Joint Species Distribution Modelling

Joint Species Distribution Modelling (JSDM) is a fast-developing field and promises to revolutionise how data on ecological communities are analysed and interpreted. Written for both readers with a limited statistical background and those with statistical expertise, this book provides a comprehensive account of JSDM. It enables readers to integrate data on species abundances, environmental covariates, species traits, phylogenetic relationships and the spatio-temporal context in which the data have been acquired. Step-by-step coverage of the full technical detail of statistical methods is provided, as well as advice on interpreting results of statistical analyses in the broader context of modern community ecology theory. With the advantage of numerous example R-scripts, this is an ideal guide to help graduate students and researchers learn how to conduct and interpret statistical analyses in practice with the R-package Hmsc, providing a fast-starting point for applying JSDM to their own data.

OTSO OVASKAINEN is Professor of Mathematical Ecology at the University of Helsinki, Finland. He has conducted research in mathematical, statistical and empirical ecology, with a particular focus on metapopulation ecology, movement ecology, population genetics, molecular species identification and community ecology.

NEREA ABREGO is a postdoctoral researcher at the University of Helsinki, Finland. After obtaining her PhD in fungal ecology, she expanded her research to general community ecology. She has conducted research in empirical, theoretical and statistical ecology, including recent developments in JSDM.

ECOLOGY, BIODIVERSITY AND CONSERVATION

The world's biological diversity faces unprecedented threats. The urgent challenge facing the concerned biologist is to understand ecological processes well enough to maintain their functioning in the face of the pressures resulting from human population growth. Those concerned with the conservation of biodiversity and with restoration also need to be acquainted with the political, social, historical, economic and legal frameworks within which ecological and conservation practice must be developed. The new Ecology, Biodiversity, and Conservation series will present balanced, comprehensive, up-to-date, and critical reviews of selected topics within the sciences of ecology and conservation biology, both botanical and zoological, and both 'pure' and 'applied'. It is aimed at advanced final-year undergraduates, graduate students, researchers, and university teachers, as well as ecologists and conservationists in industry, government and the voluntary sectors. The series encompasses a wide range of approaches and scales (spatial, temporal, and taxonomic), including quantitative, theoretical, population, community, ecosystem, landscape, historical, experimental, behavioural and evolutionary studies. The emphasis is on science related to the real world of plants and animals rather than on purely theoretical abstractions and mathematical models. Books in this series will, wherever possible, consider issues from a broad perspective. Some books will challenge existing paradigms and present new ecological concepts, empirical or theoretical models, and testable hypotheses. Other books will explore new approaches and present syntheses on topics of ecological importance.

Ecology and Control of Introduced Plants
Judith H. Myers and Dawn Bazely

Invertebrate Conservation and Agricultural Ecosystems
T. R. New

Risks and Decisions for Conservation and Environmental Management
Mark Burgman

Ecology of Populations
Esa Ranta, Per Lundberg and Veijo Kaitala

Nonequilibrium Ecology
Klaus Rohde

The Ecology of Phytoplankton
C. S. Reynolds

Systematic Conservation Planning
Chris Margules and Sahotra Sarkar

Large-Scale Landscape Experiments: Lessons from Tumut
David B. Lindenmayer

Assessing the Conservation Value of Freshwaters: An international perspective
Philip J. Boon and Catherine M. Pringle

Insect Species Conservation
T. R. New

Bird Conservation and Agriculture
Jeremy D. Wilson, Andrew D. Evans and Philip V. Grice

Cave Biology: Life in darkness
Aldemaro Romero

Biodiversity in Environmental Assessment: Enhancing ecosystem services for human well-being
Roel Slootweg, Asha Rajvanshi, Vinod B. Mathur and Arend Kolhoff

Mapping Species Distributions: Spatial inference and prediction
Janet Franklin

Decline and Recovery of the Island Fox: A case study for population recovery
Timothy J. Coonan, Catherin A. Schwemm and David K. Garcelon

Ecosystem Functioning
Kurt Jax

Spatio-Temporal Heterogeneity: Concepts and analyses
Pierre R. L. Dutilleul

Parasites in Ecological Communities: From interactions to ecosystems
Melanie J. Hatcher and Alison M. Dunn

Zoo Conservation Biology
John E. Fa, Stephan M. Funk and Donnamarie O'Connell

Marine Protected Areas: A multidisciplinary approach
Joachim Claudet

Biodiversity in Dead Wood
Jogeir N. Stokland, Juha Siitonen and Bengt Gunnar Jonsson

Landslide Ecology
Lawrence R. Walker and Aaron B. Shiels

Nature's Wealth: The economics of ecosystem services and poverty
Pieter J.H. van Beukering, Elissaios Papyrakis, Jetske Bouma and Roy Brouwer

Birds and Climate Change: Impacts and conservation responses
James W. Pearce-Higgins and Rhys E. Green

Marine Ecosystems: Human impacts on biodiversity, functioning and services
Tasman P. Crowe and Christopher L. J. Frid

Wood Ant Ecology and Conservation
Jenni A. Stockan and Elva J. H. Robinson

Detecting and Responding to Alien Plant Incursions
John R. Wilson, F. Dane Panetta and Cory Lindgren

Conserving Africa's Mega-Diversity in the Anthropocene: The Hluhluwe-iMfolozi Park story
Joris P. G. M. Cromsigt, Sally Archibald and Norman Owen-Smith

National Park Science: A century of research in South Africa
Jane Carruthers

Plant Conservation Science and Practice: The role of botanic gardens
Stephen Blackmore and Sara Oldfield

Habitat Suitability and Distribution Models: With applications in R
Antoine Guisan, Wilfried Thuiller and Niklaus E. Zimmermann

Ecology and Conservation of Forest Birds
Grzegorz Mikusiński, Jean-Michel Roberge and Robert J. Fuller

Species Conservation: Lessons from islands
Jamieson A. Copsey, Simon A. Black, Jim J. Groombridge and Carl G. Jones

Soil Fauna Assemblages: Global to local scales
Uffe N. Nielsen

Curious About Nature
Tim Burt and Des Thompson

Comparative Plant Succession among Terrestrial Biomes of the World
Karel Prach and Lawrence R. Walker

Ecological-Economic Modelling for Biodiversity Conservation
Martin Drechsler

Joint Species Distribution Modelling

With Applications in R

OTSO OVASKAINEN
University of Helsinki

NEREA ABREGO
University of Helsinki

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108716789

DOI: 10.1017/9781108591720

First published 2020

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data
Names: Ovaskainen, Otso, author.
Title: Joint species distribution modelling : with applications in R / Otso Ovaskainen, University of Helsinki, Nerea Abrego, University of Helsinki.
Description: Cambridge, United Kingdom ; New York, NY : Cambridge University Press, 2020. | Series: Ecology, biodiversity and conservation | Includes bibliographical references and index.
Identifiers: LCCN 2019042580 (print) | LCCN 2019042581 (ebook) | ISBN 9781108492461 (hardback) | ISBN 9781108716789 (paperback) | ISBN 9781108591720 (epub)
Subjects: LCSH: Biogeography–Statistical methods. | Ecology–Statistical methods. | R (Computer program language)
Classification: LCC QH84 .O93 2020 (print) | LCC QH84 (ebook) | DDC 577.2/2–dc23
LC record available at https://lccn.loc.gov/2019042580
LC ebook record available at https://lccn.loc.gov/2019042581

ISBN 978-1-108-49246-1 Hardback
ISBN 978-1-108-71678-9 Paperback

Contents

Preface *page* xi
Acknowledgements xiv

Part I Introduction to Community Ecology: Theory and Methods 1

1 Historical Development of Community Ecology 3
1.1 What Is Community Ecology? 3
1.2 What Is an Ecological Community? 4
1.3 Early Community Ecology: A Descriptive Science 6
1.4 Emergence of the First Theories 9
1.5 Current Community Ecology: Search for the Unifying Theory 11

2 Typical Data Collected by Community Ecologists 19
2.1 Community Data 20
2.2 Environmental Data 23
2.3 Spatio-temporal Context 24
2.4 Trait Data 26
2.5 Phylogenetic Data 27
2.6 Some Remarks about How to Organise Data 28

3 Typical Statistical Methods Applied by Community Ecologists 30
3.1 Ordination Methods 30
3.2 Co-occurrence Analysis 33
3.3 Analyses of Diversity Metrics 34
3.4 Species Distribution Modelling 35

4 An Overview of the Structure and Use of HMSC 39
4.1 HMSC Is a Multivariate Hierarchical Generalised Linear Mixed Model 39

4.2 The Overall Structure of HMSC 41
4.3 Linking HMSC to Community Ecology Theory 45
4.4 The Overall Workflow for Applying HMSC 47

Part II Building a Joint Species Distribution Model Step by Step 51

5 Single-Species Distribution Modelling 53
5.1 How Do Species Distribution Models Link to Species Niches? 53
5.2 The Linear Model 55
5.3 Generalised Linear Models 58
5.4 Mixed Models 63
5.5 Partitioning Explained Variation among Groups of Explanatory Variables 69
5.6 Simulated Case Studies with HMSC 70
5.7 Real Data Case Study with HMSC: The Distribution of *Corvus Monedula* in Finland 92

6 Joint Species Distribution Modelling: Variation in Species Niches 104
6.1 Stacked versus Joint Species Distribution Models 104
6.2 Modelling Variation in Species Niches in a Community 107
6.3 Explaining Variation in Species Niches by Their Traits 110
6.4 Explaining Variation in Species Niches by Phylogenetic Relatedness 114
6.5 Explaining Variation in Species Niches by Both Traits and Phylogeny 117
6.6 Simulated Case Studies with HMSC 120
6.7 Real Case Study with HMSC: How Do Plant Traits Influence Their Distribution? 133

7 Joint Species Distribution Modelling: Biotic Interactions 142
7.1 Strategies for Estimating Biotic Interactions in Species Distribution Models 143
7.2 Occurrence and Co-occurrence Probabilities 144
7.3 Using Latent Variables to Model Co-occurrence 147

7.4 Accounting for the Spatio-temporal Context through Latent Variables 152
7.5 Covariate-Dependent Species Associations 156
7.6 A Cautionary Note about Interpreting Residual Associations as Biotic Interactions 159
7.7 Using Residual Species Associations for Making Improved Predictions 160
7.8 Simulated Case Studies with HMSC 165
7.9 Real Case Study with HMSC: Sequencing Data on Dead Wood-Inhabiting Fungi 172

8 Bayesian Inference in HMSC 184
8.1 The Core HMSC Model 185
8.2 Basics of Bayesian Inference: Prior and Posterior Distributions and Likelihood of Data 187
8.3 The Prior Distribution of Species Niches 188
8.4 The Prior Distribution of Species Associations 197
8.5 The Prior Distribution of Data Models 206
8.6 What HMSC Users Need and Do Not Need to Know about Posterior Sampling 207
8.7 Sampling from the Prior with HMSC 210
8.8 How Long Does It Take to Fit an HMSC Model? 215

9 Evaluating Model Fit and Selecting among Multiple Models 217
9.1 Preselection of Candidate Models 218
9.2 The Many Ways of Measuring Model Fit 219
9.3 The Widely Applicable Information Criterion (WAIC) 225
9.4 Variable Selection by a Spike and Slab Prior 228
9.5 Reduced Rank Regression (RRR) 242

Part III Applications and Perspectives 253

10 Linking HMSC Back to Community Assembly Processes 255
10.1 Simulating an Agent-Based Model of a Competitive Metacommunity 256
10.2 Statistical Analyses of the Spatial Data Collected by a Virtual Ecologist 266

10.3 Statistical Analyses of the Time-Series Data Collected by a Virtual Ecologist 288
10.4 What Did the Virtual Ecologists Learn from Their Data? 297

11 Illustration of HMSC Analyses: Case Study of Finnish Birds 300
11.1 Steps 1–5 of the HMSC Workflow 300
11.2 Measuring the Level of Statistical Support and Propagating Uncertainty into Predictions 316
11.3 Using HMSC for Conservation Prioritisation 321
11.4 Using HMSC for Bioregionalisation: Regions of Common Profile 324
11.5 Comparing HMSC to Other Statistical Methods in Community Ecology 329

12 Conclusions and Future Directions 337
12.1 The Ten Key Strengths of HMSC 337
12.2 Future Development Needs 341

Epilogue 347
References 350
Index 369

The colour plates appear between pages 336 and 337

Preface

Species distribution modelling has become one of the most widely used tools in ecology, conservation biology and wildlife management. While methods for species distribution modelling are continually being developed, it is fair to say that the field itself is well established. Thousands of research papers have developed and applied statistical methods to map how the occurrence or abundance of species depends on environmental and spatial predictors. These methods have been summarised in several influential reviews and books, some of which are part of the Ecology, Biodiversity and Conservation series of Cambridge University Press (Franklin 2009; Guisan et al. 2017). However, the largest body of species distribution modelling literature concerns single-species models in which the response variable is the occurrence or abundance of a focal species. Compared to single-species distribution modelling, the methodological advances in multiple-species distribution modelling have lagged behind. When applying single-species models to data on multiple species, a separate model needs to be developed and validated for each species, making it challenging to model real communities consisting of many species. This is particularly difficult with regard to rare species, which are inherently common in most ecological communities. Furthermore, species do not live in isolation from each other, and thus viewing a community as a set of species that respond individualistically to environmental variation represents a major simplification. From the perspective of assembly theory in community ecology, biotic filtering is ignored when treating each species independently.

While species distribution modelling is routinely applied in single-species studies, the reasons outlined above make it less ideal for modelling species-rich ecological communities. Instead, the most widely applied methods in community ecology are ordination-based methods. Ordination methods were developed to enable the patterns in community composition to be summarised along spatial and environmental gradients. This is done by simplifying the high-dimensional structure of community data into few axes that explain the dominating part of the variation. While ordinations are very powerful for summarising the patterns in complex community data, they have limitations as well. Most importantly, ordination methods have been criticised for being of descriptive rather than predictive nature.

Both species distribution models and ordination methods are used to achieve the same general aim, namely to better understand the drivers controlling biodiversity across environmental gradients, space and time. Consequently, there is no reason why these two methodological fields should continue to develop independently of each other; rather, they could each learn from each other and eventually merge to combine their strengths. In recent years, statistical ecologists have taken the first steps in this direction, by developing the so-called joint species distribution modelling (JSDM) approaches. JSDMs build more heavily on single-species distribution models than on ordinations, as they involve a single-species distribution model for each of the species comprising the community. However, they are not a mere collection of single-species models: the species are not modelled independently of each other, but jointly, as with ordination methods. The 'joint' aspect of JSDM relates to both environmental filtering and biotic filtering. The responses of the species to environmental variation (i.e. environmental filtering) are assumed to have a joint structure that can depend on e.g. species' traits or phylogenetic relationships. This is achieved by a hierarchical model structure that involves both community-level and species-level parameters. The species' responses to each other (i.e. biotic filtering) are modelled through residual association matrices that describe the co-occurrence or co-abundance patterns that are not explained by environmental filtering.

The first JSDM approaches modelled species associations separately for each pair of species, and were thus feasible only for communities with few species (the number of pairs of species – and hence model parameters – becomes otherwise too large to be estimated). To overcome this limitation, the next generation of JSDMs applied latent variable approaches, making it possible to estimate association matrices also for communities with many species. This is where joint species distribution models have approached ordination methods. Namely, the latent variable approach is used to reduce the high dimensionality of community data. In fact, it can actually be viewed as model-based ordination. Therefore, JSDMs involve both species-specific distribution models and ordinations in their machinery, bringing these two fields closer to each other.

JSDM is currently one of the fastest developing fields in statistical ecology. While several kinds of JSDMs have already been implemented and successfully applied, the field is still in its infancy, especially compared to single-species modelling and ordination techniques. Consequently, the field of JSDM is currently experiencing much turbulence, with new approaches emerging at a fast rate and parallel developments of related approaches being simultaneously undertaken by different research groups. Some of these approaches may prove to be viable in the long run, while others may become superseded by improved approaches. While the ongoing rapid turnover of JSDM provides exciting possibilities, it also makes it difficult for their users to keep track of the pros and cons of the different approaches, and to gain an adequate understanding of their underlying assumptions and limitations. For these reasons, we considered it timely to devote an entire book to joint species distribution modelling, as this provided the possibility to present the conceptual, statistical and implementation aspects of JSDM in a much more profound and collective way than would be possible in focal research papers or software tutorials.

While several JSDM approaches and software implementations have been developed over the past decade, this book develops the argument of joint species distribution modelling from the point of view of one specific framework, namely Hierarchical Modelling of Species Communities (HMSC). However, as many of the existing JSDM approaches are closely related and can be considered as different branches of the same tree, we hope that this book will help deepen the readers understanding of the fundamentals of JSDM in general. In addition to presenting the conceptual, theoretical and statistical foundations of JSDM, this book also provides 'hands on' examples of how JSDM can be applied in practice. To this end, we build heavily on the R-package Hmsc; its use is demonstrated through R-scripts, and it has also been used to generate the majority of the figures/illustrations. Furthermore, we note that writing this book motivated us to implement some new features and extensions of HMSC, so some of the material here has not been published yet in research papers. We hope that the many R-scripts presented in this book (and the related online resources at www.helsinki.fi/en/researchgroups/statistical-ecology/hmsc) will provide a convenient starting point for a reader who wishes to apply JSDM for his or her own purposes.

Acknowledgements

This book builds on the development of Hierarchical Modelling of Species Communities (HMSC) that has continued over the past 10 years. Thus, we are thankful for the many researchers who have contributed to the work. One influential event that took place in the early phase of HMSC development was a research seminar in Helsinki in 2008, in which OO presented an approach that he had developed for species co-occurrence analyses in the context of fungal interactions. In this seminar, Janne Soininen asked whether the method could also be applied if the community matrix was transposed, to model joint responses of the species to environmental covariates instead of the responses of the species to each other. This resulted in Ovaskainen and Soininen paper entitled 'Making more out of sparse data: Hierarchical modelling of species communities', from which the HMSC approach derives its name. In this early phase of HMSC development, Guillaume Blanchet joined OO's group as a postdoctoral researcher, making many valuable conceptual and technical contributions.

Another influential event was a research visit to Duke University by OO and Guillaume Blanchet in 2013, hosted by Alan Gelfand. After a seminar given by OO, David Dunson pointed out some developments in modern Bayesian statistics that could be utilised to improve the applicability and computational efficiency of HMSC. This started a critically important and still continuing collaboration, which has resulted in the implementation of latent variable approaches to HMSC, as well as many other aspects that have made HMSC applicable to much bigger data than was possible before.

In 2013, Gleb Tikhonov started as a PhD student in OO's group. Gleb quickly became a key developer of HMSC, and defended his PhD thesis on this topic in 2018, with Alan Gelfand as the opponent. In addition to his numerous conceptual and statistical developments, Gleb made the very important contribution of leading the implementation of the R-package Hmsc (Tikhonov et al. 2020b). While the early versions of

HMSC were implemented first with Mathematica and then with Matlab, it became increasingly clear that an R-implementation would be needed for most ecologists to apply the method. The existence of the R-package is mainly thanks to the major efforts by Gleb. Another PhD student who made many contributions to HMSC was Anna Norberg, who also defended her thesis in 2018. While Gleb focused on developing the statistical approaches, Anna's main focus was on applying HMSC. This greatly aided the development and software implementations. In particular, Anna made the heroic effort of comparing the predictive performances of 33 single-species and joint species distribution models (Norberg et al. 2019), thus helping researchers assess the strengths and weaknesses of the many available approaches. Another key person who contributed to both the conceptual and implementation aspects of HMSC is Øystein Opedal, who joined the development team as a postdoctoral researcher with OO. More recently, Melinda de Jonge and Jari Oksanen also took part in the development of the R-package Hmsc, with major contributions in making the software more user-friendly and improving documentation. We also thank the many participants of the HMSC courses (organised in the context of the International Biometric Society meeting in Hobart in 2015, the European Congress of Conservation Biology in 2018 in Jyväskylä, the International Statistical Ecology Conference in 2018 in St Andrews and the Kaamos Symposium in Oulu in 2019) for their feedback, which has greatly contributed to the development of the approach itself, as well as the material presented in this book.

The participation of NA in the development of HMSC would have not been possible without the encouragement and support of her advisors. Back in 2013, Panu Halme promoted the collaboration, which resulted in some of the key papers in the development of HMSC (Abrego et al. 2017a; Ovaskainen et al. 2016a). Bernt-Erik Sæther gave valuable support while NA worked on the development of HMSC during her first postdoc, especially in the context of developing a time-series version of HMSC in collaboration with Steinar Engen and Vidar Grøtan (Ovaskainen et al. 2017a). Since 2017, NA has got the unconditional support of Tomas Roslin to continue collaborating on HMSC as her 'side project'; Tomas has also contributed to the development of HMSC himself (Ovaskainen et al. 2017b).

Writing this book was greatly facilitated by the support that we received from the publisher. Michael Usher, the series editor to Ecology, Biodiversity and Conservation, originally approached us about our interest in writing this book and encouraged us to do so. The senior commissioning editor Dominic Lewis helped us to develop the more detailed plan for the book. We are especially grateful to the editorial assistant Aleksandra Serocka, who made the process of writing the book as painless as possible by providing clear instructions and being very pragmatic and supportive during the entire process.

During the writing process, we received excellent feedback from many of our colleagues. In particular, we would like to thank Laura Antão, Melinda de Jonge, Janet Franklin, Elina Kaarlejärvi, Tuomas Kankaanpää, Janne Koskinen, Øystein Opedal, Iñaki Odriozola, Isabella Palorinne, Federico Riva, Bernt-Erik Sæther, Panu Somervuo, Tomas Roslin, Marjaana Toivonen, Jarno Vanhatalo and Helena Wirta for their invaluable comments on earlier versions of the book. Furthermore, we thank Aleksi Lehikoinen for providing the bird community data used as a case study in Chapter 11, and Atte Moilanen for performing the Zonation analyses presented in Section 11.3. We are especially grateful to Bess Hardwick for her major assistance in formatting the figures of this book and to Jacquelin De Faveri for her excellent work in revising the English language.

The writing of this book coincided with an especially busy period of our lives. To this end, we are most thankful to Argia Abrego, without whom this book would not have happened. Argia allowed us to focus on the work by having a great time with our baby for those three critically important weeks during which we wrote the first draft.