of extinction–colonization dynamics is referred to as a dynamic occupancy modeling framework (MacKenzie et al. 2003). This framework was extended from earlier efforts that focused on colonization and extinction estimation but ignored the problem of

detectability (Erwin et al. 1998; Clark and Rosenzweig 1994; Erwin et al. 1998). Here, we focus less on the problem of observation error, because it has been covered extensively elsewhere (MacKenzie et al. 2003, 2006; Royle and Kery 2007); rather, we use this general framework because of its flexibility when making inferences from ecological data.

The dynamics of species occupancy can be described and estimated with time-series data by assuming a first-order Markov process, where z_i at time t is contingent on z_i at time $t-1$, as well as local colonization, γ, and local extinction, ε, processes. If we define $\phi = 1 - \varepsilon$, then:

$$z_{i,t} = \text{Bernoulli}\left(z_{i,t-1}\phi_{i,t-1} + (1 - z_{i,t-1})\gamma_{i,t-1}\right). \tag{10.19}$$

In the above equation, when $z_{i,t-1} = 0$, then the first term on the right side of the equation goes to zero and $z_{i,t}$ comes from a Bernoulli distribution with a probability of $\gamma_{i,t-1}$. When $z_{i,t-1} = 1$, the second term on the right side of the equation goes to zero, such that $z_{i,t}$ now comes from is a Bernoulli distribution with a probability $1 - \varepsilon_{i,t-1}$, or 1 minus the probability of local extinction. In this formulation, the observation process states that we may not perfectly estimate z based on detectability issues, p, such that our observed data, $\mathbf{Y}$, is a function of detectability, where $\mathbf{Y}_t \sim z_t p$. Detectability can vary over time and/or across sites.

To implement this model, we need to estimate z at $t = 1$ and subsequent colonization and extinction parameters. Covariates can be included for z, γ, and ε (and p), providing a great deal of flexibility in the modeling process and interpretation of colonization–extinction dynamics. This framework has also been implemented to generalize the incidence function model (Risk et al. 2011).

10.2.4.3 Spatial Population Matrix Models

There is a huge literature on population matrix models; interested readers should consult Caswell (2001). We briefly illustrate some ways in which these models can be applied to spatially structured populations. First, consider a one-site, two-stage model (a stage is a class of the life cycle of a species where demographic rates are assumed to be similar among individuals) that incorporates a juvenile j and adult a life stage. This scenario can be written in matrix notation as:

$$N_{t+1} = \begin{pmatrix} N_{j,t} \\ N_{a,t} \end{pmatrix} \begin{pmatrix} \beta S_j & \beta S_a \\ S_j & S_a \end{pmatrix}, \tag{10.20}$$

where N_{t+1} is the number of individuals at time $t + 1$, S_j is the survival of juveniles, S_a is the survival of adults, and β is the fecundity of adults (i.e., the number of young an individual produces at time t; often the focus is on the number of females produced by adult females). This 2×2 demographic transition matrix in Eq. (10.20) is often referred to as $\mathbf{A}$. This formulation is based on a "pre-breeding

census," where we survey the population just prior to breeding. Note that for some species, "states" are more continuous (e.g., plant size), which can be honored through the use of integral projection models. The matrix in Eq. 10.20 can be rewritten as a set of difference equations to better interpret how the matrix model works:

$$\begin{aligned} N_{j,t+1} &= N_{j,t}\left(\beta S_j\right) + N_{a,t}(\beta S_a) \\ N_{a,t+1} &= N_{j,t}\left(S_j\right) + N_{a,t}(S_a). \end{aligned} \tag{10.21}$$

A simple extension of this model to capture a situation where a species resides in a two-patch system that potentially has source–sink dynamics can be described as (Stevens 2009):

$$N_{t+1} = \begin{pmatrix} N_t^1 \\ N_t^2 \end{pmatrix} \begin{pmatrix} S_a^1 + \beta^1 S_j^1 & M_{12} \\ M_{21} & S_a^2 + \beta^2 S_j^2 \end{pmatrix}, \tag{10.22}$$

where M_{12} is the migration of individuals from site 2 to site 1 (Fig. 10.5). In the above matrix, we assume a pre-breeding census.

Such models can be extended to multiple locations by expanding out the matrix notation (Caswell 2001) into four sub-matrices. In the above example, we can rewrite our matrix model to include two sites, 1 and 2, with no movement between them as:

$$A = \begin{pmatrix} \beta_1 S_1^j & \beta_1 S_1^a & 0 & 0 \\ S_1^j & S_1^a & 0 & 0 \\ 0 & 0 & \beta_2 S_2^j & \beta_2 S_2^a \\ 0 & 0 & S_2^j & S_2^a \end{pmatrix}. \tag{10.23}$$

We can account for movement between sites as:

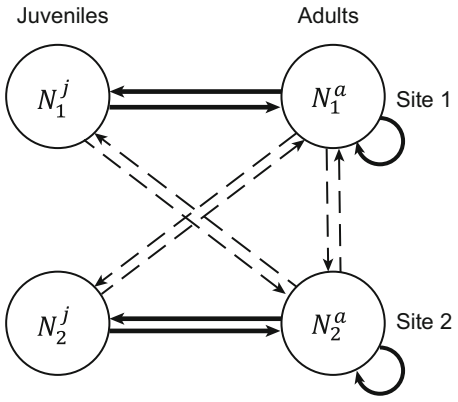


Fig. 10.5 Spatial demography as a life-cycle graph for a simple, stage-structured population with two age classes, juveniles, j, and adults, a, at two sites. Solid lines show local recruitment and survival, while dashed lines show effect of migration/movement between sites 1 and 2. Note that for adults, survival and fidelity can occur (a self loop)

$$A = \begin{pmatrix} \beta_1 S_{11}^j & \beta_1 S_{11}^a & \beta_2 S_{21}^j & \beta_2 S_{21}^a \\ S_{11}^j & S_{11}^a & S_{21}^j & S_{21}^a \\ \beta_1 S_{12}^j & \beta_1 S_{12}^a & \beta_2 S_{22}^j & \beta_2 S_{22}^a \\ S_{12}^j & S_{12}^a & S_{22}^j & S_{22}^a \end{pmatrix}. \tag{10.24}$$

In this latter parameterization, we keep track of whether individuals survive and stay in their location (S_{ii}) versus survive and move (S_{ij}). Other parameterizations of spatially structured matrix models have also been developed, including those that decompose matrices into local dynamics and movement, models that formulate different types of movement such as diffusion and patch-specific movement rates, and models that incorporate dispersal kernels directly (Lebreton 1996; Lebreton et al. 2000; Neubert and Caswell 2000; Caswell et al. 2003; Ozgul et al. 2009).

Spatial population matrix models are powerful approaches to interpret spatially structured population dynamics and the factors that govern them. For example, Cattau et al. (2016) used these approaches to interpret how an invasion of non-native prey was impacting the population growth rate of the endangered snail kite (*Rostrhamus sociabilis*) in the United States, finding that the invasive prey increased survival and altered movement, leading to increases in population growth. These types of models have been fruitfully applied to interpret the spread of invasive species as well (Jongejans et al. 2011). Despite this utility, these approaches are less frequently used than other spatial models due to the large amount of information needed to parameterize them.

10.3 Examples in R

10.3.1 Packages in R

In R, there are a few libraries that can be used for spatial population dynamics. Non-spatial demographic models can be constructed with the `popbio` and `popdemo` packages (Stubben and Milligan 2007; Stott et al. 2018). `MetaLandSim` has some functionality for fitting some types of stochastic patch occupancy models and exploring their predictions (Mestre et al. 2016). `secr` provides functions for spatial capture–recapture models (Efford 2018). `unmarked` provides functions for estimating extinction–colonization dynamics (Fiske and Chandler 2011). Population synchrony can be estimated with the `ncf` or `synchrony` packages (Bjørnstad and Falck 2001; Gouhier and Guichard 2014).

10.3.2 The Data

To illustrate spatial population dynamics, we consider a spatial time-series of patch dynamics for the wind-dispersed orchid, *Lepanthes rupestris*, over a 5-year period (Acevedo et al. 2015). The dynamics of this species can be understood in a metapopulation context, because *L. rupestris* lives in spatially discrete, ephemeral habitats, where colonization–extinction dynamics have been observed (Tremblay et al. 2006; Kindlmann et al. 2014). Moreover, many epiphytic and epilithic orchids are subject to colonization and extinction dynamics due to their small population sizes and stochastic reproductive success driven, in part, by dispersal and pollinator limitation (Ackerman et al. 1996; Tremblay 1997; Tremblay and Ackerman 2001).

Lepanthes rupestris is a small, wind-dispersed orchid (leaves 1.3–4.3 cm, shoots 15 cm in height and flowers of <6 mm) commonly found along the riverbeds of the Luquillo Mountains in Puerto Rico. *L. rupestris* anchors its roots in patches of moss growing on trees (epiphytic) or boulders (epilithic) along streams. Rocks tend to have larger population sizes and higher occupancy rates than trees (Tremblay et al. 2006). The mean dispersal distance of seeds is approximately 4.8 m (Tremblay 1997).

A permanent plot for the study of the metapopulation dynamics of *L. rupestris* was established in Quebrada Sonadora in the Luquillo Experimental Forest (18° 18′ N, 65° 47′ W) in 1999 (Tremblay et al. 2006). This permanent plot is composed of 1000 occupied and unoccupied boulders and tree trunks (patches hereafter; Tremblay et al. 2006). Initially, a total of 250 occupied patches (165 boulders and 85 trees) were identified. A patch was considered occupied if at least one living individual was present. Unoccupied patches were identified by randomly selecting three suitable patches (of any phorophyte with moss cover) with no individuals of any stage inside a 5 m radius of an initially occupied patch (Tremblay et al. 2006). These patches are spatially configured as a linear network following the river topography along a steep elevation gradient (Fig. 10.6). Most patches (975) were mapped to within approximately 10 cm (x, y, z positions) relative to the center of each patch using metal rulers, a sighting compass and a clinometer (Fig. 10.6). The presence–absence of *L. rupestris* was surveyed twice a year (at the beginning of the year and in the summer); here, we focus on data from 2000 to 2004 where all spatial information was available (i.e., coordinates, patch size). Patch size was estimated as the total moss area on the phorophyte, measured using a 150 cm^2 grid.

In this sampling design, the primary sampling periods are years (2000–2004, $n = 5$). The secondary sampling periods consisted of two censuses that were performed within each year. One census was conducted at the beginning of the year (between January and February) and the second in the summer (between July and August). This sampling formulation allows the system to be open to colonizations and extinctions during the wet season. Tropical storms are common during the wet season causing flash floods, which may be responsible for most local extinctions and anomalous strong winds, which may increase the magnitude of dispersal events potentially resulting in more local colonizations (Acevedo et al. 2015).

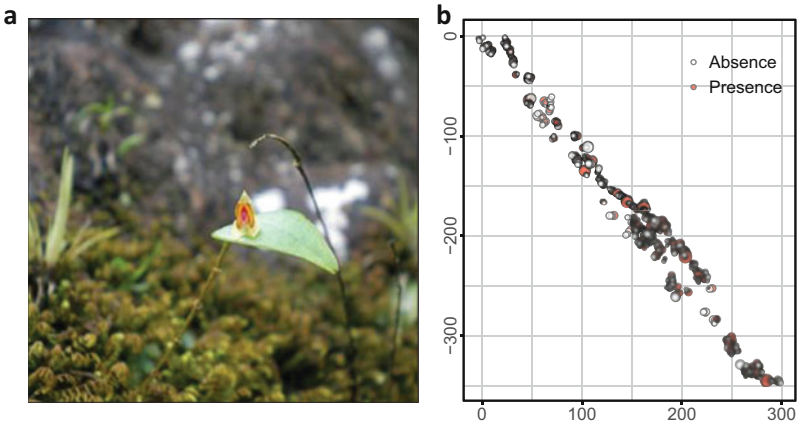


Fig. 10.6 (**a**) The wind-dispersed orchid, *Lepanthes rupestris,* exhibits colonization–extinction dynamics. (**b**) A map of the patches considered along a river, where 841 patches occurred, 266 were observed to be occupied in at least 1 year (Cancel et al. 2014)

We first import the data and explore the temporal variation in abundance and occurrence. The `orchid.csv` file contains population counts at each of 841 patches over ten surveys.

```
> surveys <- read.csv("orchid.csv")
> names(surveys)

##
[1] "siteID" "x" "y" "z" "area" "phorophyte"
[7] "primary_period" "secondary_period" "survey_number" "date"
[11] "N"
```

This file contains patch coordinates and patch sizes ("`area`"), along with ten surveys in a long format. There are five primary periods and two secondary survey periods (ten total surveys per patch). We will assume that within primary periods the population is closed, while between primary periods the population is open to colonization–extinction dynamics. This type of sampling design is frequently referred to as a "robust design" or "Polluck's robust design" (Pollock 1982).

First, we create a new column to truncate abundance to 0–1 data:

```
> surveys$presence <- ifelse(surveys$N > 0, 1, 0)
```

Then, we use the `reshape2` package (Wickham 2007) to reformat the data to a wide format. A wide format is useful for both interpreting colonization–extinction

dynamics and population synchrony. We do this separately for all sampling periods and for only primary periods (years), where we summarize both abundance and detection–non-detection data.

```
> library(reshape2)
#reshape for all/secondary periods
> surveys.occ <- dcast(surveys, siteID + x + y + z + phorophyte
  + area ~ survey_number, value.var = "presence")

> surveys.ab <- dcast(surveys, siteID + x + y + z + phorophyte +
area ~ survey_number , value.var = "N")

#reshape for primary periods
> surveys.pri.occ <- dcast(surveys, siteID + x + y + z +
  phorophyte + area ~ primary_period, value.var = "presence",
  max, fill = 0)
> surveys.pri.ab <- dcast(surveys, siteID + x + y + z +
  phorophyte + area ~ primary_period, value.var = "N", max, fill = 0)
```

Note that for the reshaping based on primary periods, we take the maximum value when aggregating data. If we do not specify the `fill` option, there will be a warning with aggregating by the maximum value (although the aggregation will still be correct). Another option could be taking the mean; the sum would not make sense because we would potentially be double counting individuals.

We can plot the network. We first determine which patches were occupied at least once over the time period and then map the patches with occupancy highlighted (Fig. 10.6):

```
#get occupancy of patches across entire period
> occ.total <- apply(surveys.occ[,7:16], 1, max)
> Nsites <- length(occ.total)
> Noccupied <- sum(occ.total)

#plot
> occ.color <- c("white", "red")
> plot(surveys.occ$x, surveys.occ$y,
  pch = 21, bg = occ.color[as.factor(occ.total)],
  cex = log(surveys.occ$area + 1) / 4)
```

The above code creates a map of patches, with point size proportional to the log of patch area and patches that were observed to be occupied at least once are in red while other patches are in white. Note the summary shows that out of 841 patches, 266 were observed to be occupied at least once.

10.3.3 Spatial Correlation and Synchrony

Understanding spatial correlation in population parameters is essential for spatially structured populations, both in terms of understanding mechanistically why spatial dynamics occur and for assessing population viability (Bjørnstad et al. 1999; Koenig 1999; Walter et al. 2017). For viability, it may be sufficient to quantify spatial dependence between pairs of sites or specific locations of interest (Morris and Doak 2002). For understanding and predicting spatial dynamics, we may want to quantify spatial dependence as a function of distance or other aspects of geography.

We consider spatial synchrony in time series data. Spatial synchrony can be considered by understanding the spatiotemporal correlation of population size, occurrence, or demographic rates as a function of geographic or effective distance, sometimes referred to as a spatial correlation function (Bjørnstad and Bascompte 2001). Typically, spatial synchrony analyses focus on either time series of population size (e.g., $\log(N_t)$) or population growth rates over space (e.g., $\log(N_{t+1}/N_t)$, or $\log(N_{t+1}) - \log(N_t)$) (Bjørnstad et al. 1999). Population change as a response variable can be useful because it reduces effects of population trends on estimating spatial synchrony. Synchrony in other processes, such as spatial synchrony in colonization and extinction rates, have also been considered (Sutherland et al. 2012).

Spatial synchrony is frequently quantified with correlograms (Bjørnstad et al. 1999), although we note that more recent applications focus on a wider variety of flexible models (e.g., wavelets; Walter et al. 2017). The `ncf` package provides a correlogram function `Sncf` specifically for interpreting spatial synchrony in time-series data (Bjørnstad and Falck 2001). This approach uses a Mantel correlogram: a multivariate correlogram that is a direct analog to univariate correlograms (see Chap. 5) (Borcard and Legendre 2012). Recall that a correlogram can be described using a Moran's I correlation coefficient for each distance category d as:

$$I(d) = \frac{n}{\mathbf{W}(\mathbf{d})} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(d)\left(z_i - \bar{z}\right)\left(z_j - \bar{z}\right)}{\sum_{i=1}^{n} \left(z_i - \bar{z}\right)^2}, \tag{10.25}$$

where $\mathbf{W}$ is a weight matrix that describes the dependency between locations i and j (e.g., whether two locations are within a specified distance category), and z is the response variable. This idea can be extended to multivariate data (e.g., time-series) using the Mantel correlation. Mantel correlations focus on quantifying the correlation between two distance matrices (Dale and Fortin 2014). A Mantel correlation for a distance category d is defined as:

$$r_M(d) = \frac{1}{\frac{n(n-1)}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{x_{ij} - \bar{x}}{s_x}\right)\left(\frac{y_{ij} - \bar{y}}{s_y}\right), \tag{10.26}$$

where s_x is the standard deviation from matrix x, and n is the number of observations in the matrix. Note that Mantel correlation values should not be interpreted as

equivalent to Pearson correlation values; rather focus should be on the shape of the correlogram values as a function of distance (Borcard and Legendre 2012). The interpretation of these correlograms is the following. When r_M is significantly positive, the similarity of values in the time series across sites of distance d is higher than what would be expected by chance, while when r_M is significantly negative, similarity is lower than what would be expected by chance.

In this situation, we pass the x–y coordinates of patches along with a site-by-time matrix of abundance/occurrence observations for each site. Here, we focus on the raw abundance data summarized at the primary period time scale. We first subset the number of patches to remove patches that had no observations over the entire time period and then log-transform the data. We log transform to stabilize variance and to interpret population change ($\lambda = \log(N_{t+1}/N_t) = \log(N_{t+1}) - \log(N_t)$).

```
> surveys.log.ab <- surveys.pri.ab
> surveys.log.ab$Total <- rowSums(surveys.log.ab[,7:11])
> surveys.log.ab <- subset(surveys.log.ab, Total > 0)
> surveys.log.ab[,7:11] <- log(surveys.log.ab[,7:11] + 1)

#population change
> surveys.log.ab$g2001 <- surveys.log.ab$y2001 -surveys.log.ab$y2000
> surveys.log.ab$g2002 <- surveys.log.ab$y2002 -surveys.log.ab$y2001
> surveys.log.ab$g2003 <- surveys.log.ab$y2003 -surveys.log.ab$y2002
> surveys.log.ab$g2004 <- surveys.log.ab$y2004 -surveys.log.ab$y2003
```

With this information, we quantify the spatial correlation. We consider the spatial dependence for the first year of the time series using a univariate spline correlogram (see Chap. 5) and then consider the entire time series of log abundance and population growth.

```
> library(ncf)
> abund.synchrony1 <- spline.correlog(x = surveys.log.ab$x,
  y = surveys.log.ab$y,
  z = surveys.log.ab[,7],
  xmax =200, resamp = 200)

#time series: log abundance
> abund.synchrony <- Sncf(x = surveys.log.ab$x,
  y = surveys.log.ab$y,
  z = surveys.log.ab[,7:11],
  xmax = 200, resamp = 200)
#time series: growth
> growth.synchrony <- Sncf(x = surveys.log.ab$x,
  y = surveys.log.ab$y,
  z = surveys.log.ab[,13:16],
  xmax = 200, resamp = 200)
```

```
> summary(abund.synchrony)$Regional.synch

##
[1] 0.05048885

> summary(abund.synchrony)$Squantile

##
 0% 2.5% 25% 50% 75% 97.5% 100%
0.01418591 0.01829298 0.03760531 0.04755823 0.06126896 0.08696007 0.10989193
```

With the default plotting for the time-series data, the horizontal lines show regional mean correlations (dotted lines) and zero synchronies (dashed lines). This analysis suggests that there is significant, but weak, spatial synchrony in abundance ($r = 0.05$, 95% CI: 0.02–0.09) and population growth ($r = 0.04$, 95% CI: 0.01–0.07). The correlograms also show only very weak evidence of spatial synchrony (Fig. 10.7). We can also check for potential anisotropy (directionality) in synchrony with the `Sncf2D` function.

Is this surprising? Spatial synchrony is often observed in plants and animals; however, the scale at which it operates is often assumed to be relatively large (e.g., across a species' range). Nonetheless, this example provides some insight as to how and why we model spatial synchrony. Beyond general estimates of spatial synchrony, there is increasing focus on understanding spatially explicit variation in synchrony, which may be vary due to spatial heterogeneity across landscapes and regions that alter dispersal and environmental variation of relevance to synchrony (Defriez and Reuman 2017a, b; Walter et al. 2017).

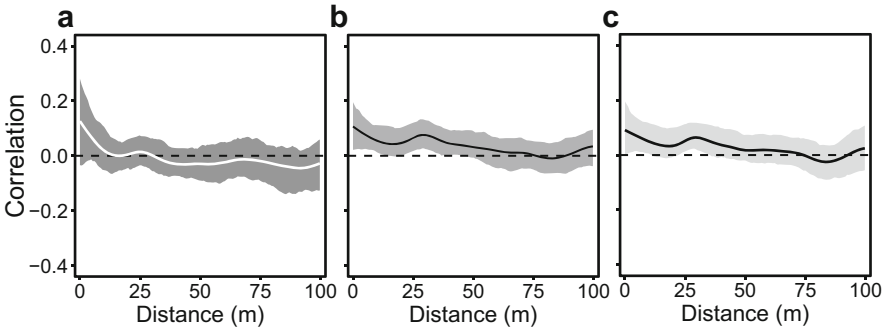


Fig. 10.7 Spatial dependence and synchrony in *Lepanthes rupestris*. (**a**) A Moran's correlogram for abundance in the first year of sampling. (**b**) A Mantel correlogram testing for spatial synchrony over time based on the log(abundance) observed on patches. (**c**) A Mantel correlogram testing for spatial synchrony over time based on the change in log(abundance), λ, observed on patches

10.3.4 Metapopulation Metrics

To interpret spatiotemporal variation in populations, we consider some classic metrics from the metapopulation area-isolation paradigm and related metrics for patch-level characteristics (Moilanen and Nieminen 2002). These metrics focus on patch area and different ways in which patch isolation can be quantified. From our `surveys.occ` data frame, we use *x*-*y* coordinates and patch area to create these metrics:

```
#patch area
> area <- surveys.occ$area

#distance matrix
> dist.matrix <- as.matrix(dist(cbind(surveys.occ$x, surveys.occ$y,
  surveys.occ$z)))
```

Note that the `dist` function can readily take coordinate data in three dimensions to calculate distances. We could consider effective distance metrics that incorporate the matrix like those described in Chap. 9, or other issues, such as wind direction (Acevedo et al. 2015), but for here we simply consider Euclidean distance. We can calculate *S* for each patch as a metric of patch connectivity (Eq. 10.12). In the below code, we set α to equal the mean dispersal distance from previous work on *L. rupestris*.

```
#define alpha
> meandist <- 4.8
> alpha <- 1/meandist

#S, ignoring occupancy
> g <- exp(-alpha * dist.matrix)
> diag(g) <- 0
> g.sweep <- sweep(g, 2, area, "*")
> S <- rowSums(g.sweep)
```

The above code breaks *S* into its components, where we first create a new matrix from the distance matrix, where linkages are $\exp(-\alpha d_{ij})$, and for the diagonal, $d_{ii} = 0$. We then multiple each value by the area *j* using the `sweep` function (similar to `apply` function) and sum across *j* for each patch *i* using `rowSums` function.

The above function ignores occupancy. That might be helpful in situations where information on occupancy for every patch is not available. However, this metric was originally envisioned as capturing potential immigrants from nearby patches, such that unoccupied patches should be ignored. To incorporate occupancy, we can calculate a time-specific variant of *S* using a "naïve" measure of occupancy (one that does not formally account for observation error). To do so, we take the `abund.primary` matrix and only consider patches where abundance >0 for each

primary period. We can then update our calculation by only considering occupied patches in the summation of *S*. An example for the first year is:

```
> Socc1 <- rowSums(g.sweep[,surveys.pri.occ$y2000 > 0])
```

In this case, patch connectivity metrics using naïve occupancy are highly correlated over this time period ($r \geq 0.93$).

A simpler metric that is often used in landscape ecology is the amount of habitat in the surrounding landscape. This metric lies at the heart of the habitat amount hypothesis of Fahrig (2013). In this case, we can use the mean dispersal distance to define a relevant scale for landscape context (Moilanen and Nieminen 2002) and calculate a buffer metric that describes the proportion of habitat in the surrounding landscape. To do so, we first create a binary matrix that describes whether distances between patches are less than or greater than the mean dispersal distance. Note, that we code this as 0 for distances $>$ the mean dispersal distance, and 1 for those $<$ mean distance. In that way, we can use this matrix as an indicator matrix of pairs of patches that are within the dispersal distance of *L. rupestris*.

```
> dist.binary <- ifelse(dist.matrix > meandist, 0, 1)
> diag(dist.binary) <- 0
> buffer <- rowSums(sweep(dist.binary, 2, area, "*"))
```

The rationale for the above code is similar for how *S* was calculated; here we are simply using `rowSums` to add up the patch areas within the mean dispersal distance of each patch *i*. With these metrics we can proceed to understand spatial dynamics based on the area-isolation paradigm. We combine these covariates into a data frame for further processing below:

```
> site.cov <- data.frame(siteID = surveys.occ$siteID, area, S, buffer)
```

For analyses below, we also center and scale *S* and buffer variables (with the `scale` function; see Chap. 6), while we take the log of area. Taking the log of area not only improves our ability to fit models to the data (as does centering and scaling), but importantly, it makes biological sense. We expect that the effect of a 1 unit increase in a 10 cm^2 moss patch is going to have a much bigger effect on occupancy and colonization–extinction dynamics than a 1 unit increase on a 1000 cm^2 patch.

10.3.5 Estimating Colonization–Extinction Dynamics

We can fit first-order Markov models for colonization–extinction dynamics with the `colext` function in the `unmarked` package in R when we have a robust-design framework (Pollock 1982). A robust design framework is often used in mark–recapture analyses, where we distinguish primary and secondary sampling periods.

We assume that the population is open between primary periods, such that colonization and extinction can occur, but that it is closed within primary periods (across secondary samples). The latter is generally necessary for formally accounting for detectability in colonization–extinction dynamics (but see Dail and Madsen 2013). An example would be surveys that occur each year (primary periods) for monitoring trends in populations, but repeated surveys are conducted within years (secondary samples).

Before considering colonization–extinction dynamics, we can use these data to focus solely on occupancy with an "implicit-dynamics" approach (MacKenzie et al. 2006). In this approach, we ignore colonization–extinction dynamics and rather focus on estimating effects of covariates on average patch occupancy over time. Note that this is akin to modeling repeated measures of occupancy. Ideally, we would formally account for repeated measures through the use of random effects or other approaches (e.g., generalized estimating equations; see Chap. 6); however, `unmarked` cannot easily accommodate random effects. Accounting for such temporal dependence is often accommodated using Bayesian inference (e.g., Rota et al. 2011), which is beyond the scope of our example. So here we will assume that site occupancy data each year is an independent sample and proceed with caution. Note, we could also potentially include primary periods as fixed effects, rather than random effects to at least partially address this issue, which we do below.

To model occupancy with an implicit-dynamics approach, we need to format data in a long format, where primary visits are repeated over time:

```
> surveys.sec.occ <- dcast(surveys, siteID + x + y + z +
  phorophyte + primary_period ~ secondary_period, value.var =
  "presence")

#merge with site covariates
> surveys.sec.occ <- merge(site.cov, surveys.sec.occ, by =
  "siteID")
```

In this model formulation, we can have site-level covariates (e.g., patch area), as well as observation-level covariates (e.g., date of survey). Note that the latter covariates can only be incorporated into the detection process rather than the occupancy process (because occupancy is not assumed to change within primary periods). We need to create covariates with a similar long format as the detection data.

For survey-specific detection, we focus on day of the year. To manipulate date information it is useful to convert date objects into `POSIX` objects. Below, we convert our date information, get the day of the year (`yday` for a `POSIX` object), center and scale the variable, and reformat the data into a detection history format.

```
> surveys$day <- as.POSIXlt(strptime(surveys$date, "%m/%d/%Y"))
> surveys$julian <- scale(surveys$day$yday)[,1]
> date <- dcast(surveys, siteID + primary_period ~
  secondary_period, value.var = "julian")
```

With these data, we create a data frame for occupancy data that unmarked can interpret:

```
> occ.data <- unmarkedFrameOccu(y = surveys.sec.occ[,10:11],
  siteCovs = surveys.sec.occ[,c(2:4,8)],
  obsCovs = list(date=date[,3:4]))
```

We can use the summary function to inspect the data object created to make sure that unmarked is interpreting the data appropriately. With this data frame, we can run a variety of models to interpret average occupancy. Here, we first consider whether date and area explain detectability.

```
> occ.p.int <- occu(~ 1 ~ 1, occ.data)
> occ.p.date <- occu(~ date ~ 1, occ.data)
> occ.p.area <- occu(~ area ~ 1, occ.data)
> occ.p.datearea <- occu(~ date + area ~ 1, occ.data)
```

In this formulation, the first formula (tilde) describes the detection process, while the second tilde describes the occupancy process. We can contrast models with model selection criteria and retrieve parameter estimates with the summary function. To create a model selection table, we first pass the model objects into a list and then use the modSel function in unmarked.

```
> model.p.list <- fitList( "p.null" = occ.p.int,
  "p.area" = occ.p.area,
  "p.date" = occ.p.date,
  "p.datearea" = occ.p.datearea)
> modSel(model.p.list)

##
 nPars AIC delta AICwt cumltvWt
p.datearea 4 6435.15 0.00 1.0e+00 1.00
p.area 3 6449.43 14.29 7.9e−04 1.00
p.date 3 6523.12 87.97 7.9e−20 1.00
p.null 2 6542.74 107.59 4.3e−24 1.00
```

Based on AIC, there is overwhelming support for both date and area explaining detectability.

```
> summary(occ.p.datearea)

##
Call:
occu(formula = ~date + area ~ 1, data = occ.data)

Occupancy (logit-scale):
 Estimate SE z P(>|z|)
 −0.797 0.0414 −19.3 1.19e−82
```

```
Detection (logit-scale):
 Estimate SE z P(>|z|)
(Intercept) −1.100 0.1753 −6.27 3.58e−10
date 0.229 0.0577 3.97 7.09e−05
area 0.704 0.0458 15.37 2.74e−53

AIC: 6435.148
Number of sites: 4205
optim convergence code: 0
optim iterations: 30
Bootstrap iterations: 0
```

These estimates suggest that as date and area increase, detectability increases. Now, with this information about detectability, we consider if area and isolation explain occupancy.

```
> occ.area.p.datearea <- occu(~ date + area ~ area, occ.data)
> occ.S.p.datearea <- occu(~ date + area ~ S, occ.data)
> occ.buffer.p.datearea <- occu(~ date + area ~ buffer,
  occ.data)
> occ.areaS.p.datearea <- occu(~ date + area ~ area + S,
  occ.data)
> occ.areabuffer.p.datearea <- occu(~ date + area ~ area +
  buffer, occ.data)

> model.occ.list <- fitList( "null" = occ.p.datearea,
  "area" = occ.area.p.datearea,
  "S" = occ.S.p.datearea,
  "buffer" = occ.buffer.p.datearea,
  "area+S" = occ.areaS.p.datearea,
  "area+buffer" = occ.areabuffer.p.datearea)
> modSel(model.occ.list)

##
 nPars AIC delta AICwt cumltvWt
area+S 6 6064.64 0.00 1.0e+00 1.00
area+buffer 6 6082.49 17.86 1.3e−04 1.00
area 5 6088.34 23.70 7.1e−06 1.00
S 5 6425.41 360.77 4.6e−79 1.00
buffer 5 6429.92 365.28 4.8e−80 1.00
null 4 6435.15 370.51 3.5e−81 1.00
phoro 5 6436.89 372.25 1.5e−81 1.00
```

In the above, we first add the log area to explain occupancy, followed by *S* and the metric of habitat area (buffer). The most supported model includes both patch area and the *S* metric that reflects patch isolation effects (as *S* increases, patch connectivity increases). We inspect the most supported model:

```
> summary(occ.areaS.p.datearea)

##
Call:
occu(formula = ~date + area ~ area + S, data = occ.data)

Occupancy (logit-scale):
 Estimate SE z P(>|z|)
(Intercept) −2.899 0.1187 −24.42 9.99e−132
area 0.562 0.0307 18.31 6.58e−75
S 0.185 0.0365 5.09 3.68e−07

Detection (logit-scale):
 Estimate SE z P(>|z|)
(Intercept) 1.303 0.2184 5.97 2.39e−09
date 0.300 0.0660 4.55 5.46e−06
area 0.137 0.0544 2.52 1.19e−02

AIC: 6064.639
Number of sites: 4205
optim convergence code: 0
optim iterations: 48
Bootstrap iterations: 0
```

There are several interesting results here. First, as both patch area and connectivity increase, occupancy increases (Fig. 10.8). Based on the ratio of the estimates to the SEs (z-value), the effect of patch area appears to be stronger than for patch isolation. Also, note that now the effect of patch area on detectability is much smaller than previously when we were exploring detectability alone. Patch area still explains some variation in detectability, but it seems like much of the variation we initially observed was actually due to occupancy. This is not too surprising: while covariate

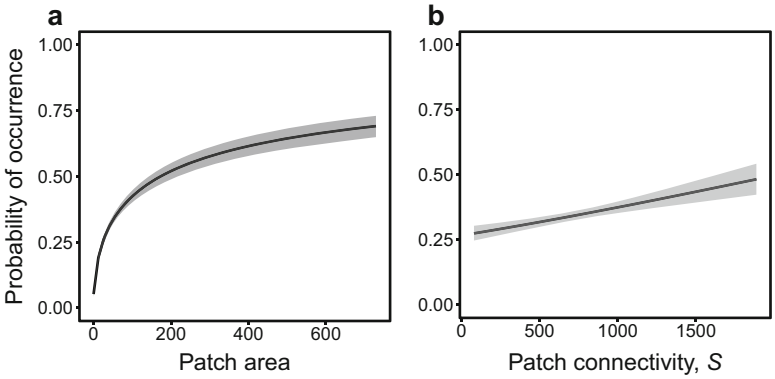


Fig. 10.8 Predictions from an "implicit-dynamics" occupancy model, where colonization and extinction are not formally estimated. Shown are partial relationships for the effect of (**a**) patch area and (**b**) patch connectivity for the probability of occurrence of *Lepanthes rupestris*

effects on detectability and occupancy are not typically confounded, the model does use some shared information during the estimation process. As a consequence, we need to be careful with interpreting covariate effects that may be relevant to both detectability and occupancy.

While the above modeling is useful for understanding variation in site occupancy, it is unclear whether this is driven by lower extinction rates with increasing patch area, or greater colonization rates (or both). To understand these dynamics, we can explicitly model these effects with dynamic occupancy models.

In a dynamic occupancy model, we can consider covariates that may explain variation in occupancy at time 1, subsequent colonization and extinction, and detectability. Covariates can be defined at the site/patch level with no temporal variation (`siteCovs` in `unmarked`) or with temporal variation at the annual (or primary) level (`yearlySiteCovs`), or for individual surveys within patches (temporal variation at survey level; `obsCovs`). In this situation, site covariates include patch area, buffer area, and S when incidence is ignored. `YearlySiteCovs` include time-dependent S that incorporates incidence. For observation-level covariates, we simply use a time category (each survey).

We first create a data frame that includes all relevant data using the `unmarkedMultFrame` function:

```
#format
> Nprimary <- length(levels(surveys$primary_period))
> date.wide <- dcast(surveys, siteID ~ survey_number, value.var
  = "julian")
> primary.cov <- data.frame(Socc = scale(c(Socc1, Socc2, Socc3,
  Socc4, Socc5))[,1])

#create unmarked object
> DO.data <- unmarkedMultFrame(y = occ.matrix,
  siteCovs = site.cov,
  obsCovs = list(date =
  date.wide[,2:11]),
   yearlySiteCovs = primary.cov,
   numPrimary = Nprimary)
```

With the `summary` function, we note that in these data, there are 1969 detections and 6641 non-detections across 841 patches, 266 of which had at least one detection over time. To illustrate some simple comparisons, we consider a few different models. First, we contrast a simple model with no covariates for each parameter and contrast it to a model that includes date and area as covariates on detection probability:

```
> DO.psi.int.col.int.eps.int.p.int <- colext(psiformula = ~1,
  gammaformula = ~ 1, epsilonformula = ~ 1, pformula = ~ 1, DO.data)
> DO.psi.int.col.int.eps.int <- colext(psiformula = ~1,
  gammaformula = ~ 1, epsilonformula = ~ 1, pformula = ~ date +
  area, DO.data)
```

Comparing these models again suggests that date and area are useful covariates for estimating variation in detection probability. Note that in the summary, all parameters are on the link (i.e., logit) scale. We can back-transform them to the probability scale using the `plogis` function; alternatively, `unmarked` has a nice function for doing so:

```
> backTransform(DO.psi.int.col.int.eps.int.p.int, type = "det")

##
Backtransformed linear combination(s) of Detection estimate(s)

 Estimate SE LinComb (Intercept)
 0.843 0.00814 1.68 1

Transformation: logistic
```

Overall, estimates suggest high (though not perfect: $p = 0.843 + 0.008$ SE) survey-specific detectability of this species. Because we had two surveys in this study, the probability of detection across the secondary sampling period is $1 - (1 - 0.843)^2 = 0.98$. Colonization and extinction rates can also be back-transformed with this function (using `type = "col"` and `type = "ext"`, respectively). In addition, rates of colonization and extinction are low on average, with patch extinction occurring slightly more frequently than colonization.

Using this general approach, we explore several models for effects of area and isolation on colonization–extinction dynamics (code not shown). The summary of these models is:

```
> modSel(model.DO.list)

##
 nPars AIC delta AICwt cumltvWt
psi-area+S,ext-area 9 3504.27 0.00 5.1e−01 0.51
psi-area+S,ext-area+S 10 3505.57 1.30 2.6e−01 0.77
psi-area+S,ext-area, col-S 10 3505.88 1.61 2.3e−01 1.00
psi-area+S,null 8 3520.49 16.23 1.5e−04 1.00
psi-area+S, col-S 9 3522.08 17.81 6.9e−05 1.00
psi-constant,ext-area 7 3590.02 85.76 1.2e−19 1.00
psi-constant,ext-area+S 8 3591.29 87.02 6.4e−20 1.00
psi-constant,ext-area, col-S 8 3591.39 87.12 6.1e−20 1.00
psi-constant,null 6 3606.40 102.13 3.4e−23 1.00
psi-constant, col-S 7 3607.71 103.45 1.8e−23 1.00
psi-constant,p-null 4 3631.58 127.32 1.1e−28 1.00

> summary(DO.psi.areaS.col.int.eps.area)

##
Call:
```

```
colext(psiformula = ~area + S, gammaformula = ~1, epsilonformula = ~area,
 pformula = ~date + area, data = DO.data)

Initial (logit-scale):
 Estimate SE z P(>|z|)
(Intercept) −2.881 0.2610 −11.04 2.59e−28
area 0.560 0.0676 8.28 1.24e−16
S 0.189 0.0809 2.33 1.98e−02

Colonization (logit-scale):
 Estimate SE z P(>|z|)
 -4.14 0.176 -23.5 3.65e-122

Extinction (logit-scale):
 Estimate SE z P(>|z|)
(Intercept) −1.21 0.439 −2.75 6.01e−03
area −0.62 0.136 −4.56 5.19e−06

Detection (logit-scale):
 Estimate SE z P(>|z|)
(Intercept) 1.278 0.2004 6.38 1.81e−10
date 0.257 0.0606 4.24 2.21e−05
area 0.111 0.0492 2.25 2.44e−02

AIC: 3504.267
Number of sites: 841
optim convergence code: 0
optim iterations: 97
Bootstrap iterations: 0
```

Based on model comparisons, there is some evidence that patch area and connectivity, *S*, have strong positive effects on occupancy at time 1, similar to the implicit dynamics model. Assuming constant initial occupancy, such that the patch area and connectivity effects are captured entirely in the colonization and extinction parameters, is not well supported. Patch area also has a strong negative effect on patch extinction (Fig. 10.9). Patch connectivity, *S*, is not a strong predictor of colonization. We can plot partial predictions from these models in ways similar to previous chapters (e.g., see Chap. 7) by creating a new data sets for predictions and using the `predict` function. Below we show an example for the patch area effect on extinction (Fig. 10.9):

```
#create new data set
> Smean <- mean(site.cov$S)
> Srange <- seq(min(site.cov$S), max(site.cov$S), length = 20)
> Arearange <- seq(min(site.cov$area), max(site.cov$area),length = 20)

> Areamean <- mean(site.cov$area)
> Datemean <- mean(as.matrix(date.wide[,2:11]))
```

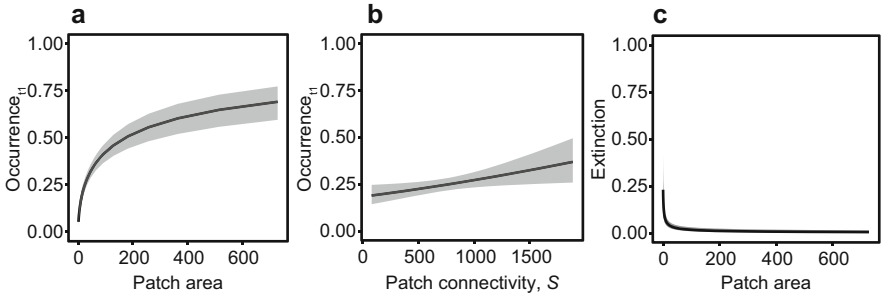


Fig. 10.9 Predictions from a dynamic occupancy model, where colonization and extinction are formally estimated. Shown are partial relationships for the effect of (**a**) patch area and (**b**) patch connectivity for the probability of occurrence of *Lepanthes rupestris* during the first time period, and (**c**) the effect of patch area on extinction probability

```
> newdata.iso <- expand.grid(area = Areamean, S = Srange, date =
  Datemean)
> newdata.area <- expand.grid(area = Arearange, S = Smean, date =
  Datemean)

#extinction
> ext.pred <- predict(DO.psi_areaS.col_int.eps_area,
  newdata.area, type = 'ext')
> newdata.ext <- cbind(newdata.area,ext.pred)
> newdata.ext$areaback <- exp(newdata.ext$area) - 1


> plot(newdata.ext$areaback, newdata.ext$Predicted, ylim = c(0, 0.5))
> lines(newdata.ext$areaback, newdata.ext$lower)
> lines(newdata.ext$areaback, newdata.ext$upper)
```

When contrasting these results to that of the implicit dynamics approach, it is clear that the area effect on average occupancy was driven partially by extinction dynamics. Given that area explains occupancy in time 1 as well, dynamics prior to surveys (before year 2000) likely contribute to occupancy, presumably from changes in extinction rates.

10.3.6 Projecting Dynamics

Estimates of colonization–extinction dynamics can be used to project potential dynamics (see Sect. 10.2.3). Accounting for uncertainty is more complex, although Bayesian models have been developed that can track appropriate uncertainty (Chandler et al. 2015). Here, we provide a simple example to illustrate projections.

With patch-specific colonization and extinction probabilities, we project forward starting with an initial estimate of patch occupancy. Recall that occupancy in

time t can be predicted from colonization and extinction probabilities coupled with occupancy in time $t-1$. Here, we pass vectors of initial occupancy, the time period of interest for projections, and patch-specific estimates of colonization and extinction probabilities taken from the dynamic occupancy model.

```
#predict onto site covariates to get site-specific col-ext
> newdata.fit <- data.frame(site.cov, date = Datemean)
> ext.fit <- predict(DO.psi.areaS.col.int.eps.area, newdata.fit,
  type = 'ext')
> col.fit <- predict(DO.psi.areaS.col.int.eps.area, newdata.fit,
  type = 'col')
> psi.fit <- predict(DO.psi.areaS.col.int.eps.area, newdata.fit,
  type = 'psi')

> col.pred <- col.fit$Predicted
> ext.pred <- ext.fit$Predicted
> psi.pred <- psi.fit$Predicted

#simulation parameters
> timeframe <- 100 #time steps to simulate
> reps <- 20 #realizations
```

With these parameters, we use a function that draws binomial realizations from the colonization and extinction probabilities to create an output array of simulated scenarios. This function creates a three-dimensional array for output, where the first dimension is the realization, the second is the sites, and the third is time:

```
#function for simulating
> colext.sim <- function(occ.int, col, ext, timeframe, reps)
{
 Nsite <- length(occ.int)
 colext.out <- array(NA,dim = c(reps, Nsite, timeframe))

 #z:time 1
 colext.out[,,1] <- rbinom(Nsite, prob = occ.int, size = 1)

 #z:time 2-T
 for(j in 1:reps){
 for(t in 2:timeframe){
 prob.jt <- colext.out[j,,t - 1] * (1 - ext) + (1 -
colext.out[j,,t - 1]) * col
 colext.out[j,,t] <- rbinom(Nsite, prob = prob.jt, size = 1)
 }#end i
 }#end j

 return(colext.out)
}
```

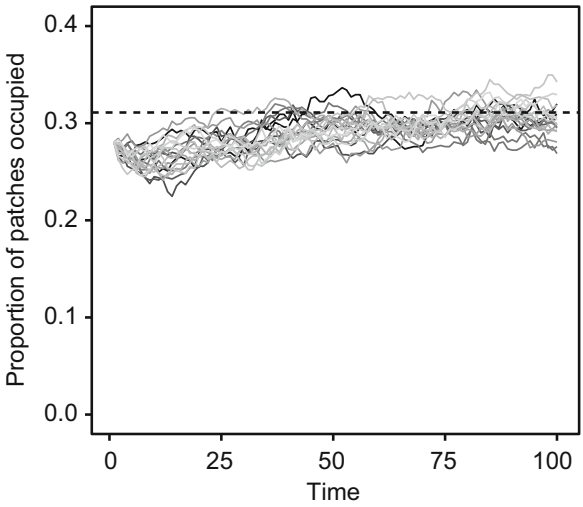


Fig. 10.10 Projecting the colonization–extinction dynamics of *Lepanthes rupestris* based on estimates from the dynamic occupancy model. Dashed line shows the estimated equilibrium occupancy

With this function we can simulate dynamics over time. These dynamics can be summarized in a variety of ways, such as visualizing patch-specific trajectories or the fraction of sites occupied over time.

```
> colext.proj <- colext.sim(occ.int = psi.pred, col = col.pred,
  ext = ext.pred, timeframe = timeframe, reps = reps)

#site-specific means across realizations
> colext.patch.mean <- apply(colext.proj, 1:2, mean)
> colext.patch.mean <- t(colext.patch.mean)

#landscape mean over time
> colext.land.mean <- apply(colext.proj, 1, mean)
```

Interestingly, even though patch extinction rates tend to be higher than colonization rates based on our modeling, these projections suggest a slight increase in the frequent of patches occupied over time (Fig. 10.10). How is this so?

Recall that in the Levins' model, persistence was only possible if $c > e$. In this example, $\gamma < \varepsilon$. However, these estimated extinction and colonization rates differ for at least two reasons. First, estimates vary by patch, whereas Levins' model assumes that rates do not differ across patches. Second, and more importantly, the parameters in the Levins' model are not equivalent to the colonization–extinction parameters here. In Levins' model, c is the rate of individuals emigrating from occupied patches, whereas for the dynamic occupancy model γ is the probability that an empty patch becomes colonized over one primary period. Overall, the long-term average occupancy in this context should converge on the predicted equilibrium occupancy value for a patch (MacKenzie et al. 2003; Ferraz et al. 2007):

$$\psi_i^* = \frac{\gamma_i}{\gamma_i + \varepsilon_i}. \tag{10.27}$$

When using the estimates of colonization and extinction, we find that $\psi^* = 0.31$, which is relatively consistent with the long-term projections from these simulations.

10.3.7 Metapopulation Viability and Environmental Change

A frequently used metric to interpret the potential viability of metapopulations is the metapopulation capacity (Hanski and Ovaskainen 2000). This metric can be used as a relative metric to contrast different landscapes and/or when landscapes experience environmental change. The metapopulation capacity only relies on information regarding patch area and isolation. It can be calculated as:

```
> meta.cap <- function(A, distmat)
{
M <- outer(A, A) * distmat
tmp <- eigen(M)
vector.M <- Re(tmp$vector[, 1]^2) #patch-specific eigenvectors
lambda.M <- Re(tmp$value[1]) #metapopulation capacity
return(list(lambda.M, vector.M))
}
```

In this function, we pass a vector of patch areas along with a square distance matrix for the patches. The function then calculates a connectivity matrix **M** and takes the leading eigenvalue of the matrix as the metapopulation capacity. It also takes the leading eigenvectors of the matrix as a measure of the contribution of patches to the overall metapopulation capacity.

With this metric, we can start to interpret the potential effects of habitat loss, fragmentation, and restoration on metapopulation capacity (e.g., Fig. 10.4). For instance, we can randomly remove patches in the orchid network, calculating metapopulation capacity with each patch removal. To do so, we remove 10% of the patches at each iteration.

```
> reps <- 20
> Npatches <- length(surveys.occ$siteID)
> metacap.rand <- matrix(0,10,reps)

> for (z in 1:reps){

 rand <- sample(surveys.occ$siteID) #shuffles patches
```

```
#randomly remove patches, 10%/time step
for (i in 1:9){
removal <- i * 10
N.removal <- round(removal * length(rand) / 100, 0)
remove <- rand[1:N.removal]
patch.i <- surveys.occ[!(surveys.occ$siteID%in%remove),]
area.i <- patch.i$area
dist.i <- as.matrix(dist(cbind(patch.i$x, patch.i$y, patch.i$z)))
metacap.i <- meta.cap(A = area.i, distmat = dist.i)
metacap.rand[i,z] <- metacap.i[[1]]
} #for i
print(z)
} #for z
```

In this case, we are simulating the random, cumulative removal of patches, calculating metapopulation capacity after each % patch removal. To do so, we randomly shuffle the patch IDs, and then remove these patches cumulatively from our vector of patch area and recalculate a distance matrix for each patch removal. On average, the decline in metapopulation capacity in this situation is nearly linear (i.e., there does not appear to be any distinct thresholds or tipping points where metapopulation capacity rapidly declines) (Fig. 10.11). This general framework can be used to address a wide variety of perturbations, such as altering patch area, isolation, and restoration (Hanski and Ovaskainen 2000).

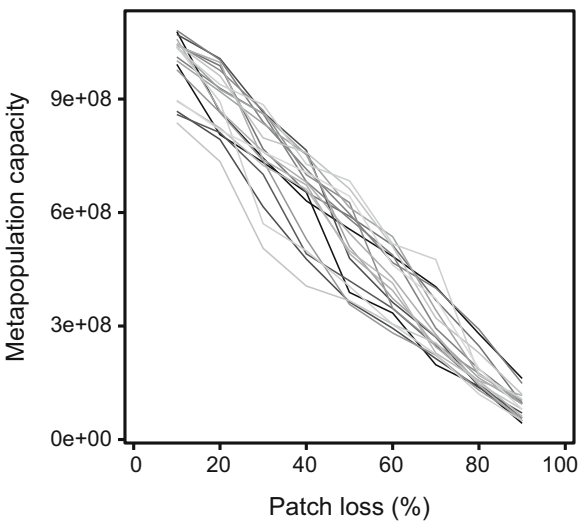


Fig. 10.11 Metapopulation capacity of *Lepanthes rupestris* with increasing levels of habitat loss. Shown are 20 realizations of random patch loss

10.4 Next Steps and Advanced Issues

10.4.1 Spatial Population Matrix Models

Spatial matrix models are often used for answering questions in landscape demography. However, these models can be challenging to fit because they require detailed, spatially explicit information on demographic rates (Ozgul et al. 2009; Cattau et al. 2016). When mark–recapture data are available, one approach to quantify movement and survival components for spatial matrix models is the use of multistate mark–recapture models (Hestbeck et al. 2010; Bailey et al. 2010). In addition, reverse-time mark–recapture models can be used to partition survival, recruitment, and movement across sites (Nichols et al. 2000; Sanderlin et al. 2012). Such modeling frameworks can be useful for interpreting a variety of issues, such as source–sink dynamics (Runge et al. 2006).

10.4.2 Diffusion and Spatial Dynamics

There is a long history in population ecology to formulating spatial dynamics using diffusion-based models (Skellam 1951). Diffusion models have been widely applied to the problem of invasive spread (Neubert et al. 2000; Fagan et al. 2002). More recently, diffusion models have been extended to capture spatial heterogeneity, such as edge or boundary effects, and predict connectivity in population dynamics (Ovaskainen 2004, 2008; Ovaskainen et al. 2008; Reeve et al. 2008). The utility of these models is that random dispersal can be integrated into models in an elegant way that may provide generalizable conclusions regarding spatial dynamics. These models require a strong foundation in calculus and require customizing models for specific problems.

10.4.3 Agent-Based Models

In spatial ecology and conservation, agent-based, or individual-based, spatially explicit models are increasingly used. These models require detailed information on the environment and how species use the environment. These types of models are useful for scenario modeling and interpreting specific effects on management and related land-use changes (Pulliam et al. 1992). Standard protocols exist for the development and reporting of these complex models (Grimm et al. 2005). These models are infrequently implemented in R directly because of computational limitations of R for handling these types of models. One notable exception is the `simecol` package, which can fit certain types of individual-based models (Petzoldt and Rinke 2007). Alternatively, models can interface R with other software, the

`RNetlogo` package (Thiele et al. 2012), which interfaces with a popular agent-based modeling software, NetLogo (Sklar 2007). Other solutions include writing models in lower-level programming languages (e.g., C++).

10.4.4 Integrated Population Models

Another approach to integrating demographic rates and spatial aspects of movement is the use of integrated population models. In integrated models, different types of data are coupled to estimate population parameters, such as mark–recapture data, count data, and data on reproduction (Schaub et al. 2007; Abadi et al. 2010; Zipkin and Saunders 2018). This type of approach can be particularly useful in increasing precision and reducing bias in population modeling. The complexity and detail of these types of models often requires customized Bayesian approaches, which can be run through R with various wrapper packages.

10.5 Conclusions

Spatial issues have been of long interest to population biologists. Spatial issues were critical in theoretical developments of metapopulation ecology and population genetics (Wright 1943, 1951; Hanski 1999), and much of the work on species distribution, colonization, dispersal, and connectivity (e.g., Chaps. 7–9) falls within the realm of spatially structured populations. Furthermore, understanding spatially structured population dynamics is highly relevant for many conservation issues, including source–sink dynamics, habitat management, and the application of population viability analysis.

Two major paradigms for understanding spatially structured populations include the metapopulation paradigm and the spatial demography (or landscape demography) paradigm (Hanski 1999; Gurevitch et al. 2016). The questions and problems that each of these paradigms address vary based on the spatial and temporal scales being captured and the state variables of interest. Understanding and quantifying colonization–extinction dynamics can help aid in understanding metapopulation viability and dynamics at broad spatial (and temporal) scales. Spatial demography is often used to understand dynamics across smaller spatial scales, where the focus is on variation in demographic vital rates, such as survival and reproduction—parameters central for estimating spatial variation in population size and growth rate. Each of these paradigms uses different quantitative modeling approaches. We expect that these paradigms will grow together over time with increases in data availability and the use of data fusion and integrative modeling approaches.

References

Abadi F, Gimenez O, Arlettaz R, Schaub M (2010) An assessment of integrated population models: bias, accuracy, and violation of the assumption of independence. Ecology 91(1):7–14. https://doi.org/10.1890/08-2235.1

Acevedo MA, Fletcher RJ Jr, Tremblay RL, Melendez-Ackerman EJ (2015) Spatial asymmetries in connectivity influence colonization-extinction dynamics. Oecologia 179(2):415–424. https://doi.org/10.1007/s00442-015-3361-z

Ackerman JD, Sabat A, Zimmerman JK (1996) Seedling establishment in an epiphytic orchid: an experimental study of seed limitation. Oecologia 106(2):192–198. https://doi.org/10.1007/bf00328598

Adler FR, Nuernberger B (1994) Persistence in patchy irregular landscapes. Theor Popul Biol 45 (1):41–75

Baguette M (2004) The classical metapopulation theory and the real, natural world: a critical appraisal. Basic Appl Ecol 5(3):213–224. https://doi.org/10.1016/j.baae.2004.03.001

Bailey LL, Converse SJ, Kendall WL (2010) Bias, precision, and parameter redundancy in complex multistate models with unobservable states. Ecology 91(6):1598–1604

Barve N, Barve V, Jimenez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberon J, Villalobos F (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol Model 222(11):1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Bascompte J, Sole RV (1996) Habitat fragmentation and extinction thresholds in spatially explicit models. J Anim Ecol 65(4):465–473

Begon M, Mortimer M, Thompson DJ (2009) Population ecology: a unified study of animals and plants, 3rd edn. Blackwell Science Ltd., Oxford, UK

Beissinger SR, Westphal MI (1998) On the use of demographic models of population viability in endangered species management. J Wildl Manag 62(3):821–841

Bjørnstad ON, Bascompte J (2001) Synchrony and second-order spatial correlation in host-parasitoid systems. J Anim Ecol 70(6):924–933. https://doi.org/10.1046/j.0021-8790.2001.00560.x

Bjørnstad ON, Falck W (2001) Nonparametric spatial covariance functions: estimation and testing. Environ Ecol Stat 8(1):53–70. https://doi.org/10.1023/a:1009601932481

Bjørnstad ON, Ims RA, Lambin X (1999) Spatial population dynamics: analyzing patterns and processes of population synchrony. Trends Ecol Evol 14(11):427–432. https://doi.org/10.1016/s0169-5347(99)01677-8

Borcard D, Legendre P (2012) Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study. Ecology 93(6):1473–1481

Brook BW, O'Grady JJ, Chapman AP, Burgman MA, Akcakaya HR, Frankham R (2000) Predictive accuracy of population viability analysis in conservation biology. Nature 404 (6776):385–387. https://doi.org/10.1038/35006050

Bulman CR, Wilson RJ, Holt AR, Bravo LG, Early RI, Warren MS, Thomas CD (2007) Minimum viable metapopulation size, extinction debt, and the conservation of a declining species. Ecol Appl 17(5):1460–1473. https://doi.org/10.1890/06-1032.1

Cancel JGG, Meléndez-Ackerman EJ, Olaya-Arenas P, Merced A, Flores NP, Tremblay RL (2014) Associations between Lepanthes rupestris orchids and bryophyte presence in the Luquillo Experimental Forest, Puerto Rico. Caribb Nat 6:1–14

Carroll C, Noss RE, Paquet PC, Schumaker NH (2003) Use of population viability analysis and reserve selection algorithms in regional conservation plans. Ecol Appl 13(6):1773–1789. https://doi.org/10.1890/02-5195

Caswell H (2001) Matrix population models: construction, analysis, and interpretation. Sinauer, Sunderland, MA

Caswell H, Etter RJ (1993) Ecological interactions in patchy environments: from patch-occupancy models to cellular automata. In: Levin SA (ed) Patch dynamics. Springer, New York

Caswell H, Lensink R, Neubert MG (2003) Demography and dispersal: life table response experiments for invasion speed. Ecology 84(8):1968–1978. https://doi.org/10.1890/02-0100
Cattau CE, Fletcher RJ Jr, Reichert BE, Kitchens WM (2016) Counteracting effects of a non-native prey on the demography of a native predator culminate in positive population growth. Ecol Appl 26:1952–1968
Caughley G (1994) Directions in conservation biology. J Anim Ecol 63(2):215–244
Cavanaugh KC, Siegel DA, Raimondi PT, Alberto F (2014) Patch definition in metapopulation analysis: a graph theory approach to solve the mega-patch problem. Ecology 95(2):316–328. https://doi.org/10.1890/13-0221.1
Chan KMA, Shaw MR, Cameron DR, Underwood EC, Daily GC (2006) Conservation planning for ecosystem services. PLoS Biol 4(11):2138–2152. https://doi.org/10.1371/journal.pbio.0040379
Chandler RB, Muths E, Sigafus BH, Schwalbe CR, Jarchow CJ, Hossack BR (2015) Spatial occupancy models for predicting metapopulation dynamics and viability following reintroduction. J Appl Ecol 52(5):1325–1333. https://doi.org/10.1111/1365-2664.12481
Channell R, Lomolino MV (2000) Dynamic biogeography and conservation of endangered species. Nature 403(6765):84–86. https://doi.org/10.1038/47487
Clark CW, Rosenzweig ML (1994) Extinction and colonization processes: parameter estimates from sporadic surveys. Am Nat 143(4):583–596. https://doi.org/10.1086/285621
Clark JS, Silman M, Kern R, Macklin E, HilleRisLambers J (1999) Seed dispersal near and far: patterns across temperate and tropical forests. Ecology 80(5):1475–1494
Cole LC (1954) The population consequences of life history phenomena. Q Rev Biol 29 (2):103–137. https://doi.org/10.1086/400074
Connor EF, McCoy ED (1979) Statistics and biology of the species-area relationship. Am Nat 113 (6):791–833. https://doi.org/10.1086/283438
Cronin JT (2007) From population sources to sieves: the matrix alters host-parasitoid source-sink structure. Ecology 88(12):2966–2976
Dail D, Madsen L (2013) Estimating open population site occupancy from presence-absence data lacking the robust design. Biometrics 69(1):146–156. https://doi.org/10.1111/j.1541-0420.2012.01796.x
Dale MRT, Fortin MJ (2014) Spatial analysis: a guide for ecologists, 2nd edn. Cambridge University Press, Cambridge
de Kroon H, van Groenendael J, Ehrlen J (2000) Elasticities: a review of methods and model limitations. Ecology 81(3):607–618. https://doi.org/10.1890/0012-9658(2000)081[0607:earoma]2.0.co;2
DeAngelis DL, Mooij WM (2005) Individual-based modeling of ecological and evolutionary processes. Annu Rev Ecol Evol Syst 36:147–168. https://doi.org/10.1146/annurev.ecolsys.36.102003.152644
Defriez EJ, Reuman DC (2017a) A global geography of synchrony for marine phytoplankton. Glob Ecol Biogeogr 26(8):867–877. https://doi.org/10.1111/geb.12594
Defriez EJ, Reuman DC (2017b) A global geography of synchrony for terrestrial vegetation. Glob Ecol Biogeogr 26(8):878–888. https://doi.org/10.1111/geb.12595
den Boer PJ (1968) Spreading of risk and stabilization of animal numbers. Acta Biotheor 18:165–194
Diamond JM (1972) Biogeographic kinetics: estimation of relaxation times for avifaunas of southwest pacific islands. Proc Nat Aca Sci USA 69(11):3199–3203
Diekmann O (1993) An invitation to structured (meta)population models. In: Levin SA (ed) Patch dynamics. Springer, New York
Diffendorfer JE (1998) Testing models of source-sink dynamics and balanced dispersal. Oikos 81 (3):417–433
Doncaster CP, Clobert J, Doligez B, Gustafsson L, Danchin E (1997) Balanced dispersal between spatially varying local populations: an alternative to the source-sink model. Am Nat 150 (4):425–445. https://doi.org/10.1086/286074

Drechsler M (2009) Predicting metapopulation lifetime from macroscopic network properties. Math Biosci 218(1):59–71. https://doi.org/10.1016/j.mbs.2008.12.004
Drechsler M, Johst K (2010) Rapid viability analysis for metapopulations in dynamic habitat networks. Proc R Soc B 277(1689):1889–1897. https://doi.org/10.1098/rspb.2010.0029
Dunning JB, Stewart DJ, Danielson BJ, Noon BR, Root TL, Lamberson RH, Stevens EE (1995) Spatially explicity population models: current forms and future uses. Ecol Appl 5(1):3–11. https://doi.org/10.2307/1942045
Dytham C (1995) The effect of habitat destruction pattern on species persistence: a cellular model. Oikos 74(2):340–344. https://doi.org/10.2307/3545665
Efford MG (2018) secr: spatially explicit capture-recapture models. R package version 3.1.6
Ehrlich PR, Daily GC (1993) Population extinction and saving biodiversity. Ambio 22(2–3):64–68
Erwin RM, Nichols JD, Eyler TB, Stotts DB, Truitt BR (1998) Modeling colony-site dynamics: a case study of gull-billed terns (Sterna nilotica) in coastal Virginia. Auk 115(4):970–978
Ezard THG, Bullock JM, Dalgleish HJ, Millon A, Pelletier F, Ozgul A, Koons DN (2010) Matrix models for a changeable world: the importance of transient dynamics in population management. J Appl Ecol 47(3):515–523. https://doi.org/10.1111/j.1365-2664.2010.01801.x
Fagan WF, Lewis MA, Neubert MG, van den Driessche P (2002) Invasion theory and biological control. Ecol Lett 5(1):148–157. https://doi.org/10.1046/j.1461-0248.2002.0_285.x
Fahrig L (1997) Relative effects of habitat loss and fragmentation on population extinction. J Wildl Manag 61(3):603–610
Fahrig L (2001) How much habitat is enough? Biol Conserv 100(1):65–74
Fahrig L (2013) Rethinking patch size and isolation effects: the habitat amount hypothesis. J Biogeogr 40(9):1649–1663. https://doi.org/10.1111/jbi.12130
Ferraz G, Nichols JD, Hines JE, Stouffer PC, Bierregaard RO, Lovejoy TE (2007) A large-scale deforestation experiment: effects of patch area and isolation on Amazon birds. Science 315 (5809):238–241. https://doi.org/10.1126/science.1133097
Fiske IJ, Chandler RB (2011) Unmarked: an R Package for fitting hierarchical models of wildlife occurrence and abundance. J Stat Softw 43(10):1–23
Flather CH, Hayward GD, Beissinger SR, Stephens PA (2011) Minimum viable populations: is there a 'magic number' for conservation practitioners? Trends Ecol Evol 26(6):307–316. https://doi.org/10.1016/j.tree.2011.03.001
Fleishman E, Ray C, Sjogren-Gulve P, Boggs CL, Murphy DD (2002) Assessing the roles of patch quality, area, and isolation in predicting metapopulation dynamics. Conserv Biol 16(3):706–716
Frank K, Wissel C (1998) Spatial aspects of metapopulation survival - from model results to rules of thumb for landscape management. Landsc Ecol 13(6):363–379. https://doi.org/10.1023/a:1008054906030
Frank K, Wissel C (2002) A formula for the mean lifetime of metapopulations in heterogeneous landscapes. Am Nat 159(5):530–552. https://doi.org/10.1086/338991
Gause GF (1932) Experimental studies on the struggle for existence. J Exp Biol 9:389–402
Gonzalez A, Holt RD (2002) The inflationary effects of environmental fluctuations in source-sink systems. Proc Natl Acad Sci U S A 99(23):14872–14877
Gotelli NJ (1991) Metapopulation models: the rescue effect, the propagule rain, and the core-satellite hypothesis. Am Nat 138(3):768–776. https://doi.org/10.1086/285249
Gotelli NJ (2008) A primer of ecology, 4th edn. Sinauer Associates, Inc, Sunderland, MA
Gotelli NJ, Kelley WG (1993) A general model of metapopulation dynamics. Oikos 68(1):36–44. https://doi.org/10.2307/3545306
Gouhier TC, Guichard F (2014) Synchrony: quantifying variability in space and time. Methods Ecol Evol 5(6):524–533. https://doi.org/10.1111/2041-210x.12188
Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke HH, Weiner J, Wiegand T, DeAngelis DL (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310(5750):987–991. https://doi.org/10.1126/science.1116681
Gurevitch J, Fox GA, Fowler NL, Graham CH (2016) Landscape demography: population change and its drivers across spatial scales. Q Rev Biol 91(4):459–485

Hanski I (1994) A practical model of metapopulation dynamics. J Anim Ecol 63(1):151–162
Hanski I (1998) Metapopulation dynamics. Nature 396(6706):41–49
Hanski I (1999) Metapopulation ecology. Oxford University Press, Oxford
Hanski I (2004) Metapopulation theory, its use and misuse. Basic Appl Ecol 5:225–229
Hanski I, Ovaskainen O (2000) The metapopulation capacity of a fragmented landscape. Nature 404 (6779):755–758
Hanski I, Ovaskainen O (2002) Extinction debt at extinction threshold. Conserv Biol 16 (3):666–673
Hanski I, Moilanen A, Gyllenberg M (1996) Minimum viable metapopulation size. Am Nat 147 (4):527–541. https://doi.org/10.1086/285864
Harrison S (1991) Local extinction in a metapopulation context: an empirical evaluation. Biol J Linn Soc 42(1–2):73–88
Heino M, Kaitala V, Ranta E, Lindstrom J (1997) Synchronous dynamics and rates of extinction in spatially structured populations. Proc R Soc B 264(1381):481–486. https://doi.org/10.1098/rspb.1997.0069
Hestbeck JB, Nichols JD, Malecki RA (1991) Estimates of movement and site fidelity using mark resight data of wintering Canada geese. Ecology 72(2):523–533
Hokit DG, Stith BM, Branch LC (2001) Comparison of two types of metapopulation models in real and artificial landscapes. Conserv Biol 15(4):1102–1113. https://doi.org/10.1046/j.1523-1739.2001.0150041102.x
Holt RD (1985) Population dynamics in 2-patch environments: some anomalous consequences of an optimal habitat distribution. Theor Popul Biol 28(2):181–208. https://doi.org/10.1016/0040-5809(85)90027-9
Holt RD (1992) A neglected facet of island biogeography: the role of internal spatial dynamics in area effects. Theor Popul Biol 41(3):354–371. https://doi.org/10.1016/0040-5809(92)90034-q
Huffaker CB (1958) Experimental studies on predation: dispersion factors and predator-prey oscillations. Hilgardia 27:343–383
Hylander K, Ehrlen J (2013) The mechanisms causing extinction debts. Trends Ecol Evol 28 (6):341–346. https://doi.org/10.1016/j.tree.2013.01.010
Jackson ST, Sax DF (2010) Balancing biodiversity in a changing environment: extinction debt, immigration credit and species turnover. Trends Ecol Evol 25(3):153–160. https://doi.org/10.1016/j.tree.2009.10.001
Johst K, Drechsler M (2003) Are spatially correlated or uncorrelated disturbance regimes better for the survival of species? Oikos 103(3):449–456
Johst K, Drechsler M, van Teeffelen AJA, Hartig F, Vos CC, Wissel S, Watzold F, Opdam P (2011) Biodiversity conservation in dynamic landscapes: trade-offs between number, connectivity and turnover of habitat patches. J Appl Ecol 48(5):1227–1235. https://doi.org/10.1111/j.1365-2664.2011.02015.x
Jongejans E, Shea K, Skarpaas O, Kelly D, Ellner SP (2011) Importance of individual and environmental variation for invasive species spread: a spatial integral projection model. Ecology 92(1):86–97. https://doi.org/10.1890/09-2226.1
Kallimanis AS, Kunin WE, Halley JM, Sgardelis SP (2005) Metapopulation extinction risk under spatially autocorrelated disturbance. Conserv Biol 19(2):534–546
Keddy PA (1981) Experimental demography of the sand-dune annual, cakile-edentula, growing along an environmental gradient in Nova Scotia. J Ecol 69(2):615–630. https://doi.org/10.2307/2259688
Keddy PA (1982) Population ecology on an environmental gradient: *Cakile edentula* on a sand dune. Oecologia 52(3):348–355. https://doi.org/10.1007/bf00367958
Kindlmann P, Melendez-Ackerman EJ, Tremblay RL (2014) Disobedient epiphytes: colonization and extinction rates in a metapopulation of Lepanthes rupestris (Orchidaceae) contradict theoretical predictions based on patch connectivity. Bot J Linn Soc 175(4):598–606. https://doi.org/10.1111/boj.12180

Kininmonth S, Drechsler M, Johst K, Possingham HP (2010) Metapopulation mean life time within complex networks. Mar Ecol Prog Ser 417:139–149. https://doi.org/10.3354/meps08779
Koenig WD (1998) Spatial autocorrelation in California land birds. Conserv Biol 12(3):612–620. https://doi.org/10.1046/j.1523-1739.1998.97034.x
Koenig WD (1999) Spatial autocorrelation of ecological phenomena. Trends Ecol Evol 14 (1):22–26. https://doi.org/10.1016/s0169-5347(98)01533-x
Koenig WD, Liebhold AM (2016) Temporally increasing spatial synchrony of North American temperature and bird populations. Nat Clim Chang 6(6):614. https://doi.org/10.1038/nclimate2933
Kuussaari M, Bommarco R, Heikkinen RK, Helm A, Krauss J, Lindborg R, Ockinger E, Partel M, Pino J, Roda F, Stefanescu C, Teder T, Zobel M, Steffan-Dewenter I (2009) Extinction debt: a challenge for biodiversity conservation. Trends Ecol Evol 24(10):564–571
Lebreton JD (1996) Demographic models for subdivided populations: the renewal equation approach. Theor Popul Biol 49(3):291–313. https://doi.org/10.1006/tpbi.1996.0015
Lebreton JD, Khaladi M, Grosbois V (2000) An explicit approach to evolutionarily stable dispersal strategies: no cost of dispersal. Math Biosci 165(2):163–176. https://doi.org/10.1016/s0025-5564(00)00016-x
Levin SA (1976) Population dynamic models in heterogeneous environments. Annu Rev Ecol Syst 7:287–310. https://doi.org/10.1146/annurev.es.07.110176.001443
Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. Bull Entomol Soc Am 15:237–240
Liebhold A, Koenig WD, Bjørnstad ON (2004) Spatial synchrony in population dynamics. Annu Rev Ecol Evol Syst 35:467–490. https://doi.org/10.1146/annurev.ecolsys.34.011802.132516
Liu JG, Hull V, Morzillo AT, Wiens JA (eds) (2011) Sources, sinks and sustainability. Cambridge University Press, Cambridge
Lotka AJ (1927) Fluctuations in the abundance of a species considered mathematically. Nature 119:12–12
Lundberg P, Ranta E, Ripa J, Kaitala V (2000) Population variability in space and time. Trends Ecol Evol 15(11):460–464
Mace GM, Collar NJ, Gaston KJ, Hilton-Taylor C, Akcakaya HR, Leader-Williams N, Milner-Gulland EJ, Stuart SN (2008) Quantification of extinction risk: IUCN's system for classifying threatened species. Conserv Biol 22(6):1424–1442. https://doi.org/10.1111/j.1523-1739.2008.01044.x
MacKenzie DI, Nichols JD, Hines JE, Knutson MG, Franklin AB (2003) Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84 (8):2200–2207
MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE (2006) Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier, Amsterdam
Malthus TR (1798) An essay on the principle of population. Annoynmous
Martin J, Nichols JD, McIntyre CL, Ferraz G, Hines JE (2009) Perturbation analysis for patch occupancy dynamics. Ecology 90(1):10–16. https://doi.org/10.1890/08-0646.1
Matter SF (2001) Synchrony, extinction, and dynamics of spatially segregated, heterogeneous populations. Ecol Model 141(1–3):217–226. https://doi.org/10.1016/s0304-3800(01)00275-7
Matthews DP, Gonzalez A (2007) The inflationary effects of environmental fluctuations ensure the persistence of sink metapopulations. Ecology 88(11):2848–2856. https://doi.org/10.1890/06-1107.1
McKay JK, Christian CE, Harrison S, Rice KJ (2005) "How local is local?" - a review of practical and conceptual issues in the genetics of restoration. Restor Ecol 13(3):432–440. https://doi.org/10.1111/j.1526-100X.2005.00058.x
McPeek MA, Holt RD (1992) The evolution of dispersal in spatially and temporally varying environments. Am Nat 140(6):1010–1027. https://doi.org/10.1086/285453

Mestre F, Canovas F, Pita R, Mira A, Beja P (2016) An R package for simulating metapopulation dynamics and range expansion under environmental change. Environ Model Softw 81:40–44. https://doi.org/10.1016/j.envsoft.2016.03.007
Moilanen A (1999) Patch occupancy models of metapopulation dynamics: efficient parameter estimation using implicit statistical inference. Ecology 80(3):1031–1043. https://doi.org/10.2307/177036
Moilanen A (2000) The equilibrium assumption in estimating the parameters of metapopulation models. J Anim Ecol 69(1):143–153
Moilanen A (2002) Implications of empirical data quality to metapopulation model parameter estimation and application. Oikos 96(3):516–530
Moilanen A (2004) SPOMSIM: software for stochastic patch occupancy models of metapopulation dynamics. Ecol Model 179(4):533–550. https://doi.org/10.1016/j.ecolmodel.2004.04.019
Moilanen A, Hanski I (1995) Habitat destruction and coexistence of competitors in a spatially realistic metapopulation model. J Anim Ecol 64(1):141–144. https://doi.org/10.2307/5836
Moilanen A, Hanski I (1998) Metapopulation dynamics: effects of habitat quality and landscape structure. Ecology 79(7):2503–2515
Moilanen A, Nieminen M (2002) Simple connectivity measures in spatial ecology. Ecology 83 (4):1131–1145
Morris WF, Doak DF (2002) Quantitative conservation biology: theory and practice of population viability analysis. Sinauer Associates, Inc, Sunderland, MA
Mortelliti A, Amori G, Boitani L (2010) The role of habitat quality in fragmented landscapes: a conceptual overview and prospectus for future research. Oecologia 163(2):535–547. https://doi.org/10.1007/s00442-010-1623-3
Neubert MG, Caswell H (2000) Demography and dispersal: calculation and sensitivity analysis of invasion speed for structured populations. Ecology 81(6):1613–1628. https://doi.org/10.1890/0012-9658(2000)081[1613:Dadcas]2.0.Co;2
Neubert MG, Kot M, Lewis MA (2000) Invasion speeds in fluctuating environments. Proc R Soc B 267(1453):1603–1610. https://doi.org/10.1098/rspb.2000.1185
Nichols JD, Hines JE, Lebreton JD, Pradel R (2000) Estimation of contributions to population growth: a reverse-time capture-recapture approach. Ecology 81(12):3362–3376
Norris K (2004) Managing threatened species: the ecological toolbox, evolutionary theory and declining-population paradigm. J Appl Ecol 41(3):413–426
Ovaskainen O (2004) Habitat-speclfic movement parameters estimated using mark-recapture data and a diffusion model. Ecology 85(1):242–257
Ovaskainen O (2008) Analytical and numerical tools for diffusion-based movement models. Theor Popul Biol 73(2):198–211. https://doi.org/10.1016/j.tpb.2007.11.002
Ovaskainen O, Sato K, Bascompte J, Hanski I (2002) Metapopulation models for extinction threshold in spatially correlated landscapes. J Theor Biol 215(1):95–108. https://doi.org/10.1006/jtbi.2001.2502
Ovaskainen O, Luoto M, Ikonen I, Rekola H, Meyke E, Kuussaari M (2008) An empirical test of a diffusion model: predicting clouded apollo movements in a novel environment. Am Nat 171 (5):610–619. https://doi.org/10.1086/587070
Ozgul A, Oli MK, Armitage KB, Blumstein DT, Van Vuren DH (2009) Influence of local demography on asymptotic and transient dynamics of a yellow-bellied marmot metapopulation. Am Nat 173(4):517–530. https://doi.org/10.1086/597225
Petzoldt T, Rinke K (2007) simecol: an object-oriented framework for ecological modeling in R. J Stat Softw 22(9):1–31
Pianka ER (1970) R-selection AND K-selection. Am Nat 104(940):592. https://doi.org/10.1086/282697
Pollock KH (1982) A capture-recapture design robust to unequal probability of capture. J Wildl Manag 46(3):752–757. https://doi.org/10.2307/3808568
Pulliam HR (1988) Sources, sinks, and population regulation. Am Nat 132(5):652–661
Pulliam HR (2000) On the relationship between niche and distribution. Ecol Lett 3(4):349–361

Pulliam HR, Dunning JB, Liu JG (1992) Population dynamics in complex landscapes: a case study. Ecol Appl 2(2):165–177. https://doi.org/10.2307/1941773

Ranta E, Kaitala V, Lundberg P (1997) The spatial dimension in population fluctuations. Science 278(5343):1621–1623. https://doi.org/10.1126/science.278.5343.1621

Reed JM, Mills LS, Dunning JB, Menges ES, McKelvey KS, Frye R, Beissinger SR, Anstett MC, Miller P (2002) Emerging issues in population viability analysis. Conserv Biol 16(1):7–19

Reeve JD, Cronin JT, Haynes KJ (2008) Diffusion models for animals in complex landscapes: incorporating heterogeneity among substrates, individuals and edge behaviours. J Anim Ecol 77 (5):898–904. https://doi.org/10.1111/j.1365-2656.2008.01411.x

Risk BB, de Valpine P, Beissinger SR (2011) A robust-design formulation of the incidence function model of metapopulation dynamics applied to two species of rails. Ecology 92(2):462–474

Rockwood LL (2009) Introduction to population ecology. Wiley, Chichester

Roff DA (1974) Analysis of a population model demonstrating importance of dispersal in a heterogeneous environment. Oecologia 15(3):259–275. https://doi.org/10.1007/bf00345182

Rota CT, Fletcher RJ Jr, Evans JM, Hutto RL (2011) Does accounting for detectability improve species distribution models. Ecography 34:659–670

Roy M, Holt RD, Barfield M (2005) Temporal autocorrelation can enhance the persistence and abundance of metapopulations comprised of coupled sinks. Am Nat 166(2):246–261

Royle JA, Kery M (2007) A Bayesian state-space formulation of dynamic occupancy models. Ecology 88(7):1813–1823. https://doi.org/10.1890/06-0669.1

Runge JP, Runge MC, Nichols JD (2006) The role of local populations within a landscape context: defining and classifying sources and sinks. Am Nat 167(6):925–938. https://doi.org/10.1086/503531

Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA, Baughman S, Cabin RJ, Cohen JE, Ellstrand NC, McCauley DE, O'Neil P, Parker IM, Thompson JN, Weller SG (2001) The population biology of invasive species. Annu Rev Ecol Syst 32:305–332. https://doi.org/10.1146/annurev.ecolsys.32.081501.114037

Sanderlin JS, Waser PM, Hines JE, Nichols JD (2012) On valuing patches: estimating contributions to metapopulation growth with reverse-time capture-recapture modelling. Proc R Soc B 279 (1728):480–488. https://doi.org/10.1098/rspb.2011.0885

Saura S, Pascual-Hortal L (2007) A new habitat availability index to integrate connectivity in landscape conservation planning: comparison with existing indices and application to a case study. Landsc Urban Plan 83(2–3):91–103. https://doi.org/10.1016/j.landurbplan.2007.03.005

Saura S, Rubio L (2010) A common currency for the different ways in which patches and links can contribute to habitat availability and connectivity in the landscape. Ecography 33(3):523–537. https://doi.org/10.1111/j.1600-0587.2009.05760.x

Schaub M, Gimenez O, Sierro A, Arlettaz R (2007) Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. Conserv Biol 21(4):945–955. https://doi.org/10.1111/j.1523-1739.2007.00743.x

Schrott GR, With KA, King AW (2005) Demographic limitations of the ability of habitat restoration to rescue declining populations. Conserv Biol 19(4):1181–1193. https://doi.org/10.1111/j.1523-1739.2005.00205.x

Schwartz MW (2008) The performance of the endangered species act. Annu Rev Ecol Evol Syst 39:279–299. https://doi.org/10.1146/annurev.ecolsys.39.110707.173538

Shaw MW (1995) Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance. Proc R Soc B 259(1356):243–248. https://doi.org/10.1098/rspb.1995.0036

Skellam JG (1951) Random dispersal in theoretical populations. Biometrika 28:196–218

Sklar E (2007) Software review: NetLogo, a multi-agent simulation environment. Artif Life 13 (3):303–311. https://doi.org/10.1162/artl.2007.13.3.303

Stevens MHH (2009) A primer of ecology with R. Springer, New York

Stott I, Hodgson D, Townley S (2018) popdemo: demographic modelling using projection matrices. R package version 1.3-0

Stubben C, Milligan B (2007) Estimating and analyzing demographic models using the popbio package in R. J Stat Softw 22(11):1–23
Sutherland C, Elston DA, Lambin X (2012) Multi-scale processes in metapopulations: contributions of stage structure, rescue effect, and correlated extinctions. Ecology 93(11):2465–2473
Talluto MV, Boulangeat I, Vissault S, Thuiller W, Gravel D (2017) Extinction debt and colonization credit delay range shifts of eastern North American trees. Nat Ecol Evol 1(7). https://doi.org/10.1038/s41559-017-0182
Thiele JC, Kurth W, Grimm V (2012) RNetLogo: an R package for running and exploring individual-based models implemented in NetLogo. Methods Ecol Evol 3(3):480–483. https://doi.org/10.1111/j.2041-210X.2011.00180.x
Thomas CD, Kunin WE (1999) The spatial structure of populations. J Anim Ecol 68(4):647–657. https://doi.org/10.1046/j.1365-2656.1999.00330.x
Tilman D, May RM, Lehman CL, Nowak MA (1994) Habitat destruciton and the extinction debt. Nature 371(6492):65–66
Tilman D, Lehman CL (1997) Habitat destruction and species extinctions. In: Tilman D, Kareiva P (eds) Spatial ecology: the role of space in population dynamics and interspecific interactions. Princeton University Press, Princeton, NJ
Traill LW, Bradshaw CJA, Brook BW (2007) Minimum viable population size: a meta-analysis of 30 years of published estimates. Biol Conserv 139(1–2):159–166. https://doi.org/10.1016/j.biocon.2007.06.011
Traill LW, Brook BW, Frankham RR, Bradshaw CJA (2010) Pragmatic population viability targets in a rapidly changing world. Biol Conserv 143(1):28–34. https://doi.org/10.1016/j.biocon.2009.09.001
Tremblay RL (1997) Distribution and dispersion patterns of individuals in nine species of Lepanthes (Orchidaceae). Biotropica 29(1):38–45. https://doi.org/10.1111/j.1744-7429.1997.tb00004.x
Tremblay RL, Ackerman JD (2001) Gene flow and effective population size in Lepanthes (Orchidaceae): a case for genetic drift. Biol J Linn Soc 72(1):47–62. https://doi.org/10.1006/bijl.2000.0485
Tremblay RL, Melendez-Ackerman E, Kapan D (2006) Do epiphytic orchids behave as metapopulations? Evidence from colonization, extinction rates and asynchronous population dynamics. Biol Conserv 129(1):70–81. https://doi.org/10.1016/j.biocon.2005.11.017
Walter JA, Sheppard LW, Anderson TL, Kastens JH, Bjørnstad ON, Liebhold AM, Reuman DC (2017) The geography of spatial synchrony. Ecol Lett 20(7):801–814. https://doi.org/10.1111/ele.12782
Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol 15(6):1419–1439. https://doi.org/10.1111/j.1365-294X.2006.02890.x
Wickham H (2007) Reshaping data with the reshape package. J Stat Softw 21(12):1–20
Wiegand T, Revilla E, Moloney KA (2005) Effects of habitat loss and fragmentation on population dynamics. Conserv Biol 19(1):108–121. https://doi.org/10.1111/j.1523-1739.2005.00208.x
Wilson RJ, Thomas CD, Fox R, Roy DB, Kunin WE (2004) Spatial patterns in species distributions reveal biodiversity change. Nature 432(7015):393–396. https://doi.org/10.1038/nature03031
Wright S (1943) Isolation by distance. Genetics 28(2):114–138
Wright S (1951) The genetical structure of populations. Ann Eugenics 15(4):323–354
Zipkin EF, Saunders SP (2018) Synthesizing multiple data types for biological conservation using integrated population models. Biol Conserv 217:240–250. https://doi.org/10.1016/j.biocon.2017.10.017

Chapter 11
Spatially Structured Communities



11.1 Introduction

Biodiversity is the variety of life. It is fundamental to all aspects of ecology and conservation biology. Biodiversity can be measured at different levels of organization and at different scales (Noss 1990; Magurran 2003). For instance, the number species in a local area, or species richness, is commonly measured in field investigations (Myers et al. 2000; Gotelli and Colwell 2001), whereas genetic diversity is a fundamental concept in evolutionary biology (Ellstrand and Elam 1993; Keller and Waller 2002). Over the last two decades, there has also been a great deal of interest in functional and phylogenetic diversity (Petchey and Gaston 2002; Cadotte et al. 2009; Cavender-Bares et al. 2009; Devictor et al. 2010; Mouchet et al. 2010).

To understand patterns of biodiversity over space and time and to effectively implement biodiversity conservation, concepts from community ecology are essential. Community ecology focuses on the importance of interactions among species and how communities assemble to drive variation in biodiversity (Mittelbach 2012). The importance of space in the outcomes of species interactions and community assembly has long been emphasized (Huffaker 1958; Diamond 1975). As a consequence, modern theory and concepts for community ecology emphasize the role of space (Vellend 2010; Leibold and Chase 2017), and many problems in conservation emphasize spatial issues for protecting or maintaining biodiversity (Moilanen et al. 2009; Rands et al. 2010).

Here, we provide an overview of how space influences biological communities, why space is important for biodiversity conservation, and we illustrate some common approaches for modeling communities over space and time. Space is generally important for understanding biodiversity for at least three reasons. First, some measures of diversity, such as beta diversity, are inherently spatial, focusing on the change in diversity across spatial and/or environmental gradients (Soininen et al. 2007; Anderson et al. 2011). Second, space can provide a mechanism for biodiversity patterns by altering community assembly and disassembly processes (Leibold et al. 2004). Third, incorporating spatial

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_11

issues into conservation strategies aimed at promoting biodiversity can provide new insight and can help solve some problems (Karp et al. 2012). We illustrate these issues through the use of spatial modeling of biological communities.

11.2 Key Concepts and Approaches

11.2.1 Spatial Community Concepts

Spatially structured communities can be described in several ways. For instance, diversity can be measured based on different types of variation (e.g., species-level, genetic) and it can be measured at different scales. In addition, there have been a wide range of concepts and theoretical developments on spatially structured communities. We first provide some terms and definitions regarding the ways in which diversity and communities are quantified across space. We then briefly provide an overview of some key spatial ecology concepts for communities, starting with early work on species–area relationships and moving to more contemporary concepts regarding metacommunities and hierarchies in community assembly.

11.2.1.1 A Diversity of Diversities

Species diversity is a major component of biodiversity. Species diversity is often partitioned into three types: alpha diversity, beta diversity, and gamma diversity (Fig. 11.1 and Table 11.1). *Alpha diversity* is the number of species residing at a locality and is often referred to as species richness. *Beta diversity* has been contextualized in a variety of ways, but it generally focuses on the turnover or change in

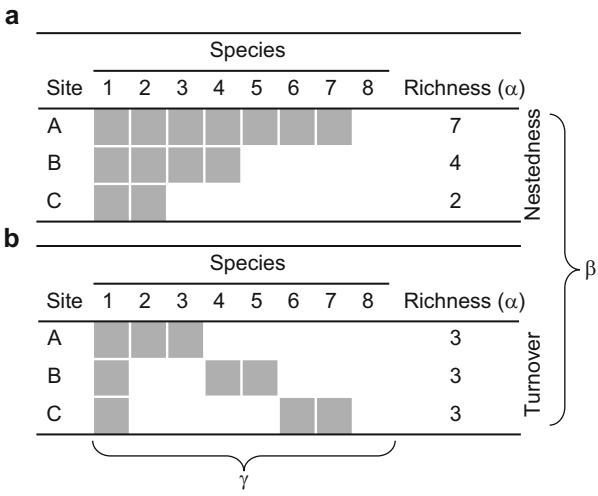


Fig. 11.1 A diversity of diversities illustrated through species-site matrices for two scenarios. In both, eight species occur in the species pool ($\gamma = 8$). In (**a**), species richness (α diversity) across sites varies, with beta diversity showing nestedness. In (**b**), species richness is constant across sites, but spatial turnover occurs. Adapted from Baselga (2010)

Table 11.1 Terms and concepts frequently considered in spatial community ecology

Term	Definition
Alpha diversity	The number of species residing at a locality.
Beta diversity	The change in species across environmental, spatial, or temporal gradients.
Competition-colonization tradeoff	When species that are good colonizers are poor competitors, and species that are good competitors are poor colonizers. This tradeoff can facilitate coexistence across landscapes.
Community assembly rules	Rules that make predictions for what species will occur in a location, given the species pool.
Gamma diversity	The total number of species in a region.
Habitat diversity hypothesis	One proposed mechanism for the species–area relationship. This hypothesis predicts that variation in the environment (habitat, resources, etc.) increases with area, leading to potentially more species.
Limiting similarity	A category of community assembly rules whereby a community is assembled via interspecific competition, such that species that coexist have low similarity in key traits of relevance to local resources.
Mass effect	When community structure is driven by species source-sink dynamics.
Metacommunity	A set of local communities linked by dispersal.
Ordination	Analyses that aim to summarize the variation of multivariate data into reduced dimensions or space by arranging species and/or locations along gradients. Typically based on eigen-analysis.
Nestedness	The change in species between two (or more) locations that arises from the loss of certain species.
Neutral community models	Models that assume community assembly is driven entirely by stochastic forces.
Passive sampling	A mechanism for species–area relationships that is based solely on sampling effort. As sample effort increases in larger patches or islands, the number of species detected increases.
Patch dynamics	When community assembly is viewed through the lens of variation in colonization-extinction dynamics, wherein better dispersers are frequently assumed to be poorer competitors.
Rank-abundance curve	Relative abundance of species plotted as a function of their rank abundance (most abundant species = rank 1). Used to visualize skewness in community composition.
Species–area relationship	The pattern whereby the number of species increases with habitat area.
Species pool	The total number of potential species in a region that are capable of colonizing localities.
Species sorting	When community structure over space is driven by individual species responses to environmental heterogeneity, such that certain local conditions may favor certain species and not others. In this concept, dispersal is not assumed to be a limiting factor.
Target effect	A mechanism for species–area relationships, whereby immigration is greater to larger habitats simply because larger areas are bigger targets for dispersers. This relationship typically scales with the linear diameter of habitats rather than the area per se.
Turnover	The change in species between two (or more) locations based on species replacement.

species across environmental, spatial, or temporal gradients (Anderson et al. 2011). Beta diversity is sometimes split into its nestedness and turnover components (Baselga 2010). *Nestedness* between two locations refers to the change in species based on the loss of species, where one location may be "nested" within another; that is, it is a nested subset of the location that contains more species (Wright and Reeves 1992). The idea of nestedness in biological communities has received a great deal of interest over the years (Wright et al. 1998; Mac Nally and Lake 1999; Kerr et al. 2000; Fernandez-Juricic 2002; Driscoll 2008), in part because it was a fundamental issue involved in the SLOSS (single-large versus several small) debate (Wright and Reeves 1992). *Turnover*, on the other hand, refers to a change in species via species replacement (not loss) (Williams 1996). *Gamma diversity* typically refers to the *species pool* in the region, or the species that are potentially available for colonizing local sites or communities (Karger et al. 2016). Understanding the interplay and dependence of each of these components of species diversity is of long-standing interest to community ecologists (Ricklefs 1987; Partel et al. 1996; Caley and Schluter 1997; McPeek and Brown 2000; Koleff and Gaston 2002; Podani and Schmera 2011; Lessard et al. 2012; Fukami 2015).

11.2.1.2 Species–Area Relationships

One of the few laws in ecology is the *species–area relationship* (Lawton 1999): the number of species increases with area (island area, patch area, etc.) (Fig. 11.2). This relationship has been documented throughout the world. There is a long history of exploring why this relationship exists and using this relationship to forecast changes in species diversity with ongoing environmental change (Gonzalez 2000; Seabloom et al. 2002; Thomas et al. 2004; Dobson et al. 2006).

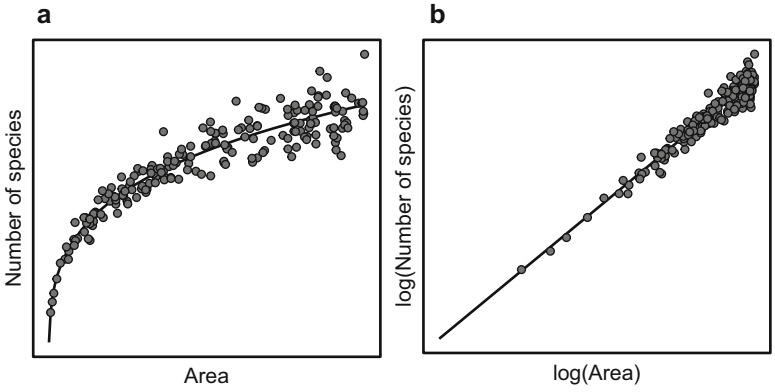


Fig. 11.2 The species–area relationship, shown on the (**a**) raw (original) scale, and (**b**) a log–log scale ($\log_{10}$ scale)

Arrhenius (1921) was one of the first scientists to formally quantify the species–area relationship, describing the relationship as a power function. Preston (1962) further developed this idea. He defined this relationship as:

$$S = cA^z, \tag{11.1}$$

where S is the number of species, A is area, and c and z are constants that describe the shape of the relationship of species with area. This relationship describes a pattern where the number of species quickly increases with area and then the rate of change slows (a power function relationship). It can be linearized when transformed to a log–log (base 10) scale as:

$$\log(S) = \log(c) + z\log(A). \tag{11.2}$$

There has been interest in understanding variation in z, because it describes the magnitude of the species–area relationship. Often z values tend to range from 0.10 to 0.25 (Drakare et al. 2006). We note that in practice, there are actually several types of species–area relationships that have been documented, where different types of sampling designs and functional forms (e.g., power, logistic) have been used to interpret species–area relationships (Scheiner 2003).

Given the ubiquity of the species–area relationship, the immediate question that arises is why this relationship occurs. Understanding why the relationship occurs is essential for understanding the importance of this pattern. Several hypotheses have been put forward to explain SAR; here we focus on a few common ones. First, the *habitat diversity* hypothesis suggests that as area increases, habitat diversity increases, such that the increase in number of species simply reflects an increase in the diversity of habitat or resources. Second, the *target effect* hypothesis posits that larger areas are more likely to be colonized, even simply by chance (or passive diffusion) due to an increasing circumference of the area (Bowman et al. 2002). Third, the *passive sampling* hypothesis states that SAR relationships are simply a reflection of greater sampling effort as area increases (Coleman et al. 1982), such that 10, 10 ha sites would yield the same number of species as 1, 100 ha site. Consequently, this hypothesis implies that there is nothing special about habitat area per se, and that the number of species per unit area sampled will not increase with increasing patch or island area. This hypothesis has similar rationale to the habitat amount hypothesis (Fahrig 2013). Finally, a great deal of interest and effort has focused on the hypothesis that SAR relationships arise from a balance of immigration and extinction effects that may change as a function of area. This hypothesis underlies the Equilibrium Theory of Island Biogeography (MacArthur and Wilson 1967).

11.2.1.3 Equilibrium Theory of Island Biogeography

Arguably, the most important conceptual development to our understanding of communities across space was the development of the Equilibrium Theory of Island

Biogeography (ETIB). MacArthur and Wilson (1967, 1963) developed this theory in detail to understand and predict the number of species residing on an island and the turnover rate of species on islands. The underlying premise of this development is that the number of species in an area is a balance between recurrent immigration of new species and recurrent extinction of species in a local area. When immigration and extinction are balanced, the number of species and the rate of species turnover is at equilibrium. This model is a neutral model (Caswell 1976), in the sense that species identity does not inform the model and expectations in the model are driven entirely by stochastic forces.

Immigration rates of new species per unit time are assumed to decline as the number of species increases on the island, eventually reaching 0 when the number of species on the island is equal to P, the species pool, or the mainland source pool of species. Simply put, as fewer species are able to immigrate from the mainland, immigration rates should decline. Often, this immigration curve is drawn as being non-linear (~exponential decline), to reflect the idea that some species might be better at dispersing than others, where good dispersers will immigrate rapidly, while poor dispersers will be slow to immigrate. Extinction rates are assumed to increase as the number of species on the island increases, where the abscissa is zero (no extinctions can occur when no species inhabit the island). Extinction rates are assumed to occur stochastically, such that extinction rates increase with the number of species simply because there is a greater number of potential species to go extinct. Again, this relationship is also frequently drawn as a non-linear (exponential) relationship, where interspecific competition may increase the extinction rate as the number of species increases (a linear rate would assume that all species behave independently of each other).

This theory received the most attention when MacArthur and Wilson invoked island area and isolation from mainland as critical factors that may alter immigration and extinction rates. MacArthur and Wilson (1967) assumed that as island area increased, the extinction rate should decline relative to smaller islands. The rationale for this assumption is that larger islands will harbor larger populations, where population size is proportional to island size (note: density is assumed to be constant, or in some cases decline; MacArthur 1972), such that demographic stochasticity may play a smaller role in extinction risk of individual species. This component of ETIB leads to specific predictions regarding species–area relationships—that species number increases with area and that turnover rates decline. MacArthur and Wilson also assumed that island isolation would alter immigration rates, where increasing isolation should reduce immigration. Note that since their seminal work, area effects have also been considered to alter immigration rates, where target effects occur (larger islands lead to greater immigration rates) (Lomolino 1990), and isolation has also been considered to alter extinction rates, where less isolated islands are expected to have lower extinction rates via rescue effects (immigration rates preventing extinction) (Brown and Kodric-Brown 1977). While much of the focus of this work has been on area and isolation, MacArthur and Wilson (1967) also developed a variety of related issues, such as the role of corridors, stepping stones, and island aggregation on expected numbers of species.

This theory has been instrumental in community ecology and conservation (Whitcomb et al. 1976; Whittaker et al. 2005). Nonetheless, it is now known that this framework does not capture many of the pressing issues influencing biodiversity across space and over time, such as edge and matrix effects, landscape complementarity, species interactions, and situations where no "mainlands" occur (Haila 2002; Laurance 2008). Importantly, this theory does not predict the distribution of individual species nor species identity (and related traits) in the community. Since this seminal work, several extensions have been made to accommodate some of these issues (Holt 1992; Cook et al. 2002; Gravel et al. 2011). One major advancement has been the development of metacommunity theory (Holyoak et al. 2005).

11.2.1.4 Metacommunities

The metacommunity concept extends ideas from metapopulation ecology (Chap. 10) and community ecology to explicitly understand variation in communities across space (Wilson 1992; Leibold et al. 2004; Holyoak et al. 2005; Leibold and Chase 2017). This concept aims to unite several processes that have been hypothesized to be critical to community structure over space (Vellend 2010). At its core, a *metacommunity* consists of local communities (i.e., communities residing at a particularly locality, such as a patch) that are linked spatially through dispersal. Leibold et al. (2004) identified four paradigms that have been applied to understanding metacommunities: the patch-dynamics paradigm, the species-sorting paradigm, the mass effects paradigm, and the neutral paradigm.

The *patch-dynamics* paradigm is a direct extension of two-species metapopulation models to N species. This paradigm emphasizes that species diversity may be limited by species interactions (e.g., competition) and dispersal. The focus is on colonization-extinction dynamics of N species, where it is often assumed that patches are similar and each patch is capable of containing populations of each species. In this paradigm, competition-colonization tradeoffs among species are often assumed (Levins and Culver 1971). In the *competition-colonization tradeoff*, it is assumed that poor dispersers are dominant competitors, while good dispersers tend to be poor competitors, which has been observed when contrasting some annual and perennial plants. This tradeoff provides a stabilizing mechanism for species coexistence across landscapes or regions. Tilman et al. (1994) popularized this general framework when modeling communities under scenarios of habitat destruction (see also Neuhauser 1998). This paradigm generally emphasizes that community structure is limited by variation in dispersal limitations.

In contrast to the patch-dynamics paradigm, the *species-sorting* paradigm emphasizes that environmental gradients drive variation in species diversity, while dispersal is less of a limiting force but rather dispersal allows species to track variation in resource gradients across landscapes. It assumes that diversity is driven by spatial niche separation above and beyond spatial dynamics arising from variation in dispersal and colonization (Holyoak et al. 2005).

The *mass-effects* paradigm is largely an extension of source-sink population dynamics (Chap. 10) to community assembly. In this paradigm, variation in immigration and emigration rates across landscapes and their impact on local population dynamics are emphasized. Variation in immigration and emigration rates can generate rescue effects (Brown and Kodric-Brown 1977) and can thereby offset competitive exclusion. In this paradigm, the role of dispersal is emphasized in being a key factor driving variation in local densities and it is assumed that patches vary in their suitability, leading to variation in immigration/emigration rates.

Finally, the *neutral* paradigm assumes that all species are similar in the competitive abilities, dispersal abilities, and fitness. This paradigm assumes that stochastic processes of species loss and gain drive variation in diversity. One of the first popular neutral models for species diversity was the Equilibrium Theory of Island Biogeography (MacArthur and Wilson 1967). The neutral paradigm has been emphasized by Hubbell and colleagues (Hubbell 2001) to explain community structure. Thus, dispersal and spatial dynamics are highly relevant to the neutral paradigm, although these dynamics are assumed to be driven by stochastic forces (Economo and Keitt 2008; Lowe and McPeek 2014; Guichard 2017).

11.2.1.5 Hierarchies from Regional Pools to Local Assemblages

There has been a great deal of interest in scaling from regional species pools to local assemblages by using community assembly rules to interpret how species may coexist. *Community assembly rules* are rules that make predictions for what species will occur in a location, given the regional species pool (Keddy 1992). Diamond (1975) was the first to consider the problem of assembly rules by considering how species traits (e.g., body size) could explain species composition of birds on islands. Predictions for assembly have also been made based on *limiting similarity* of key species' traits and for the role of environmental filtering in wetlands (Van der Valk 1981). *Environmental filtering* occurs when local environmental (abiotic) conditions "filter out" species from the regional pool (i.e., the environment selects against certain species), such that some species do not occur at certain localities due to the poor environmental conditions for that species (Cadotte and Tucker 2017). At its core, the assumption for environmental filtering is that species absence is not driven by biotic interactions (Kraft et al. 2015).

Poff (1997) took the general concept of environmental filtering and applied it in a hierarchical, landscape context (Fig. 11.3). In this framework, environmental filters operate at different scales, placing constraints on local communities. Different spatial constraints (e.g., spatial isolation, resource heterogeneity) operate at different spatial scales, and species traits (e.g., dispersal mode, foraging breadth) will lead to selective filtering of certain species based on these traits. It is often envisioned that environmental filters operate at relatively broad scales, while biotic interactions govern constraints at local scales (akin to ideas in species distribution modeling; see Chap. 7). This idea comes out of applications of hierarchy theory in landscape

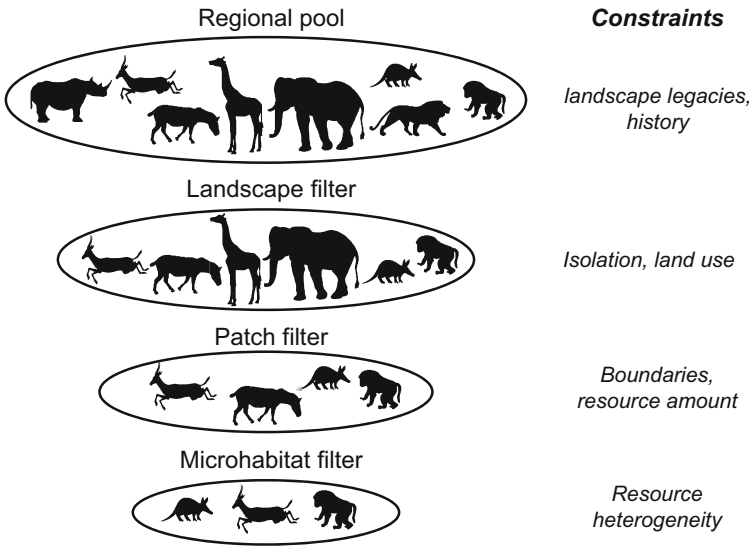


Fig. 11.3 Space-time hierarchies of environmental filtering. Shown are hierarchical filters and some examples of constraints and factors that operate at each scale to drive community structure

ecology (O'Neill et al. 1989; Urban et al. 1987), with Poff (1997) placing an emphasis specifically on local community assembly.

11.2.1.6 Communities and Conservation

Components of biodiversity are often used as targets for conservation. For instance, species richness is frequently considered as a key indicator of biodiversity across landscapes, albeit an imperfect one. Beta diversity is increasingly emphasized in conservation (Karp et al. 2012; Socolar et al. 2016), in part because of concerns of *biotic homogenization*: where environmental change causes communities to be more similar across space due to an increase in generalist and exotic species (Olden and Rooney 2006). At a larger scale, Identifying bioregions, or biographic regions that harbor similar communities, has been helpful for interpreting ecological dynamics and developing broad-scale conservation strategies (Vilhena and Antonelli 2015). At a global scale, identifying and mapping biodiversity hotspots across the planet has been central to conservation initiatives (Myers et al. 2000; Brooks et al. 2002; Orme et al. 2005).

These components of diversity are frequently integrated in spatial conservation planning through the identification of sites with high local diversity and how a collection of sites or protected areas combine to reach conservation goals through several conservation concepts (Kukkala and Moilanen 2013). For instance, *comprehensiveness* refers to the objective of capturing the full spectrum of biodiversity in the region of interest while *representativeness* describes the extent to which a

collection of sites (or protected areas) meets that goal (Kukkala and Moilanen 2013). Concepts that directly capture alpha and beta diversity are irreplaceability and complementarity. *Irreplacibility* describes the importance of a potential site for conservation, in terms of its unique contribution to the overall biodiversity goal, such that if the site is lost the ability to reach conservation goals is hampered (Ferrier et al. 2000). *Complementarity* in conservation planning represents the degree to a site (or group of sites) contributes to unrepresented features—typically species—to an existing set of protected sites (Margules and Pressey 2000). A site has higher complementarity when it contains species not protected by existing sites. Thus, when high turnover occurs between one or more protected areas and another site being considered for conservation, there is high complementarity for that site.

The role of species interactions across space is also increasingly considered in conservation strategies. In particular, certain types of interactions, such as plant-pollinator interactions are important for maintaining ecosystem services, and such interactions can vary across landscapes (Winfree et al. 2009). Trophic interactions are also important in some conservation planning, particularly in marine and freshwater environments (Baskett et al. 2007; Decker et al. 2017). As a consequence, spatial modeling of communities and related ecosystem services has increased over the years and is essential for these types of conservation efforts (Brosi et al. 2008; Moilanen et al. 2009; Kaiser-Bunbury and Bluthgen 2015).

11.2.2 Common Approaches to Understanding Community–Environment Relationships

Predicting and mapping communities over space is challenging. There are several types of modeling frameworks for communities. Frameworks that focus on environmental filtering and species sorting as the primary drivers of (meta) communities are most commonly used, likely in part because these frameworks are more feasible to implement than other frameworks that emphasize other metacommunity processes (e.g., variation in dispersal). Ferrier and Guisan (2006) classified community-level models into three categories: (1) predict first, assemble later; (2) assemble first, predict later; and (3) assemble and predict simultaneously (see also D'Amen et al. 2017). Here, we follow this categorization to illustrate several themes regarding the spatial modeling of communities.

11.2.2.1 Predict First, Assemble Later

One way in which models for communities have been developed is simply to model each species separately (see, e.g., Chaps. 6 and 7) and then with this multi-species

information, aggregate or pool across species to predict communities across space. In the species distribution modeling literature, this general approach is commonly referred to as "stacked species distribution models", or S-SDM (Guisan and Rahbek 2011). This approach implicitly emphasizes that species may respond individualistically to environmental relationships, a "Gleasonian" perspective for community structure.

Predictions for models of individual species can be combined in a variety of ways. For example, probabilities of occurrence can be truncated to expected presence–absence and then summed across species to derive species richness. Alternatively, individual model outputs could be used to interpret spatial variation in community composition by applying predictions to similarity or distance-based metrics (see below). Consequently, this approach uses model predictions as inputs for community classification and summary metrics, rather than the raw data.

11.2.2.2 Assemble First, Predict Later

In this approach, communities are first summarized in some way without reference to the environment. For instance, species richness may be quantified, community types (e.g., number of foraging guilds) may be summarized, or community (dis)similarity may be estimated.

Species richness can be estimated in a variety of ways. Typically, the raw count of species is a biased estimator of species richness. Instead, community ecologists attempt to adjust raw counts in at least three different ways. First, rarefaction is commonly used (Gotelli and Colwell 2001). Rarefaction acknowledges that the number of species observed will be a function of the number of individuals detected: as the number of individuals detected increases, we expect that that number of species detected will also increase. This relationship is typically asymptotic, such that rarefaction curves can be used to interpret the point at which sampling for the community was sufficient for interpreting species richness. When using rarefaction curves estimated at different localities, the species richness estimate is often truncated to the site with the lowest number of individuals, thereby allowing less biased comparison among locations regarding species richness. We note that rarefaction approaches have also been extended to account for spatial dependence in species data (Bacaro et al. 2016). Second, some estimators adjust counts of species based on the number of "singletons" (i.e., number of species detected once) or "doubletons" (i.e., the number of species detected twice) in the data (Palmer 1990; Nichols et al. 1998). This is sometimes referred to as species richness estimation through extrapolation, rather than through truncation, as in rarefaction (Colwell and Coddington 1994). The idea here is that if singletons and/or doubletons are rare in the data, then it is likely that few species have been missed. In contrast, if singletons and/or doubletons are frequent, then it is likely that many species have been missed and sampling was not sufficient. The jackknife estimator and Chao estimators for species richness

are both based on this general idea (Palmer 1990). The third approach is to formally estimate species-specific detectability and subsequently derive species richness once species-specific detectability is estimated. This approach is an extension of occupancy modeling (MacKenzie et al. 2002), termed "multi-species occupancy modeling" (Dorazio et al. 2006; Royle and Dorazio 2008; Kery and Royle 2016).

Summarizing community composition typically involves the use of (dis)similarity matrices. These matrices quantify the pairwise (dis)similarity between all sampling locations (i.e., they are square matrices). Similarity can be quantified in several ways (Koleff et al. 2003; Barwell et al. 2015). For abundance data, a Bray–Curtis index is frequently used:

$$\beta_{ij} = \frac{B + C}{2A + B + C}, \tag{11.3}$$

where A is the sum of the minimum abundance of species between site i and j (i.e., the number of individuals occurring at both sites), B is the number of individuals unique to site i and C is the number of individuals unique to site j. For binary data of species occurrence, a common approach is to use the Sørenson dissimilarity index (another common measure is the Jaccard Index):

$$\beta_{\text{sor},ij} = \frac{b + c}{2a + b + c}, \tag{11.4}$$

where a is the number of species common to sites i and j, b is the number of species in site i that are not in site j and c is the number of species in site j that do not occur in site i.

The Sørenson and Bray–Curtis indices are functionally very similar but work with binary and count data, respectively. Both of these metrics range from 0 to 1. Dissimilarity is simply 1 − similarity and can sometimes be considered a distance metric (note, however, that some dissimilarity matrices do not satisfy the "triangle inequality" and are thus not measures of ecological distance). With this approach, we may be interested in only considering the nestedness and turnover components of beta diversity (Fig. 11.1; Baselga 2010). For the Sørenson index, turnover between two sites is:

$$\beta_{\text{turn},ij} = \frac{\min(b, c)}{a + \min(b, c)}. \tag{11.5}$$

Nestedness can then be described as the fraction of $\beta_{\text{sor},i,j}$ not explained by $\beta_{\text{turn},i,j}$:

$$\beta_{\text{nest},ij} = \beta_{\text{sor},ij} - \beta_{\text{turn},ij}. \tag{11.6}$$

With these newly assembled matrices and community summary statistics, we can then proceed to predict changes in communities across space.

11.2.2.3 Predict and Assemble Together

Rather than treating the assembly of communities and their prediction over space as separate components, several modeling approaches integrate these steps formally into a single modeling framework. Multivariate regression (Ovaskainen et al. 2010), constrained gradient ordination techniques (e.g., canonical correspondence analysis) (Palmer 1993), multi-species occupancy modeling (Dorazio et al. 2006; Iknayan et al. 2014), and joint community models (Warton et al. 2015a) are just some techniques that approach the problem in this way. In this case, modeling frameworks typically provide predictions for each species, thereby honoring species identity, such that species richness is typically a derived parameter from this modeling framework (blurring the lines between predict first and assemble later approaches and predict and assemble together).

11.2.3 Spatial Models for Communities

Depending on the framework considered for community modeling, there are a variety of modeling approaches that could be considered. Here, we provide a brief overview of those approaches that are commonly used, with a focus on approaches that have not been considered elsewhere in the book. We first describe these common approaches and then explicitly address how the problem of space can be accommodated with these models and related issues for communities.

Spatial community models typically either work with summaries of species, such as species richness (Rahbek and Graves 2001), distance-based matrices regarding similarity in community composition (Ferrier et al. 2007), or work directly with species-level variation in occurrence or abundance (Rahbek and Graves 2001; Ovaskainen et al. 2010). Recently, it has been argued that the latter shows better properties than using distance-based summary statistics (Warton et al. 2012), because distance-based analyses can conflate dispersion versus location effects (Fig. 11.4). Some have organized these approaches into algorithmic models and statistical, model-based approaches (Warton et al. 2015b). Algorithmic models are those that are defined based on a set of algorithmic steps taken to interpret the data and these typically do not take fully into account the statistical properties of the data; examples include several techniques based on *ordination* (see below). Model-based approaches focus on explicit, multivariate statistical models that attempt to capture the statistical properties of the data (Warton et al. 2015b); most approaches are extensions of the generalized linear model.

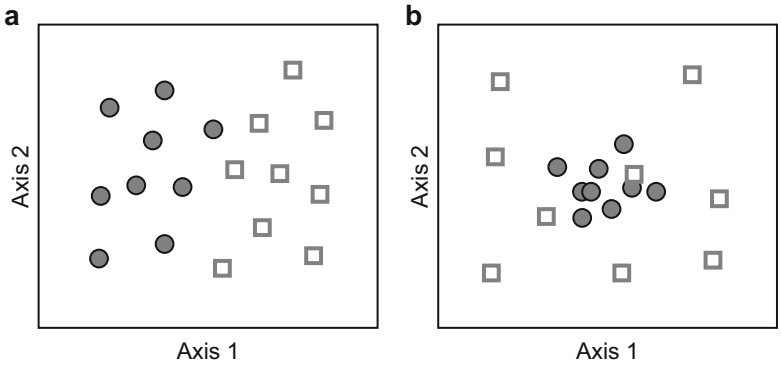


Fig. 11.4 An illustration of (**a**) location versus (**b**) dispersion effects for two groups observed in ordination techniques that use distance-based approaches. Adapted from Anderson et al. (2008) and Warton et al. (2012)

11.2.3.1 Multivariate Regression Analysis

Regression models can be extended to simultaneously model multiple species in a community. In this case, there are multiple response variables and as such, these models are referred to as multivariate regression techniques (rather than "multiple regression," which refers to situations where there is >1 explanatory variable).

Multivariate regression can be implemented in a variety of ways (Legendre 1993; Lichstein 2007; Wang et al. 2012). Traditionally, this approach used distance matrices of response and explanatory variables and used permutation tests to assess significance, because of the lack of independence of site pairs in the matrix formulation. In this approach, matrix regression can be described as:

$$d_{ij} = \alpha + \beta |x_i - x_j|, \tag{11.7}$$

where d_{ij} is the distance (e.g., compositional dissimilarity) between locations i and j and $|x_i - x_j|$ is the absolute value of the difference in environmental variable x between locations.

More recently, generalized linear models (GLMs) have been advocated (Wang et al. 2012; Warton et al. 2012) as a way to analyze community data. This GLM approach may be useful for community data because it can be applied to non-normal response variables typically used in community-level modeling without resorting to summarizing species composition based on distance matrices (Warton et al. 2015a). In this case, GLMs are fit to each species separately, similar to a "predict first-assemble later" strategy; however, multi-species (community-level) inference is made based on the suite GLM models fit. Statistical tests have been developed to account for correlations among species (via permutation tests) as well as using combined summary statistics (e.g., sums of squares across models) to make community-wide inference.

The GLM approach can also be extended to generalized linear mixed model (GLMM) formulations, allowing to account for potential dependencies between species. GLMMs can also reduce the need for permutation tests for inferences (as used in some multivariate regression approaches). In Chap. 6, we specified a generalized linear model for the presence–absence of a single species as:

$$\text{logit}(p_i) = \alpha + \beta x_i, \tag{11.8}$$

where p_i is the expected value for the probability of occurrence for sampling unit i, α is the intercept, β_1 is the slope (coefficient), x_i is the explanatory variable measured at i. Multivariate GLMMs extend this idea to K species as:

$$\text{logit}(p_{ik}) = \alpha + \beta x_i + \gamma_k + \delta_k x_i, \tag{11.9}$$

where now the response variable is the presence–absence of species k at location i. We can account for different species prevalence by adding a species-level random intercept, γ_k, and for variation environment relationships among species through the use of species-specific random coefficients (aka random slopes), δ_k, for an environmental variable x (Bates et al. 2015; Warton et al. 2015a) (Fig. 11.5). In these cases, random effects are typically assumed to be distributed as $\sim N(0, \sigma^2)$. This general approach lies at the heart of several advances in community-level analyses and has been extended to account for imperfect detection (Dorazio et al. 2006), metacommunity colonization-extinction dynamics (Dorazio et al. 2010), hierarchical spatial scaling effects on communities (Ovaskainen et al. 2016a), the potential for

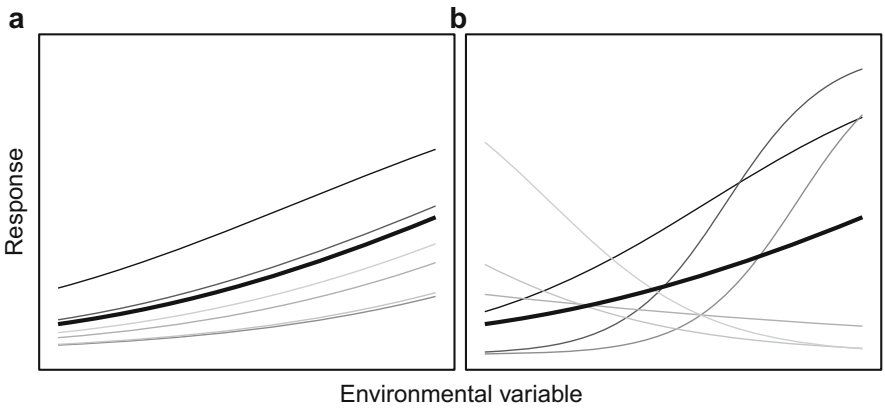


Fig. 11.5 An illustration of the difference between (**a**) random intercepts and (**b**) random coefficients (or random slopes) in generalized linear mixed models. Note that for (**b**), both random intercept and coefficients are shown. Grey lines are species-specific responses, while black line is the average response across species

biotic interactions between pairs of species by altering the variance–covariance matrices of random effects (Ovaskainen et al. 2010), as well as trait-based dependencies (Dorazio and Connor 2014).

11.2.3.2 Canonical Ordination: Redundancy and Canonical Correspondence Analysis

Direct gradient analysis, also known as canonical ordination or constrained ordination analysis, is often used by community ecologists to interpret how communities respond to environmental gradients. In this case, rather than only considering the species community in the *ordination* (as in principal components analysis, PCA, and correspondence analysis, CA; see Legendre and Legendre 2012), the community is related to environmental and/or spatial data in the context of the ordination. Two of the most common approaches are redundancy analysis (RDA) and canonical correspondence analysis (CCA).

Redundancy analysis is a method that effectively combines regression-like techniques with ordination (specifically, PCA). The general idea of RDA is that it is a multivariate linear regression where the fitted values are then subjected to PCA, which provides eigenvectors of the fitted values (Borcard et al. 2011). RDA then takes these eigenvectors and computes new orthogonal (i.e., independent) axes that are linear combinations of all explanatory variables, where the first axis explains the most variation in the response variables, the second axis explains the next most, and so on. This aspect of RDA is reflected in the decrease in eigenvalues for each axis (similar to PCA). The analysis can then be summarized based on species scores, site/location scores (summarizing species scores for each site), and site constraints (the linear combinations of environmental variables for each site). This approach is appropriate when one expects linear environmental relationships, although similar to linear regression (see Chaps. 6 and 7), polynomial terms can be added when warranted to capture some types of non-linear relationships.

Canonical correspondence analysis is similar to RDA but rather than using PCA in its formulation, it uses correspondence analysis (CA). It captures Gaussian relationships of species responses to environmental gradients. Because niche theory often envisions species responses across environmental gradients as hump-shaped, Gaussian curves, CCA has had major appeal since its introduction in the 1980s (Ter Braak 1987). However, there are known limitations of CCA, particularly its use of a χ^2 distance among sites. This distance measure is known to be a poor distance metric for community composition analyses. Consequently, there is currently a greater focus on the use of RDA for direct gradient ordination analyses (Borcard et al. 2011), and we focus on RDA below.

11.2.3.3 Generalized Dissimilarity Modeling

Generalized dissimilarity modeling (GDM) is increasingly used to understand and predict beta diversity across space for ecology and conservation problems (Ferrier et al. 2007; Thimassen et al. 2011; Fitzpatrick and Keller 2015; Jewitt et al. 2016; Rose et al. 2016). This approach is a non-linear extension of multivariate regression, where the response variables are measures of community dissimilarity, and predictors often include spatial (e.g., distance matrices) and environmental factors.

The GDM approach was derived to accommodate two forms of non-linearity in community modeling. First, because dissimilarity is constrained to the 0–1 scale, non-linearities of the response variable occur. This non-linearity is addressed by formulating the problem as a generalized linear model with a custom link function and error distribution (Ferrier et al. 2007). The link function, η, used is:

$$\eta = -\log(1 - \mu). \tag{11.10}$$

where μ is the expected value. Note that a beta distribution could also be used, which is a continuous distribution bounded to the 0–1 scale. Second, the rate of turnover at different locations on environmental gradients is expected to be non-linear. To address this issue, GDM fits non-linear, monotonic functions directly to the environmental variables, which are referred to as I-spline basis functions (Ferrier et al. 2007). I-splines are similar to the splines discussed in Chaps. 6 and 7, with the general difference being that they are constrained to be non-decreasing functions. This constraint makes sense in this case because we expect a priori that turnover rates should increase with increasing distance across environmental gradients. Similar to standard matrix regression (see above), significance is inferred through permutation tests.

11.2.3.4 The Problem of Space

Most of the above mentioned approaches only indirectly account for spatial dependence in community modeling. Spatial dependence in community modeling is often overlooked (Urban et al. 2002), but Dray et al. (2012) argued that it may alter inferences in our understanding and conservation of communities. Furthermore, they argued that spatial dependence only needs to occur in a portion of the community for it to potentially impact inferences.

Partial ordinations have long been used to account for potential spatial dependence via the inclusion of a geographic distance matrix in modeling (Borcard et al. 1992). Such matrices could be based on Euclidean or some other (effective) distance metric (see Chap. 9). The distance matrix (a square matrix of pair-wise distances between sites) is frequently used as a predictor or "controlling" variable (Borcard et al. 1992). Partial ordination is often used to then partition variance based on

different spatial and environmental factors (Cushman and McGarigal 2002). Yet partitioning generally assumes only additivity in the explanatory variables and it can yield negative components of variance due to interactions between variables. As such, partitioning should be used with caution.

Partial Mantel tests have also been frequently used to account for space. Mantel tests are statistical tests of the correlation between two distance matrices of the same rank (i.e., the same dimensions). Mantel tests calculate a correlation coefficient between the two matrices, and significance is inferred via permutation tests. These matrices are symmetric, distance-based matrices, so the number of distances is $n(n-1)/2$, or the number of observations in the upper (or lower) triangle of the matrices. Typically, the Pearson correlation coefficient is used (see Chap. 5). To assess significance of the Mantel correlation, the rows and columns of one the matrices are shuffled many times and the Mantel correlation is calculated on these randomized matrices. Significance is then inferred based on the proportion of times the observed correlation is higher than that of the correlations from the randomized matrices.

In a spatial context, Mantel tests can provide a single global test of autocorrelation for community data when comparing a spatial distance matrix (e.g., geographic distance) the community dissimilarity matrix. It is important to note here that the implicit assumption is that autocorrelation is linear gradient (i.e., the Mantel test typically uses a linear correlation coefficient). Also, there has been some criticism of this approach for a variety of reasons (Guillot and Rousset 2013; Legendre et al. 2015).

The Mantel correlogram is a multivariate extinction of the correlogram described in Chap. 5, which quantifies spatial autocorrelation as a function of distance (Bjørnstad and Falck 2001; Borcard and Legendre 2012). For each distance bin, the Mantel correlogram simply calculates a normalized correlation coefficient based on comparing the species dissimilarity matrix to a binary matrix, where sites within the distance bin are 0 and all others are 1. Stringing these correlation coefficients together results in a Mantel correlogram. Significance is inferred via permutation in the same manner as with a standard Mantel test.

Multivariate variograms can also be used. Wagner (2003) pioneered the application of multivariate variograms to community data in ecology. In this application, she derived the *variogram matrix*, $\mathbf{C}(d)$, for communities, where the diagonal is the semivariance for species i at distance class d (see Chap. 5) and the off-diagonals represent the pairwise cross-variograms for species i and j at distance class d. Cross-variograms are similar to variograms except that they quantify the distance-dependent covariance between two types of observations; in this case, two species. A cross-variogram for species i and j can be quantified as:

$$\gamma_{i,j}(d) = \frac{1}{2n_d} \sum \big(z(x_i) - z(x_i + d)\big)\big(z(x_j) - z(x_j + d)\big), \tag{11.11}$$

where γ is a measure of covariance, n is the number of observations at distance bin d, and z is the observation at location x_i. The variogram matrix can be used in a variety

of ways to interpret spatial dependence of communities. Beyond the species-specific variogram for spatial dependence and pairwise cross-variogram for spatial covariance between pairs of species, Wagner (2003) emphasized two other properties. First, the sum of the diagonal of $\mathbf{C}(d)$, she referred to as the empirical variogram of complementarity, or the spatial complementary of species composition at locations. Second, the sum of $\mathbf{C}(d)$ (diagonal + off-diagonals) can be considered the empirical variogram for sample-level species richness.

Multivariate variograms can be extended to ordination techniques, commonly referred to as multiscale ordination. The idea is similar to that described above for species composition and richness. In a nutshell, the $\mathbf{C}(d)$ matrices are summed across distance classes to create a global matrix $\mathbf{C}$ of empirical variance–covariance. This matrix is then subjected to ordination techniques, typically either PCA (Wagner 2003) or CA (Wagner 2004). Eigenvalues from the ordination can then be partitioned among distance classes and plotted as a function of distance, providing an empirical variogram of ordination axes that describe the spatial covariance of complementary in the species assemblage.

While geographic distance matrices are frequently used in Mantel tests and related analyses (e.g., GDM), the use of geographic distance matrices for inferring and controlling for spatial dependence in community-level modeling may be limited (Dray et al. 2012), due to the difficulty of proper interpretation (Legendre et al. 2015) and potential low power in detecting spatial structures (Legendre et al. 2005). Yet Borcard and Legendre (2012) contrasted multivariate variograms and Mantel correlograms using simulations, finding that under the simulated conditions, the power of these multivariate approaches was high and similar to univariate approaches.

An alternative to the use of distance matrices is using spatial weighting matrices, which come in several forms (Dray et al. 2012). Spatial eigenvector mapping (Dray et al. 2006) described in Chaps. 5 and 6 is one technique that is based upon spatial weighting matrices. Like a distance matrix, a spatial weighting matrix is a site-by-site matrix (i.e., a square matrix) that describes the potential pairwise linkages between sites. Weights can be binary or weighted (continuous, non-negative). This weighting matrix can also be directed (links between i and $j \neq j$ and i) to account for directed flows across landscapes (Blanchet et al. 2008). The subsequent incorporation of spatial weighting matrices can often occur in ways similar to the inclusion of geographic distances in the methods described above. This general approach provides great flexibility in formally capturing the role of space on communities.

11.3 Examples in R

11.3.1 Packages in R

In R, there are a few libraries that can be used for community-related models. Some common packages include the `vegan` package for ordination techniques

(Dixon 2003), `betapart` for interpreting beta diversity metrics (Baselga and Orme 2012), `gdm` for fitting generalized dissimilarity models (Manion et al. 2018), `mvnabund` and `VGAM` for multivariate GLM models for abundance and occurrence (Wang et al. 2012).

11.3.2 The Data

We return to the data shown in Chaps. 6 and 7 regarding bird distribution in Montana and Idaho, USA (Hutto and Young 2002). Sampling locations consist of point counts (100-m radius), along a transect (10 points/transect; transects are approximately 3 km long), with transects randomly selected within USFS Forest Regions across Montana and Idaho. Previously, we considered only one species; here, we extend our questions and analysis to the community. To do so, we only consider species adequately sampled by point counts (e.g., we remove waterfowl, raptors, and nocturnal species). In this example, we pool data across 3 years (2000, 2002, and 2004) for each point location. We consider three covariates used in Chap. 7: elevation, precipitation, and canopy cover.

11.3.3 Modeling Communities and Extrapolating in Space

We first illustrate common ways to approach modeling without explicit focus on incorporating space into the analysis. We then extend these ideas to formal accounting of space.

To begin, we will import a raster layers of elevation and canopy cover, as well as data on species detections at points, using the `raster` package.

```
> library(raster)

> Elev <- raster("elev.gri") #elevation layer (km)
> Canopy <- raster("cc2.gri") #linear gradient, from PCA
> Precip <- raster("precip.gri") #precipitation (cm)

#convert precipitation to meters
> Precip <- Precip / 100
> layers <- stack(Canopy, Elev, Precip)
> names(layers) <- c("canopy", "elev", "precip")

#species data
> birds <- read.csv("birdcommunity.csv")
```

These community data come in a format that is common for data entry purposes, where each row of data reflects a detection of a species at a site. We need to re-format

the data to produce a species by site data frame, where the columns are species and the rows are sites (Fig. 11.1). Also, note that the coordinates for the site data are in WGS84, which is not the same coordinate reference system as the raster data.

We first convert these data to a `SpatialPointsDataFrame` and transform the data to the projection of the raster data.

```
> birds.latlong <- data.frame(x = birds$LONG_WGS84, y =
  birds$LAT_WGS84)
> birds.attributes <- data.frame(transect = birds$TRANSECT,
  point = birds$STOP, species = birds$SPECIES, pres =
  birds$PRES)

#define CRS
> crs.latlong <- CRS("+proj=longlat +datum=WGS84")
> crs.layers <- CRS("+proj=aea +lat_1=46 +lat_2=48 +lat_0=44
  +lon_0=-109.5 +x_0=600000 +y_0=0 +ellps=GRS80 +datum=NAD83
  +units=m +no_defs")

#create SpatialPointsDataFrame
> birds.spdf <- SpatialPointsDataFrame(birds.latlong, data =
  birds.attributes, proj4string = crs.latlong)

#transform CRS for sites to layers CRS
> birds.spdf <- spTransform(birds.spdf, crs.layers)

#data frame with new x,y coordinates
> birds.df <- data.frame(birds.spdf@data, x =
  coordinates(birds.spdf)[,1], y = coordinates(birds.spdf)[,2])

> head(birds.df, 2)

##
 transect point species pres x y
1 452511619 5 AMDI 0 59142.22 173151.8
2 452511619 6 AMDI 0 58834.36 173185.7
```

Now, we reformat the data to a wide format to create a format of a site $\times$ species data frame with the `reshape2` package (Wickham 2007).

```
> library(reshape2)
> species.site <- dcast(birds.df, transect + point + x + y ~
  species, value.var = "pres")

#no attributes (species only)
> spp.matrix <- species.site[, -c(1:4)]
```

Finally, we will check for very rare species, creating a vector of names of species with detections on $>$20 points:

```
#subset based on frequency of occurrence
> prevalence <- colSums(spp.matrix)
> prevalence.20 <- prevalence[prevalence > 20]
> species.20 <- names(prevalence.20)
```

We use this list of species to subset the data for species with detections on >20 points (out of the 1145 points). To do so, we use the `%in%` command to select columns with column names matching our vector of species to retain:

```
> species.matrix <- spp.matrix[,colnames(spp.matrix) %in% species.20]
```

We can summarize this species matrix in a variety of ways. For instance, we can calculate the observed number of species per point (see below), as well as how prevalent species are across the points sampled (Fig. 11.6). In this case, observed (uncorrected) species richness varies considerably, with 12.8 species being observed on average. The prevalence of most species is low: only a few species are commonly observed across most points (Fig. 11.6b). Community composition and similar can also be calculated. For instance, the Sorenson index for community dissimilarity can be calculated as:

```
> sorenson <- vegdist(species.matrix, method = "bray")
> sorenson.mat <- as.matrix(sorenson)
```

Where the method `"bray"` reduces to the Sørenson with binary data (as is the case here). With these data, we extract relevant environmental information from the points and check for correlations among environmental variables:

```
> site.cov <- extract(layers, species.site[,c("x","y")])
> cor(site.cov)
```

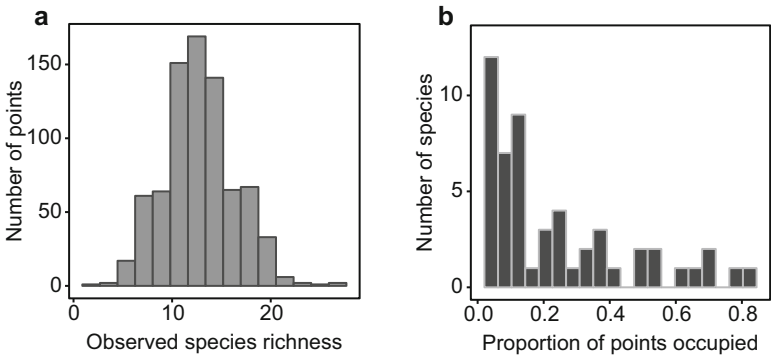


Fig. 11.6 Summarizing community-data for the Landbird Monitoring Program. Data come from 782 points sampled during 3 years (2000, 2002, and 2004). Shown are point-level observations of (**a**) species richness (for species >20% point locations) and (**b**) species prevalence

```
##
 canopy elev precip
canopy 1.00 0.01 0.13
elev 0.01 1.00 −0.04
precip 0.13 −0.04 1.00
```

These environmental variables are not strongly correlated at sampling locations, so in modeling we can consider each of these variables without any substantial concerns regarding collinearity (Dormann et al. 2013). With this information, we can now proceed to spatial modeling of communities.

11.3.3.1 Predict First, Assemble Later

First, we model each species separately and combine them to make predictions across the region, sometimes referred to as "stacked species distribution models," or S-SDMs (Dubuis et al. 2011; D'Amen et al. 2015). To do so, we will illustrate with similar approaches as described in Chaps. 6 and 7, where we use a logistic regression framework to model each species individually, and store relevant output for each species in lists for post-processing. Note, if we were interested in community-level inferences, such as if the community overall changed with elevation or canopy cover, we could use the `mvabund` package (Wang et al. 2018) to automatically model each species as a function of covariates (using the `manyglm` function). The primary benefit of that package is that there are summary inference-related tests that can be applied to the species-specific logistic regression models. Here, we illustrate applying these models manually, which could provide more flexibility on the types of models considered for each species.

```
> pred.map <- list() #stores map predictions
> pred.coef <- list() #stores coefficients
```

Storing in a list format can be particularly helpful when each species has different amounts of summary statistics, such as if different covariates are used for each species (e.g., species-specific model selection). Below we consider each of the environmental covariates and include a potential non-linear effect of elevation on species distribution.

```
> Nspecies <- ncol(species.matrix)
> Nsites <- nrow(species.matrix)

#create covariate vectors for simpler processing
> canopy <- site.cov[,"canopy"]
> elev <- site.cov[,"elev"]
> precip <- site.cov[,"precip"]

#Run a GLM for each species
> for (i in Nspecies){
```

```
 species.i <- glm(species.matrix[,i] ~ canopy + poly(elev,2) +
  precip, family = binomial)

 #coefficients from model
 pred.coef[[i]] <- coef(species.i)

 #predictions for mapping
 logit.pred <- predict(model = species.i, object = layers, fun = predict)
 prob.pred <- exp(logit.pred) / (1 + exp(logit.pred))
 pred.map[[i]] <- prob.pred
}

#convert list to a multi-layered raster stack
> prob.map.stack <- stack(pred.map)
> names(prob.map.stack) <- colnames(species.matrix)
```

Note that initially predictions are on the link (logit) scale, but we back-transform predictions to the probability scale for mapping. We can check out maps for any species and the estimated coefficients. Here, we check for the first species in the species matrix, the American robin (*Turdus migratorius;* map not shown):

```
# Plot prediction for species 1
> plot(prob.map.stack$AMRO, xlab = "Long", ylab =
  "Lat",
  main="AMRO - Predict first, assemble later")
> pred.coef[[1]]

##
 (Intercept) canopy poly(elev, 2)1 poly(elev, 2)2 precip
 −1.4362196 −0.1011538 1.3322123 6.8103904 0.3742213
```

Now, with these species-specific maps, we can assemble the predicted community in a variety of ways. We could convert the probability maps to binary maps of presence–absence based on some sort of threshold (see Chap. 7) (Algar et al. 2009; Dubuis et al. 2011). For example, Dubuis et al. (2011) created binary predictions by selecting a species-specific threshold that maximizes the sum of sensitivity and specificity (D'Amen et al. 2015). With models where predictions are probabilities of occurrence, a more natural way may be to create binary maps based on realizations from the binomial distribution because we assume in the logistic model that observations come from this distribution. Here, we illustrate the use of random deviates (realizations) from the binomial distribution and contrast this to using a simple threshold of species prevalence, which has been shown to be a useful threshold technique for single species models (Liu et al. 2005).

We first illustrate one realization from the binomial distribution. To do so, we create a function for generating binary maps from the predicted probabilities:

```
> binary.map <- function(map){
  values.i <- values(map)
```

```
  binom.i <- rbinom(length(values.i),prob = values.i, size = 1)
  map <- setValues(map, binom.i)
  return(map)
}
```

This function takes a single map, extracts the values on the map and uses `rbinom` to generate one realization (random deviate) based on the predicted probability. We can then implement this function on the raster stack, where the function will execute on each layer of the stack individually:

```
> binary.map.stack <- binary.map(prob.map.stack)
```

With these predicted, binary maps, we can assemble a variety of summaries at the community level. For instance, we calculate species richness as:

```
#Species richness from predictions
> spp.binomial.map <- sum(binary.map.stack)
```

This map illustrates how using a single random deviate from a binomial distribution based on the predicted probability of occurrence can lead to a great deal of noise in the predictions, where the map shows little spatial pattern. If we do this several times and then plot the mean or median predicted richness, we get a different perspective, where we observe spatial pattern in predicted richness across the region based on the covariates considered (Fig. 11.7a). Below we repeat the `binary.map` function, and with each realization from the binomial distribution we add it to our raster stack using the `addLayer` function (this loop is relatively slow to run):

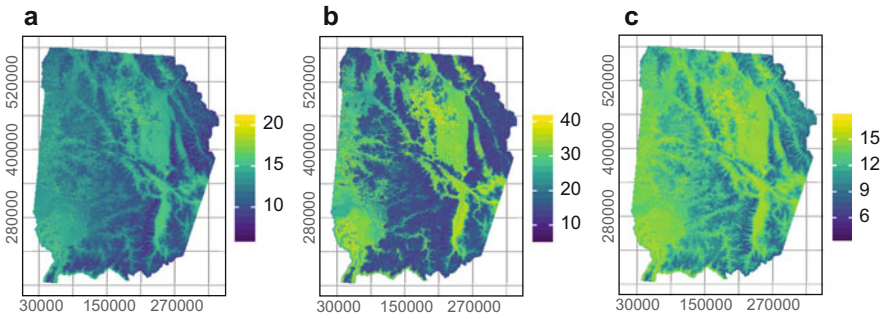


Fig. 11.7 Contrasting maps for species richness based on predict then assemble and assemble then predict approaches. For predict then assemble, maps differ when using (**a**) random deviates from the binomial distribution, or (**b**) species-specific thresholds based on prevalence. (**c**) Map based on assemble then predict, which tends to predict lower overall species richness

```
#richness from 19 total random deviates:
> for (i in 1:18){
  binary.map.i <- binary.map(prob.map.stack)
  richness.i <- sum(binary.map.i)
  spp.binomial.map <- addLayer(spp.binomial.map, richness.i)
  print(i)
}

#summarize distributions
> spp.mean.map <- mean(spp.binomial.map)
> plot(spp.mean.map)
```

We can contrast the above approach to the use of thresholding probabilities, which is more commonly applied. We illustrate the application of thresholds by using the prevalence of each species to truncate probabilities to 0, 1 (Fig. 11.7b).

```
> spp.t.map <- pred.map
> for (i in 1:Nspecies){
  thresh.i <- sum(species.matrix[,i]) / Nsites
  spp.t.map[[i]][which(spp.t.map[[i]][] > thresh.i)] =
  1
   spp.t.map[[i]][which(spp.t.map[[i]][] <= thresh.i)] =
  0
  }
```

Note that other approaches could be used to stack predictions from models. For example, probabilities from species-specific models could be summed across species (Distler et al. 2015). However, when data used in S-SDMs are presence-only data (e.g., herbarium collections), summing probabilities is problematic, as these probabilities are not probabilities of occurrence, but are instead relative probabilities assumed to be proportional to occurrence (Hastie and Fithian 2013; Yackulic et al. 2013). Pearson et al. (2004) argued that in presence-only situations, false negative error rates (i.e., omission errors) should be minimized, such that thresholds are selected that minimize predicting absences (or unsuitable habitat) where observed presence locations occurred. For instance, using presence-only data, Newbold et al. (2009) used a threshold that resulted in a sensitivity to 95% (see also Mateo et al. 2012). More recently, Liu et al. (2013) argued that maximizing the sum of sensitivity and specificity is most appropriate. Other approaches that have been used with presence-only data include selecting thresholds that maximize agreement with independent data on species richness or other variants of sensitivity thresholds (Pineda and Lobo 2009, 2012; Milanovich et al. 2012; Zhang et al. 2016).

11.3.3.2 Assemble First, Predict Later

In contrast to modeling each species separately and then compiling model predictions, we can assemble the community first and then model it. The simplest scenario is to compile the total number of species detected/site (or something similar,

such as the use of rarefaction) and then model species richness directly using a GLM like before, but in this case it would be a count-based GLM, such as Poisson regression. Other approaches include assembling measures of species composition and similarity.

Modeling Species Richness. Poisson regression is a natural GLM for count-based data. Poisson regression assumes that the data come from a Poisson distribution, or integer data ≥ 0. However, the Poisson distribution assumes that the mean = variance, which often is not the case. Instead, it is more common in ecological data to observe that the variance increases with the mean. When this pattern occurs, GLM models based on the Poisson distribution can be over-dispersed, which can lead to inferences being too liberal (i.e., more likely to commit a type I error; Zeileis et al. 2008). In such cases, quasi-Poisson or negative binomial regression are natural alternatives. Quasi-Poisson models tend to estimate the same coefficients as a Poisson, but will also estimate a scale parameter from the data, which is then used to adjust for over-dispersion in inferences. Quasi-Poisson models can be used with the `glm` function, but specifying "`family = quasipoisson`". More commonly, negative binomial regression is used when data are over-dispersed. Negative binomial regression does not make the assumption that the mean = variance, but rather estimates an additional scaling parameter, typically referred to as theta. Negative binomial models can be run with the `glm.nb` function in the `MASS` package (Venables and Ripley 2002). Quasi-Poisson and negative binomial models have the same number of parameters but make different assumptions regarding the relationship of the variance as a function of the mean: the quasi-Poisson assumes a linear relationship, while the negative binomial assumes a quadratic relationship (Hoef and Boveng 2007).

We can diagnose the potential for over-dispersion in several ways. A crude approach is to look at the ratio of the residual deviance to the degrees of freedom in the model. If this ratio, c, is $\gg 1$, that suggests over-dispersion. An alternative is to use the `dispersiontest` function in the `AER` package (Kleiber and Zeileis 2008). This test is based on asking whether the mean equals the variance in the Poisson model (Cameron and Trivedi 1990).

We fit the Poisson model and then determine if there is any evidence for over-dispersion. We first assemble richness data and then fit a Poisson GLM.

```
> richness <- rowSums(species.matrix])

> pois.rich <- glm(richness ~ canopy + poly(elev,2) + precip,
family = poisson)

> summary(pois.rich)

##
glm(formula = richness ~ canopy + poly(elev, 2) + precip, family =
poisson)

Deviance Residuals:
 Min 1Q Median 3Q Max
-3.6603 -0.6006 -0.0088 0.6149 2.9355
```

```
Coefficients:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) 2.72432 0.03549 76.770 < 2e-16 ***
canopy -0.19808 0.02889 -6.856 7.06e-12 ***
poly(elev, 2)1 -1.52544 0.29129 -5.237 1.63e-07 ***
poly(elev, 2)2 -1.14501 0.29472 -3.885 0.000102 ***
precip -0.18532 0.04015 -4.615 3.92e-06 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

 Null deviance: 775.20 on 781 degrees of freedom
Residual deviance: 658.86 on 777 degrees of freedom
AIC: 4077.3
```

For this model, the residual deviance is 658.86 on 777 df, or $c = 0.85$, which suggests over-dispersion is absent. We can test this more formally with the `AER` package:

```
> dispersiontest(pois.rich, trafo = 1)
```

This tests confirms that over-dispersion is absent ($p = 1.0$). If there was a signal for over-dispersion, we could contrast this model to the quasi-Poisson and negative binomial models, such as:

```
> qpois.rich <- glm(richness ~ canopy + poly(elev,2) + precip,
  family = quasipoisson)
> nb.rich <- glm.nb(richness ~ canopy + poly(elev,2) + precip)
```

We contrast predictive maps from these "assemble first" models to the prior map created based on "assemble later":

```
#map the Poisson model
> pois.raster <- predict(pois.rich, layers)
> spp.raster <- exp(pois.raster) #back-transform to count scale
```

To highlight spatial variability between models, we can map their differences:

```
> spp.diff <- spp.mean.map - spp.raster
```

There are several important differences between these approaches. In general, using S-SDMs tend to overpredict species richness (Dubuis et al. 2011; D'Amen et al. 2015), particularly when using species-specific thresholding (Fig. 11.7b). Yet these approaches do tend to be correlated (Newbold et al. 2009). We can check this as:

```
> richness.stack <- stack(spp.mean.map, spp.binomial.thres.map,
  spp.raster)
> names(richness.stack) <- c("binom-rich", "thres-rich", "pois-rich")
> richness.map.corr <- layerStats(richness.stack, 'pearson',
  na.rm = T)
> richness.map.corr

##
$'pearson correlation coefficient'
 binom.rich thres.rich pois.rich
binom.rich 1.0000000 0.4150215 0.8118022
thres.rich 0.4150215 1.0000000 0.5685640
pois.rich 0.8118022 0.5685640 1.0000000

$mean
binom.rich thres.rich pois.rich
 6.811837 24.379847 6.784896
```

In this case, there is a much stronger positive correlation when using binomial realizations from the logistic model than when thresholding predictions. More fundamentally, by modeling species individually, S-SDMs do not put explicit constraints on the number of species occurring at a site (e.g., S-SDMs assume biotic interactions and available energy are not limiting locally) and implicitly suggests a Gleasonian and species-sorting perspective to community assembly.

Modeling species richness could also be done with subsets of the total number of species, such as the number of endemics or the number of species in a functional group. Related metrics, such as calculating Simpson's or Shannon diversity (Magurran 2004), could also be modeled directly using a similar approach, although the type of GLM used might differ, depending on the distribution of the response variable being considered. Ordination metrics have also been modeled directly in this way (Faith et al. 2003; Chang et al. 2004). For example, we might extract ordination axes from our community data (see below) and then model those axes directly.

Dissimilarity Modeling. Another application of assemble first is focused on interpreting beta diversity by first assembling species into a species dissimilarity matrix. With that information, one can use generalized dissimilarity modeling (GDM) (Ferrier et al. 2007), Mantel tests (Legendre et al. 2005), or distance-based redundancy analysis (Legendre and Anderson 1999) to model beta diversity across space. Here, we focus on GDM; see below for some applications of Mantel tests and redundancy analysis.

The `gdm` package can fit generalized dissimilarity models (GDMs). There are several ways in which data can be formatted for the `gdm` package, which is typically accomplished with the `formatsitepairs` function prior to implementation of the GDM algorithm. We illustrate one format that most closely aligns with the data formats used above. Briefly, we take a site-by-species data matrix for the response variables and a site-by-covariates matrix for the explanatory variables. It requires a

site id column, then *x*–*y* coordinates can be specified for each site (coordinates can also be passed into the model function separately), and the remaining site-by-species data. Other formats gdm accepts include list formats and passing dissimilarity matrices that have been previously created rather than the raw input data.

```
> library(gdm)
> siteID <- 1:nrow(species.matrix)
> site.utm <- data.frame(x = species.site$x, y = species.site$y)

> gdm.species.matrix <- data.frame(cbind(siteID, site.utm,
  species.matrix))
> gdm.site.matrix <- data.frame(cbind(siteID, site.cov))
```

The gdm package uses a dissimilarity matrix of species composition between sites as the response variables. We use a Sørenson dissimilarity matrix. We format the data with the formatsitepair function:

```
#get gdm formatted object
> gdm.data <- formatsitepair(gdm.species.matrix, bioFormat = 1,
  dist = "bray", abundance = F, XColumn = "x", YColumn = "y",
  siteColumn = "siteID", predData = gdm.site.matrix)
```

Note that gdm passes the raw species data to the vegan package to calculate a dissimilarity matrix. In this case, we specify the Bray–Curtis dissimilarity metric and because abundance = F, such that this metric collapses to the Sørenson index. With this newly formatted object, we can run the GDM algorithm with the gdm function:

```
> gdm.dist <- gdm(gdm.data, geo = T)
> summary(gdm.dist)
```

The geo = T command tells the function that our geographic distance matrix should be used as a covariate in the model. The summary function provides several key results. First, it provides information on the deviance explained by the model. This can be thought of as a metric similar to metrics of the variation explained (R^2). Note that if we re-fit the above model without our geographic information (using geo = F), the proportion of deviance explained drops only from 13.2% to 12.7% such that geographic effects explain little variation in dissimilarity overall in this dataset. For each covariate, the summary function also provides information on the coefficients fit to the explanatory variables. Recall that one aspect of GDM is the use of I-splines: non-linear, monotonic (non-decreasing) splines for interpreting turnover across environmental or spatial gradients. The gdm function defaults to using three knots (see Chap. 6 for discussion of the use of knots in splines) to create I-splines and the summary output provides information on the spline fitting. The user can manually alter the number of knots and their location with splines and knots

commands. It is also straightforward to interpret the estimated environmental relationships from the model object that is created. For example, we can plot the partial response plots as:

```
> plot(gdm.dist , plot.layout = c(3,2))
```

These plots (Fig. 11.8) provide several insights. First, the maximum height of each spline describes the total magnitude of change along the gradient for the explanatory variable (Manion et al. 2018). In this case, elevation captures the largest change, while the geographic distance captures the least. The shape of the spline provided information on the rate of change (turnover) in the community and where the rate of change is greatest.

We can also make spatial predictions from a `gdm` model object in several ways, but care should be taken. First, we can use the `predict` function to assess model fit, that is, plot the predicted dissimilarity as a function of the observed dissimilarity (output not shown):

```
> gdm.fit <- predict(gdm.dist, gdm.data)
> plot(gdm.data$distance, gdm.fit, xlim = c(0,1), ylim = c(0,1),
  lines(c(0,1),c(0,1)))
```

In this case, the model is a relatively poor fit to the data. This is not too surprising, given the low amount of deviance explained by the model. We can also predict over space. There are several steps to do so. First, we need to transform the raster layers based on the GDM model:

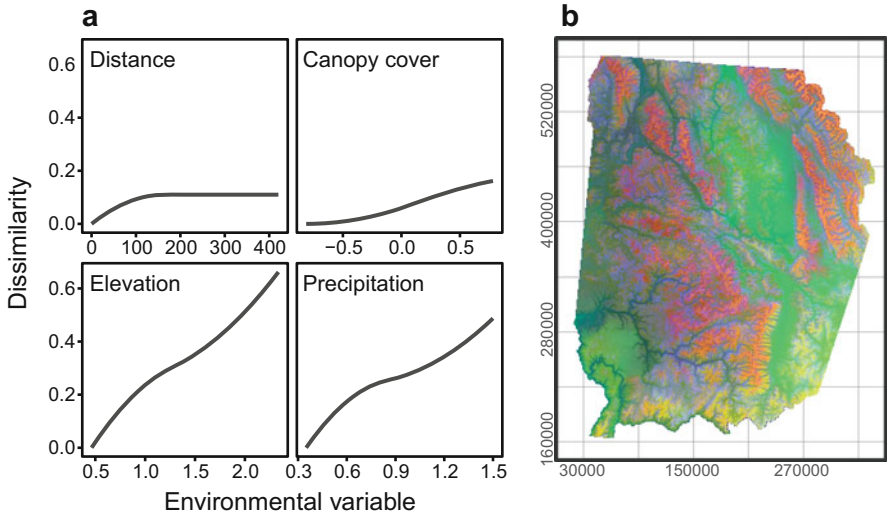


Fig. 11.8 (**a**) GDM partial plots and predictions of dissimilarity. (**b**) Mapping dissimilarity, where similar colors represent similar communities

```
> gdm.trans.data <- gdm.transform(gdm.dist, layers)
> plot(gdm.trans.data)
```

The `gdm.transform` function takes a `gdm` object and transforms the raster layers for further analysis into "biological space" (Manion et al. 2018), such that the values represent the dissimilarity along each gradient where the minimum value for the covariate is zero and the maximum reflects the maximum dissimilarity across the gradient. Consequently, the result of this function is a prediction of dissimilarity for each environmental gradient. Note that the order of the raster layers must be in the same order as specified in the GDM model. To make an overall prediction of dissimilarity based on all variables, we need to combine predictions for each covariate (raster layer). We can either: (1) scale and sum each individual variable when the number of variables are small; or (2) use principal components analysis (PCA) when the number of variables is large (see vignette in Manion et al. 2018). Here we show an example using PCA. PCA is an ordination technique for reducing the dimensionality of multivariate data, wherein PCA produces new variables (i.e., principal components) from multivariate data that are linear combinations of the original multivariate data. We will not focus on the details of how PCA works here; interested readers should see Legendre and Legendre (1998). For relatively large rasters, we can first sample the raster values and then use these values in a PCA.

```
> sample.trans <- sampleRandom(gdm.trans.data, 10000)
> sample.pca <- prcomp(sample.trans)

#inspect
> summary(sample.pca)

##
Importance of components:
 PC1 PC2 PC3 PC4 PC5
Standard deviation 0.1556 0.08892 0.04303 0.02779 0.01519
Proportion of Variance 0.6924 0.22600 0.05292 0.02208 0.00659
Cumulative Proportion 0.6924 0.91841 0.97133 0.99341 1.00000

> round(sample.pca$rotation,2) #eigenvectors

##
 PC1 PC2 PC3 PC4 PC5
xCoord 0.04 −0.13 0.07 0.26 −0.95
yCoord −0.02 0.10 0.03 0.96 0.25
canopy 0.04 0.01 1.00 −0.04 0.06
elev 0.77 −0.63 −0.03 0.04 0.13
precip 0.64 0.76 −0.03 −0.04 −0.09
```

Here, the first two principal components explain 92% of the variation (taken from the "Importance of components" table). The eigenvectors provide information on the linear combinations of the original data that make up the new PC variables. In this case, the first two components are elevation and precipitation effects, while the third focuses on canopy cover. This result makes sense, given the partial predictions (Fig. 11.8a) and correlations among variables. With this PCA, we can then predict the pc scores onto the data. Here, we just focus on the first three principal components, using the index=1:3 command.

```
> gdm.pca <- predict(gdm.trans.data, sample.pca, index = 1:3)
```

This will provide predictions for each PC. One way to create an integrated prediction is via rescaling the PCs to a 0–1 scale and summing them by using the plotRGB function from the raster package (see vignette in Manion et al. 2018).

```
#scale to 0-1 range
> gdm.pca <- (gdm.pca - minValue(gdm.pca)) /
 (maxValue(gdm.pca) - minValue(gdm.pca))

> plotRGB(gdm.pca, r = 1, g = 2, b = 3, scale = 1)
```

In the above plot, we specify that the first PC reflects the red channel, the second the green channel, etc.; scale = 1 notifies the plotRGB function that the maximum values are 1 for the PC data. Note that by scaling the PCs in a standardized way and then using the plotRGB command, we are implicitly weighting each PC similarly. Taken together, this predictive map (Fig. 11.8b) reflects differences in community composition across space, with similar values reflecting similar communities.

11.3.3.3 Assemble and Predict Together

Increasingly, community models assemble and predict together. There are several advantages of doing so. Some frequently used techniques include some types of constrained ordination (Guisan et al. 1999) and multivariate GLM-like models, sometimes referred to as "joint species distribution models" (Dorazio et al. 2006; Ovaskainen et al. 2010; Wang et al. 2012).

Direct Gradient Analysis. Direct gradient analysis, such as RDA and CCA, is frequently used in community modeling. The vegan package can implement RDA and CCA; here, we focus on RDA. These models use a site-by-species matrix as the response variables. Blanchet et al. (2014) noted that the case of using binary data in an RDA is equivalent to the use of a distance-based RDA (see below; Legendre and Anderson 1999) when dissimilarity is quantified based on the simple matching coefficient:

$$\left(1 - \frac{a+d}{a+b+c+d}\right)^{0.5}, \tag{11.12}$$

where a, b, and c are defined as in Eq. (11.4), and d is the number of sites where both species are absent. In this way, a simple RDA can be fit as:

```
> rda.bird <- rda(species.matrix ~ canopy + poly(elev,2) + precip)
> rda.bird

##
Call: rda(formula = species.matrix ~ canopy + poly(elev, 2) + precip)

 Inertia Proportion Rank
Total 7.1866 1.0000
Constrained 0.8163 0.1136 4
Unconstrained 6.3703 0.8864 53
Inertia is variance

Eigenvalues for constrained axes:
 RDA1 RDA2 RDA3 RDA4
0.4356 0.2517 0.0769 0.0522

Eigenvalues for unconstrained axes:
 PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8
0.3771 0.3384 0.3102 0.2809 0.2553 0.2274 0.2111 0.2033
(Showed only 8 of all 53 unconstrained eigenvalues)
```

The output provided by typing `rda.bird` shows the proportion of the total inertia (~variance) explained by the constraining axes. It also provides the eigenvalues for the constraining axes along with eigenvalues for the top unconstrained axes. What does this really mean? Recall that RDA can be thought of as a multivariate regression where the fitted values from separate regressions for each species are then subjected to PCA (Borcard et al. 2011). RDA computes new axes that are linear combinations of all explanatory variables. There will be as many constrained axes as there are explanatory variables, though the first axis will explain more variation than the second, the second more variation than the third, etc. Eigenvalues for each axis are proportional to the variation explained based on the new axes, while eigenvectors reflect the weight of each explanatory variable for explaining that axis. The eigenvalues of the unconstrained (PCA) axes represent the amount of residual variation that is not captured by the explanatory covariates. The `summary` function provides site and species scores, along with the summary of inertia described above. Specifically, this shows the `'species scores'`, and the `'site constraints'`, which reflect where the species and sites (point count locations) fall in this constrained multivariate ordination space. We can extract the site and species scores for plotting and further interpretation with the `scores` function (output not shown):

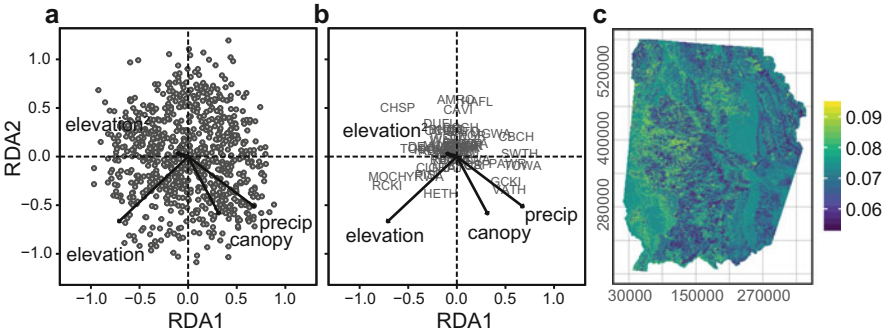


Fig. 11.9 Redundancy analysis on the bird community. (**a**) biplot based on site scores, (**b**) biplot with species scores (shown are four-letter species codes based on the American Ornithological Society's standardized codes), and (**c**) mapping predictions of species scores for the varied thrush

```
> scores(rda.bird, choices = 1:2, display = "sites")
> scores(rda.bird, choices = 1:2, display = "species")
```

These scores can be plotted using "biplots" with the `plot` function, which is a common way to visualize ordination data (Fig. 11.9a, b).

Permutation tests, where rows of the community matrix are shuffled, to assess the significance of the RDA axes can be accomplished with the `anova` function:

```
> anova(rda.bird)

##
Permutation test for rda under reduced model
Permutation: free
Number of permutations: 999

Model: rda(formula = species.matrix ~ canopy + poly(elev, 2) + precip)
 Df Variance F Pr(>F)
Model 4 0.8163 24.893 0.001 ***
Residual 777 6.3703
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The term "anova" is a bit of a misnomer here—it is not an analysis of variance test. However, the output of the permutation test can be summarized in a similar way and provided like a standard ANOVA table. In this case, despite the fact that the RDA axes explain little variation in variance (inertia), these axes are considered significant based on the permutation test. Each covariate can also be tested sequentially (`by = 'term'`) or at the same time:

```
> anova(rda.bird, by = 'mar')#marginal tests
```

```
##
Permutation test for rda under reduced model
Marginal effects of terms
Permutation: free
Number of permutations: 999

Model: rda(formula = species.matrix ~ canopy + poly(elev, 2) + precip)
 Df Variance F Pr(>F)
canopy 1 0.1291 15.752 0.001 ***
poly(elev, 2) 2 0.4043 24.658 0.001 ***
precip 1 0.2427 29.607 0.001 ***
Residual 777 6.3703
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results from this permutation show that each explanatory variable explains significant variation in the species data. There are several other ways to summarize RDA and we will not cover all of them here. See Borcard et al. (2011) for more details.

Finally, we can map elements of our RDA model across the region of interest (Fig. 11.9). In this case, we can make predictions based on site scores (the linear combinations of environmental covariates) or species scores as:

```
#first convert raster to data frame
> layers.df <- as.data.frame(layers, xy = T, na.rm = T)

#predict site scores onto new data frame
> rda.site.pred <- predict(rda.bird, layers.df, type = "lc")

#predict species scores onto new data frame
> rda.species.pred <- predict(rda.bird, layers.df, type = "response")
```

Predictions can be mapped in space. In Fig. 11.9, we illustrate mapping the species scores for the varied thrush (to compare with mapping in Chap. 7). Species scores could be truncated to 0, 1 data based on optimal thresholds that split occurrence observations (Liu et al. 2005), and then summed for predictions of species richness, similar to the S-SDM approach explained above.

In contrast to using a species occurrence matrix, we could instead use a dissimilarity matrix in RDA, termed distance-based redundancy analysis, or dbRDA (Legendre and Anderson 1999). In that case, we would be assembling first (community dissimilarity) and then interpreting dissimilarity between locations, analogous to generalized dissimilarity modeling. This form of RDA can be accomplished with the `capscale` function in the `vegan` package. See Blanchet et al. (2014) for discussion on the relationship of dbRDA to RDA.

Multivariate Regression. Several univariate modeling approaches have been extended to model multiple species simultaneously, sometimes referred to as joint species distribution models (jSDM) (Clark et al. 2014; Pollock et al. 2014;

Ovaskainen et al. 2016b). These include multivariate logistic regression models, multi-species occupancy models, multivariate machine learning methods (e.g., neural networks), and multivariate adaptive regression splines. These methods can be very complex; here, we illustrate some simpler implementations that illustrate these types of models.

We start with a multivariate logistic regression that is formulated through the use of species-level random effects. In Chap. 6, we used random effects as a potential means of accounting for spatial dependence. Specifically, we fit multilevel models by adding random effects (e.g., transects, grids, or watersheds) that capture spatial hierarchies in the data. In doing so, we were adding "random-intercepts" to the modeling framework. Another type of random effect is a "random coefficient" (aka a "random slope"). This type of random effect allows for an effect of a predictor variable to change with the random effect (Fig. 11.5). If we add species as a random coefficient, then this would allow us to consider that species may respond differently to environmental covariates.

To implement a multivariate logistic model, we need to use community data in a long, rather than wide, format. To do so, we will merge the `species.matrix` and `site.cov` objects and then use the `melt` function with the `reshape2` package.

```
> species.matrix.df <- data.frame(cbind(site.cov,
  species.matrix))
> sp.multi <- melt(species.matrix.df, id.vars = c("elev",
  "canopy", "precip"), variable.name = "SPECIES", value.name =
  "pres")
```

With this format, we can start with a random intercept model with the `glmer` function in the `lme4` package:

```
> library(lme4)
> multi.int <- glmer(pres ~ canopy + elev + I(elev^2) + precip
  +(1|SPECIES), family = "binomial", data = sp.multi,
  glmerControl(optimizer = "bobyqa"))
> summary(multi.int)

##
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [
glmerMod]
 Family: binomial ( logit )
Formula: pres ~ canopy + elev + I(elev^2) + precip + (1 | SPECIES)
 Data: sp.multi
Control: glmerControl(optimizer = "bobyqa")

 AIC BIC logLik deviance df.resid
 35404.4 35456.1 −17696.2 35392.4 41440
```

```
Scaled residuals:
 Min 1Q Median 3Q Max
−2.4474 −0.4686 −0.2841 −0.1620 6.7555

Random effects:
 Groups Name Variance Std.Dev.
 SPECIES (Intercept) 1.751 1.323
Number of obs: 41446, groups: SPECIES, 53

Fixed effects:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) −1.59635 0.24292 −6.572 4.98e−11 ***
canopy −0.35606 0.03881 −9.175 < 2e−16 ***
elev 0.89779 0.23139 3.880 0.000104 ***
I(elev^2) −0.42630 0.08412 −5.068 4.02e−07 ***
precip −0.32776 0.05357 −6.118 9.47e−10 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
 (Intr) canopy elev I(l^2)
canopy 0.041
elev −0.623 −0.048
I(elev^2) 0.595 0.045 −0.986
precip −0.200 −0.121 0.014 −0.006
```

In this case, we are modeling the presence–absence of each species and including canopy, elevation and precipitation as predictors for each species. We include species as a random intercept, but no other effect of species. This model formulation would then account for variation in species occurrence (i.e., prevalence), but it assumes that all species respond to environmental variables similarly. Consequently, the above model is not very helpful in most situations. However, we can extend this model by adding random coefficients of species and their relationships with canopy and elevation:

```
#random coefficient model by species
> multi.coef <- glmer(pres ~ elev +I(elev^2) + canopy + precip +
  (1|SPECIES)+(0 + elev|SPECIES) +(0 + I(elev^2)|SPECIES) + (0 +
  canopy|SPECIES) + (0 + precip|SPECIES), family = "binomial",
  data = sp.multi)
```

In this case, we also include random coefficients for the species effect with environmental variables (e.g., `(0 + elev|SPECIES))`. The syntax for random effects structure in `lme4` can be a bit daunting (Bates et al. 2015). Above, the 0 tells `lme4` that there is no random intercept of species (because we have already specified it separately) but `elev|SPECIES` states that there is a random coefficient of species with elevation. This allows species to respond differently to these covariates.

However, it does assume that these random coefficients come from the same normal distribution. In doing so, it implicitly assumes species respond somewhat similarly to these covariates and for species with very little information, it will pull those species' coefficients toward the mean across species (Dorazio et al. 2010). That assumption can be relaxed to some degree using Bayesian methods, which can account for phylogenetic dependence among species and the potential for species interactions altering outcomes (Ovaskainen and Soininen 2011), but we do not cover those methods here. Note that we may want to scale variables (using the `scale` function) prior to fitting this model to help improve model convergence. We can view output from this model as:

```
> summary(multi.coef)

##
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [
glmerMod]
 Family: binomial ( logit )
Formula: pres ~ elev + I(elev^2) + canopy + precip + (1 | SPECIES) + (0 +
 elev | SPECIES) + (0 + I(elev^2) | SPECIES) + (0 + canopy |
 SPECIES) + (0 + precip | SPECIES)
 Data: sp.multi
Control: glmerControl(optimizer = "bobyqa")

 AIC BIC logLik deviance df.resid
 32514.0 32600.3 -16247.0 32494.0 41436

Scaled residuals:
 Min 1Q Median 3Q Max
-8.1046 -0.4232 -0.2506 -0.0462 25.7761

Random effects:
 Groups Name Variance Std.Dev.
 SPECIES (Intercept) 5.0330 2.2434
 SPECIES.1 elev 2.7689 1.6640
 SPECIES.2 I(elev^2) 0.5695 0.7547
 SPECIES.3 canopy 0.9091 0.9534
 SPECIES.4 precip 3.3542 1.8314
Number of obs: 41446, groups: SPECIES, 53

Fixed effects:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.2549 0.3623 -6.224 4.84e-10 ***
elev 2.2226 0.3625 6.131 8.73e-10 ***
I(elev^2) -1.0261 0.1493 -6.871 6.37e-12 ***
canopy -0.5030 0.1398 -3.598 0.000321 ***
precip -0.5416 0.2609 -2.076 0.037882 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Correlation of Fixed Effects:
 Intr) elev I(l^2) canopy
elev −0.381
I(elev^2) 0.331 −0.547
canopy 0.010 −0.008 0.005
precip −0.041 −0.003 0.002 −0.008
```

Parameters estimated from the model can be viewed as (output not shown):

```
> fixef(multi.coef)
> ranef(multi.coef)
> coef(multi.coef) #combines fixed and random effects
> coef(multi.coef)$SPECIES[,"CC2"]
```

Interestingly, in some cases when random coefficients are added, the effect may no longer be significant (when without the random coefficient they are). Why would that be? In general, if there is a great deal of variation in effects of an explanatory variable across species, then, once captured, the marginal effect may vanish. Predictions from the model can be compiled similar to above as:

```
> glmm.map <- list() #stores map predictions

# extract maps for all species
> for (i in 1:Nspecies){
 logit.raster <- coef(multi.coef)$SPECIES[i,1]+
 coef(multi.coef)$SPECIES[i,"elev"] * Elev +
 coef(multi.coef)$SPECIES[i,"canopy"] * Canopy
 prob.raster <- exp(logit.raster) / (1 + exp(logit.raster))
 glmm.map[[i]] <- prob.raster
}
```

Note that here we are grabbing the species-specific random coefficients to make predictions. We can also make partial predictions of environmental relationships for each species, similar to what we did in Chap. 6. Partial predictions (or response curves) provide a means to visualize how the response variable changes with a covariate, while holding all other covariates constant (typically at their mean value). To do so, we look more closely at the output from this model and contrast it to the individual models we created using the S-SDM approach. Below, we contrast outputs for the black-capped chickadee (*Poecile atricapillus*):

```
> coef(multi.coef)$SPECIES[2,] #GLMM coefficients

##
 (Intercept) elev I(elev^2) canopy precip
BCCH −0.2874097 2.089603 −1.322244 −0.2010582 −1.359818
```

```
> pred.coef[[2]] #GLM coefficients from S-SDM
 (Intercept) canopy poly(elev, 2)1 poly(elev, 2)2 precip
 0.09001473 −0.17531427 −15.42057383 −5.72431042 −1.43894282
```

We find that the multivariate model generally estimates similar coefficients for the variables, except for elevation. The different coefficients can be partly explained by the way in which elevation was modeled using the `poly` function for the S-SDM (which re-scales the variable to ensure each polynomial term is orthogonal) but with the `I` function for the multivariate regression. We can generate partial plots by first creating a new data set for predictions. Below, we focus on elevation effects:

```
> site.cov.df <- data.frame(site.cov)
> elev.range <- seq(min(site.cov.df$elev),
  max(site.cov.df$elev), length = 20)
> precip.mean <- mean(site.cov.df$precip)
> canopy.mean <- mean(site.cov.df$canopy)

> newdata.glmm <- data.frame(expand.grid(SPECIES = species.20,
  precip = precip.mean, elev = elev.range, canopy =
  canopy.mean))
```

Note that we set the other variables to their mean values and then use the `expand.grid` function to create a new data frame for predictions. With these data we can use the `predict` function to generate partial plots for each species:

```
> pred <- predict(multi.coef, newdata.glmm, type = "response")
> glmm.pred <- cbind(newdata.glmm, pred)
```

These predictions show that most species are relatively rare and do not respond strongly to elevation and precipitation in the landscape, whereas some of the more common species respond both positively and negatively to these variables (Fig. 11.10). We can use this model in a similar way to our S-SDM to predict

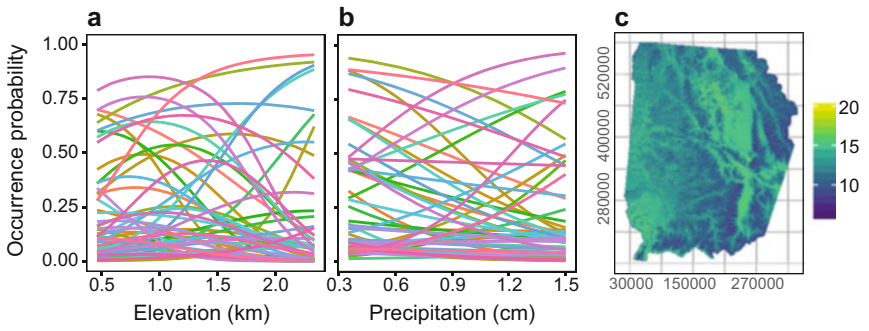


Fig. 11.10 Partial predictions for elevation effects for each of the 53 species from the random coefficient model for (**a**) elevation and (**b**) precipitation, and the (**c**) resulting prediction for species richness across the region

species richness across the region. We could use a simple thresholding technique or obtain binomial realizations from the model using the `binary.map` function to derive species richness from the `glmm.map` output in an identical way as shown in Sect. 11.3.3.1. This model leads to similar predictions for species richness than the approach taken in the S-SDM code above (compare Fig.11.10c to Fig. 11.7a).

11.3.4 Spatial Dependence in Communities

While the above modeling frameworks provide a means for predicting and mapping communities across space, most of what we have illustrated does not account directly for spatial dependence (Dray et al. 2012).

We consider several approaches for addressing the problem of space. First, we consider the spatial dependence of the community data via multivariate correlograms and variograms. Multivariate correlograms can be fit with several packages. We start with a Mantel test on geographic distance and follow it with a Mantel correlogram, which uses a distance matrix regarding community composition as the response variable:

```
#calculate distance matrix
> dist.matrix <- as.matrix(dist(site.utm))

#Mantel test
> mantel(sorenson, dist.matrix, method = "pearson", permutations =
  999)

##
Mantel statistic based on Pearson's product-moment correlation

Call:
mantel(xdis = sorenson, ydis = dist.matrix, method = "pearson",
permutations = 999)

Mantel statistic r: 0.1141
 Significance: 0.001

Upper quantiles of permutations (null model):
 90% 95% 97.5% 99%
0.0147 0.0191 0.0221 0.0265
Permutation: free
Number of permutations: 999
```

The Mantel test finds significant ($p < 0.001$), but weak ($r = 0.11$), correlation of community dissimilarity with geographic distance, consistent with the geographic effects observed in the GDM (Fig. 11.8). More broadly, we can calculate a Mantel

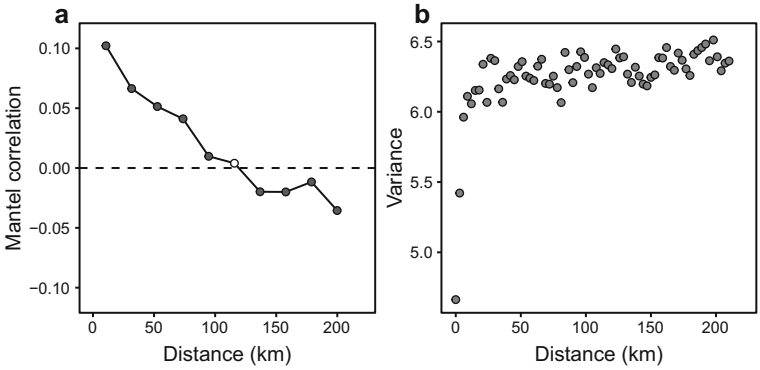


Fig. 11.11 Interpreting community-level spatial dependence. (**a**) The Mantel correlogram based on species dissimilarity, and (**b**) the multivariate variogram from residuals in the RDA

correlogram with the `vegan` package to better understand spatial dependence in species dissimilarity:

```
#correlogram
> mantel.corr <- mantel.correlog(sorenson, XY = site.utm, cutoff
  = T, r.type = "pearson", nperm = 99)
```

Here, we find modest evidence for spatial dependence in the community data (Fig. 11.11a). Consequently, we may wish to revisit some of the above modeling approaches to explicitly account for spatial dependence.

11.3.5 Community Models with Explicit Accounting for Space

We can extend direct gradient ordination methods to account for space. In this case, we are doing a partial RDA, in which geographic distance is conditioned or "partialled out" before considering the environmental covariates. Note that it is not possible to pass the entire geographic matrix into a partial RDA; instead, we need to summarize the spatial structure in some way. Consequently, we create a spatial weights matrix from a Euclidean distance matrix. With this distance matrix, we can use a principal coordinates analysis on a truncated distance matrix to capture spatial structure, as we did in Chap. 6. In this case, we can use the `pcnm` function in the `vegan` package:

```
#Principal Components on Neighborhood Matrices
> pcnm.dist <- pcnm(dist.matrix)
```

The pcnm function defaults to setting the truncation to the minimum distance that provides a connected network using the minimum spanning tree of the distance matrix. We then fit a partial RDA, controlling for space.

```
#partial RDA
> rda.partial <- rda(species.matrix ~ canopy + elev + precip +
  Condition(scores(pcnm.dist, choices = 1:10)))
```

In the above model, we arbitrarily use the first ten PCNM axes from the PCNM analysis. We could more formally screen pcnm variables for their effects using the permutation test described above and then incorporate variables that explain variation in species occurrence:

```
> rda.distance <- rda(species.matrix ~ (scores(pcnm.dist)))
> rda.distance
> anova(rda.distance, by = 'axis', permutations = 200)
```

We find that the first two axes explain the majority of the spatial variation. We then re-fit the partial RDA with only those pcnm variables of importance. In this case, once we partial out spatial effects, the environmental covariates are still considered important based on permutation tests. This is consistent with results from the GDM showing that turnover could largely be explained by environmental, rather than geographic, variation.

To understand the spatial dependence in the residuals of community ordination models, we can use "multiscale ordination," or a multivariate variogram on the residuals of a RDA (or CCA) model (Wagner 2004).

```
> mso(rda.bird, site.utm, grain = 1000, perm = 200)

#variogram based on residuals
> plot(mso.rda$vario$H, mso.rda$vario$CA)
```

In this function, grain refers to the bin size for distance classes, so here we choose a 1-km grain (Fig. 11.11b). This analysis finds evidence for spatial dependence at distances <3 km, consistent with what we found in Chap. 6, but it identifies a smaller range in spatial dependence than the correlogram based on raw dissimilarity.

We have already seen how space can be accounted for in GDMs. In addition, Chap. 6 shows how space could be accounted for in S-SDMs through the use of spatial regression models fit to each species. To date, spatial dependence for multivariate regression techniques has been less explored than for univariate spatial regression approaches. In principle, adding spatial weighting functions, such as eigenvector mapping, could be straightforward to implement. Other approaches covered in Chap. 6 may be more difficult to implement.

11.4 Next Steps and Advanced Issues

11.4.1 Decomposition of Space–Environment Effects

Isolating the role of space relative to environmental effects is important for understanding the mechanisms of community assembly and factors that may limit community structure. There has been a long tradition in some areas of community ecology to decompose the spatial effect from environmental effects using variance partitioning methods (Borcard et al. 1992; Peres-Neto et al. 2006). In this approach, several models are fit with and without key covariates (e.g., geographic distance in and out of the model) (Cushman and McGarigal 2002). By quantifying changes in the variance explained (or inertia explained in some modeling approaches, such as RDA), isolating the role of space and environmental effects can be quantified. While such approaches can be helpful, caution must be used because variance partitioning relies on some implicit assumptions, such as the assumption that interactions do not occur between covariates and that biologically these variables are conceptually and empirically independent. Variance partitioning can be implemented several ways in R; see the `varpart` function in the `vegan` package for one approach.

11.4.2 Accounting for Dependence Among Species

There is increasing interest in explicitly modeling the dependence among species in spatial modeling of communities, what is frequently referred to as joint species distribution models (Clark et al. 2014; Pollock et al. 2014; Warton et al. 2015a). Such dependence can arise for a variety of reasons, including effects of species interactions, phylogenetic dependence, and the fact that species may use similar environmental gradients, such that the distribution of one species may help predict the distribution of another. The generalized linear mixed model approach and its extensions are common ways to address these issues. In the above code, we provided the simplest implementation where we assumed that species came from a common distribution. However, more complex dependences can be addressed. In most of these situations, Bayesian hierarchical models are used to capture potential latent variables that may capture the species dependence in a formal way. See the `boral` and `HMSC` packages that can fit such models (Hui 2016; Ovaskainen et al. 2017).

11.4.3 Spatial Networks

Communities are often described as networks of interacting species (Bascompte 2007; Ings et al. 2009). In Chap. 9, we considered spatial networks, where nodes were patches and links were reflected movement or flow. In a community context,

nodes typically represent species and links represent interspecific interactions. The network approach to communities has revealed several insights into community structure and stability. In particular, a network approach can potentially capture indirect effects in communities, such as diffuse species interactions and indirect interactions, as well as providing a means to quantify and interpret emergent structure in communities. Increasingly, there has been interest in applying this general approach in space as a means to interpret metacommunity dynamics and related spatial issues (Araújo et al. 2011; Gonzalez et al. 2011). For example, spatial networks of species interactions have been used to determine spatial beta diversity (Poisot et al. 2012, 2017). Community networks are starting to be used to tackle conservation problems as well (Kaiser-Bunbury and Bluthgen 2015).

11.5 Conclusions

Spatial modeling of communities is an important and rapidly advancing topic in spatial ecology and conservation (Dray et al. 2012; Warton et al. 2015b; D'Amen et al. 2017). Much of this work has focused on predicting and mapping community structure, including species richness and beta diversity across space. This information is often used in conservation planning both locally and globally (Myers et al. 2000; Brooks et al. 2002; Wilson et al. 2006; Gray et al. 2016; Cardinale et al. 2018).

Despite these rapid advances, there are many challenges to understanding and modeling spatially structured communities. While there has been rapid growth in the theory of spatially structured communities (Gravel et al. 2011; Leibold and Chase 2017), often data are limited for interpreting how different factors govern community structure across space. The use of species co-occurrence data for these questions has a long history (e.g., Diamond 1975), but it can be unclear the extent to which co-occurrence can provide information on limiting factors, such as dispersal limitation and species interactions (Borthagaray et al. 2014; Freilich et al. 2018). Current challenges for spatial modeling of communities include better capturing dependencies among species, spatial dependence, and the potential for dispersal limitation to impact observed outcomes (Cumming et al. 2010; Rota et al. 2016; Clark et al. 2017; Tikhonov et al. 2017). Furthermore, better integration of community theory with empirical modeling of communities (e.g., Dorazio et al. 2010) is needed to interpret why communities assemble and better predict how they may change over time.

References

Algar AC, Kharouba HM, Young ER, Kerr JT (2009) Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. Ecography 32(1):22–33. https://doi.org/10.1111/j.1600-0587.2009.05832.x

Anderson M, Gorley RN, Clarke RK (2008) Permanova+ for primer: Guide to software and statisticl methods. Primer-E Limited

Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, Sanders NJ, Cornell HV, Comita LS, Davies KF, Harrison SP, Kraft NJB, Stegen JC, Swenson NG (2011) Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. Ecol Lett 14 (1):19–28. https://doi.org/10.1111/j.1461-0248.2010.01552.x

Araújo MB, Rozenfeld A, Rahbek C, Marquet PA (2011) Using species co-occurrence networks to assess the impacts of climate change. Ecography 34(6):897–908. https://doi.org/10.1111/j.1600-0587.2011.06919.x

Arrhenius O (1921) Species and area. J Ecol 9:95–99

Bacaro G, Altobelli A, Cameletti M, Ciccarelli D, Martellos S, Palmer MW, Ricotta C, Rocchini D, Scheiner SM, Tordoni E, Chiarucci A (2016) Incorporating spatial autocorrelation in rarefaction methods: implications for ecologists and conservation biologists. Ecol Indic 69:233–238. https://doi.org/10.1016/j.ecolind.2016.04.026

Barwell LJ, Isaac NJB, Kunin WE (2015) Measuring beta-diversity with species abundance data. J Anim Ecol 84(4):1112–1122. https://doi.org/10.1111/1365-2656.12362

Bascompte J (2007) Networks in ecology. Basic Appl Ecol 8(6):485–490. https://doi.org/10.1016/j.baae.2007.06.003

Baselga A (2010) Partitioning the turnover and nestedness components of beta diversity. Glob Ecol Biogeogr 19(1):134–143. https://doi.org/10.1111/j.1466-8238.2009.00490.x

Baselga A, Orme CDL (2012) betapart: an R package for the study of beta diversity. Methods Ecol Evol 3(5):808–812. https://doi.org/10.1111/j.2041-210X.2012.00224.x

Baskett ML, Micheli F, Levin SA (2007) Designing marine reserves for interacting species: insights from theory. Biol Conserv 137(2):163–179. https://doi.org/10.1016/j.biocon.2007.02.013

Bates D, Machler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67(1):1–48

Bjørnstad ON, Falck W (2001) Nonparametric spatial covariance functions: estimation and testing. Environ Ecol Stat 8(1):53–70. https://doi.org/10.1023/a:1009601932481

Blanchet FG, Legendre P, Borcard D (2008) Modelling directional spatial processes in ecological data. Ecol Model 215(4):325–336. https://doi.org/10.1016/j.ecolmodel.2008.04.001

Blanchet FG, Legendre P, Bergeron JAC, He FL (2014) Consensus RDA across dissimilarity coefficients for canonical ordination of community composition data. Ecol Monogr 84 (3):491–511. https://doi.org/10.1890/13-0648.1

Borcard D, Legendre P (2012) Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study. Ecology 93(6):1473–1481

Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. Ecology 73(3):1045–1055

Borcard D, Gillet F, Legendre P (2011) Numerical ecology with R. Springer, New York

Borthagaray AI, Arim M, Marquet PA (2014) Inferring species roles in metacommunity structure from species co-occurrence networks. Proc R Soc B 281(1792). https://doi.org/10.1098/rspb.2014.1425

Bowman J, Cappuccino N, Fahrig L (2002) Patch size and population density: the effect of immigration behavior. Conserv Ecol 6(1):9

Brooks TM, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Rylands AB, Konstant WR, Flick P, Pilgrim J, Oldfield S, Magin G, Hilton-Taylor C (2002) Habitat loss and extinction in the hotspots of biodiversity. Conserv Biol 16(4):909–923. https://doi.org/10.1046/j.1523-1739.2002.00530.x

Brosi BJ, Armsworth PR, Daily GC (2008) Optimal design of agricultural landscapes for pollination services. Conserv Lett 1(1):27–36. https://doi.org/10.1111/j.1755-263X.2008.00004.x

Brown JH, Kodric-Brown A (1977) Turnover rates in insular biogeography: effect of immigration on extinction. Ecology 58(2):445–449

Cadotte MW, Tucker CM (2017) Should environmental filtering be abandoned? Trends Ecol Evol 32(6):429–437. https://doi.org/10.1016/j.tree.2017.03.004

Cadotte MW, Cavender-Bares J, Tilman D, Oakley TH (2009) Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. PLoS One 4(5). https://doi.org/10.1371/journal.pone.0005695

Caley MJ, Schluter D (1997) The relationship between local and regional diversity. Ecology 78 (1):70–80

Cameron AC, Trivedi PK (1990) Regression-based tests for overdispersion in the Poisson model. J Econ 46(3):347–364. https://doi.org/10.1016/0304-4076(90)90014-k

Cardinale BJ, Gonzalez A, Allington GRH, Loreau M (2018) Is local biodiversity declining or not? A summary of the debate over analysis of species richness time trends. Biol Conserv 219:175–183. https://doi.org/10.1016/j.biocon.2017.12.021

Caswell H (1976) Community structure: neutral model analysis. Ecol Monogr 46(3):327–354. https://doi.org/10.2307/1942257

Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW (2009) The merging of community ecology and phylogenetic biology. Ecol Lett 12(7):693–715. https://doi.org/10.1111/j.1461-0248.2009.01314.x

Chang CR, Lee PF, Bai ML, Lin TT (2004) Predicting the geographical distribution of plant communities in complex terrain - a case study in Fushian Experimental Forest, northeastern Taiwan. Ecography 27(5):577–588. https://doi.org/10.1111/j.0906-7590.2004.03852.x

Clark JS, Gelfand AE, Woodall CW, Zhu K (2014) More than the sum of the parts: forest climate response from joint species distribution models. Ecol Appl 24(5):990–999. https://doi.org/10.1890/13-1015.1

Clark JS, Nemergut D, Seyednasrollah B, Turner PJ, Zhang S (2017) Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. Ecol Monogr 87(1):34–56. https://doi.org/10.1002/ecm.1241

Coleman BD, Mares MA, Willig MR, Hsieh YH (1982) Randomness, area, and species richness. Ecology 63(4):1121–1133. https://doi.org/10.2307/1937249

Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. Phil Trans R Soc B 345(1311):101–118. https://doi.org/10.1098/rstb.1994.0091

Cook WM, Lane KT, Foster BL, Holt RD (2002) Island theory, matrix effects and species richness patterns in habitat fragments. Ecol Lett 5(5):619–623. https://doi.org/10.1046/j.1461-0248.2002.00366.x

Cumming GS, Bodin O, Ernstson H, Elmqvist T (2010) Network analysis in conservation biogeography: challenges and opportunities. Divers Distrib 16(3):414–425. https://doi.org/10.1111/j.1472-4642.2010.00651.x

Cushman SA, McGarigal K (2002) Hierarchical, multi-scale decomposition of species-environment relationships. Landsc Ecol 17(7):637–646. https://doi.org/10.1023/a:1021571603605

D'Amen M, Dubuis A, Fernandes RF, Pottier J, Pellissier L, Guisan A (2015) Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. J Biogeogr 42(7):1255–1266. https://doi.org/10.1111/jbi.12485

D'Amen M, Rahbek C, Zimmermann NE, Guisan A (2017) Spatial predictions at the community level: from current approaches to future frameworks. Biol Rev 92(1):169–187. https://doi.org/10.1111/brv.12222

Decker E, Linke S, Hermoso V, Geist J (2017) Incorporating ecological functions in conservation decision making. Ecol Evol 7(20):8273–8281. https://doi.org/10.1002/ece3.3353

Devictor V, Mouillot D, Meynard C, Jiguet F, Thuiller W, Mouquet N (2010) Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. Ecol Lett 13(8):1030–1040. https://doi.org/10.1111/j.1461-0248.2010.01493.x

Diamond JM (1975) Assembly of species communities. In: Cody ML, Diamond JM (eds) Ecology and evolution of communities. Harvard University Press, Cambridge, MA

Distler T, Schuetz JG, Velasquez-Tibata J, Langham GM (2015) Stacked species distribution models and macroecological models provide congruent projections of avian species richness under climate change. J Biogeogr 42(5):976–988. https://doi.org/10.1111/jbi.12479

Dixon P (2003) VEGAN, a package of R functions for community ecology. J Veg Sci 14 (6):927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Dobson A, Lodge D, Alder J, Cumming GS, Keymer J, McGlade J, Mooney H, Rusak JA, Sala O, Wolters V, Wall D, Winfree R, Xenopoulos MA (2006) Habitat loss, trophic collapse, and the decline of ecosystem services. Ecology 87(8):1915–1924. https://doi.org/10.1890/0012-9658(2006)87[1915:hltcat]2.0.co;2

Dorazio RM, Connor EF (2014) Estimating abundances of interacting species using morphological traits, foraging guilds, and habitat. PLoS One 9(4). https://doi.org/10.1371/journal.pone.0094323

Dorazio RM, Royle JA, Soderstrom B, Glimskar A (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. Ecology 87(4):842–854. https://doi.org/10.1890/0012-9658(2006)87[842:esraab]2.0.co;2

Dorazio RM, Kery M, Royle JA, Plattner M (2010) Models for inference in dynamic metacommunity systems. Ecology 91(8):2466–2475. https://doi.org/10.1890/09-1033.1

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, Garcia Marquez JR, Gruber B, Lafourcade B, Leitao PJ, Muenkemueller T, McClean C, Osborne PE, Reineking B, Schroeder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36(1):27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Drakare S, Lennon JJ, Hillebrand H (2006) The imprint of the geographical, evolutionary and ecological context on species-area relationships. Ecol Lett 9(2):215–227. https://doi.org/10.1111/j.1461-0248.2005.00848.x

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol Model 196(3–4):483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

Dray S, Pelissier R, Couteron P, Fortin MJ, Legendre P, Peres-Neto PR, Bellier E, Bivand R, Blanchet FG, De Caceres M, Dufour AB, Heegaard E, Jombart T, Munoz F, Oksanen J, Thioulouse J, Wagner HH (2012) Community ecology in the age of multivariate multiscale spatial analysis. Ecol Monogr 82(3):257–275. https://doi.org/10.1890/11-1183.1

Driscoll DA (2008) The frequency of metapopulations, metacommunities and nestedness in a fragmented landscape. Oikos 117(2):297–309. https://doi.org/10.1111/j.2007.0030-1299.16202.x

Dubuis A, Pottier J, Rion V, Pellissier L, Theurillat JP, Guisan A (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. Divers Distrib 17(6):1122–1131. https://doi.org/10.1111/j.1472-4642.2011.00792.x

Economo EP, Keitt TH (2008) Species diversity in neutral metacommunities: a network approach. Ecol Lett 11(1):52–62. https://doi.org/10.1111/j.1461-0248.2007.01126.x

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. Annu Rev Ecol Syst 24:217–242. https://doi.org/10.1146/annurev.es.24.110193.001245

Fahrig L (2013) Rethinking patch size and isolation effects: the habitat amount hypothesis. J Biogeogr 40(9):1649–1663. https://doi.org/10.1111/jbi.12130

Faith DP, Carter G, Cassis G, Ferrier S, Wilkie L (2003) Complementarity, biodiversity viability analysis, and policy-based algorithms for conservation. Environ Sci Pol 6(3):311–328. https://doi.org/10.1016/s1462-9011(03)00044-3

Fernandez-Juricic E (2002) Can human disturbance promote nestedness? A case study with breeding birds in urban habitat fragments. Oecologia 131(2):269–278. https://doi.org/10.1007/s00442-002-0883-y

Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. J Appl Ecol 43 (3):393–404. https://doi.org/10.1111/j.1365-2664.2006.01149.x

Ferrier S, Pressey RL, Barrett TW (2000) A new predictor of the irreplaceability of areas for achieving a conservation goal, its application to real-world planning, and a research agenda for

further refinement. Biol Conserv 93(3):303–325. https://doi.org/10.1016/s0006-3207(99)00149-4
Ferrier S,Manion G, Elith J, Richardson K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Divers Distrib 13(3):252–264. https://doi.org/10.1111/j.1472-4642.2007.00341.x
Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecol Lett 18(1):1–16. https://doi.org/10.1111/ele.12376
Freilich MA, Wieters E, Broitman BR, Marquet PA, Navarrete SA (2018) Species co-occurrence networks: can they reveal trophic and non-trophic interactions in ecological communities? Ecology 99(3):690–699. https://doi.org/10.1002/ecy.2142
Fukami T (2015) Historical contingency in community assembly: integrating niches, species pools, and priority effects. Annu Rev Ecol Evol Syst 46:1–23. https://doi.org/10.1146/annurev-ecolsys-110411-160340
Gonzalez A (2000) Community relaxation in fragmented landscapes: the relation between species richness, area and age. Ecol Lett 3(5):441–448. https://doi.org/10.1046/j.1461-0248.2000.00171.x
Gonzalez A, Rayfield B, Lindo Z (2011) The disentangled bank: how loss of habitat fragments and disassembles ecological networks. Am J Bot 98(3):503–516. https://doi.org/10.3732/ajb.1000424
Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecol Lett 4(4):379–391. https://doi.org/10.1046/j.1461-0248.2001.00230.x
Gravel D, Massol F, Canard E, Mouillot D, Mouquet N (2011) Trophic theory of island biogeography. Ecol Lett 14(10):1010–1016. https://doi.org/10.1111/j.1461-0248.2011.01667.x
Gray CL, Hill SLL, Newbold T, Hudson LN, Borger L, Contu S, Hoskins AJ, Ferrier S, Purvis A, Scharlemann JPW (2016) Local biodiversity is higher inside than outside terrestrial protected areas worldwide. Nat Commun 7:12306. https://doi.org/10.1038/ncomms12306
Guichard F (2017) Recent advances in metacommunities and meta-ecosystem theories. F1000 Res 6:610
Guillot G, Rousset F (2013) Dismantling the Mantel tests. Methods Ecol Evol 4(4):336–344. https://doi.org/10.1111/2041-210x.12018
Guisan A, Rahbek C (2011) SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. J Biogeogr 38(8):1433–1444. https://doi.org/10.1111/j.1365-2699.2011.02550.x
Guisan A, Weiss SB, Weiss AD (1999) GLM versus CCA spatial modeling of plant species distribution. Plant Ecol 143(1):107–122
Haila Y (2002) A conceptual genealogy of fragmentation research: from island biogeography to landscape ecology. Ecol Appl 12(2):321–334. https://doi.org/10.2307/3060944
Hastie T, Fithian W (2013) Inference from presence-only data; the ongoing controversy. Ecography 36(8):864–867. https://doi.org/10.1111/j.1600-0587.2013.00321.x
Hoef JMV, Boveng PL (2007) Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? Ecology 88(11):2766–2772. https://doi.org/10.1890/07-0043.1
Holt RD (1992) A neglected facet of island biogeography: the role of internal spatial dynamics in area effects. Theor Popul Biol 41(3):354–371. https://doi.org/10.1016/0040-5809(92)90034-q
Holyoak M, Leibold MA, Holt RD (2005) Metacommunities: spatial dynamics and ecological communities. University of Chicago Press, Chicago
Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, NJ
Huffaker CB (1958) Experimental studies on predation: dispersion factors and predator-prey oscillations. Hilgardia 27:343–383

Hui FKC (2016) Boral - Bayesian ordination and regression analysis of multivariate abundance data in r. Methods Ecol Evol 7(6):744–750. https://doi.org/10.1111/2041-210x.12514
Hutto RL, Young JS (2002) Regional landbird monitoring: perspectives from the Northern Rocky Mountains. Wildl Soc Bull 30(3):738–750
Iknayan KJ, Tingley MW, Furnas BJ, Beissinger SR (2014) Detecting diversity: emerging methods to estimate species diversity. Trends Ecol Evol 29(2):97–106. https://doi.org/10.1016/j.tree.2013.10.012
Ings TC, Montoya JM, Bascompte J, Bluthgen N, Brown L, Dormann CF, Edwards F, Figueroa D, Jacob U, Jones JI, Lauridsen RB, Ledger ME, Lewis HM, Olesen JM, van Veen FJF, Warren PH, Woodward G (2009) Ecological networks - beyond food webs. J Anim Ecol 78 (1):253–269. https://doi.org/10.1111/j.1365-2656.2008.01460.x
Jewitt D, Goodman PS, O'Connor TG, Erasmus BFN, Witkowski ETF (2016) Mapping landscape beta diversity of plants across KwaZulu-Natal, South Africa, for aiding conservation planning. Biodivers Conserv 25(13):2641–2654. https://doi.org/10.1007/s10531-016-1190-y
Kaiser-Bunbury CN, Bluthgen N (2015) Integrating network ecology with applied conservation: a synthesis and guide to implementation. Aob Plants 7. https://doi.org/10.1093/aobpla/plv076
Karger DN, Cord AF, Kessler M, Kreft H, Kuehn I, Pompe S, Sandel B, Cabral JS, Smith AB, Svenning JC, Tuomisto H, Weigelt P, Wesche K (2016) Delineating probabilistic species pools in ecology and biogeography. Glob Ecol Biogeogr 25(4):489–501. https://doi.org/10.1111/geb.12422
Karp DS, Rominger AJ, Zook J, Ranganathan J, Ehrlich PR, Daily GC, Cornell H (2012) Intensive agriculture erodes beta-diversity at large scales. Ecol Lett 15(9):963–970. https://doi.org/10.1111/j.1461-0248.2012.01815.x
Keddy PA (1992) Assembly and response rules: two goals for predictive community ecology. J Veg Sci 3(2):157–164. https://doi.org/10.2307/3235676
Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends Ecol Evol 17 (5):230–241. https://doi.org/10.1016/s0169-5347(02)02489-8
Kerr JT, Sugar A, Packer L (2000) Indicator taxa, rapid biodiversity assessment, and nestedness in an endangered ecosystem. Conserv Biol 14(6):1726–1734. https://doi.org/10.1046/j.1523-1739.2000.99275.x
Kery M, Royle JA (2016) Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS. Academic, San Diego, CA
Kleiber C, Zeileis A (2008) Applied econometrics with R. Springer, New York
Koleff P, Gaston KJ (2002) The relationships between local and regional species richness and spatial turnover. Glob Ecol Biogeogr 11(5):363–375. https://doi.org/10.1046/j.1466-822x.2002.00302.x
Koleff P, Gaston KJ, Lennon JJ (2003) Measuring beta diversity for presence-absence data. J Anim Ecol 72(3):367–382. https://doi.org/10.1046/j.1365-2656.2003.00710.x
Kraft NJB, Adler PB, Godoy O, James EC, Fuller S, Levine JM (2015) Community assembly, coexistence and the environmental filtering metaphor. Funct Ecol 29(5):592–599. https://doi.org/10.1111/1365-2435.12345
Kukkala AS, Moilanen A (2013) Core concepts of spatial prioritisation in systematic conservation planning. Biol Rev 88(2):443–464. https://doi.org/10.1111/brv.12008
Laurance WF (2008) Theory meets reality: how habitat fragmentation research has transcended island biogeographic theory. Biol Conserv 141(7):1731–1744. https://doi.org/10.1016/j.biocon.2008.05.011
Lawton JH (1999) Are there general laws in ecology? Oikos 84(2):177–192. https://doi.org/10.2307/3546712
Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? Ecology 74(6):1659–1673
Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecol Monogr 69(1):1–24. https://doi.org/10.1890/0012-9615(1999)069[0001:dbratm]2.0.co;2
Legendre P, Legendre L (1998) Numerical ecology. Elsevier, Amsterdam

Legendre P, Legendre L (2012) Numerical ecology, 3rd edn. Elsevier, Amsterdam
Legendre P, Borcard D, Peres-Neto PR (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. Ecol Monogr 75(4):435–450. https://doi.org/10.1890/05-0549
Legendre P, Fortin M-J, Borcard D (2015) Should the Mantel test be used in spatial analysis? Methods Ecol Evol 6(11):1239–1247. https://doi.org/10.1111/2041-210x.12425
Leibold MA, Chase JM (2017) Metacommunity ecology. Princeton University Press, Princeton, NJ
Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman D, Loreau M, Gonzalez A (2004) The metacommunity concept: a framework for multi-scale community ecology. Ecol Lett 7(7):601–613. https://doi.org/10.1111/j.1461-0248.2004.00608.x
Lessard JP, Belmaker J, Myers JA, Chase JM, Rahbek C (2012) Inferring local ecological processes amid species pool influences. Trends Ecol Evol 27(11):600–607. https://doi.org/10.1016/j.tree.2012.07.006
Levins R, Culver D (1971) Regional coexistence of species and competition between rare species. Proc Natl Acad Sci U S A 68(6):1246. https://doi.org/10.1073/pnas.68.6.1246
Lichstein JW (2007) Multiple regression on distance matrices: a multivariate spatial analysis tool. Plant Ecol 188(2):117–131. https://doi.org/10.1007/s11258-006-9126-3
Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28(3):385–393
Liu C, White M, Newell G (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. J Biogeogr 40(4):778–789. https://doi.org/10.1111/jbi.12058
Lomolino MV (1990) The target area hypothesis: the influence of island area on immigration rates of non-volant mammals. Oikos 57(3):297–300. https://doi.org/10.2307/3565957
Lowe WH, McPeek MA (2014) Is dispersal neutral? Trends Ecol Evol 29(8):444–450. https://doi.org/10.1016/j.tree.2014.05.009
Mac Nally R, Lake PS (1999) On the generation of diversity in archipelagos: a re-evaluation of the Quinn-Harrison 'saturation index'. J Biogeogr 26(2):285–295. https://doi.org/10.1046/j.1365-2699.1999.00268.x
MacArthur RH (1972) Geographical ecology: patterns in teh distribution of species. Princeton University Press, Princeton, NJ
Macarthur RH, Wilson EO (1963) Equilibrium theory of insular zoogeography. Evolution 17 (4):373. https://doi.org/10.2307/2407089
MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press, Princeton, NJ
MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. Ecology 83(8):2248–2255
Magurran AE (2003) Measuring biological diversity. Wiley, Chichester
Magurran AE (2004) Measuring biological diversity. Blackwell Science Ltd, Malden, MA
Manion G, Lisk M, Ferrier S, Nieto-Lugilde D, Mokany K, Fitzpatrick MC (2018) gdm: generalized dissimilarity modeling. R package 1.3.11
Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405(6783):243–253. https://doi.org/10.1038/35012251
Mateo RG, Felicisimo AM, Pottier J, Guisan A, Munoz J (2012) Do stacked species distribution models reflect altitudinal diversity patterns? PLoS One 7(3). https://doi.org/10.1371/journal.pone.0032586
McPeek MA, Brown JM (2000) Building a regional species pool: diversification of the Enallagma damselflies in eastern North America. Ecology 81(4):904–920
Milanovich JR, Peterman WE, Barrett K, Hopton ME (2012) Do species distribution models predict species richness in urban and natural green spaces? A case study using amphibians. Landsc Urban Plan 107(4):409–418. https://doi.org/10.1016/j.landurbplan.2012.07.010
Mittelbach GG (2012) Comunity ecology. Sinauer Associates, Sunderland, MA

Moilanen A, Wilson KA, Possingham H (eds) (2009) Spatial conservation prioritization: quantitative methods and computational tools. Oxford University Press, Oxford

Mouchet MA, Villeger S, Mason NWH, Mouillot D (2010) Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. Funct Ecol 24(4):867–876. https://doi.org/10.1111/j.1365-2435.2010.01695.x

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403(6772):853–858. https://doi.org/10.1038/35002501

Neuhauser C (1998) Habitat destruction and competitive coexistence in spatially explicit models with local interactions. J Theor Biol 193(3):445–463

Newbold T, Gilbert F, Zalat S, El-Gabbas A, Reader T (2009) Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. J Biogeogr 36 (11):2085–2095. https://doi.org/10.1111/j.1365-2699.2009.02140.x

Nichols JD, Boulinier T, Hines JE, Pollock KH, Sauer JR (1998) Inference methods for spatial variation in species richness and community composition when not all species are detected. Conserv Biol 12(6):1390–1398

Noss RF (1990) Indicators for monitoring biodiversity: a hierarchical approach. Conserv Biol 4 (4):355–364

O'Neill RV, Johnson AR, King AW (1989) A hierarchical framework for the analysis of scale. Landsc Ecol 3(3–4):193–205. https://doi.org/10.1007/bf00131538

Olden JD, Rooney TP (2006) On defining and quantifying biotic homogenization. Glob Ecol Biogeogr 15(2):113–120. https://doi.org/10.1111/j.1466-822x.2006.00214.x

Orme CDL, Davies RG, Burgess M, Eigenbrod F, Pickup N, Olson VA, Webster AJ, Ding TS, Rasmussen PC, Ridgely RS, Stattersfield AJ, Bennett PM, Blackburn TM, Gaston KJ, Owens IPF (2005) Global hotspots of species richness are not congruent with endemism or threat. Nature 436(7053):1016–1019. https://doi.org/10.1038/nature03850

Ovaskainen O, Soininen J (2011) Making more out of sparse data: hierarchical modeling of species communities. Ecology 92(2):289–295. https://doi.org/10.1890/10-1251.1

Ovaskainen O, Hottola J, Siitonen J (2010) Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. Ecology 91(9):2514–2521. https://doi.org/10.1890/10-0173.1

Ovaskainen O, Abrego N, Halme P, Dunson D (2016a) Using latent variable models to identify large networks of species-to-species associations at different spatial scales. Methods Ecol Evol 7 (5):549–555. https://doi.org/10.1111/2041-210x.12501

Ovaskainen O, Roy DB, Fox R, Anderson BJ (2016b) Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. Methods Ecol Evol 7(4):428–436. https://doi.org/10.1111/2041-210x.12502

Ovaskainen O, Tikhonov G, Norberg A, Blanchet FG, Duan L, Dunson D, Roslin T, Abrego N (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. Ecol Lett 20(5):561–576. https://doi.org/10.1111/ele.12757

Palmer MW (1990) The estimation of species richness by extrapolation. Ecology 71(3):1195–1198. https://doi.org/10.2307/1937387

Palmer MW (1993) Putting things in even better order: the advantages of canonical correspondence analysis. Ecology 74(8):2215–2230. https://doi.org/10.2307/1939575

Partel M, Zobel M, Zobel K, vanderMaarel E (1996) The species pool and its relation to species richness: evidence from Estonian plant communities. Oikos 75(1):111–117. https://doi.org/10.2307/3546327

Pearson RG, Dawson TP, Liu C (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography 27(3):285–298. https://doi.org/10.1111/j.0906-7590.2004.03740.x

Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology 87(10):2614–2625. https://doi.org/10.1890/0012-9658(2006)87[2614:vposdm]2.0.co;2

Petchey OL, Gaston KJ (2002) Functional diversity (FD), species richness and community composition. Ecol Lett 5(3):402–411. https://doi.org/10.1046/j.1461-0248.2002.00339.x

Pineda E, Lobo JM (2009) Assessing the accuracy of species distribution models to predict amphibian species richness patterns. J Anim Ecol 78(1):182–190. https://doi.org/10.1111/j.1365-2656.2008.01471.x

Pineda E, Lobo JM (2012) The performance of range maps and species distribution models representing the geographic variation of species richness at different resolutions. Glob Ecol Biogeogr 21(9):935–944. https://doi.org/10.1111/j.1466-8238.2011.00741.x

Podani J, Schmera D (2011) A new conceptual and methodological framework for exploring and explaining pattern in presence-absence data. Oikos 120(11):1625–1638. https://doi.org/10.1111/j.1600-0706.2011.19451.x

Poff NL (1997) Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. J N Am Benthol Soc 16(2):391–409. https://doi.org/10.2307/1468026

Poisot T, Canard E, Mouillot D, Mouquet N, Gravel D (2012) The dissimilarity of species interaction networks. Ecol Lett 15(12):1353–1361. https://doi.org/10.1111/ele.12002

Poisot T, Gueveneux-Julien C, Fortin MJ, Gravel D, Legendre P (2017) Hosts, parasites and their interactions respond to different climatic variables. Glob Ecol Biogeogr 26(8):942–951. https://doi.org/10.1111/geb.12602

Pollock LJ, Tingley R, Morris WK, Golding N, O'Hara RB, Parris KM, Vesk PA, McCarthy MA (2014) Understanding co-occurrence bymodelling species simultaneously with a Joint Species DistributionModel (JSDM). Methods Ecol Evol 5(5):397–406. https://doi.org/10.1111/2041-210x.12180

Preston FW (1962) The canonical distribution of commonness and rarity: part I. Ecology 43 (2):185–215, 431–432

Rahbek C, Graves GR (2001) Multiscale assessment of patterns of avian species richness. Proc Natl Acad Sci U S A 98(8):4534–4539. https://doi.org/10.1073/pnas.071034898

Rands MRW, Adams WM, Bennun L, Butchart SHM, Clements A, Coomes D, Entwistle A, Hodge I, Kapos V, Scharlemann JPW, Sutherland WJ, Vira B (2010) Biodiversity conservation: challenges beyond 2010. Science 329(5997):1298–1303. https://doi.org/10.1126/science.1189138

Ricklefs RE (1987) Community diversity: relative roles of local and regional processes. Science 235(4785):167–171. https://doi.org/10.1126/science.235.4785.167

Rose PM, Kennard MJ, Sheldon F, Moffatt DB, Butler GL (2016) A data-driven method for selecting candidate reference sites for stream bioassessment programs using generalised dissimilarity models. Mar Freshw Res 67(4):440–454. https://doi.org/10.1071/mf14254

Rota CT, Wikle CK, Kays RW, Forrester TD, McShea WJ, Parsons AW, Millspaugh JJ (2016) A two-species occupancy model accommodating simultaneous spatial and interspecific dependence. Ecology 97(1):48–53. https://doi.org/10.1890/15-1193.1

Royle JA, Dorazio RM (2008) Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations, and communities. Academic, San Diego, CA

Scheiner SM (2003) Six types of species-area curves. Glob Ecol Biogeogr 12(6):441–447. https://doi.org/10.1046/j.1466-822X.2003.00061.x

Seabloom EW, Dobson AP, Stoms DM (2002) Extinction rates under nonrandom patterns of habitat loss. Proc Natl Acad Sci U S A 99(17):11229–11234. https://doi.org/10.1073/pnas.162064899

Socolar JB, Gilroy JJ, Kunin WE, Edwards DP (2016) How should beta-diversity inform biodiversity conservation? Trends Ecol Evol 31(1):67–80. https://doi.org/10.1016/j.tree.2015.11.005

Soininen J, McDonald R, Hillebrand H (2007) The distance decay of similarity in ecological communities. Ecography 30(1):3–12. https://doi.org/10.1111/j.2006.0906-7590.04817.x

Ter Braak CJF (1987) The analysis of vegetation-environment relationships by canonical correspondence analysis. Vegetatio 69(1–3):69–77. https://doi.org/10.1007/bf00038688

Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, de Siqueira MF, Grainger A, Hannah L, Hughes L, Huntley B, van Jaarsveld AS, Midgley GF,

Miles L, Ortega-Huerta MA, Peterson AT, Phillips OL, Williams SE (2004) Extinction risk from climate change. Nature 427(6970):145–148. https://doi.org/10.1038/nature02121
Thomassen HA, Fuller T, Buermann W, Mila B, Kieswetter CM, Jarrin P, Cameron SE, Mason E, Schweizer R, Schlunegger J, Chan J, Wang O, Peralvo M, Schneider CJ, Graham CH, Pollinger JP, Saatchi S, Wayne RK, Smith TB (2011) Mapping evolutionary process: a multi-taxa approach to conservation prioritization. Evol Appl 4(2):397–413. https://doi.org/10.1111/j.1752-4571.2010.00172.x
Tikhonov G, Abrego N, Dunson D, Ovaskainen O (2017) Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context. Methods Ecol Evol 8(4):443–452. https://doi.org/10.1111/2041-210x.12723
Tilman D, May RM, Lehman CL, Nowak MA (1994) Habitat destruciton and the extinction debt. Nature 371(6492):65–66
Urban DL, Oneill RV, Shugart HH (1987) Landscape ecology. Bioscience 37(2):119–127
Urban D, Goslee S, Pierce K, Lookingbill T (2002) Extending community ecology to landscapes. Ecoscience 9(2):200–212
van der Valk AG (1981) Succession in Wetlands: A Gleasonian Appraoch. Ecology 62 (3):688–696
Vellend M (2010) Conceptual synthesis in community ecology. Q Rev Biol 85(2):183–206. https://doi.org/10.1086/652373
Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York
Vilhena DA, Antonelli A (2015) A network approach for identifying and delimiting biogeographical regions. Nat Commun 6. https://doi.org/10.1038/ncomms7848
Wagner HH (2003) Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. Ecology 84(4):1045–1057. https://doi.org/10.1890/0012-9658(2003)084[1045:scipci]2.0.co;2
Wagner HH (2004) Direct multi-scale ordination with canonical correspondence analysis. Ecology 85(2):342–351. https://doi.org/10.1890/02-0738
Wang Y, Naumann U, Wright ST, Warton DI (2012) mvabund - an R package for model-based analysis of multivariate abundance data. Methods Ecol Evol 3(3):471–474. https://doi.org/10.1111/j.2041-210X.2012.00190.x
Wang Y, Naumann U, Eddelbuettel D, Warton D (2018) mvabund: statistical methods for analysing multivariate abundance data. R version 3.13.1
Warton DI, Wright ST, Wang Y (2012) Distance-based multivariate analyses confound location and dispersion effects. Methods Ecol Evol 3(1):89–101. https://doi.org/10.1111/j.2041-210X.2011.00127.x
Warton DI, Blanchet FG, O'Hara RB, Ovaskainen O, Taskinen S, Walker SC, Hui FKC (2015a) So many variables: joint modeling in community ecology. Trends Ecol Evol 30(12):766–779. https://doi.org/10.1016/j.tree.2015.09.007
Warton DI, Foster SD, De'ath G, Stoklosa J, Dunstan PK (2015b) Model-based thinking for community ecology. Plant Ecol 216(5):669–682. https://doi.org/10.1007/s11258-014-0366-3
Whitcomb RF, Lynch JF, Opler PA, Robbins CS (1976) Island biogeography and conservation: strategy and limitations. Science 193(4257):1030–1032
Whittaker RJ, Araújo MB, Paul J, Ladle RJ, Watson JEM, Willis KJ (2005) Conservation biogeography: assessment and prospect. Divers Distrib 11(1):3–23. https://doi.org/10.1111/j.1366-9516.2005.00143.x
Wickham H (2007) Reshaping data with the reshape package. J Stat Softw 21(12):1–20
Williams PH (1996) Mapping variations in the strength and breadth of biogeographic transition zones using species turnover. Proc R Soc B 263(1370):579–588. https://doi.org/10.1098/rspb.1996.0087
Wilson DS (1992) Complex interactions in metacommunities, with implications for biodiversity and higher levels of selection. Ecology 73(6):1984–2000. https://doi.org/10.2307/1941449
Wilson KA, McBride MF, Bode M, Possingham HP (2006) Prioritizing global conservation efforts. Nature 440(7082):337–340. https://doi.org/10.1038/nature04366

Winfree R, Aguilar R, Vazquez DP, LeBuhn G, Aizen MA (2009) A meta-analysis of bees' responses to anthropogenic disturbance. Ecology 90(8):2068–2076

Wright DH, Reeves JH (1992) On the meaning and measurement of nestedness of species assemblages. Oecologia 92(3):416–428. https://doi.org/10.1007/bf00317469

Wright DH, Patterson BD, Mikkelson GM, Cutler A, Atmar W (1998) A comparative analysis of nested subset patterns of species composition. Oecologia 113(1):1–20

Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Grant EHC, Veran S (2013) Presence-only modelling using MAXENT: when can we trust the inferences? Methods Ecol Evol 4 (3):236–243. https://doi.org/10.1111/2041-210x.12004

Zeileis A, Kleiber C, Jackman S (2008) Regression models for count data in R. J Stat Softw 27 (8):1–25. https://doi.org/10.18637/jss.v027.i08

Zhang MG, Slik JWF, Ma KP (2016) Using species distribution modeling to delineate the botanical richness patterns and phytogeographical regions of China. Sci Rep 6. https://doi.org/10.1038/srep22400

Chapter 12
What Have We Learned? Looking Back and Pressing Forward



12.1 The Impact of Spatial Ecology and Conservation

Nearly 25 years ago, Peter Kareiva emphasized that space was the final frontier for ecology (Kareiva 1994). To understand the importance of space, the field of spatial ecology has emerged as a central focus for many topics in ecology and conservation (see Chap. 1). Looking back at the last two decades, it is impressive just how much the problem of space has intrigued the minds of ecologists and conservation biologists. This focus has arisen from both theoretical and applied developments, most notably the growth of landscape ecology, spatial analysis, and the increased availability of spatial data. This growth has made huge impacts on the field, and we expect those impacts to increase with time. Several key insights have emerged from spatial ecology, many of which have been emphasized throughout this book.

Ecological and spatial theories are now well developed that emphasize the role of space on populations and communities (Hastings and Gross 2012). For populations, metapopulation theory has made several general insights (Hanski 1999; Hanski and Ovaskainen 2003), many of which are supported through empirical investigations (e.g., Hames et al. 2001; but see Baguette 2004). In recent years, metacommunity theory has developed rapidly and is still growing (Leibold and Chase 2017), with ongoing tests of the utility of this theoretical development in explaining community structure (Logue et al. 2011). Metaecosystem theory has also emerged, borrowing ideas from metacommunity theory and placing them squarely in the context of ecosystem ecology (Loreau et al. 2003; Loreau and Holt 2004; Massol et al. 2011; Gravel et al. 2016).

The role of spatial and temporal scale is now widely recognized and integrated into most ecological and conservation problems (Levin 1992; Hein et al. 2006; Boyd et al. 2008; Hastings 2010; Power 2010). Spatial scale can alter conclusions on the drivers of species and communities, and conservation decisions are impacted by the scale at which problems are viewed. Multiscale and multilevel modeling of species–

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1_12

environment relationships are becoming standard (Boyce 2006; Mayor et al. 2009; Nichols et al. 2008; McGarigal et al. 2016).

Spatial dependence and autocorrelation are widely recognized as being pervasive (Bini et al. 2009), being both nuisance issues that researchers must account for but also opportunities for a better understanding of ecological processes and more accurate predictions of ecological patterns (Legendre 1993). A wide variety of approaches are now used to diagnose and account for spatial dependence (Dale and Fortin 2009; Beale et al. 2010; Bardos et al. 2015), and in some cases ignoring this autocorrelation has proved dangerous (Crase et al. 2014).

Space use, resource selection, and species distribution are related, fundamental topics for ecology and these approaches now primarily focus on a spatially explicit understanding of how and why distribution varies (Guisan and Zimmermann 2000; Wilson et al. 2004; Guisan and Thuiller 2005; Aarts et al. 2012; Wisz et al. 2013). Several insights have emerged from these efforts regarding the role of abiotic, biotic, and movement limitations in driving distribution (Soberón 2007). Biotic interactions are often thought to drive distributions at local scales, movement limitations at moderate scales, and abiotic factors at large scales (Pulliam 2000; Soberón 2010; Barve et al. 2011). These efforts have also made insights regarding how habitat quality and distribution are related in space and time.

Insights from spatial ecology are now making major impacts for conservation. Spatial concepts are at the forefront of information used for conservation problems and decision-making. Ongoing issues, such as the land-sharing versus land-sparing debate for limiting impacts of agriculture on biodiversity (Perfecto and Vandermeer 2010), are firmly grounded in spatial ecology (Fischer et al. 2014). Many conservation planning priorities, such as facilitating species complementary and representation in reserve design (Margules and Pressey 2000; Kukkala and Moilanen 2013), rest on spatial ecology concepts. The use of indicators for conservation decisions also increasingly acknowledges spatial ecology concepts (Lindenmayer et al. 2000). Several policies around the world have integrated elements of spatial ecology. For example, the Conservation Reserve Enhancement Program in the USA uses spatial information regarding riparian buffer widths in their allocations to farmers.

Spatial modeling tools are now well integrated into conservation. These tools have been used for prioritizing lands for future acquisition (e.g., the delineation of new protected areas). Tools are commonly used for forecasting the effects of climate change, such as the use of species distribution models, and identifying climate adaptation strategies where conservation is needed most (Heller and Zavaleta 2009; Bellard et al. 2012; Guisan et al. 2013). Ecosystem services are increasingly viewed from a spatial perspective and spatial mapping of such services is now common (Naidoo et al. 2008; Kandziora et al. 2013; Schagner et al. 2013; Verhagen et al. 2017). These types of exercises are now used in policy decision-marking (Maes et al. 2012). Connectivity modeling is increasingly used to identifying limitations for populations with land-use and climate change (Pascual-Hortal and Saura 2006; Crooks and Sanjayan 2006; Rudnick et al. 2012; Lawler et al. 2013).

12.2 Looking Forward: Frontiers for Spatial Ecology and Conservation

With the advancement of spatial ecology concepts and tools, and their illustration in this book, it is useful to identify new directions and ongoing challenges for the field. Here, we provide an idiosyncratic discussion of prospects.

Throughout this book, we have focused primarily on the spatial dimension of ecology and conservation. This emphasis was inspired by many key developments in ecology and conservation, such as the realization of the importance of spatial scale and the developments of metapopulation, metacommunity, and metaecosystem ecology. Yet this spatial emphasis always plays out in the context of time (Hastings 2010). Understanding spatiotemporal dynamics are increasingly emphasized being necessary to interpret ecological dynamics and to better contextualize conservation. For instance, understanding movement ecology and accurately predicting connectivity are contingent on the time scale being considered (Zeigler and Fagan 2014). A variety of spatiotemporal modeling approaches are being developed to predict ecological patterns and processes (Gelfand 2007; Chen et al. 2011; Thorson et al. 2016; Martensen et al. 2017; Wittemyer et al. 2017).

In ecology, it is increasingly recognized that ecological and evolutionary processes can operate simultaneously and feedback on each other (Schoener 2011). Furthermore, evolutionary processes are often relevant to conservation. Yet spatial eco-evolutionary interactions are only beginning to be explored. For example, despite decades of research on habitat loss and fragmentation, and the fact that environmental change can have strong selection pressures on dispersal (Cheptou et al. 2008), empirical work on the spatial effects of land-use and related environmental change on eco-evolutionary processes is limited (Farkas et al. 2015; Legrand et al. 2017).

While the importance of scale is now well known and many tools exist for diagnosing scaling effects (Keitt and Urban 2005; Chandler and Hepinstall-Cymerman 2016), integrating patterns and processes across levels of organization and scales remains an important frontier in ecology (Peters et al. 2007; Chave 2013; Carmona et al. 2016). Such integration will require formal linking of different types of data operating at different spatial and temporal scales (Talluto et al. 2016).

In applied ecology and conservation, coupled human-natural systems is a framework that emphasizes feedbacks of humans and human decision-making on biodiversity and how biodiversity may then feedback into human decision-making (Liu et al. 2007; Turner et al. 2007). Several examples exist, each of which highlights the complex interplay of humans and their environment (Hull et al. 2015; Liu et al. 2015). While spatial concepts and modeling tools are often used to explore such topics, further methodological and conceptual advancements are needed. In particular, modeling frameworks that honor feedbacks and potential mismatches of sampling and process scales are needed. All these modeling approaches will require ecological data to parameterize them.

Forecasting changes in biodiversity in space and time is increasingly needed for guiding decision-making and for accurately predicting the effects of environmental change (Clark et al. 2001; Botkin et al. 2007; Barnosky et al. 2012). Accurate forecasting over the near-term and the long-term can be incredibly challenging, however. Understanding the time frame for which spatiotemporal forecasts can be usefully made, termed the "ecological forecast horizon," is an ongoing priority (Petchey et al. 2015). To make useful predictions, embracing different types of uncertainty is required. Uncertainty can arise from a variety of sources (Regan et al. 2002; Nichols et al. 2011), and oftentimes either uncertainty is ignored or is only partially addressed.

Finally, the use of "big" data is increasingly embraced in ecology and conservation (Kelling et al. 2009). Big spatial data are now being collected in a variety of ways. These types of data include remotely sensed data (Kerr and Ostrovsky 2003; Turner et al. 2003), data from citizen science programs (Dickinson et al. 2010), and field data where automated data collection occurs (Tomkiewicz et al. 2010). For example, drones are now being used to obtain high-resolution data on plants and animals (Anderson and Gaston 2013). Acoustic arrays capture spatial variation in activity of animals and their communication, as well as other aspects of the "soundscape" (Blumstein et al. 2011; Mennill et al. 2012; Merchant et al. 2015). The use of high-resolution GPS telemetry data has exploded in recent years (Cagnacci et al. 2010). In each of these cases, ecologists will need to use new analytical and modeling tools for interpreting ecological patterns and processes with big data (Hochachka et al. 2007; Olden et al. 2008; Kampichler et al. 2010).

12.3 Where to Go from Here for Advanced Spatial Modeling?

Throughout this book, we have provided an introduction to a broad suite of topics and problems in spatial ecology and conservation. Our aim for this book was to provide a foundation and springboard rather than an in-depth, advanced account of each of these problems. There are several excellent books dedicated to specific topics covered here, such as species distributions (Franklin 2009; Peterson et al. 2011; Guisan et al. 2017), resource selection and animal movement (Hooten et al. 2017), connectivity (Crooks and Sanjayan 2006), point patterns (Baddeley et al. 2015), geostatistics and spatial statistics more broadly (Cressie 1993; Cressie and Wikle 2011; Dale and Fortin 2014), spatial population biology (Hanski 1999; Ovaskainen et al. 2016), metacommunity ecology (Holyoak et al. 2005), and spatial conservation planning (Moilanen et al. 2009). Some of these books focus more on ecological modeling (Franklin 2009), while others provide a detailed accounting of concepts and theory (Hanski 1999). There are also several excellent books also specifically focus on the use of R for general spatial analysis and modeling outside of ecology and conservation (Bivand et al. 2013; Baddeley et al. 2015: Brunsdon and Comber 2015). Finally, for landscape

ecology problems, Gergel and Turner (2017) provided a useful set of exercises that illustrate a variety of software that can be used.

There are also several modeling approaches and tools in R that we did not cover. Some of them focus on using advanced statistics, such as Bayesian spatial modeling, while others cover more specific topics, such as point pattern analysis on 1D surfaces. In general, R is rapidly advancing for spatial modeling and analysis. We encourage the reader to check out: http://rspatial.org. Another useful link for a summary on various R packages for spatial data is: https://cran.r-project.org/web/views/Spatial.html.

The topics covered in this book are increasingly merged together for more complex problems. For example, systematic conservation planning is often implemented in conservation biology (Margules and Pressey 2000). In this general approach, spatially explicit information on land use, biodiversity, connectivity, and socioeconomic costs are developed, often using spatial concepts described in this book (Loiselle et al. 2003; Rondinini et al. 2006; Carvalho et al. 2016; Gunton et al. 2017). With this information, optimization routines are often used to identify (near) optimal decisions for land prioritization and key areas under threat from environmental change (Pressey et al. 2007). From a modeling perspective, many of the inputs required can be implemented with tools described here, although for optimization, R has not been fully developed for integrating this information. Rather, programs such as Marxan and Zonation are frequently used (Moilanen 2007; Watts et al. 2009).

Another advanced approach for spatial modeling that is commonly used is agent-based (or individual-based) modeling (DeAngelis and Mooij 2005; Grimm et al. 2005). Agent-based modeling can be useful for capturing highly detailed complexity of variation in the spatial environment and how individuals respond to the environment (be it individual plants or animals). Agent-based models are often used to gain a detailed understanding of scenarios that may impact a location, such as climate change, management strategies (DeAngelis et al. 1998), or disturbance regimes. They are also increasingly used to interpret population viability (Pulliam et al. 1992; Schumaker et al. 2014). We did not cover agent-based modeling in this book because it is often very case specific and frequently R is not the best platform for agent-based modeling. Rather, agent-based modeling is commonly performed with lower-level programming languages, such as C++. One common software designed for agent-based modeling is NetLogo (Sklar 2007). This software can now be implemented through R with the `RNetLogo` package (Thiele and Grimm 2010; Thiele et al. 2012).

12.4 Beyond R

Throughout this book, we have used R to illustrate spatial ecological concepts and models. R is incredibly useful for many problems in spatial ecology and conservation, and it is excellent for teaching these principles. Nonetheless, R does have its

limits for spatial modeling. R has limits in its ability to use, manipulate, and model spatial data. In particular, R typically accesses objects and data stored in memory. As a consequence, there is a hard limit on the size of data that can be used in R, where limits are based on the memory on the user's computing platform (see `memory.size` and `memory.limits` functions). This limit makes working with large spatial data cumbersome and impractical in some situations. R can also be awkward for viewing spatial data: it is straightforward to zoom into locations on maps (see for example Chap. 2), but the ability to rapidly alter the visualization is impractical in most situations. In these cases, there are at least two solutions. First, there are several ways to link R with other types of spatial software. Second, one may wish to exit R and work with other software.

R can be coupled with other useful packages for spatial analysis. QGIS (http://qgis.org/) and GRASS (https://grass.osgeo.org/) GIS are both open-source GIS software that can interface with R. QGIS is more of an all-purpose GIS, while GRASS is specifically tailored for spatial modeling. In R, QGIS can be called through the recently developed `RQGIS` package (Muenchow et al. 2017), which can also provide a means to access some GRASS functions. Within GRASS, R can be called with the `rgrass7` package (Bivand 2017). There is also some limited linkages of ArcGIS extensions with R (e.g., the geospatial modeling environment) and ArcGIS now has a toolbox extension that allows users to call R from within ArcGIS.

Other software can used for spatial modeling. These include other GIS software, such as ArcGIS, QGIS, GRASS, and PostGIS. ArcGIS is useful for mapping, but is limited for spatial modeling (particularly statistical models). QGIS and GRASS GIS have several spatial modeling functions that make them useful alternatives to R. PostGIS is increasingly used for big data, such as GPS telemetry, because of its ability to manipulate large amounts of spatial data. Other spatial analysis programs include Programitta for point pattern analysis (Wiegand and Moloney 2014), and Passage (Rosenberg and Anderson 2011) for several types of analyses. More all-purpose software is also sometimes used, such as Python, Matlab, and C++. In particular, Python is frequently used (e.g., Etherington et al. 2015), in part because of the fact that ArcGIS uses the Python programming language.

12.5 Conclusions

We have come a long way in our understanding of when space is (and is not!) critical for ecology, and how spatial ecology can help guide management and conservation decisions. Spatial issues are now foundational bedrock for most ecological questions and conservation problems. Looking forward, we expect these issues will be even more embraced with the flood of spatial data now available. The steps ahead to take as much as possible advantage of the large amount of data are to develop: (1) methodologies that combine data with different spatial and temporal resolutions; and (2) meta-models that couple ecological, evolutionary, and human processes from

one scale to another (Talluto et al. 2016; Ferguson et al. 2017). Such advancements will help ecologists to formally link spatial issues with other key issues to better understand how biodiversity and ecosystem function will respond to environmental change, such that they will play an increasingly essential role in ecology and conservation.

References

Aarts G, Fieberg J, Matthiopoulos J (2012) Comparative interpretation of count, presence-absence and point methods for species distribution models. Methods Ecol Evol 3(1):177–187. https://doi.org/10.1111/j.2041-210X.2011.00141.x

Anderson K, Gaston KJ (2013) Lightweight unmanned aerial vehicles will revolutionize spatial ecology. Front Ecol Environ 11(3):138–146. https://doi.org/10.1890/120150

Baddeley A, Rubak E, Turner R (2015) Spatial point patterns: methodology and applications with R. CRC Press, Boca Raton, FL

Baguette M (2004) The classical metapopulation theory and the real, natural world: a critical appraisal. Basic Appl Ecol 5(3):213–224. https://doi.org/10.1016/j.baae.2004.03.001

Bardos DC, Guillera-Arroita G, Wintle BA (2015) Valid auto-models for spatially autocorrelated occupancy and abundance data. Methods Ecol Evol 6(10):1137–1149. https://doi.org/10.1111/2041-210x.12402

Barnosky AD, Hadly EA, Bascompte J, Berlow EL, Brown JH, Fortelius M, Getz WM, Harte J, Hastings A, Marquet PA, Martinez ND, Mooers A, Roopnarine P, Vermeij G, Williams JW, Gillespie R, Kitzes J, Marshall C, Matzke N, Mindell DP, Revilla E, Smith AB (2012) Approaching a state shift in Earth's biosphere. Nature 486(7401):52–58. https://doi.org/10.1038/nature11018

Barve N, Barve V, Jimenez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberon J, Villalobos F (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol Model 222(11):1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Beale CM, Lennon JJ, Yearsley JM, Brewer MJ, Elston DA (2010) Regression analysis of spatial data. Ecol Lett 13(2):246–264. https://doi.org/10.1111/j.1461-0248.2009.01422.x

Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F (2012) Impacts of climate change on the future of biodiversity. Ecol Lett 15(4):365–377. https://doi.org/10.1111/j.1461-0248.2011.01736.x

Bini LM, Diniz JAF, Rangel T, Akre TSB, Albaladejo RG, Albuquerque FS, Aparicio A, Araujo MB, Baselga A, Beck J, Bellocq MI, Bohning-Gaese K, Borges PAV, Castro-Parga I, Chey VK, Chown SL, de Marco P, Dobkin DS, Ferrer-Castan D, Field R, Filloy J, Fleishman E, Gomez JF, Hortal J, Iverson JB, Kerr JT, Kissling WD, Kitching IJ, Leon-Cortes JL, Lobo JM, Montoya D, Morales-Castilla I, Moreno JC, Oberdorff T, Olalla-Tarraga MA, Pausas JG, Qian H, Rahbek C, Rodriguez MA, Rueda M, Ruggiero A, Sackmann P, Sanders NJ, Terribile LC, Vetaas OR, Hawkins BA (2009) Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography 32(2):193–204. https://doi.org/10.1111/j.1600-0587.2009.05717.x

Bivand R (2017) rgrass7: interface between GRASS 7 geographic information systems and R. R package version 0.1-10

Bivand RS, Pebesma EJ, Gomez-Rubio V (2013) Applied spatial data analysis with R. Use R! 2nd edn. Springer, New York

Blumstein DT, Mennill DJ, Clemins P, Girod L, Yao K, Patricelli G, Deppe JL, Krakauer AH, Clark C, Cortopassi KA, Hanser SF, McCowan B, Ali AM, Kirschel ANG (2011) Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological

considerations and prospectus. J Appl Ecol 48(3):758–767. https://doi.org/10.1111/j.1365-2664.2011.01993.x

Botkin DB, Saxe H, Araujo MB, Betts R, Bradshaw RHW, Cedhagen T, Chesson P, Dawson TP, Etterson JR, Faith DP, Ferrier S, Guisan A, Hansen AS, Hilbert DW, Loehle C, Margules C, New M, Sobel MJ, Stockwell DRB (2007) Forecasting the effects of global warming on biodiversity. Bioscience 57(3):227–236. https://doi.org/10.1641/b570306

Boyce MS (2006) Scale for resource selection functions. Divers Distrib 12(3):269–276. https://doi.org/10.1111/j.1366-9516.2006.00243.x

Boyd C, Brooks TM, Butchart SHM, Edgar GJ, da Fonseca GAB, Hawkins F, Hoffmann M, Sechrest W, Stuart SN, van Dijk PP (2008) Spatial scale and the conservation of threatened species. Conserv Lett 1(1):37–43. https://doi.org/10.1111/j.1755-263X.2008.00002.x

Brunsdon C, Comber L (2015) An introduction to R for spatial analysis and mapping. Sage Publications, Inc, London

Cagnacci F, Boitani L, Powell RA, Boyce MS (2010) Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. Phil Trans R Soc B 365 (1550):2157–2162. https://doi.org/10.1098/rstb.2010.0107

Carmona CP, de Bello F, Mason NWH, Leps J (2016) Traits without borders: integrating functional diversity across scales. Trends Ecol Evol 31(5):382–394. https://doi.org/10.1016/j.tree.2016.02.003

Carvalho SB, Goncalves J, Guisan A, Honrado JP (2016) Systematic site selection for multispecies monitoring networks. J Appl Ecol 53(5):1305–1316. https://doi.org/10.1111/1365-2664.12505

Chandler R, Hepinstall-Cymerman J (2016) Estimating the spatial scales of landscape effects on abundance. Landsc Ecol 31(6):1383–1394. https://doi.org/10.1007/s10980-016-0380-z

Chave J (2013) The problem of pattern and scale in ecology: what have we learned in 20 years? Ecol Lett 16:4–16. https://doi.org/10.1111/ele.12048

Chen QW, Han R, Ye F, Li WF (2011) Spatio-temporal ecological models. Eco Inform 6(1):37–43. https://doi.org/10.1016/j.ecoinf.2010.07.006

Cheptou PO, Carrue O, Rouifed S, Cantarel A (2008) Rapid evolution of seed dispersal in an urban environment in the weed *Crepis sancta*. Proc Natl Acad Sci U S A 105(10):3796–3799. https://doi.org/10.1073/pnas.0708446105

Clark JS, Carpenter SR, Barber M, Collins S, Dobson A, Foley JA, Lodge DM, Pascual M, Pielke R, Pizer W, Pringle C, Reid WV, Rose KA, Sala O, Schlesinger WH, Wall DH, Wear D (2001) Ecological forecasts: an emerging imperative. Science 293(5530):657–660. https://doi.org/10.1126/science.293.5530.657

Crase B, Liedloff A, Vesk PA, Fukuda Y, Wintle BA (2014) Incorporating spatial autocorrelation into species distribution models alters forecasts of climate-mediated range shifts. Glob Chang Biol 20(8):2566–2579. https://doi.org/10.1111/gcb.12598

Cressie NAC (1993) Statistics for spatial data. Wiley, Chichester

Cressie N, Wikle CK (2011) Statistics for spatio-temporal data. Wiley, Chichester

Crooks KR, Sanjayan M (eds) (2006) Connectivity conservation. Cambridge University Press, New York

Dale MRT, Fortin MJ (2009) Spatial autocorrelation and statistical tests: some solutions. J Agric Biol Environ Stat 14(2):188–206. https://doi.org/10.1198/jabes.2009.0012

Dale MRT, Fortin MJ (2014) Spatial analysis: a guide for ecologists, 2nd edn. Cambridge University Press, Cambridge

DeAngelis DL, Mooij WM (2005) Individual-based modeling of ecological and evolutionary processes. Annu Rev Ecol Evol Syst 36:147–168. https://doi.org/10.1146/annurev.ecolsys.36.102003.152644

DeAngelis DL, Gross LJ, Huston MA, Wolff WF, Fleming DM, Comiskey EJ, Sylvester SM (1998) Landscape modeling for everglades ecosystem restoration. Ecosystems 1(1):64–75. https://doi.org/10.1007/s100219900006

Dickinson JL, Zuckerberg B, Bonter DN (2010) Citizen science as an ecological research tool: challenges and benefits. Annu Rev Ecol Evol Syst 41:149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636

Etherington TR, Holland EP, O'Sullivan D (2015) NLMpy: a PYTHON software package for the creation of neutral landscape models within a general numerical framework. Methods Ecol Evol 6(2):164–168. https://doi.org/10.1111/2041-210x.12308

Farkas TE, Hendry AP, Nosil P, Beckerman AP (2015) How maladaptation can structure biodiversity: eco-evolutionary island biogeography. Trends Ecol Evol 30(3):154–160. https://doi.org/10.1016/j.tree.2015.01.002

Ferguson JM, Reichert BE, Fletcher RJ, Jager HI (2017) Detecting population-environmental interactions with mismatched time series data. Ecology 98(11):2813–2822. https://doi.org/10.1002/ecy.1966

Fischer J, Abson DJ, Butsic V, Chappell MJ, Ekroos J, Hanspach J, Kuemmerle T, Smith HG, von Wehrden H (2014) Land sparing versus land sharing: moving forward. Conserv Lett 7 (3):149–157. https://doi.org/10.1111/conl.12084

Franklin J (2009) Mapping species distributions: spatial inference and prediction. Cambridge University Press, Cambridge, UK

Gelfand AE (2007) Guest Editorial: Spatial and spatio-temporal modeling in environmental and ecological statistics. Environ Ecol Stat 14(3):191–192. https://doi.org/10.1007/s10651-007-0026-z

Gergel SE, Turner MG (eds) (2017) Learning landscape ecology: a practical guide to concepts and techniques, 2nd edn. Springer, New York

Gravel D, Massol F, Leibold MA (2016) Stability and complexity in model meta-ecosystems. Nat Commun 7. https://doi.org/10.1038/ncomms12457

Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke HH, Weiner J, Wiegand T, DeAngelis DL (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310(5750):987–991. https://doi.org/10.1126/science.1116681

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8(9):993–1009

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135(2–3):147–186

Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, Martin TG, Rhodes JR, Maggini R, Setterfield SA, Elith J, Schwartz MW, Wintle BA, Broennimann O, Austin M, Ferrier S, Kearney MR, Possingham HP, Buckley YM (2013) Predicting species distributions for conservation decisions. Ecol Lett 16(12):1424–1435. https://doi.org/10.1111/ele.12189

Guisan A, Thuiller W, Zimmermann NE (2017) Habitat suitability and distribution models: applications with R. Cambridge University Press, Cambridge, UK

Gunton RM, Marsh CJ, Moulherat S, Malchow AK, Bocedi G, Klenke RA, Kunin WE (2017) Multicriterion trade-offs and synergies for spatial conservation planning. J Appl Ecol 54 (3):903–913. https://doi.org/10.1111/1365-2664.12803

Hames RS, Rosenberg KV, Lowe JD, Dhondt AA (2001) Site reoccupation in fragmented landscapes: testing predictions of metapopulation theory. J Anim Ecol 70(2):182–190

Hanski I (1999) Metapopulation ecology. Oxford University Press, Oxford

Hanski K, Ovaskainen O (2003) Metapopulation theory for fragmented landscapes. Theor Popul Biol 64(1):119–127. https://doi.org/10.1016/s0040-5809(03)00022-4

Hastings A (2010) Timescales, dynamics, and ecological understanding. Ecology 91 (12):3471–3480. https://doi.org/10.1890/10-0776.1

Hastings A, Gross L (eds) (2012) Encyclopedia of theoretical ecology. UC Press, Berkeley

Hein L, van Koppen K, de Groot RS, van Ierland EC (2006) Spatial scales, stakeholders and the valuation of ecosystem services. Ecol Econ 57(2):209–228. https://doi.org/10.1016/j.ecolecon.2005.04.005

Heller NE, Zavaleta ES (2009) Biodiversity management in the face of climate change: a review of 22 years of recommendations. Biol Conserv 142(1):14–32. https://doi.org/10.1016/j.biocon.2008.10.006

Hochachka WM, Caruana R, Fink D, Munson A, Riedewald M, Sorokina D, Kelling S (2007) Data-mining discovery of pattern and process in ecological systems. J Wildl Manag 71 (7):2427–2437. https://doi.org/10.2193/2006-503

Holyoak M, Leibold MA, Holt RD (2005) Metacommunities: spatial dynamics and ecological communities. University of Chicago Press, Chicago

Hooten MB, Johnson DS, McClintock BT, Morales JM (2017) Animal movement: statistical models for telemetry data. CRC Press, Boca Raton, FL

Hull V, Tuanniu MN, Liu JG (2015) Synthesis of human-nature feedbacks. Ecol Soc 20(3). https://doi.org/10.5751/es-07404-200317

Kampichler C, Wieland R, Calme S, Weissenberger H, Arriaga-Weiss S (2010) Classification in conservation biology: a comparison of five machine-learning methods. Ecol Inform 5 (6):441–450. https://doi.org/10.1016/j.ecoinf.2010.06.003

Kandziora M, Burkhard B, Muller F (2013) Mapping provisioning ecosystem services at the local scale using data of varying spatial and temporal resolution. Ecosyst Serv 4:47–59. https://doi.org/10.1016/j.ecoser.2013.04.001

Kareiva P (1994) Space: the final frontier for ecological theory. Ecology 75(1):1. https://doi.org/10.2307/1939376

Keitt TH, Urban DL (2005) Scale-specific inference using wavelets. Ecology 86(9):2497–2504. https://doi.org/10.1890/04-1016

Kelling S, Hochachka WM, Fink D, Riedewald M, Caruana R, Ballard G, Hooker G (2009) Data-intensive science: a new paradigm for biodiversity studies. Bioscience 59(7):613–620. https://doi.org/10.1525/bio.2009.59.7.12

Kerr JT, Ostrovsky M (2003) From space to species: ecological applications for remote sensing. Trends Ecol Evol 18(6):299–305. https://doi.org/10.1016/s0169-5347(03)00071-5

Kukkala AS, Moilanen A (2013) Core concepts of spatial prioritisation in systematic conservation planning. Biol Rev 88(2):443–464. https://doi.org/10.1111/brv.12008

Lawler JJ, Ruesch AS, Olden JD, McRae BH (2013) Projected climate-driven faunal movement routes. Ecol Lett 16(8):1014–1022. https://doi.org/10.1111/ele.12132

Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? Ecology 74(6):1659–1673

Legrand D, Cote J, Fronhofer EA, Holt RD, Ronce O, Schtickzelle N, Travis JMJ, Clobert J (2017) Eco-evolutionary dynamics in fragmented landscapes. Ecography 40(1):9–25. https://doi.org/10.1111/ecog.02537

Leibold MA, Chase JM (2017) Metacommunity ecology. Princeton University Press, Princeton, NJ

Levin SA (1992) The problem of pattern and scale in ecology. Ecology 73(6):1943–1967. https://doi.org/10.2307/1941447

Lindenmayer DB, Margules CR, Botkin DB (2000) Indicators of biodiversity for ecologically sustainable forest management. Conserv Biol 14(4):941–950. https://doi.org/10.1046/j.1523-1739.2000.98533.x

Liu JG, Dietz T, Carpenter SR, Alberti M, Folke C, Moran E, Pell AN, Deadman P, Kratz T, Lubchenco J, Ostrom E, Ouyang Z, Provencher W, Redman CL, Schneider SH, Taylor WW (2007) Complexity of coupled human and natural systems. Science 317(5844):1513–1516. https://doi.org/10.1126/science.1144004

Liu JG, Hull V, Luo JY, Yang W, Liu W, Vina A, Vogt C, Xu ZC, Yang HB, Zhang JD, An L, Chen XD, Li SX, Ouyang ZY, Xu WH, Zhang HM (2015) Multiple telecouplings and their complex interrelationships. Ecol Soc 20(3). https://doi.org/10.5751/es-07868-200344

Logue JB, Mouquet N, Peter H, Hillebrand H, Metacommunity Working Group (2011) Empirical approaches to metacommunities: a review and comparison with theory. Trends Ecol Evol 26 (9):482–491. https://doi.org/10.1016/j.tree.2011.04.009

Loiselle BA, Howell CA, Graham CH, Goerck JM, Brooks T, Smith KG, Williams PH (2003) Avoiding pitfalls of using species distribution models in conservation planning. Conserv Biol 17 (6):1591–1600. https://doi.org/10.1111/j.1523-1739.2003.00233.x

Loreau M, Holt RD (2004) Spatial flows and the regulation of ecosystems. Am Nat 163 (4):606–615. https://doi.org/10.1086/382600

Loreau M, Mouquet N, Holt RD (2003) Meta-ecosystems: a theoretical framework for a spatial ecosystem ecology. Ecol Lett 6(8):673–679. https://doi.org/10.1046/j.1461-0248.2003.00483.x

Maes J, Egoh B, Willemen L, Liquete C, Vihervaara P, Schagner JP, Grizzetti B, Drakou EG, La Notte A, Zulian G, Bouraoui F, Paracchini ML, Braat L, Bidoglio G (2012) Mapping ecosystem services for policy support and decision making in the European Union. Ecosyst Serv 1 (1):31–39. https://doi.org/10.1016/j.ecoser.2012.06.004

Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405(6783):243–253. https://doi.org/10.1038/35012251

Martensen AC, Saura S, Fortin MJ (2017) Spatio-temporal connectivity: assessing the amount of reachable habitat in dynamic landscapes. Methods Ecol Evol 8(10):1253–1264. https://doi.org/10.1111/2041-210x.12799

Massol F, Gravel D, Mouquet N, Cadotte MW, Fukami T, Leibold MA (2011) Linking community and ecosystem dynamics through spatial ecology. Ecol Lett 14(3):313–323. https://doi.org/10.1111/j.1461-0248.2011.01588.x

Mayor SJ, Schneider DC, Schaefer JA, Mahoney SP (2009) Habitat selection at multiple scales. Ecoscience 16(2):238–247. https://doi.org/10.2980/16-2-3238

McGarigal K, Wan HY, Zeller KA, Timm BC, Cushman SA (2016) Multi-scale habitat selection modeling: a review and outlook. Landsc Ecol 31(6):1161–1175

Mennill DJ, Battiston M, Wilson DR, Foote JR, Doucet SM (2012) Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour. Methods Ecol Evol 3(4):704–712. https://doi.org/10.1111/j.2041-210X.2012.00209.x

Merchant ND, Fristrup KM, Johnson MP, Tyack PL, Witt MJ, Blondel P, Parks SE (2015) Measuring acoustic habitats. Methods Ecol Evol 6(3):257–265. https://doi.org/10.1111/2041-210x.12330

Moilanen A (2007) Landscape Zonation, benefit functions and target-based planning: unifying reserve selection strategies. Biol Conserv 134(4):571–579. https://doi.org/10.1016/j.biocon.2006.09.008

Moilanen A, Wilson KA, Possingham H (eds) (2009) Spatial conservation prioritization: quantitative methods and computational tools. Oxford University Press, Oxford

Muenchow J, Schratz P, Brenning A (2017) RQGIS: integrating R with QGIS for Statistical Geocomputing. R Journal 9(2):409–428

Naidoo R, Balmford A, Costanza R, Fisher B, Green RE, Lehner B, Malcolm TR, Ricketts TH (2008) Global mapping of ecosystem services and conservation priorities. Proc Natl Acad Sci U S A 105(28):9495–9500. https://doi.org/10.1073/pnas.0707823105

Nichols JD, Bailey LL, O'Connell AF, Talancy NW, Grant EHC, Gilbert AT, Annand EM, Husband TP, Hines JE (2008) Multi-scale occupancy estimation and modelling using multiple detection methods. J Appl Ecol 45(5):1321–1329. https://doi.org/10.1111/j.1365-2664.2008.01509.x

Nichols JD, Koneff MD, Heglund PJ, Knutson MG, Seamans ME, Lyons JE, Morton JM, Jones MT, Boomer GS, Williams BK (2011) Climate change, uncertainty, and natural resource management. J Wildl Manag 75(1):6–18. https://doi.org/10.1002/jwmg.33

Olden JD, Lawler JJ, Poff NL (2008) Machine learning methods without tears: a primer for ecologists. Q Rev Biol 83(2):171–193. https://doi.org/10.1086/587826

Ovaskainen O, De Knegt HJ, del Mar Delgado M (2016) Quantitative ecology and evolutionary biology: integrating models with data. Oxford University Press, Oxford

Pascual-Hortal L, Saura S (2006) Comparison and development of new graph-based landscape connectivity indices: towards the priorization of habitat patches and corridors for conservation. Landsc Ecol 21(7):959–967. https://doi.org/10.1007/s10980-006-0013-z

Perfecto I, Vandermeer J (2010) The agroecological matrix as alternative to the land-sparing/agriculture intensification model. Proc Natl Acad Sci U S A 107(13):5786–5791. https://doi.org/10.1073/pnas.0905455107

Petchey OL, Pontarp M, Massie TM, Kefi S, Ozgul A, Weilenmann M, Palamara GM, Altermatt F, Matthews B, Levine JM, Childs DZ, McGill BJ, Schaepman ME, Schmid B, Spaak P, Beckerman AP, Pennekamp F, Pearse IS (2015) The ecological forecast horizon, and examples of its uses and determinants. Ecol Lett 18(7):597–611. https://doi.org/10.1111/ele.12443

Peters DPC, Bestelmeyer BT, Turner MG (2007) Cross-scale interactions and changing pattern-process relationships: consequences for system dynamics. Ecosystems 10(5):790–796. https://doi.org/10.1007/s10021-007-9055-6

Peterson AT, Soberon J, Pearson RG, Anderson RP, Martinez-Mery E, Nakamura M, Araújo MB (2011) Ecological niches and geographic distributions. Princeton University Press, Princeton, NJ

Power AG (2010) Ecosystem services and agriculture: tradeoffs and synergies. Phil Trans R Soc B 365(1554):2959–2971. https://doi.org/10.1098/rstb.2010.0143

Pressey RL, Cabeza M, Watts ME, Cowling RM, Wilson KA (2007) Conservation planning in a changing world. Trends Ecol Evol 22(11):583–592. https://doi.org/10.1016/j.tree.2007.10.001

Pulliam HR (2000) On the relationship between niche and distribution. Ecol Lett 3(4):349–361

Pulliam HR, Dunning JB, Liu JG (1992) Population dynamics in complex landscapes: a case study. Ecol Appl 2(2):165–177. https://doi.org/10.2307/1941773

Regan HM, Colyvan M, Burgman MA (2002) A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecol Appl 12(2):618–628. https://doi.org/10.2307/3060967

Rondinini C, Wilson KA, Boitani L, Grantham H, Possingham HP (2006) Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecol Lett 9(10):1136–1145. https://doi.org/10.1111/j.1461-0248.2006.00970.x

Rosenberg MS, Anderson CD (2011) PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. Methods Ecol Evol 2(3):229–232. https://doi.org/10.1111/j.2041-210X.2010.00081.x

Rudnick DA, Ryan SJ, Beier P, Cushman SA, Dieffenbach F, Epps CW, Gerber LR, Hartter J, Jenness JS, Kintsch J, Merenlender AM, Perkl RM, Preziosi DV, Trombulak SC (2012) The role of landscape connectivity in planning and implementing conservation and restoration priorities. Issues Ecol, Report Number16, Ecological Society of America, Washington, DC.

Schagner JP, Brander L, Maes J, Hartje V (2013) Mapping ecosystem services' values: current practice and future prospects. Ecosyst Serv 4:33–46. https://doi.org/10.1016/j.ecoser.2013.02.003

Schoener TW (2011) The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. Science 331(6016):426–429. https://doi.org/10.1126/science.1193954

Schumaker NH, Brookes A, Dunk JR, Woodbridge B, Heinrichs JA, Lawler JJ, Carroll C, LaPlante D (2014) Mapping sources, sinks, and connectivity using a simulation model of northern spotted owls. Landsc Ecol 29(4):579–592. https://doi.org/10.1007/s10980-014-0004-4

Sklar E (2007) Software review: NetLogo, a multi-agent simulation environment. Artif Life 13(3):303–311. https://doi.org/10.1162/artl.2007.13.3.303

Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. Ecol Lett 10(12):1115–1123. https://doi.org/10.1111/j.1461-0248.2007.01107.x

Soberón JM (2010) Niche and area of distribution modeling: a population ecology perspective. Ecography 33(1):159–167. https://doi.org/10.1111/j.1600-0587.2009.06074.x

Talluto MV, Boulangeat I, Ameztegui A, Aubin I, Berteaux D, Butler A, Doyon F, Drever CR, Fortin MJ, Franceschini T, Lienard J, McKenney D, Solarik KA, Strigul N, Thuiller W, Gravel D (2016) Cross-scale integration of knowledge for predicting species ranges: a metamodelling framework. Glob Ecol Biogeogr 25(2):238–249. https://doi.org/10.1111/geb.12395

Thiele JC, Grimm V (2010) NetLogo meets R: linking agent-based models with a toolbox for their analysis. Environ Model Softw 25(8):972–974. https://doi.org/10.1016/j.envsoft.2010.02.008

Thiele JC, Kurth W, Grimm V (2012) RNetLogo: an R package for running and exploring individual-based models implemented in NetLogo. Methods Ecol Evol 3(3):480–483. https://doi.org/10.1111/j.2041-210X.2011.00180.x

Thorson JT, Ianelli JN, Larsen EA, Ries L, Scheuerell MD, Szuwalski C, Zipkin EF (2016) Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. Glob Ecol Biogeogr 25(9):1144–1158. https://doi.org/10.1111/geb.12464

Tomkiewicz SM, Fuller MR, Kie JG, Bates KK (2010) Global positioning system and associated technologies in animal behaviour and ecological research. Phil Trans R Soc B 365 (1550):2163–2176. https://doi.org/10.1098/rstb.2010.0090

Turner W, Spector S, Gardiner N, Fladeland M, Sterling E, Steininger M (2003) Remote sensing for biodiversity science and conservation. Trends Ecol Evol 18(6):306–314. https://doi.org/10.1016/s0169-5347(03)00070-3

Turner BL, Lambin EF, Reenberg A (2007) The emergence of land change science for global environmental change and sustainability. Proc Natl Acad Sci U S A 104(52):20666–20671. https://doi.org/10.1073/pnas.0704119104

Verhagen W, Kukkala AS, Moilanen A, van Teeffelen AJA, Verburg PH (2017) Use of demand for and spatial flow of ecosystem services to identify priority areas. Conserv Biol 31(4):860–871. https://doi.org/10.1111/cobi.12872

Watts ME, Ball IR, Stewart RS, Klein CJ, Wilson K, Steinback C, Lourival R, Kircher L, Possingham HP (2009) Marxan with zones: software for optimal conservation based land- and sea-use zoning. Environ Model Softw 24(12):1513–1521. https://doi.org/10.1016/j.envsoft.2009.06.005

Wiegand T, Moloney K (2014) Handbook of spatial point-pattern analysis in ecology. Chapman & Hall, CRC Applied Environmental Statistics, Boca Raton, FL

Wilson RJ, Thomas CD, Fox R, Roy DB, Kunin WE (2004) Spatial patterns in species distributions reveal biodiversity change. Nature 432(7015):393–396. https://doi.org/10.1038/nature03031

Wisz MS, Pottier J, Kissling WD, Pellissier L, Lenoir J, Damgaard CF, Dormann CF, Forchhammer MC, Grytnes JA, Guisan A, Heikkinen RK, Hoye TT, Kuhn I, Luoto M, Maiorano L, Nilsson MC, Normand S, Ockinger E, Schmidt NM, Termansen M, Timmermann A, Wardle DA, Aastrup P, Svenning JC (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biol Rev 88 (1):15–30. https://doi.org/10.1111/j.1469-185X.2012.00235.x

Wittemyer G, Keating LM, Vollrath F, Douglas-Hamilton I (2017) Graph theory illustrates spatial and temporal features that structure elephant rest locations and reflect risk perception. Ecography 40(5):598–605. https://doi.org/10.1111/ecog.02379

Zeigler SL, Fagan WF (2014) Transient windows for connectivity in a changing world. Movement Ecol 2(1):1. https://doi.org/10.1186/2051-3933-2-1

Appendix A: An Introduction to R

Introduction

In this book, we use R to illustrate several important concepts and analyses in spatial ecology and conservation. In the context of spatial ecology and conservation, R is an impressive, do-it-all software, where we can seamlessly go back and forth from spatial data to models, mapping, and analysis, all while visualizing the data and results that we find. For spatial analysis relevant to spatial ecology and conservation, R is particularly strong, both in terms of depth and breadth, which makes it a very useful software for the topic of this book.

R is a high-level, command-driven language. While there is sometimes a steep learning curve to using R, there are several benefits of using R. First and foremost, the use of commands (rather than GUI/menu-driven approaches) allows for "scripts" to be written, which facilitates reproducible science. This benefit cannot be overstated. The ability to go back and repeat data summaries and analyses is incredibly important in science, and is one important element of reproducible science (Casadevall and Fang 2010; Munafo et al. 2017). It may also make your life easier if you notice an error in data or realize that you need to add some more data to your analysis. Second, by using scripts, we can potentially automate mundane tasks that we have to repeatedly do. This benefit helps free some of our time to focus on more interesting and important issues. Third, R is an open software, meaning that users, like yourself, can contribute to its development. Finally, R is free. This is perhaps our students' favorite reason for using R.

Here, we provide a brief overview on the use of R, particularly focused on key issues that come up throughout the book. While there are several detailed and helpful books on the use of R in general, and two specifically on spatial data in R (Bivand et al. 2013; Brunsdon and Comber 2015), we intend this Appendix to provide an introduction that will allow readers to understand the main chapters of our book without having to go to other books for interpreting R-specific details of programming that are presented in the main chapters. To illustrate the use of R, we use the data described in Chap. 6.

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1

R Beginnings: Before Any Analysis

We assume that many readers of this book may be new users of R. For instance, in graduate courses at the University of Florida, we find that about ⅓ to ½ of students have not used R or do not feel comfortable using R when they begin the course on this topic. Beginning R can be a daunting task, but one that is ultimately worthwhile.

Often new users of R jump right into data analysis and modeling without having a solid foundation of how R works. This can lead to frustration. Here we provide some key basics that will hopefully limit those frustrations so that users can focus on the more interesting and fun topics of spatial ecology and conservation.

R Packages

Much of the heavy lifting done in R is through various R packages (or libraries). We use a variety of R packages throughout the book, but we will really just scratch the surface of R packages. There are literally thousands of R packages. R comes with several packages installed (e.g., `MASS`) and enabled (e.g., `stats`, `graphics`), but most of the ones we use we in the book will need to be downloaded from the CRAN website manually and loaded for use. For R packages relevant to spatial data analysis, see: http://cran.r-project.org/web/views/Spatial.html. It is important to keep in mind that while many packages, particularly those that come installed with R, have been well vetted, users can download and use packages that have not been well vetted. Consequently, when you first open R, you will see on the console, "R is free software and comes with ABSOLUTELY NO WARRANTY". Through the years the R base system version changed and it now at version 3.4.3 at the time of this writing; some R packages were developed for this version, while others were developed for previous versions of R.

Editors for R

To facilitate writing R scripts an advanced editor can help. Editors allow you to more easily interpret code than using the script prompt in R, because different commands and statements will be coded in different colors and fonts. While this may seem like not a big deal, in practice this can be an enormous help.

R comes with an editor, but it is very minimalistic. For Mac users, many find that the editor that comes with R is sufficient, but Windows users typically like to download other editors rather than using the editor provided in the R download. Some common editors people use include: RStudio, Tinn-R, Notepad++, Vim (and Vim-R), and Emacs Speaks Statistics. RStudio has quickly become perhaps the most common editor for running R. See http://rstudio.org/. It is generally easy to use and

has some really nice functionality. This is what we typically use for teaching. It has several advanced features that make it helpful. Prior to the development of RStudio, Tinn-R was a popular editor for R. Notepad++ (with `NppToR` package for sending code to R) is a general editor (not specific to R), but it is very intuitive and easy to use. Finally, Emacs Speaks Statistics (ESS; http://ess.r-project.org/) is an editor that works well on Macs.

The R Prompt, Console, and Editor

Once you have downloaded and installed R (and perhaps an editor for R, such as RStudio), you can fire it up. We can work in R in two general ways. First, we can work directly on the command line, which looks like:

```
>
```

On the line we can type a variety of code and hit enter. For example, you type:

```
> 3 + 7
```

You will get:

```
##
[1] 10
```

So R can be used as a calculator. But that is not its main power. While you can type directly on the command line, most of the work you will do will be writing code on the Editor and then passing parts (or all) of the code to the console to run analyses.

In the code we provide, you will often see statements that start with *#*. These statements are comments regarding the code in R. R ignores them. For example, if you write.

```
#R is a calculator
> 3 + 7
```

And hit enter, R ignores everything after *#* and provides:

```
##
[1] 10
```

We can also make and store new variables or summary statistics from analyses:

```
> d <- 3 +7
```

or

```
> d = 3 + 7
```

The = sign may be used in place of the <- operator for the purpose of assignment, but there are some situations where the former does not work for assignment, whereas the latter always does. In this book, we use <- throughout. Note that you must then type d and hit enter to retrieve the result:

```
> d

##
[1] 10
```

If you see a "+" symbol on the R prompt, it indicates that R expects more code. If you want to escape this line and start over, simply hit the Esc button on your computer. *Importantly, R is case-sensitive, so "Ecology" and "ecology" are NOT considered the same!*

Getting Help in R

One common criticism of R is that the documentation is not as useful as some other software programs. However, once you understand the way in which R summarizes help information, it can be quite useful. For any function, you can retrieve some useful information on various functions with the following code, which focuses on using the lm function (linear model function) in R:

```
> ?lm
```

Or equivalently:

```
> help(lm)
```

Both commands provide information on linear models in R. Note that this command takes you to a standardized window, which provides details on the various options for the function, shows you what the default options are (under Usage), what the function returns (e.g., the results of a linear regression), and it provides examples of its implementation. Examples are designed such that you can copy the code to your console and they should work without any additional code or information needed. This standardized output can be very helpful. You can also get help on packages by either typing the specific function in the package (as above), or asking for general help:

```
> help(package="MASS")
```

The above command will provide you with information on the `MASS`package, which comes with the R program. Many packages come with vignettes that provide examples on the use of the package. There are also a variety of message boards and list-serves online that provide a means to have questions answered by other R users. These forums have an impressively large number of examples, questions, and answers that can be found by using search terms in search engines on the Internet. One important list-serve for spatial analysis is the R-sig-geo mailing list (https://stat.ethz.ch/mailman/listinfo/r-sig-geo).

R Classes

Understanding how R works requires understanding R classes. There are several types of R classes—these are structures used for storing and handling data. You need to understand these to fully understand R and leverage its capabilities. There are general types of classes used in R as well classes specific to spatial data (see below) and objects that can be created from various analyses done in R. Common R classes include: variables, vectors, matrices, arrays, data frames, and lists.

We have already seen the use of variables above. For instance, we can create a variable, *x*, and give it a value of 4:

```
> x <- 4
```

Vectors of data are commonly used. A vector is typically a string of numbers (think a column of data in a spreadsheet). We can create a vector in a variety of ways:

```
> x.vector <- vector(mode = "numeric", length = 4)
> x.vector <- 1:10
> x.vector <- c(0,4,0,1)
```

In the first line above, we explicitly name a vector and define it as "numeric," which means the vector can hold numeric data that are real numbers (positive or negative, with decimals). The second line creates a string from 1 to 10 (in integers). The third example uses the `c` function (or concatenate function) to string together the four numbers that follow `c`.

We can access information from a vector in a variety of ways. If we simply type the vector name, the entire vector will appear. Alternatively, we can call elements of the vector using a variety of commands. Below, we first view the entire vector, then grab elements 1–3, and finally access all elements that are >1.

```
#accessing data from a vector
> x.vector
> x.vector[2]
> x.vector[1:3]
> x.vector[x.vector > 1]
```

We can expand on the vector concept by creating matrices and arrays. A matrix is essentially a vector in 2-dimesions, typically thought of as rows and columns in a spreadsheet. Arrays extend matrices in >2 dimensions. These are a bit harder to visualize but can be very helpful in some applications, such as if we want to store several matrices of the same dimensions. It is important to note that matrices and arrays must have the same type of data in each dimension. Below, we illustrate making a square, 5×5 matrix, and subsequently making an array, that is essentially composed of 2, 5×5 matrices.

```
#making a matrix
> x.matrix <- matrix(0, nrow = 5, ncol = 5)

#making an array
> x.array <- array(0, dim = c(5, 5, 2))
```

Similar to vectors, we can access elements of matrices in a variety of ways. Note that in this case, we use single brackets with two elements, where the first element represents the rows in the matrix and the second element represents the columns.

```
#accessing data from a matrix
> x.matrix
> x.matrix[2,2] <- 5
> x.matrix[,2] <- 2
> x.matrix[3,] <- 3
> diag(x.matrix)<- 1
```

Data frames are commonly used. The constraint mentioned above regarding matrices and arrays requiring the same type of data is relaxed with data frames. Data frames can be thought of as a two-dimensional matrix where different columns of data can have different data types. For instance, one column could have characters (e.g., names of sites), while another column might include numeric values. Data frames are often the type of R class used when going from a spreadsheet to R. We can populate new data frames in a variety of ways. Below is an example with three columns of data, one with character information, a second with numeric information, and the third we population with NA (no data):

```
> x.df <- data.frame(site = c("North", "South", "Mid"), x = 1:3, y = NA)
```

Finally, lists are very handy and are commonly used to store R output. A list is a set of any of the above classes. For instance, one component of a list could be a vector, another an array and even another a data frame. This flexibility is why so much output in R is stored as a list. Here is an example where we create a list of a vector and a matrix.

```
#making a list that is a combination of data objects above
> x.list <- list(x.vector, x.matrix)
```

To access information elements from lists, we use a double-bracket notation for the main elements (rather than single brackets as used for vectors, matrices, and arrays). Single brackets can be used to access lower-level information. For instance, in this example we can use double brackets to view the vector and matrix, but combine double bracket and single-bracket notation to access specific elements of the matrix.

```
> x.list[[1]]
> x.list[[2]]
> x.list[[2]][1, 2]
```

Getting Data Into and Out of R

If we save spreadsheets or database files as a .txt or .csv file type, we can then import the .txt file into R. We will import two data files, which are used in Chap. 6 on spatial prediction. The first is a data file that includes point-count data from Montana as part of a landbird monitoring program. In this monitoring program, transects were placed across the region, with 10 point counts on each transect, placed approximately 300 m apart. It includes presence–absence data (or detection, non-detection data) on one species, the varied thrush. The second file includes potential covariates that might be of interest, in terms of explaining thrush distribution.

Before we import these data, it is useful to set the working directory. The working directory is a folder on your computer where R will default to looking at when importing data and it will also be the default location for exporting or saving data. This can be done in at least two different ways. First, we can use the `setwd` function:

```
> setwd('C:/Desktop/') #your working directory may be different!
  #Or equivalently:
> setwd('C:\\Desktop\\')
```

If you use the above commands, make sure that the path is correct! One way to check is to open up Windows Explorer and find where the folder is. You can right click on the folder and select "Properties." In the pop-up window on the General tab, you will find the path after "Location." The alternative way to ensure the directory is correct is to use drop-down menus. In *R*: use the menu options (File>Change dir). In *RStudio*, use the menu options to do the same (>Session>Set Working Directory>Choose Directory).

Once we have set the working directory, we can import the data files using the `read.table` or `read.csv` functions:

```
> landbird <- read.csv("vath_2004.csv", header = T)
```

The `header = T` statement (`T` is for "True") tells R that your file has a header line. *Note that header labels should not include spaces* (if they do, R will alter the names). The `read.table` function defaults to storing the data as a data frame, which can be manipulated to some degree (e.g., you can add columns of data). You can then call columns (vectors) of data with the header label (more on this below). Also note that in the above code, we did not specify a path/directory for where the data are stored, so R will look in the working directory. The same file could be accessed by providing the entire path:

```
> landbird <- read.csv("C:/ Spatial Ecology Book/Data/vath_2004.csv",
  header = T)
```

If you are having problems with the `read.table` statement, it is likely because the working directory is not set to the correct folder. Consider opening the file in Windows Explorer (or Mac equivalent) and copying-and-pasting the directory path to reduce the likelihood of typing errors. Another way to accomplish this goal is to use `file.choose()`:

```
> landbird <- read.table(file.choose(), header = T)
```

The `file.choose` statement will let you use point-and-click windows methods for finding and importing your file. Here, we also provide the covariate data as a .csv:

```
> landbird_cov <-read.csv("vath_covariates.csv", header = T)
```

We can also export data from R in a straightforward way. For example, there are several write functions (`write.table`, `write.csv`, `writeRaster`, etc.). To export the array described above we can use the `write.table` or `write.csv` function:

```
> write.table(x.array, file = "array.txt", sep = " ")
> write.csv(x.array, file = "array.csv")
```

Finally, you can save all of your work in R. This is accomplished by:

```
> save.image("Appendix_results.RData")
```

Note that you need to specify the extension `.RData`. You can then reload this information later, which can be very helpful when restarting an R session (so that analyses and models will not need to be rerun).

```
> load("Appendix_results.RData")
```

This is particularly useful when you have done analyses that take a considerable amount of time. It also provides a means to have an organized workspace that you can share with others.

Functions in R

Much of what you will work on in R involves the use of functions. Commands such as `load`, `write.table`, and `data.frame` are all functions that perform one or more series of operations.

One very important and useful component of R is the ability to write your own functions. The primary reason why you might want to write your own functions is to help automate some sort of operation(s). For example, we could make a function that calculates the mean of a series of numbers (a vector):

```
> mean.function <- function(x){
`N <- length(x)
`mean.x <- sum(x)/N
`return(mean.x)
}
> mean.function(x.vector)
```

So, how does this work? The function created requires that we pass an R class (`x`). Note that we are not formally declaring in this case whether `x` is a scalar (single number), a vector, or a matrix, so we have to be careful in this way. Then, within the curly brackets we first determine the number of records using the `length` function. With that we can then calculate the mean by summing all of the values and dividing by `N`. `return` tells R what it should provide when the function is run. If we have not included return, R would provide the last calculation (in this case, it would be fine).

If you write functions, a common way to make them available in any r script is through the use of the `source` command. To do so, you would want to have a file that contains the relevant functions that you might want to use, say it is called `myfunctions.R`. Then in any R script, you can call the functions by stating:

```
> source("myfunctions.R")
```

In this example, `myfunctions.R` is stored in the working directory. If your functions are not stored in the working directory, the entire path should be explicitly stated.

Data Access, Management, and Manipulation in R

Accessing Data

We can access data stored in R objects in a variety of ways, particularly for data frames. To see the headers of a data frame, simply type:

```
> names(landbird)
```

Other useful functions for getting a feel for your data include:

```
> head(landbird)
> tail(landbird, 3)
```

The `str` function is particularly useful. It shows how the R object is structured.

```
> str(landbird)
```

Now we can immediately do some simple summaries, by calling the data we imported. For example, if you want to access a vector in the data set, you can use:

```
#three ways to access the first column of data
> landbird[ , 1]
> landbird [,"SURVEYID"]
> landbird$SURVEYID

#accessing row 1 of data
> landbird[1, ]
```

The above commands provide us with a means to grab different parts of data in a streamlined way. If you want to know the numbers of columns and rows of your data (the dimensions), type:

```
> dim(landbird)
```

Or just the rows or columns:

```
> nrow(landbird)
> ncol(landbird)
```

If you want to know the unique values observed for a given variable, such as the values of `POINT`, type:

```
> unique(landbird$POINT)
> levels(landbird$POINT)
```

Note that `levels` is useful for factor variables, while `unique` is applicable to any variable.

Merging, Appending, and Removing

A common issue in data management is combining two different data sets that have the same unique identifier. For example, let us say you have data on bird presence–absence in one spreadsheet and you have another spreadsheet that has covariate (e.g., vegetation) information that was measured at these locations. The `merge` function is helpful for combining data sets:

```
> landbird < -merge(landbird, landbird_cov, by = "SURVEYID", all=T)
```

Above we over-wrote the landbird data frame to add the covariates to the data frame. Now take a look at this new data frame, by using `str(landbird)`. What do you see? You can also merge based on >1 variable with the `c` function. For instance, if we had different years of data, we could use a statement within the `merge` function like `by = c("SURVEYID", "YEAR")`.

You can append data by simply using the `rbind` function. This function combines rows of data (`cbind` combines columns of data). Note that the data you append must have the same format (including the same headers when combining data frames) as the existing data.

Data Subsetting and Summaries

Often, we may want to acquire subsets of data. In R, there are a few ways to subset data. If we want a subset of the data, such as the points where varied thrush were detected, we could type:

```
> landbird.pres <- landbird[landbird$VATH == 1, ]

#or with subset:
> landbird.pres <- subset(landbird, VATH == 1)

#subset based on two variables
> landbird.pres.mesic <- subset(landbird, VATH == 1 & Mesic == 1)
```

Summary statistics can be calculated using several approaches. You can also make simple calculations easily on the vectors of data. For example, you can get summary statistics, such as the mean value for elevation (`Elev`), the standard deviation, min, or a specific quantile:

```
> mean(landbird$Elev)
> sd(landbird$Elev)
> min(landbird$Elev)
> quantile(landbird$Elev, probs = 0.05)#lower 5% quantile
```

Let us say you want to calculate one of these functions (e.g., the mean) on a subset of the data. Simply type:

```
> tapply(landbird$Elev, landbird$VATH, mean)
```

Or equivalently,

```
> tapply(X= landbird$Elev, INDEX = landbird$VATH, FUN = mean)
  #two variables:
> tapply(landbird$Elev, list(landbird$VATH, landbird$Mesic), FUN
  = mean)
```

The above calculates the mean elevation value for locations where varied thrush were detected and locations where it was not detected (it calculates the function on the first variable based on the levels of the second variable). If we wanted to simultaneously calculate summary statistics for many variables, we could use the `sapply` or `lapply` functions, which calculate the means for multiple variables for *all* the data (not a subset, like `tapply`), or the `summary` function:

```
> summary(landbird)
```

Three commonly used packages for manipulating and summarizing data are the `reshape2`, `plyr`, and `dplyr` packages, created by Hadley Wickham (2007). The `plyr` package provides a straightforward way to do summaries of data, such as means and standard deviations, similar in some respects to the `tapply` function, through the use of the `ddply` function:

```
#summarize by VATH presence/absence
> ddply(landbird, .(VATH), summarize,
`elev.mean = mean(Elev),
`elev.sd=sd(Elev))
```

This function returns a data frame that summarizes the data. In this case, it will provide the mean and standard deviation of elevation for locations where varied thrush were present and absent. Similarly, we could get means as a function of two variables by altering the first line of data:

```
> ddply(landbird, .(VATH, Mesic), summarize,
`elev.mean = mean(Elev),
`elev.sd = sd(Elev))
```

The dplyr package has largely replaced the plyr package, because it has more functionality. It requires a bit more background for its use, but once one becomes familiar with the syntax it is very powerful. It is beyond the scope to cover the dplyr package, but the above approach with ddply can be re-created in the dplyr package as:

```
> landbird %>% #pipe
`group_by(VATH, Mesic) %>% #categories for grouping
`summarise(elev.mean= mean(Elev),
`elev.sd=sd(Elev))
```

The use of dplyr partially rests on the idea of making an all-purpose language for data manipulation. In this case, the use of %>% is a "pipe" operator, where we are calling a data frame (landbird) and then can make other calls to functions in the code; that is, it allows for the output of one function to be the input for another. group_by provides a means to then get summaries of the data by the categories mention in the group.

Reformatting Data

There are several ways to create new variables, such as changing continuous data to binary (0/1) data or to a set of categories. If you want to convert continuous data into binary data (i.e., 0/1 data) or categorical data, we can do this in at least two ways. For this example, we can convert the continuous value of elevation to a new factor variable, of low and high elevation, based on the median value.

```
> elev.median <- median(landbird$Elev)

> landbird$Elev_cat <- "low"
> landbird$Elev_cat[landbird$Elev > elev.median] <- "high"

#or
> landbird$Elev_cat2 <- as.factor(ifelse(landbird$Elev > elev.median,
  "high", "low"))
```

If we take a look at the structure of the data again (using str(landbird)), we find the new columns, but that R still considers the first variable we created to be a character, rather than a factor. To change the structure to categorical:

```
> landbird$Elev_cat <- factor(landbird$Elev_cat)
```

Note that if >2 categories need to be created, the cut function can accommodate this task.

a

	Species		
Location	2014	2015	2016
A	36	35	38
B	23	18	17
C	44	48	46

b

Location	Year	Species
A	2014	36
A	2015	35
A	2016	38
B	2014	23
B	2015	18
…	…	…

Fig. A.1 (**a**) Wide versus (**b**) long versus formats for hypothetical data on species richness at different sampling locations (A, B, C) over time

The `reshape2`package focuses on the task of restructuring data, or "reshaping" data. Oftentimes, data can generally be considered to be in "wide" or "long" formats (Fig. A.1). The distinction is clear when working with data on repeated measures, such as repeated surveys at a location over time. In a wide format, repeated samples for a given location will be represented by additional columns of data, such as each year of sampling, and only one row occurs for each sample location. In contrast, in a long format, there may be only one column, "Year," and each sampling location may have several rows of data, one for each year (Fig. A.1). The `reshape2` package makes it straightforward to go from wide to long format using the `melt` function, and going from long to wide format using the `cast` function.

As an example, we take the landbird data, which has each point (POINT) within each transect in a long format, and convert it to a wide format using the `dcast` function.

```
> transect.vath <- dcast(landbird, TRANSECT ~ POINT, value.var =
  "VATH")
```

In this case, `dcast` creates a new data frame where each row is a transect and the columns are the point counts (1–10) along each transect, with the values in each column being the presence or absence of the varied thrush. Note that in the above formula, variables on the left side of the tilde (~; TRANSECT) are identification variables to keep, whereas variables on the right side of the tilde are variables that will be swung into wide format. This type of restructuring of data can be very helpful if interest is in creating "detection histories" used for mark–recapture or occupancy modeling or creating site by species matrices for community data. If we were starting with wide format data and would like to convert it to long format (which is the most common format used for statistical modeling in R), we would use the `melt` function. Here, we can melt the wide-formatted data we just created:

```
> transect.long.vath <- melt(transect.vath, id.vars = "TRANSECT",
  variable.name = "POINT", value.name = "VATH")
```

Note that this new data frame is longer than the original data frame because it includes NAs for transects where <10 points occurred. We can remove those lines of data, if we wish using a variety of approaches, such as the `complete.cases` function:

```
> transect.long.vath <-transect.long.vath[complete.cases(transect.
  long.vath[,3]),]
```

Finally, another common approach to summarizing and reformatting data is using the `apply` function. The `apply` function takes matrices or data frames and applies a function (either an existing function or a custom function, like `stderr` described above) to either the rows or the columns of data. For instance, we can use apply to sum up the detections by row for the wide-format data just created by:

```
> apply(transect.vath[,2:11], 1, sum, na.rm = T)
```

For the `apply` function, the first statement reflects the data of interest, the second statement reflects whether the function is applied to rows (1) or columns (2), the third statement provides the function (here, `sum`), and the last simply states to ignore NA values. In this case, this argument would be equivalent to the `rowSums` function:

```
> rowSums(transect.vath[,2:11], na.rm=T)
```

While the `rowSums` (and `colSums`) function works well in this regard, the power of the `apply` function is that we could just as easily use a customized function to do row or column operations. In general, using functions like `apply` that focus on vectors tend to be much computationally faster than using for loops and related functions that pass through rows or columns one row/column at a time.

Graphics in R

R is very well known for its excellent graphics. Simply put, you can make some of the best graphics producible in R, if given the time and motivation (i.e., it can make fantastic graphics, but it is not always easy to do so!). In this book, we will typically not show the code we used for making graphics for figures, but nearly all of the graphics were made using R (either with the base `graphics` package or using the `ggplot2`package). Below we provide some details regarding the interpretation of viewing figures as shown in the chapters of the book.

The main function for graphing is `plot`. It can do quite a lot. For instance, we can make a simple graph as:

```
> plot(landbird$EASTING, landbird$NORTHING)
```

The first command is displayed on the horizontal axis and the second on the vertical axis. In this case, we are plotting the UTM coordinates (Easting, Northing) of each site. To avoid confusion, you could also type:

```
> plot(x = landbird$EASTING, y = landbird$NORTHING)
```

Or

```
> plot(NORTHING ~ EASTING, data = landbird)
```

The tilde is used in many R stats packages to describe y as a function (~) of x. There are numerous arguments in the plot function that allows us to specify more control over the figure. For example, the following code adds x, y labels:

```
> plot(NORTHING ~ EASTING, data = birds, xlab = "easting (UTMs)", ylab =
  "northing (UTMs)")
```

We can also make other types of plots. For example, we can plot the pattern of elevation and mesic and Easting, which presumably should be correlated:

```
> plot(landbird$Elev, landbird$Mesic)
> plot(landbird$Elev, landbird$EASTING)
```

One can also make a multi-panel scatter plot by calling several columns (or rows) of data you want to consider:

```
> plot(landbird[ ,c(4:5,7)]
```

Or equivalently

```
> pairs(landbird[ ,c(4:5,7)])
```

This calls plots for each combination of columns 5–7 (easting, northing, elevation).

We can also calculate boxplots to summarize distributions of data.

```
> boxplot(landbird$Elev)
```

Finally, we can easily graph histograms of the data as well:

```
> hist(landbird$Elev, xlab = "Elev", main = "")
```

There are numerous other neat ways that R can provide figures, but it is beyond the scope of this introduction to provide examples. For those interested in graphics, see `ggplot2` and `lattice`.

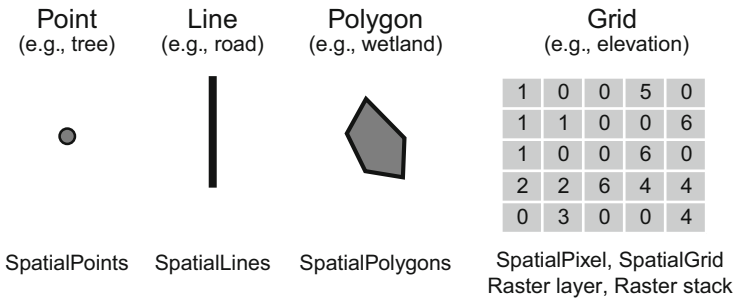


Fig. A.2 Different types of spatial data and their common associated classes in R

Spatial Data in R

Spatial data are data that have a spatial reference, that is, the data are associated with geographic locations. These locations have coordinate values and a reference system associated with these coordinates. Reference systems are representations of Earth (e.g., the shape of Earth). Locations can have associated information, frequently termed attribute data.

Spatial data come in many forms. Two general types of data include vector-based data and grid (or raster) based data (Fig. A.2). Vector-based data include data in the form of points, lines, or polygons, while raster data are based on lattice grids. Both of these data types are useful, but often (though not always) in spatial ecology and conservation we use raster data for analyses and modeling, rather than vector data. Note that we can potentially convert data from vector to raster and vice versa, as we shall see below.

Using Spatial Classes

There are a variety of classes for spatial data in R (Bivand et al. 2013). It is beyond the scope of this book to address these thoroughly (see Bivand et al. 2013), but here we cover common data types and objects used in the book.

R developers created the `sp`package to help unify different types of spatial data. This package can accommodate both vector and raster data (Fig. A.2), including lines, points, polygons, and grid data. Often this package is loaded with other spatial packages in R, so in the book we rarely call this package because it is loaded with calls to other packages.

For vector data, there are three key data classes that several packages use (e.g., `sp`, `rgeos`): `SpatialPoints`, `SpatialLines`, and `SpatialPolygons`. There can be data frames associated with each of these, or *attribute data*, which provide information for points, lines, or polygons. In such cases, there are data frames associated with the spatial objects that store information relevant to each point, line, or polygon, and are referred to as `SpatialPointsDataFrames`,

SpatialLinesDataframes, etc. At the time of publication, the sf package ("simple features") was launched and we expect sf will be increasingly used for vector data although we do not focus on it in the book.

As an example, we can take the data frame on bird data used above and create a SpatialPoints data frame. We first extract the UTM coordinates from the data frame and associated attributes when want to include with the spatial data.

```
> points.coord <- cbind(landbird$EASTING, landbird$NORTHING)
> points.attributes <-
`data.frame(transect = landbird$TRANSECT, point = landbird$POINT,
 VATH = landbird$VATH)

#create spatial points data frame
> points.spdf <- SpatialPointsDataFrame(points.coord, data = points.
  attributes)
> plot(points.spdf)
```

We can then look at the coordinates in the following way.

```
> coordinates(points.spdf)
```

For the attribute data, we can access and extraction information in similar ways to non-spatial R classes. For instance, we can look up the names of the attribute data and we can subset the spatial data we just created to only consider locations where varied thrush were detected:

```
#look up names of attribute data
> names(points.spdf)

#subset
> points.spdf.pres <- points.spdf[points.spdf$VATH == 1,]
> plot(points.spdf.pres)
```

Spatial lines and spatial polygons can be treated in similar ways. For example, we can convert the spatial points data above regarding point locations on transects into a SpatialLines data frame, with one line for each point. To do so, we use the lapply and split functions to create a list of points associated with each transect (105 transects). We can then use this list to create a SpatialLines object and then add attribute data (one row of data/line) to create a SpatialLines data frame:

```
#list of lines for each transect
> points.lines <- lapply(split(points.spdf, points.spdf$transect),
  function(x) Lines(list(Line(coordinates(x))), x$transect[1L]))
> str(points.lines)
```

When looking at the structure of this new object, it shows that we created a list, where each element of the list (e.g., points.lines[[1]]) represents the *x*–*y*

coordinates for each point on the transect and the transect ID, such that there are 105 list elements (105 transects). We then convert this to a `SpatialLines` object and then create a `SpatialLinesDataFrame`:

```
> points.sl <- SpatialLines(points.lines)

#add line attributes
> transect.data <- data.frame(transect = unique(landbird$TRANSECT))
> rownames(transect.data) = transect.data$transect #must add ids

> points.sldf <- SpatialLinesDataFrame(points.sl, data = transect.
  data)
```

We can then plot the lines across the region or zoom into one transect by subsetting the data and re-plotting:

```
> plot(points.sldf)

#subset and re-plot
> points.sldf1 <- points.sldf@lines[[1]]
> plot(points.sldf1@Lines[[1]]@coords[])
> lines(points.sldf1@Lines[[1]]@coords)
```

As an example of spatial polygon data, we consider a map of watersheds in the region. These data come from the level 8 hydrologic unit classification system of the USGS and USDA. To read these data, we use the `readOGR` function from the `rgdal` package, asking for the folder that contains a vector polygon (.shp) file:

```
> watersheds<-readOGR("water")
```

We can glean relevant summary information about this layer in a similar way to data frames.

```
> head(watersheds,4)
> summary(watersheds)
> dim(watersheds)
```

These summaries show that there are 63 watersheds in the region and that this layer contains several columns of data, most of which are not relevant for our purposes here. We can plot this watershed layer and add text to the figure:

```
#plot
> plot(watersheds, col = "gray")
> invisible(text(getSpPPolygonsLabptSlots(watersheds),
`Labels = as.character(watersheds$NAME), cex = 0.4))
```

Several standard data manipulation features can be used on SpatialPolygons. For example, we can subset the layer in a way similar to a data frame. Below we subset to one watershed, the Bitterroot:

```
#subset and plot one watershed
> huc_bitterroot <- watersheds[watersheds$NAME == "Bitterroot",]
> plot(huc_bitterroot)
> plot(points.spdf, col = "red", add = T)
```

If we want to write SpatialPolygons, SpatialPoints, or SpatialLines to a file for importing into GIS software, such as ArcGIS, we use the writeOGR function:

```
#write shp files to folder
> writeOGR(huc_bitterroot, dsn = "water", layer = "bitterroot", driver
  = "ESRI Shapefile")
```

Because these layers have several files associated with them, we call a folder for the files to be placed with the dsn = statement.

For raster data, we typically use objects from the rasterpackage, although the sppackage also handles raster data. Rasterlayer, rasterbrick, or rasterstack objects are very user-friendly and are a workhorse for spatial data analysis in ecology. Note that raster stack and brick objects are made up of several raster layers of the same grain and extent. These objects are useful when working with and manipulating several layers of geographic data. We can load raster layers with the raster function:

```
> elev <- raster("elev.asc") #elevation layer
```

To interpret this raster layer, several aspects of the layer can be readily gleaned, such as the resolution (grain) and extent of the layer, the number of cells in the layer and its dimensions.

```
> res(elev)
> extent(elev)
> dim(elev)
> ncell(elev)
> summary(elev)
```

The above example only considers one raster layer, but often times we work with several raster layers for a region. When these raster layers have the same grain and extent, we can create a raster stack (or raster brick), which is a useful way to have a raster object that contains information from all rasters. Below we load another raster layer that contains information on the presence of mesic forest cover.

```
> mesic <- raster("mesic.grd")
>plot(mesic)
```

To make a raster stack, we check to make sure this raster is comparable to the layer on elevation and then use the `stack` function to combine.

```
> compareRaster(elev, mesic)

#create raster stack
> layers <- stack(elev, mesic)
> plot(layers)
```

There are a variety of useful functions for extracting and manipulating raster data, which we illustrate throughout the book. Some manipulations include cropping rasters, extracting values from rasters, and summarizing raster values in particular regions (or zones).

Projections and Transformations

A common issue in the use of geographic/spatial data is appropriately dealing with geographic projections. Because of the spherical nature of the Earth, we need a means to take spatial information and project it onto a flat surface. For example, a common model is the WGS84, or the World Geodetic System 1984, which is an ellipsoid model for the shape of the Earth.

When working with spatial data, all maps contain some distortion (e.g., in shape, distance, direction) from taking the spherical Earth and projecting it onto a flat surface. This can be done in many ways through the use of different types of coordinate systems. We always want to be careful about the coordinate systems being used. In general, each coordinate system is defined by four components: (1) measurement framework (geographic or planimetric); (2) units of measurement; (3) map projection; and (4) reference locations (e.g., datum).

There are two kinds of measurement frameworks for coordinate systems: geographic and projected coordinate systems (or planar coordinate systems). Geographic coordinate systems (GCS) use latitude/longitude coordinates for locations on the surface of the Earth whereas projected coordinate systems use analytical transformations for mapping to a flat surface and rectangular coordinates.

A GCS includes an angular unit of measure, a prime meridian, and a datum (based on a spheroid). Meridians are lines of longitude and the prime meridian is the line that is considered the origin. A datum defines the position of the spheroid relative to the center of the Earth and is a reference from which measurements are made.

A projected coordinate system is defined on a flat, two-dimensional surface. Unlike a geographic coordinate system, a projected coordinate system has constant lengths, angles, and areas across the two dimensions. A projected coordinate system

is always based on a geographic coordinate system that is based on a sphere or spheroid.

The most important thing is to know what type of coordinate system you are working with in regard to the GIS layers you are using. However, some coordinate systems may be better than others for certain problems, depending on what you are doing and the scale that you are doing it. For global applications, geographic coordinate systems, such as WGS84, work best. For most large extent, but not global, applications (e.g., the U.S.A. or even regions of the U.S.A.), it is best to use an equal-area projection (e.g., State Plane Coordinate System), which allows for accurate delineations of polygons and distances. For smaller areas (e.g., a few watersheds), using localized projections may work best for measurements, because little to no distortion will occur. In this situation, there are different areas or zones where finely tuned projections occur. But you need to be aware that mapping/measuring across zones can be problematic.

To describe projections, spatial data in R typically use what is called a coordinate reference system (CRS). R uses a PROJ.4 formatted character string to describe the CRS. For example, for geographic data collected in the latitude–longitude coordinate system, we can define the CRS as:

```
> crs.latlong <- CRS("+proj=longlat +ellps=WGS84")
```

The raster layers we used above for elevation and mesic forest have a different projection. This projection is based on the Albers Conical Equal Area. It can be defined as:

```
> crs.layers <- CRS("+proj=aea +lat_1=46 +lat_2=48
`+lat_0=44 +lon_0=-109.5 +x_0=600000 +y_0=0
`+ellps=GRS80 +datum=NAD83 +units=m +no_defs")
```

The above definition contains all information needed to create a CRS, which was taken from metadata from the R1-VMP map (Brewer et al. 2004) used for creating the mesic layer described above. Specifically, we state that the `proj = aea` (Albers equal area), we set the standard latitude parallel 1 to 46, and the standard latitude parallel 2 to 48, the Central Meridan to −109.5, the latitude of origin to 44, the units to meters, and the ellipsoid model to Geodetic Reference System 80 (GRS80).

We can determine the current coordinate system for spatial data with the `proj4string` function:

```
> proj4string(points.spdf)
```

In this case, R has not added a coordinate system to these points. However, we know that these points are in the same coordinate system as in the R1-VMP. We can set this CRS as:

```
> proj4string(points.spdf) <- crs.layers
```

Sometimes we might want to transform projections in R of various spatial data for proper overlaying of different data sources or because we might want to make calculations (e.g., distance measurements) based on different datum. For instance, if our original spatial data is in a latitude–longitude format, where the datum is decimal degrees, we might want to transform the projection to another format, where the datum is in meters. The `sp`package has a function, `spTransform`, which is intended to accomplish this task. In this case, we can use this function to transform our projection. To transform the spatial points to WGS84, we can use this function. First, take a peek at the coordinates of this object, which are in UTMs:

```
> coordinates(points.spdf)
> points.spdf.wgs84 <- spTransform(points.spdf, crs.latlong)
```

Now take look at the new coordinates:

```
> coordinates(points.spdf.wgs84)
```

These new coordinates are now in decimal degrees, which is consistent with the WGS84 projection. For converting rasters, we can take a similar approach with the `projectRaster` function from the `raster`package.

Next Steps: Where to for Further R Mastery?

This Appendix is meant to provide a concise overview of using R, one that aims to facilitate understanding the examples provided throughout this book. Indeed, we use information provided in this Appendix to help guide students and new users of R before diving into specific topics covered in this book.

For further knowledge in using R, there are several excellent sources. For statistics, there are several books that provide more general overviews (Crawley 2007), as well as books focused on specific types of modeling, such as generalized linear mixed models (Zuur et al. 2009). For spatial analysis, Bivand et al. (2013) provided a comprehensive discussion of both handling spatial data in R and spatial analysis. For an overview of ecological models, Bolker (2008), Borcard et al. (2011), and Legendre and Legendre (2012) are excellent sources.

References

Bivand RS, Pebesma EJ, Gomez-Rubio V (2013) Applied spatial data analysis with R. Use R! 2nd edn. Springer, New York

Bolker B (2008) Ecological models and data in R. Princeton University Press, Princeton, NJ

Borcard D, Gillet F, Legendre P (2011) Numerical ecology with R. Springer, New York

Brewer CK, Berglund D, Barber JA, Bush R (2004) Northern region vegetation mapping project summary report and spatial datasets, version 42. Northern Region USFS

Brunsdon C, Comber L (2015) An introduction to R for spatial analysis and mapping. Sage Publications, Inc, London

Casadevall A, Fang FC (2010) Reproducible science. Infect Immun 78:4972–4975

Crawley MJ (2007) The R book. Wiley, Chichester

Legendre P, Legendre L (2012) Numerical ecology, 3rd edn. Elsevier, Amsterdam

Munafo MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, du Sert NP, Simonsohn U, Wagenmakers E-J, Ware JJ, Loannidis JPA (2017) A manifesto for reproducible science. Nat Hum Behav 1:0021

Wickham H (2007) Reshaping data with the reshape package. J Stat Softw 21:1–20

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) Mixed effects models and extensions in ecology with R. Springer, New York

Index

A
Allee effects, 273
Anisotropy, 135, 138, 150–152, 178, 393
Assumption
  independence, 134, 136
  stationary, 112, 160, 165
Autocovariates, 171, 172, 179, 180, 183, 194, 197, 198, 204, 205

B
Barrier
  ecological, 326
  physical, 326
Biodiversity, x, 1–3, 5, 6, 21, 23, 55, 58, 61, 62, 64, 83, 92, 101, 213, 321, 360, 419, 420, 425, 427, 476–479, 481
Biotic
  homogenization, 427
  interactions, 21, 218, 311, 426, 447, 476
Biotic abiotic movement (BAM), 218, 476
Biotope, 216–218, 220
Bivariate
  cross-correlograms, 142, 165
  cross-semivariograms, 142
Bonferroni adjustment correction, 139
Buffer sizes, 37, 38, 40, 42

C
Calibration plot
  threshold-independent, 232
Climate change, 5, 6, 92, 169, 170, 213, 217, 231, 260, 476, 479
Clustering, 88, 108, 125, 126, 134, 226, 310
Colonization-extinction dynamics, 259, 331, 370, 373, 376, 379, 380, 385, 388, 389, 395–403, 405, 409, 421, 425, 433
Community
  Gleason, 429, 447
  species assemblage, 437
Community environment relationships, 428–431
Complementarity, 6, 23, 428, 437
Connectivity
  demographic, 321–324
  flow, 323, 332–334
  functional, 322, 323, 328, 336, 338–344
  gap-crossing, 323
  immigration/colonization, 329
  integral index, 352
  least-cost corridor, 331, 341–343, 358
  limit, 326–328
  network, 333, 335, 339, 347–349, 351, 352
  patch-based, 330, 332, 334, 335, 345, 352, 355–357
  probability, 345, 346, 352, 357, 378
  propagule pressure, 329, 375
  realized, 322–324, 327, 328, 345
  resistance, 328–332, 335, 336
  stepping-stones, 326, 333, 349, 355
  structural, 322, 323, 328, 335, 336
Conservation
  corridor, 5, 61, 328, 343
  planning, 1, 6, 23, 170, 271, 335, 343, 360, 427, 428, 464, 476, 478, 479
  prioritization, 1, 5, 6, 336, 381

© Springer Nature Switzerland AG 2018

R. Fletcher, M.-J. Fortin, *Spatial Ecology and Conservation Modeling*,
https://doi.org/10.1007/978-3-030-01989-1

Correlogram
Mantel, 391, 393, 436, 437, 460, 461
multivariate, 165, 391, 436, 460
spline, 143, 146, 147, 392
Covariance, 18, 120, 133, 135, 137, 139, 141, 142, 182, 436, 437

D
Data
aggregation, 101, 105–107, 226
big, 478, 480
binomial, 40, 200
category, 149
count, 28, 105, 175, 220, 222, 409, 430, 445, 495
covariates, 45, 113, 186, 199, 496
disaggregation, 30
forest cover, 40, 43
geo-spatial, 260, 480
GPS telemetry, 277, 279, 280, 284, 286, 311, 478, 480
grid, 147, 160, 175, 505
imperfect absence, 221
independence, 134, 280
independent and identically distributed (IID), 136, 172, 173
integrating multiple, 258
mark–recapture, 331, 336, 359, 408, 409, 502
marks/marked, 103, 111, 112, 120–122, 260, 348
National Land cover Dataset (NLCD), 65, 66
non-detection observations, 221
Normalized Difference Vegetation Index (NDVI), 55
point, 101–129
presence–absence, 113, 120, 176, 220–222, 228, 231–234, 236, 241, 255, 258, 284, 388, 429, 433, 442, 456, 495
presence-only, 104, 220–224, 226, 228, 229, 231, 233–236, 255, 258, 444
projection, 33, 34, 144, 186, 439, 509–511
pseudo-absences, 220, 234
radiotelemetry, 271, 286, 295, 311, 331, 336
sample selection bias, 221
simulated, 28–31, 157
sources, 220, 236, 258, 260, 286, 287, 329, 511
species prevalence, 217, 221, 232, 259, 433, 442
state, 311, 384
time-series, 382, 383, 385, 388, 391, 393
trajectory, 282, 294, 295, 303, 305, 309, 310
Dataset
bioclim, 222, 240, 254, 255
bird distribution, 438
Breeding Bird Survey, 185
endangered Florida panther (*Puma concolor coryi*), 286, 336
endangered snail kite, 336, 345, 387
Northern Region Landbird Monitoring Program, 184, 235
Prickly pear cactus (*Opuntia humifusa*), 113, 143
varied thrush (*Ixoreus naevius*), 184, 185, 453, 454, 495, 499, 500, 502, 506
wind-dispersed orchid (*Lepanthes rupestris*), 388, 389
Design
block, 160
sampling, 20, 22, 27, 48, 83, 147, 184, 185, 191, 388, 389, 423
Detection errors, 47, 204, 205
Detection probability, 400, 401
Dispersal
distance mean, 23, 57, 334, 347, 379, 388, 394, 395
limitation, 22, 215, 258, 259, 379, 425, 464
range, 380
Disperser, immigrant, 120, 323, 325, 370, 376, 394, 421, 424, 425
Dispersion, 446
over-, 445, 446
Distance matrix
dissimilarity, 436, 448
Euclidean, 8, 9, 163, 338, 340, 341, 394, 435, 461
geographic, 436, 437, 448, 460, 461
similarity, 431
Distribution
Bernoulli, 86, 88, 173, 191, 384, 385
beta, 435
exponential, 347
gamma, 173
log-normal, 346
multinomial, 29
negative exponential, 203, 334, 345, 348, 376, 377
normal, 126, 172, 204, 279, 291, 457
Poisson, 28–30, 104–106, 125, 173, 445
Weibull, 346, 347
Diversity
alpha
evenness, 78–80, 82
Shannon, 78–80, 85, 447

beta
nestedness, 420–422, 430
turnover, 9, 420, 422, 430, 435
gamma, 420–422
habitat, 421, 423
richness, 78, 79, 420, 427, 429, 447

E
Ecological
fallacy, 19, 22
response, ix, 7
trap, 272, 274
Ecosystem
services, 5, 23, 92, 321, 428, 476
Edge effects, 57, 63, 76, 104, 109, 112, 116–119, 138, 376
Effective
area, 75, 76, 376
dispersal, 323, 325
Environmental
change, 3, 5, 6, 90, 92, 216, 261, 274, 321, 359, 360, 369, 406, 407, 422, 427, 477–479, 481
conditions, 2, 3, 6, 57, 59, 187, 215–217, 223, 228, 259, 260, 275, 284, 426
filtering, 2, 205, 426–428
stochasticity, 274, 375, 376, 424
Equilibrium
occupancy, 405
Error
distribution, 172, 173, 435
false positive, 204, 232, 233, 260
residual, 40

F
Fractal dimension, 70, 71, 73, 88, 310
Fragmentation, 55, 57, 58, 60, 62, 64, 73, 89, 110, 378, 379, 406, 477
Function
I-spline, 435, 448
Functional statistics, 105

G
Geostatistics, 22, 88, 133, 139, 140, 143, 478
Gradient
environment, 56, 59, 89, 106, 123, 133, 135, 176, 214, 216, 224, 278, 425, 434, 435, 448, 450, 463
Graph
adjacency matrix, 73, 333, 349
directed, 334, 349, 352, 437
links, 350
nodes, 226, 327, 333, 336, 345, 347, 348, 351, 463
undirected, 334, 347–349, 351
unweighted, 334, 347, 352
weighted, 334, 348–350, 354

H
Habitat
critical, 23, 271, 276, 278, 312, 321, 325, 380
disturbance, 6, 59, 101, 325, 380
fragmentation, 55, 58, 60–62, 64, 110, 378, 379, 406, 477
habitat loss, 55, 57–62, 64, 101, 378, 379, 406, 407
quality, 89, 259, 272–274, 324, 373, 376, 476
restoration, 335, 378
Home range
biased random bridge models, 310
Brownian bridge movement, 280
convex hulls, 279, 280, 293
kernel density estimation (KDE), 280
local convex hulls (LoCoH), 279, 280, 293, 295
minimum convex polygons (MCP), 279, 280, 289, 291, 298, 299
Hypothesis
habitat area, 423
tests, 183

I
Index
Bray–Curtis, 430, 448
dispersion, 80
diversity, 57, 79, 430
evenness, 79, 80
Gini, 225
Jaccard, 430
Shannon, 79, 80
Sørenson, 430, 448
Individual traits, 275, 425
Isolation-by-distance, 326
Isolation-by-resistance, 325
Isotropic, 112, 116, 117, 125, 126, 138

K
Kernel dispersal
“fat-tailed” dispersal, 346

L
Land
cover, ix, 9, 22, 26, 27, 30–47, 55–92, 158, 236, 287, 288, 290, 298, 299, 302, 309, 329, 337, 338, 342, 343
cover types, 22, 27, 29, 35, 48, 57, 58, 62–64, 66, 72, 75, 78–82, 86, 89, 287, 289, 298, 302, 307, 343
sharing, 476
sparing, 476
Landscape
composition, 64
configuration, 63, 322, 329
connectivity matrix, 378
demography, 370, 372, 382, 408, 409
Landscape feature
corridor, 58, 64, 91
matrix, 58, 64
patch, 58, 64, 91
Landscape matrix effect
friction, 329
permeability, 329
resistance, 329
Landscape metrics
class-level, 63, 65, 78
functional, 57, 58, 78
landscape-level, 63, 65, 78–83
patch-level, 63, 64
structural, 57, 58, 65
Landscape model
continuum, 58, 59, 91
gradient, 24, 59, 91, 425
habitat variegation, 58, 59
island, 56, 58, 64
mosaic, 3, 58, 91
patch, 26, 59, 64
patch-matrix-corridor, 58, 64, 91
Land-use
extensification, 56
intensification, 55, 56, 92
land use and land cover (LULC), 9, 55–57, 60, 62, 65, 91, 92
Local
local indicators of spatial association (LISA), 142, 165
spatial pattern, statistics, 105
Long-distance colonization, 377

M
Machine learning methods
Boosted Regression Trees, 225
classification and regression trees (CART), 224
classification tree analysis (CTA), 224, 225
Maximum entropy, 226–228, 246–249
Metacommunities, ix, 1, 2, 5, 7, 67, 321, 326, 421, 425, 426, 428, 433, 475, 477, 478
Metaecosystems, 1, 2, 5, 7, 475, 477
Meta-models, 8, 480
Metapopulation
capacity, 353, 378, 379, 406, 407
incidence function model (IFM), 375, 376, 378, 383
island–mainland, 381
persistence, 22, 63, 325, 373, 374, 378–380
source–sink, 380, 381, 409
Metrics
surface, 91
Migrations, 1–3, 325, 386
Model
agent-based, 8, 359, 383, 408, 409, 479
bagging, 225, 226
balanced dispersal model, 381, 382
Bayesian hierarchical model, 463
Bayesian models, 184, 198, 200, 204, 403
calibration, 231, 233, 234, 255
climate envelope models, 213
combining models, 229, 257, 258
correlative, 217, 219–222, 260
data validation, 231, 234, 237, 239, 253, 256
diffusion, 4, 8, 276, 277, 359, 387, 408
distance-based matrices, 431, 436
dynamic, 4, 8, 259, 276
dynamic occupancy, 384, 385, 400, 403–405
ecological niche models (ENMs), 213, 219
ensembles, 229, 230, 257, 258
envelopes, 106, 222, 223, 229, 240, 241
evaluation, 230, 231, 235, 236, 248, 253–257, 259
false negative errors, 204, 221, 232, 260, 384
first-order Markov chain, 384
GAP, 213
Gaussian random fields, 88, 157, 171, 182
habitat suitability models (HSMs), 213, 219
hierarchical, 24, 25, 47, 181, 463
individual-based, 359, 373, 382, 383, 408, 479
kappa, 140, 232, 257
K-fold partitioning, 231
log-likelihood, 41–44, 154, 234, 254
mechanistic, 92, 217, 219, 220, 260, 281, 310
multi-species occupancy, 430, 431, 455
neutral, 57, 88, 90, 144, 159, 424, 426
non-linear, 169, 177, 187, 224, 225, 252, 435, 441, 448

observation errors, 40, 185, 204, 221, 259, 384, 385
occupancy, 8, 204, 213, 258, 260, 375, 376, 382–385, 387, 399, 400, 402, 430, 502
over-fitting, 227
partitioning, 88, 231, 310, 463
penalized quasi-likelihood, 203
phenomenological, 3, 219, 310
predictive distribution models, 213
process-based, 2, 3, 219
profile methods, 222, 229
recursive partitioning (RP), 88, 224
second-order Markov chain, 384
species distribution models (SDMs), 178, 213–217, 219, 220, 225, 226, 228, 229, 235, 246, 255, 259, 260, 278, 281, 284, 426, 429, 441, 451, 454, 463, 476
stacked species distribution, 429, 441
state-space, 311, 384
stochastic patch occupancy models (SPOMs), 371, 373, 375, 376, 382–384, 387
threshold-dependent, 233
type I errors, 136, 178, 445
weighted regression, 178, 250
Model community-level
assemble and predict simultaneously, 428
assemble first, predict later, 428–430, 444–451
predict first, assemble later, 428, 429, 431, 432, 441–444
Model machine learning
maxent, 222, 226–228, 246–250, 253, 255, 257
random forests, 225, 249, 255
Model selection
Akaike's Information Criterion (AIC), 41, 47, 124, 152, 230, 302
Bayesian Information Criterion (BIC), 154
Modifiable areal unit problem, 22
Monte Carlo, 106, 118, 127, 138, 142, 146–148, 150, 154, 166
Morphological image processing, 91
Movement ecology
brownian motion, 88, 280
correlated random walks (CRWs), 276, 277, 281
Eulerian model, 277
fractal dimension, 310
hidden Markov approaches, 310
Lagrangian model, 277
motion capacity, 277, 322, 328
navigation capacity, 277, 322
random walk, 275–277, 310, 331, 344
state-space models, 384
step lengths, 277, 284, 310
turning angles, 277, 284, 310
Moving windows, 8, 63, 64, 79, 83–86, 165, 289
Multilevel, 18, 21, 23–25, 172, 181–183, 194, 201, 202, 205, 455, 475
Multi-scale, 18, 23–28, 31–47, 144, 164, 166, 276, 437, 475
eigenvalue methods, 437
fourier series, 159–162
Principal Coordinates of Neighbourhood Matrices (PCNM), 162, 163
spectral decomposition, 134, 159, 162–165
wavelet, Moran's Eigenvector Map (MEM), 162
Multivariate analysis
canonical correspondence analysis (CCA), 431, 434, 451, 462
canonical ordination, 434
constrained gradient ordination, 431
correspondence analysis (CA), 434
direct gradient analysis, 434, 451–454
distance-based redundancy analysis (dbRDA), 454
joint community models, 431
MANOVA, 283
Mantel tests, 436, 460
multi-species occupancy modeling, 431, 455
ordination, 431, 434, 437, 450–452
partial mantel tests, 436
partial ordination, 435
principal component analysis (PCA), 434, 437, 450, 452
redundancy analysis (RDA), 434, 451, 452, 454, 462
regression, 431–435, 452, 454–460, 462
Multivariate spatial dependence, 165

N

Neighborhood
buffer, 40, 85
habitat area index, 353
minimum spanning tree, 195
moving window, 79, 83–86, 165
proximity index, 75, 77, 329
Network
betweenness, 350
closeness centrality, 349

Network (*cont.*)
connectance, 352
eigenvector centrality, 349, 350, 356, 378
graph, 330, 332, 333, 335, 345, 347–349, 352
information centrality, 349
metrics, 47, 328, 335
minimum spanning tree, 462
modularity, 351
patch degree, 327, 328, 330, 333, 335, 336, 349, 351
spatial, 1, 9, 332, 463, 464
Neutral landscapes, 56, 57, 86–91, 144, 158, 159
Niche
Elton, 215
empty, 215, 217
fundamental, 215–218, 381
Grinnell, 215, 218
Hutchinson, 215, 216
potential, 215
realized, 215–218
Non-ideal behaviors, 273
Non-stationary inhomogeneous, 104, 141
Null envelopes, 146, 154, 155, 188, 190

O
Organization
level, 2, 18–21, 23–25, 324, 326, 327, 419, 477
Overlapping landscapes, 47

P
Package
adehabitat, 286
adehabitatHR, 286, 293, 310
adehabitatHS, 286, 298, 300
adehabitatLT, 286, 294, 303, 308
AER, 445, 446
bcpa, 310
betapart, 438
biomod2, 251, 256
boral, 463
bsam, 311
deldir, 199
dismo, 199, 235, 239, 240, 246–248, 250, 251, 253, 255
ecospat, 235, 255, 256
Ergm, 335
Fdrtool, 346
FieldSim, 88
fitdistrplus, 346
Gbm, 245
gdistance, 335, 336, 339, 344
gdm, 438, 447, 448, 450
geoR, 142, 150–152, 154–156
ggplot2, 252, 349, 503, 504
gstat, 142, 150–152, 154, 156, 157
hSDM, 205, 260
igraph, 335, 336, 345, 347, 349–352, 356
inla, 198–200, 204
latentnet, 335
lme4, 183, 201, 456
lulcc, 65
MASS, 183, 202, 445, 490, 493
MetaLandSim, 335, 387
mgcv, 183, 195, 243, 244
MigClim, 258
moveHMM, 310
mvnabund, 438
ncf, 142, 145, 146, 150, 165, 189, 387, 391
nlme, 202
PatternClass, 90
pgirmess, 142, 145, 146, 148, 150
popbio, 387
popdemo, 387
PresenceAbsence, 235, 253–255
RandomFields, 88, 157, 158
randomForest, 245
raster, 28, 33, 35, 37, 48, 65, 69, 79, 80, 83, 84, 92, 157, 162, 164, 185, 186, 189, 239, 249, 288, 298, 299, 307, 337, 338, 343, 358, 451, 508, 511
rasterVis, 67
reshape2, 161, 298, 300, 389, 439, 455, 500, 502
resource selection, 303, 338
rgdal, 33, 115, 337
rgeos, 37, 65, 75, 76, 338, 505
rgrass7, 480
rJava, 247
RNetlogo, 409, 479
ROCR, 235
RQGIS, 480
sdm, 235, 256
SDMTools, 65, 69, 70, 72–76, 78, 79, 91, 335
secr, 90, 387
simecol, 408
sp, 299, 303, 438, 505, 508, 511
spamM, 184, 200, 203
spatial, 145, 164
spatialEco, 38
spatstat, 113–116, 118–125, 128
spBayes, 184, 198
spdep, 142, 145–148, 150, 165, 183, 189, 195–197, 200, 308

statnet, 335, 349
stocc, 205, 260
stpp, 128
synchrony, 387, 390
T-LocoH, 280
unmarked, 260, 387, 395–397, 400, 401
vegan, 163, 183, 195, 437, 448, 451, 454, 461, 463
VGAM, 438
Voss, 88
waveslim, 160
Paradigm
area-isolation, 370, 376, 384, 394, 395
mass effects, neutral, 425, 426
patch dynamics, 425
species-sorting, 425
Parameter estimation, 153, 170
Partialled out, control for, 461
Patch
area, 58, 70, 71, 73, 77, 78, 122, 329, 345, 353, 354, 356, 357, 370, 376, 379, 390, 394–396, 398–400, 402, 403, 406, 407, 422
connectivity, 323, 333, 349, 350, 395, 398, 399, 402, 403
core area, 67, 68, 70, 71, 73, 75
dynamics, 7, 21, 22, 64, 388, 421, 425
edge, 68, 70, 71, 73, 75, 76, 78, 82, 91, 333, 376
isolation, 5, 56, 57, 63, 64, 67, 68, 75–77, 110, 323, 329, 370, 375, 376, 394, 398, 399, 406, 407
shape, 68, 70, 71, 73
size, 56, 57, 64, 67–70, 72, 73, 75, 325, 375, 380, 384, 388, 389
Patch delineation
4-neighbor rule, rook's rule, 67, 68, 80
8-neighbor rule, queen's rule, 67, 147, 148
16-neighbor rule, 67, 68
Path
Dijkstra's algorithm, 331
least-cost, 331, 332, 337–339, 341–344, 354, 356, 358, 360
mapping flow, 344
optimal, 331
randomized shortest, 332, 339, 340, 343, 344
selection functions, 285, 308–310, 331
shortest, 331, 338–340, 343, 344, 349, 350, 353, 354
Percolation, 73, 78, 89
Phenomenological time-series analysis, 310
Phylogenetic dependency, 457, 463
Point analysis
bivariate, 103, 111, 121
edge effect, 112, 116
envelopes, 122, 125, 129
G-function, 110, 120
marked (marks), 111, 120
multivariate, 111, 112
O-ring statistic, 110
pair correlation function, 108, 110, 118, 119
radius, 107, 108, 112, 117, 118
ring, 110, 119
Ripley's K, 107, 109, 116, 119
Ripley's L, 107–109
Point process
inhomogeneous, 104, 123–125, 221, 222, 224, 226, 228, 247, 249, 250, 260, 283, 284, 302, 303
Pollinator, 388
Population
closed, 370, 371, 389
colonization, 321, 325, 329, 370, 372, 373, 375, 376, 379, 380, 383, 388, 389, 409
density, 273, 275
density-dependence, 274, 323, 327
dynamics, 2, 4, 9, 22, 23, 276, 312, 359, 369, 426
emigration, 370, 371, 381, 383, 426
extinction, 5, 101, 137, 325, 369, 370, 372, 373, 375, 376, 378, 380, 382, 383, 389
genetics, 321, 324–326, 371, 409
growth, 216, 218, 272, 274, 323, 324, 369, 373, 381, 383, 387, 391–393, 409
immigration, 329, 370, 371, 373, 381, 383, 426
integrated population models, 409
local, 22, 216, 217, 323, 325, 370, 373, 375, 380, 426
minimum viable populations (MVP), 379
panmixia, 371
persistence, 22, 63, 137, 325, 360, 369, 373, 374, 378–380
size, 61, 273, 274, 325, 329, 370, 376, 381, 388, 391, 409, 424
spatially structured, 370, 371, 373, 381, 382, 385, 387, 391, 409
transition matrix models, 353, 385
viability analysis (PVA)
elasiticity analysis, 383
sensitivity analysis, 383
Population spatially-structured populations, 370, 371, 381, 382, 387, 391, 409

Potential immigrants, 376, 394
Predictions, 27, 39, 88, 104, 107, 136, 141, 155, 156, 169, 172, 173, 183, 188, 189, 213, 217, 219, 221, 222, 225–227, 229–234, 243, 246, 249, 251–253, 255–260, 326, 328, 340, 378, 379, 382, 387, 399, 402, 421, 424, 426, 429, 431, 441–444, 447, 449–451, 453, 454, 458–460, 476, 478, 495
Predictive accuracy
  area under the curve (AUC), 229
  Cohen's kappa, 232
  receiver operating characteristic curve (ROC), 233
  true skill statistic (TSS), 232
Process
  complete spatial randomness (CSR), 106, 116
  diffusion, 280, 359
  discrete in space, 276
  discrete in time, 384
  dispersal, 4, 20, 258, 383
  disturbance, 102, 325, 380
  endogenous, 2, 135, 176
  exogenous, 2, 135, 176
  inhomogeneous, 104, 106, 107, 123–125, 127–129, 221, 222, 224, 226, 228, 246, 249, 250, 260, 283, 284, 302, 303
  Matérn, 102, 106, 125–127
  point, 9, 103, 104, 106, 107, 112, 115, 116, 118, 123–125, 127, 128, 221, 222, 224, 226, 228, 229, 247, 249, 250, 260, 283, 284, 302, 303
  Poisson point, 104, 106, 107, 112, 116, 118, 119, 126, 127
  thinned point, 104
  Thomas, 106, 126, 127
Proportion, 27, 37, 38, 43–46, 57, 60, 61, 63, 64, 73, 78, 80, 83, 85, 86, 88, 158, 162, 173, 221, 233, 234, 236, 239, 279, 282, 290, 302, 325, 373, 374, 395, 436, 448, 452
Protected areas, 1, 6, 23, 213, 336–344, 358, 427, 476
Pseudo-replication, 136, 170

R
Randomization
  bootstrapping, 225, 226
  bootstrap sampling, 225
  Monte Carlo, 138, 142
  permutation, 154
  sampling with replacement, 225
Rarefaction
  Chao estimators, 429
  jackknife estimator, 429
Raster layers, 28–30, 33, 37, 38, 48, 66, 76, 123, 185, 186, 287, 329, 336, 338, 341, 356, 438, 449, 508, 510
Rate
  colonization, 325, 370, 372, 376, 377, 380, 400, 401, 405
  extinction, 101, 370, 374–376, 380, 391, 400, 401, 403, 405, 424
  immigration, 67, 326, 373, 374, 424
Regression
  conditional autoregressive models (CAR), 183, 194, 199, 204
  fixed effects, 172, 177, 178, 204
  generalised least square regression (GLS), 142, 182, 183, 204
  generalized additive models (GAMs), 176, 177, 183, 223, 224, 226, 228
  generalized dissimilarity modeling (GDM), 435, 462
  generalized estimating equations, 307
  generalized linear mixed models (GLMMs), 182, 194, 433
  generalized linear models (GLMs), 40, 42, 123, 142, 171–174, 179, 223, 224, 226, 228, 433, 445
  linear, 40, 135, 136, 141, 142, 165, 171–173, 179, 434, 492
  link function, 173
  local, 176, 177
  logistic, 40, 42, 43, 46, 173, 179, 185–188, 190, 192, 194, 196–198, 224, 228, 250, 283, 284, 301, 302, 307, 441, 455
  monotonic, 225
  multi-species occupancy models, 455
  multivariate, 204, 431–435, 452, 454–460, 462
  multivariate adaptive regression splines, 455
  multivariate GLMMs, 433
  multivariate logistic regression, 455
  multivariate machine learning methods, 455
  negative binomial regression, 445
  neural networks, 455
  N-mixture, 204
  Poisson, 107, 123, 173, 228, 284, 302, 445
  polynomial, 176, 177
  quasi-Poisson, 445
  random effects, 455
  random-intercepts, 455
  root mean squared error (RSME), 220

simultaneous autoregressive models (SAR), 183, 204
spatial, 9, 28, 165, 172, 175, 183, 184, 198, 235, 242, 462
spatial eigenvector mapping, 462
spatial filtering, 183
Relationship
Gaussian, 434
linear, 228, 242, 445
non-linear, 73, 169, 225, 242, 424, 434
species-area, 422, 423
Resampling
majority rule, 29
Rescue effects, 5, 325, 370, 373, 375, 424, 426
Reserve design, 476
Resolution thematic, 58, 62, 66
Resource
compositional analysis, 283
cross-walks, 272
path selection, 282, 285, 308, 309, 312
point selection, 282–284, 298–300, 302, 303
random path, 285
resource selection functions (RSFs), 281–284, 298–309
resource selection probability functions (RSPFs), 221, 272, 281
step selection, 282, 284, 285, 298, 303–307, 312
Resource selection
first-order, 278
second-order, 278
third-order, 278
fourth-order, 278
hierarchical, 278
predator–prey, 311
Response curves, 214, 247, 248, 458
Response plots, 248, 251, 449
Richness
land cover, 78, 79, 85, 86
species, 22, 419, 420, 427, 429, 431, 437, 440, 443–447, 454, 459, 460, 464, 502

S

Sample selection bias, 217, 221
Sampling
errors, 185, 204, 259, 260, 311, 384
scale, 20
units, 20, 22, 25, 27, 105, 172, 174, 185, 236, 433
Scale
analysis, 9
critical, 23
cross-scale interactions, 19, 25
effects, 27, 38, 40
focal patch, 26, 377
landscape, 26, 64, 68, 236, 323, 324, 352, 353, 355, 356
meso-scale, 323, 350–352
multilevel, 23–25, 47, 181, 182, 475
patch-scale, 24, 26, 27, 334, 349, 351
phenomenon, 19, 20, 25, 27, 48
sampling, 20
spatio-temporal, 18, 48, 278, 283, 285, 477
temporal, 21, 101, 271, 278, 286, 360, 372, 409, 475, 477
Scalogram, 135, 160–162
Second-order statistics, 104, 105, 107, 108, 129
Semivariance, 135, 138–140, 142, 152, 436
Semivariograms, 47, 137, 139, 141, 150, 191
Single-large versus several small (SLOSS), 422
Site scores, 453, 454
Smoother
locally weighted polynomial regression (LOWESS), 178
Social behavior, 101, 136, 169, 274
Software
ArcGIS, 480, 508
BIOCLIM, 222, 223, 257
BIOMOD, 235, 251, 256
Fragstats, 65, 74, 79, 80, 82, 328, 335
GIS, 27, 28, 38, 185, 223, 236, 300, 349, 480, 508, 510
GRASS, 480
Marxan, 479
Matlab, 480
Passage, 480
PostGIS, 480
Programitta, 480
QGIS, 480
winbugs, 204
Zonation, 479
Space use, 271–312, 476
Space-time analysis, 128
Spatial
autocorrelation, 4, 88, 133–166, 170, 171, 173, 174, 179, 182, 189–192, 195, 197, 198, 200, 203–205, 274, 436, 476
correlations, 142, 170, 182, 202, 203, 380, 391, 392
demography, 372, 373, 380–382, 386, 409
dependence, 28, 47, 88, 91, 133–166, 169–205, 231, 249, 375, 383, 391–393, 435, 437, 455, 460–462, 464, 476

Spatial (*cont.*)
ecology, vii–ix, 1–9, 17, 21, 28, 31, 48, 55, 60, 62, 88, 91, 92, 105, 127, 129, 271, 282, 285, 332, 371, 408, 420, 464, 475–479, 489, 490, 505
eigenvector mapping, 47, 163, 178, 195, 437, 462
heterogeneity, 1, 9, 22, 124, 369–371, 373, 376, 393, 408
lags, 25, 27, 140
legacy, 4, 8, 135
patterns, viii, ix, 2–4, 6, 8, 9, 17, 19, 20, 22, 55–57, 61, 90, 91, 101, 103, 105, 106, 111, 113, 119–122, 129, 133–136, 138–140, 157, 159, 166, 178, 274, 443
point pattern analysis, 47, 101, 102, 106, 107, 111–113, 121, 129, 139
point process models, 123, 228, 229, 247, 249, 260
quantification, 8, 55, 57, 92
synchrony, 136, 371, 373, 375, 391, 393
trend surface analysis, 172, 176
variance, 9, 22, 133, 134
Spatial interpolation
inverse distance weighting, 156
kriging, 135, 141, 155–157
spatially interpolate, 133, 140
Spatial resolution
extent, study area, 20, 73, 298
grain, 18, 22, 24, 25, 33, 59, 508
grain size, 19, 20, 29, 31
scope, 182, 505
Spatial statistics
Geary's *c*, 138, 145
Moran's *I*, 137–139, 142, 143, 145–149, 165, 166, 179, 191, 196, 391
Spatiotemporal
analysis, 128, 391
connectivity, 477
correlation, 391
dimension, 17, 18
pattern, 128
Species
competition, 2, 111, 215, 381, 421, 425
detectability, 401, 430
distribution, viii, ix, 1–3, 5, 6, 9, 21, 27, 32, 59, 83, 101, 104, 105, 111, 178, 182, 185, 213–261, 271, 278, 281, 283, 284, 311, 331, 379, 409, 426, 429, 441, 451, 454, 463, 476, 478
interactions, 2, 4, 6, 9, 17, 62, 63, 101, 176, 215, 259, 273, 326, 381, 419, 425, 428, 457, 463, 464
persistence, 5, 8, 215, 321
richness, 22, 419, 420, 427, 429–431, 437, 440, 443–447, 454, 459, 460, 464
scores, 434, 452–454
traits, 426
Species–area relationship (SAR), 423
Species–environment relationships, 39, 47, 48, 169, 215, 222, 476
Stationary homogeneous, 104, 112
Statistic first-order, 104–106
Statistic second-order, 105–108, 129
Statistical
inferences, 20, 133, 169, 170
nuisance, 142
Statistics
ANOVA, 172, 453
autocovariate, 179, 180, 197, 198
degree of freedom, 170
goodness-of-fit tests, 234
Pearson linear correlation, 137
Stochasticity
demographic, 424
environmental, 375, 376
Study area, 25–27, 78, 103–107, 109, 112–115, 118, 123, 138, 159, 174, 234

T

Territory, 21, 23, 101, 272, 275, 278, 331, 375
Theory
behavioral ecology, 273, 324
biogeography, 5, 326
circuit, 327, 331, 332, 338–340, 344
Equilibrium Theory of Island Biogeography (ETIB), 61, 423–426
foraging, 324
graph, 327, 328, 330, 332, 336, 345, 356, 358
habitat selection, 273–277
hierarchy, 21, 426
ideal free distribution (IFD), 273, 274, 382
information, 78, 324
island biogeography, 5, 7, 56, 326
landscape ecology, 325, 426
macroecology, 5
mainland, 61, 425
metacommunity, 7, 326, 421, 425, 475
metaecosystem, 7, 475
metapopulation, 214, 325, 335, 372, 378–380, 475
niche, 214, 224, 326, 434
probability, 249, 383
scale transition, 25
shifting-mosaic steady state, 4
source–sink, 325

Tobler's first law of geography, 133, 205
Trajectory patch-specific, 405
Transition matrix, 329, 331, 333, 347, 353, 356, 385
Trophic interactions, 2, 215, 375, 428

U
Uncertainty, 42, 47, 48, 136, 141, 146, 156, 170, 177, 181, 183, 205, 221, 222, 251, 302, 403, 478

V
Validation samples, 248, 253
Variables
  covariate, 45, 164, 182, 298, 395, 463, 499
  dummy, 172
  response, 7, 9, 27, 47, 135, 136, 138, 141, 145, 164, 172, 173, 176, 179, 181–183, 189, 205, 224, 231, 246, 391, 432, 434, 435, 447, 448, 451, 458, 460
Variance partitioning methods, 463
Variogram
  cross-variogram, 142, 143, 150, 165, 436, 437
  empirical, 111, 139, 140, 143, 150–152, 154, 437
  model-based, 140, 141, 143, 150, 154, 166, 182
  multivariate, 165, 436, 437, 461, 462
  nugget, 140, 141, 152, 159
  practical range, 154
  range, 140, 152, 154, 159, 166
  sill, 140, 141, 152, 154, 159
  theoretical, 140, 151, 152, 154, 155, 157

W
Wavelet analysis, 47, 134, 159–162