Predictive Species and Habitat Modeling in Landscape Ecology

C. Ashton Drew • Yolanda F. Wiersma
Falk Huettmann
Editors

Predictive Species and Habitat Modeling in Landscape Ecology

Concepts and Applications

Springer

Editors
C. Ashton Drew
North Carolina State University
Raleigh, NC
USA
cadrew@ncsu.edu

Yolanda F. Wiersma
Memorial University
St. John's, NL
Canada
ywiersma@mun.ca

Falk Huettmann
University of Alaska
Institute of Arctic Biology
Department of Biology & Wildlife
99775 Fairbanks Alaska
USA
fhuettmann@alaska.edu

ISBN 978-1-4419-7389-4 e-ISBN 978-1-4419-7390-0
DOI 10.1007/978-1-4419-7390-0
Springer New York Dordrecht Heidelberg London

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

CAD: To my son, Matthew, for his love, patience, and laughter.

YFW: To my husband John, for support, and to my sons William and Xavier, for inspiration and with hopes for beautiful landscapes for them to explore their whole lives through.

FH: With gratitude to my long-term co-workers Sergey Resvy, Nikita Chernetsov, Yuri Gerasimov, Igor Dorogoy, Sasha Solovyow, Aleksey Antonov, Misha Markovets, Olga Valchuck, Katya and Sasha Matsyna, Dima Lisityn, Andrey Koroblyov, and Pavel Ktitorov. Without your dynamic support, skill, time, effort, discussions and fun, most of this research and the underlying thoughts would not have evolved.

Foreword

Chair of the NASA-MSU Program, Former President of US-IALE
Rachel Carson Chair in Sustainability, Center for Systems Integration and Sustainability, Michigan State University
Jianguo (Jack) Liu

Landscape ecology is a relatively young but rapidly growing discipline. This phenomenal growth is partially due to the energetic "new blood" to the field – a young generation of dedicated scholars who are innovative thinkers, insightful theorists, skillful modelers and experimentalists, and enthusiastic promoters and practitioners of applying landscape ecology to natural resources management, planning, and policy-making processes.

The content of this book is an excellent example of scholarship under the leadership of some outstanding young landscape ecologists. Focusing on species distribution and habitat modeling, it contributes to landscape ecology theories, methods, and applications in many ways. The book integrates fundamental ecological theories into the model development processes, presents novel approaches to untangling the complexities of real-world systems, addresses variability and uncertainty in species–habitat associations across space and time, and discusses research frontiers. As species and habitats are important components of global biodiversity, models are particularly useful tools for predicting and explaining biodiversity distribution and dynamics. While the world is celebrating the International Year of Biodiversity (2010), and the United Nations is revising its 2010 biodiversity targets for the future, this timely book helps meet relevant global challenges, such as protecting biodiversity and providing sustainable ecosystem services for human well-being. Many valuable insights in the book have long-lasting impacts.

This book evolved from a symposium celebrating the tenth anniversary of the NASA–MSU Professional Enhancement Program. With support from the National Aeronautics and Space Administration (NASA) and Michigan State University (MSU), especially William Taylor, then the chair of the Department of Fisheries and Wildlife at MSU, I established the NASA–MSU Program to provide funds for some outstanding students to attend the 1998 meeting of US-IALE (US Regional Association of the International Association for Landscape Ecology). The awardees were selected based on their academic backgrounds, professional goals, abstracts for presentations, and financial needs. While at the meeting, they learned the latest developments in landscape ecology, contributed to the meeting by presenting and sharing their research, and established professional and personal networks for future collaborations. They were recognized at the conference banquet with award certificates. Also, a special dinner was organized for the awardees to interact with leaders in the field, such as Richard Forman, Frank Golley, Eugene Odum, Monica Turner, and John Wiens. The students benefitted enormously from interacting with those leaders and felt honored to meet scholars whose publications had influenced them. Meeting with leaders boosted students' confidence, further inspired their interest in landscape ecology, and helped their professional growth. After that meeting, John Wiens, then president of IALE, asked me to continue this program for the 1999 World Congress of Landscape Ecology that he was organizing. Since then, the NASA–MSU Program has become an integral part of the US-IALE annual meetings. The program has continued under three NASA managers (Anthony Janetos, Jon Ranson, and Garik Gutman); all of them have maintained a strong level of commitment to the program. There are now more than 270 NASA–MSU alumni around the world, and many of them have become leaders in landscape ecology and within US-IALE.

The symposium was organized by the three editors of this book – Ashton Drew, Yolanda F. Wiersma, and Falk Huettmann. They are NASA–MSU alumni and did a fantastic job in selecting an excellent group of speakers (most of whom were also NASA–MSU alumni). The room for the symposium was packed with a very engaged and enthusiastic audience. To benefit those who were not present, the symposium organizers decided to turn the symposium presentations into a book. Ten of the 13 chapters have a former NASA–MSU awardee as the lead author, and of all 28 chapter authors, 12 are NASA–MSU alumni. They are shining and rising stars of the field. With these and other promising scholars, landscape ecology continues to develop rapidly and play an increasing role in addressing important challenges to biodiversity and society around the world.

Acknowledgments

This book developed as a wide and international collaboration; consequently, it was supported by many “helpers” from all over the world. It presents the fruits from years of “thinking” and doing science. We are further grateful to all our chapter authors. Secondly, we would like to thank US-IALE for providing a fertile “breeding ground” for the development of the discipline and for the mentoring of young scientists and students. US-IALE kindly provided a professional platform and audience for our modeling symposium, and for this publication. Most of the book contributors are directly involved in IALE. We are grateful to Jack Liu for his work in creating and maintaining the NASA–MSU awards. This has created a nice cohort of Modeling Landscape Ecologists, many of whom came together at the Golley–Odum NASA–MSU symposium at the US-IALE meeting in 2007 (again, facilitated by Jack Liu). That symposium has formed the basis for the book, and many of the past NASA–MSU awardees who presented at the symposium are contributors here.

Production of this book would not be possible without the assistance of our many reviewers. All primary chapter authors provided a review of at least one, with some providing constructive feedback on as many as three chapters. In addition, many colleagues provided blind peer-review of individual chapters. Specifically, we would like to thank Thomas Albright, Ali Arab, Mike Austin, Janine Bollinger, Ray Dezzani, Alessandro Gimona, Thomas Gottschalk, Barry Grand, Joe Hightower, Jeff Hollister, Todd Jobe, Matthew Krachey, Cohn Loredo-Osti, Bruce Marcot, Jana McPherson, Rua Mordicai, Dirk Nemitz, Tom Nudds, Anatha Prasad, Sam Riffell, Kim Scribner, Mark Seamans, Darren Sleep, Erica Smithwick, John Wiens, Chris Wilke, and Miguel Zavala for providing peer-review of individual chapters. Thanks are also due to N. Dunn, who under supervision of YW and funding from a Memorial University Career Experience Program grant, assisted with copy editing and formatting of a number of chapters in this volume.

AD first recognizes the outstanding support and friendship of colleagues in the USFWS Eastern North Carolina and Southeastern Virginia Strategic Habitat Conservation Team, especially those that have served as experts and provided logistical and financial support for her postdoctoral research. She wishes to thank her supervisor J. Collazo at the NC Cooperative Fish and Wildlife Research Unit at NCSU for allowing her great flexibility to pursue her diverse academic and professional interests. Also within the Coop Unit, colleagues at the Biodiversity and

Spatial Information Center (BaSIC: A. McKerrow, S. Williams, C. Belyea, T. Earnhardt, M. Rubino, N. Tarr, A. Terando, J. White, M. Iglecia) have provided data and modeling assistance, invaluable advice, friendly reviews of very rough drafts, and generally an amazingly fun and productive work environment.

YW wishes to thank D. Urban and the group in the Landscape Ecology lab at Duke 2002–2003 for providing an excellent venue to be immersed in landscape ecology and habitat modeling. This "PhD sabbatical" was supported by the Fulbright Foundation. She is also grateful to her PhD supervisor and mentor T. Nudds for promoting thinking on philosophy of science and to colleagues M. Drever and D. Sleep for discussions on statistics. A final thanks to colleagues in the Biology Department at Memorial University, especially P. Marino for fantastic support of new faculty. Also thanks to J. Sandlos for unconditional love and support, and to my two sons for inspiration.

FH wishes to acknowledge the student power of his EWHALE lab at UAF, and his project collaborators in Alaska GAP (T. Gotthardt, S. Payare, A. Baltensberger), ArcOD (Arctic Ocean Diversity Project: B. Bluhm, C. Hopcroft, R. Hopcroft and R. Gradinger; funded by SLOAN Foundation), Global Mountain Biodiversity Project (GMBA, E. Spehn and C. Koerner; funded by DIVERSITAS), and the UAF Avian Influenza Center of Excellence (J. Rundstadler, G. Happ; funded by NIH). Further, FH likes to thank all GIS, data and modeling supporters and for brainstorming and feedback: T. Gottschalk, G. Humphries, M. Lindgren, S. Oppel, D. Nemitz, B. Dickore, E. Green, B. Best, GBIF, Eamonn O'Tuama, NCEAS, T. Peterson, B. Chen, M. Parson, T. Chapin, V. Wadley, J. Ausubel, T. Diamond, S. Cushman, S. Linke, A. Huettmann, I. Presse, Salford Systems Ltd team of D. Steinberg.

Finally, all three of the editors appreciate Springer NY Publishers, and especially J. Slobodien, for taking on this project and working with us toward completion. Final thanks to M. Higgs, our Production Manager at Springer, J. Quatela our Book Production Editor at Springer and to Marian John Paul of SPi Technology Services who oversaw production of the internal text.

Contents

Part IV Designing Models for Increased Utility

Abbreviations

AIC	Akaike's Information Criteria
AKN	Avian Knowledge Network
ANN	Artificial neural networks
AUC	Area Under Curve
BAS	British Antarctica Service
BBS	Breeding bird survey
BIC	Baysian Information Criterion
BRT	Boosted regression trees
CART	Classification and Regression Tree
DiGIR	Distributed Generic Information Retrieval
DIS	Digital Information Science
DOP	Dilution of precision
EIR	Entomological inoculation rate
EML	Ecological metadata language
FGDC	Federal Geographic Data Committee
GAM	Generalized additive model
GAP	Gap Analysis Program
GARP	Genetic Algorithms for Rule-set Production
GBIF	Global Biodiversity Information Facility
GCM	General Circulation Model
GDM	Generalized dissimilarity models
GEE	Generalized estimating equations
GIGO	"Garbage in, garbage out"
GIS	Geographic Information Systems
GLM	Generalized Linear Model
GPDD	Global Populations Dynamics Database
GPS	Global Positioning System
HSI	Habitat suitability index
IDE	Integro-Difference Equations
IPY	International Polar Year
ITIS	Integrated Taxonomic Information System
KML	Keyhole Markup Language
LTER	Long-term ecological research

MARS	Multivariate Adaptive Regression Spline
MAUP	Modifiable aerial unit
MDM	Mechanistic distribution models
MIGAP	Michigan Gap Analysis Program
NBII	National Biodiversity Information Infrastructure
NCEAS	National Centre for Ecological Analysis and Synthesis
NGO	Nongovernment organization
NIH	National Institutes of Health
NRLMP	Northern Region Landbird Monitoring Program
NSF	National Science Foundation
OBIS	Oceanic Biogeographic Information System
OOB	Out-of-bag
OPS	Overall prediction success
PCA	Principal component analysis
PCC	Percent correctly classified
PDA	Personal Digital Assistant
PDE	Partial Differential Equations
PDOP	Positional dilution of precision
PRISM	Parameter Elevation Regressions on Independent Slopes Model
RDBMS	Relational databases management system
ROC	Receiver Operator Characteristic
RPART	Recursive Partitioning
SML	Sensor Metadata Language
USFS	US Forest Service
USGS	US Geological Survey
XML	Extensible Markup Language

Chapter 1
Introduction. Landscape Modeling of Species and Their Habitats: History, Uncertainty, and Complexity

Yolanda F. Wiersma, Falk Huettmann, and C. Ashton Drew

1.1 Where Do We Come from?

From the start, the discipline of MODERN Landscape Ecology has focused on the interaction between spatial pattern and ecological processes. One area of focus has been to better understand how the patterns of environmental features, habitats, and resources (e.g., gradients, patches) influence patterns of species distribution. Some of the earliest predictive modeling papers in the journal *Landscape Ecology* dealt with predicting vegetation patterns based on topography (e.g., Bolstad et al. 1998; Ostendorm and Reynolds 1998). One of the early tools developed for wildlife management was "Habitat Evaluation Procedures" (HEP) (Schamberger 1982; Urich and Graham 1983; Mladenoff et al. 1995).

A suite of qualitative and quantitative models to predict habitat and species distributions followed. Landscape ecologists were also aware of the power of computer-based modeling to predict other ecological processes. The early 1990s saw the development of models to predict ecological processes in space and time. These included predicting forest stand dynamics (Keane et al. 1990) and the HARVEST model for predicting forest pattern as a result of disturbance (Gustafson and Crow 1996). There has been an increase in studies carried out to model species distribution and other ecological processes since the mid-1990s, largely the result of increased access to large data sets, advances in Geographic Information Systems (GIS) and remote sensing technology, new statistical models, and, most importantly, dramatic increases in computing power (see Fig. 11.1; Chap. 11). Models exist for two main reasons: to infer and explain ecological processes, and to predict future conditions or distributions and patterns in locations that have not yet been sampled. Some of the literature to date has focused on the latter function of models for prediction in space and/or time.

Landscape ecologists must make assumptions about how organisms experience and make use of the landscape. This is true whether models are being used to code

---

Y.F. Wiersma (✉)
Department of Biology, Memorial University, St. John's NL, A1B 3X9, Canada
e-mail: ywiersma@mun.ca

C.A. Drew et al. (eds.), *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*, DOI 10.1007/978-1-4419-7390-0_1,

behavioral rules for dispersal of simulated organisms through simulated landscapes, or for designing the sampling extent of field surveys and experiments in real landscapes. These convenient working postulates allow modelers to project the model through time and space, yet rarely are these predictions explicitly considered. Out of necessity, the early years of landscape ecology focused on the evolution of effective data sources, metrics, and statistical approaches that could truly capture the spatial and temporal patterns and processes of interest. Many texts already exist which deal with predictive modeling for both species distribution (Manly et al. 2002; Scott et al. 2002; Cushman and Huettmann 2010) and ecological processes, (Dale 2002) or in relation to resource management and conservation planning (Bissonette and Storch 2003; Millspaugh and Thompson 2009). The available literature offers detailed "how to" sections, with an emphasis on the statistical tools used to develop models (Franklin 2009). However, little of the available literature includes detailed information on the explicit link between underlying ecological theories and the structure of the models; or on the actual role of data, scale, and autocorrelation. Now that tools and techniques are well established, we are prompted to reflect on the ecological theories that underpin the assumptions commonly made during species distribution modeling and mapping.

This book will (1) highlight how fundamental ecological theories are being explicitly integrated into the model building processes for a robust inference approach, (2) offer practical examples of how modelers are addressing the conflict between the complexity of ecological systems and the relative simplicity of their modeled systems, and (3) present novel prediction methods and underlying philosophies to identify and quantify sources of uncertainty and variability in species–habitat associations in time and space, and to contribute progressive management applications.

1.2 Where Are We Going?

As recently completed PhDs who all had a modeling focus in their dissertations, it seemed to us that ecological theory was being largely ignored. We feel that modelers must take a moment to consider existing "first principles" of the ecology of an organism or process that they are trying to model. That motivated us to organize a special symposium at the 2007 meeting of the United States chapter of the International Society of Landscape Ecology (US-IALE) in Arizona. This book is a result of that international symposium, and many of the chapter authors were presenters at that meeting.

We set out to make the symposium and the resultant book more than a collection of individual presentations and case studies. To provide additional progress on the field of predictive modeling in space and time, we asked authors to reflect on the role of ecological theory in modeling, where they felt the field of modeling was going, and to identify some of the gaps that needed to be addressed to move the field forward. We do not intend this book to be a "how to model" guide (there are many excellent books and articles that already do this), but rather, we wanted participants in the symposium to offer their own thoughts and insights on potential pitfalls, highlight areas of confusion, and provide commentary on how they felt the field of modeling could be improved. Authors were encouraged to share potential solutions to ongoing modeling problems, and to identify those areas requiring further research. This volume is an attempt to promote thinking about the current challenges in predictive species/habitat modeling, including issues about dynamic ecological systems, stochasticity, complexity, data management, public use of models and data, and the need to include explicit ecological mechanisms.

This book is written primarily for graduate students and professionals and/or practitioners, especially those considering the use and/or application of models for the first time – although we feel that our review of ideas and concepts is of value to veteran modelers as well. For the new modeler, the field may seem daunting with its technical language, seemingly arcane debates over statistical methods, and overwhelming data sets. However, we remind readers that models are all around us. Nearly everyone interacts with models daily when they consult the local weather forecast. Weather forecasting is based on a complex set of models and carries with it uncertainties and variation in predictive accuracy. Yet, most of us feel comfortable with the uncertainty inherent in a 5-day weather projection and are content to be largely oblivious to the myriad meteorological models underlying the report we consult. On a larger scale, the International Panel on Climate Change (IPCC) is the international body that reports on climate models and makes projections regarding the impact that varying climate change scenarios may have on the earth's ecosystems and societies. These models are much less trusted by the lay public. The recent "climategate" scandal, wherein the emails of some IPCC scientists at East Anglia University were leaked to the media, highlighted the mistrust that many still have in models. While fellow scientists recognized and were largely comfortable with the uncertainty in the models expressed in the leaked emails, some elements of the media promoted the idea to the lay public that the scientific uncertainty was a sign that climate models were completely unreliable. When models are – or are perceived to be – unreliable, it can be difficult to use them for decision making, management, and policy formation.

Several key themes emerge through this volume. In some cases there is congruence between chapters, in others there is conflict. We hope the conflicts and apparent contradictions do not discourage prospective landscape modelers. Rather, we feel the disagreements in perspective between the authors of some chapters are part of the scientific discourse and reflect a healthy, dynamic, and still-growing field. These areas of disagreement or uncertainty represent diverse areas that are ripe for more research and offer up new hypotheses for testing.

1.3 Key Themes

One of the key themes, and the one which prompted us to organize the symposium that was the foundation for this book, is the issue of mechanistic versus predictive models and the tradeoffs inherent in choosing one over other. Mechanistic models have traditionally been perceived as useful for testing ecological hypotheses, but whether models need to be mechanistic to make accurate predictions is an open question.

The tradeoff between simplicity (parsimony) and complexity for increased predictive performance has emerged as a key debate in the modeling literature, and also in science as a whole. This book is focused on *predictive* habitat–species distribution models. Prediction in general is a key tenet of the scientific method; prediction allows one to generalize, and forms the very basis of statistical hypothesis testing. Thus, according to scientific standards, models (i.e., hypotheses) should yield inferences and results that can be generalized to other points in space and time; that is, they should predict well, or to the best possible degree. Principles of parsimony suggest that simplicity in models will yield better generalization (Burnham and Anderson 2002).

Thus, a key question that remains is whether the intent of models is to be explanatory, predictive, or both. While the focus of this volume has been on predictive models, and notwithstanding the importance of prediction as part of the scientific process, it is entirely possible to have accurate predictive models with no underlying biological or mechanistic explanation.

The available data also influence the models and methods of accuracy assessment – another key theme in this volume. To adequately assess predictive power, accuracy assessment, and the use of alternative data is essential. Management decisions based on models should probably not be done without quantifying model uncertainty. The available data are the essential foundation for choosing the method of accuracy assessment to use (e.g., reserving validation data vs. leave-one-out methods such as bootstrapping). When using models in management and decision-making, it is also important to consider situations where models are created based on and assessed with existing models. Errors add up, and can behave in a chaotic fashion. Such concepts are known to landscape ecologists from other applications (e.g., percolation theory, landscape metrics) but have not been widely discussed in the context of predictive habitat modeling.

Several other themes throughout this volume hinge on process and technical issues. These are areas which are still quite open to debate, but we feel that highlighting them opens up avenues for further research, provides guidelines for starting a modeling exercise, and highlights some of the important considerations managers and end-users must take into account when implementing the results of modeling exercises. These issues include statistical and model selection decisions (see Chap. 2 for an overview of available tools), and considerations around open access data sharing and metadata (see Chap. 3). Debates are not always resolved. Whereas chapter authors do not always agree on statistical tools, the authors have clarified their thought process around particular statistical choices made in their case examples so that readers may apply these considerations in their own decision-making. Most authors agree that the statistical choices should reflect the question being asked, and that suitability is somewhat predicated on predictive accuracy. That is, if you can build a model that will predict well using a standard statistical model, there is no need to apply a more recent or apparently more sophisticated model. On the other hand, there is valid argument for work to explore better models and to use data, software, and computer power to build on and further improve existing models.

A related consideration is the issue of sampling design. Sampling considerations are important for data management and accuracy assessment. Many statistical textbooks that deal with sampling considerations still assume that samples are used for frequentist-based testing. We are not aware of any texts that address sampling design for predictive modeling which employ alternative statistical methods (e.g., neural networks or random forests).

As illustrated by some of the controversy around climate modeling, model literacy is important. Increased literacy around the field of predictive species/habitat modeling will also increase the ability of the scientific community and the public to offer constructive critiques of models as they appear. Here we have identified some key elements necessary to improve the use of models (i.e., comprehensive data management, appropriate sampling and statistical methods, clear and transparent processes, robust accuracy assessment, online delivery, and effective communication of model results including model uncertainty). Models are likely to be at the forefront of legislative and judicial processes in the future; thus, transparent and robust models that are based on the best available science will be the most effective. It is our hope that this volume will contribute to more of such models.

1.4 Organization of the Book

This book is organized into four sections. The first section outlines the current state of knowledge in the field of modeling. Chapter 2 by Kerr et al. provides a perspective on the ecological issues germane to these types of models. In Chap. 3, Hooten outlines the statistical landscape, and Zuckerberg et al. offer some thoughts on the challenges of acquiring and managing large data sets in Chap. 4.

The second section, "*Integration of Ecological Theory into Modeling Practice*," outlines how we can ensure that ecological first principles remain at the forefront of modeling. Laurent et al. provide a thoughtful overview of some of the common ecological assumptions inherently made in modeling in Chap. 5. In Chap. 6, Fletcher et al. look at how predictive models can also be tests of ecological theory dealing with temporal processes.

The third section of the book, "*Simplicity, Complexity, and Uncertainty in Applied Models*," highlights some key issues that the new modeler may want to consider. This is the most "how to" section of the book, and probably of most interest to those attempting to develop models for the first time. Although the chapters are not structured as instructive "recipes," they highlight specific issues, offer suggestions and references for further reading, and through case studies, illustrate how some of these issues might be addressed. In Chap. 7, Lookingbill et al. address pattern-based issues through a discussion highlighting how focused assessment techniques can help identify scale-dependent vegetation pattern. Evans et al. deal with process-based issues and provide detail on machine-learning approaches through a treatise on "Random Forests" in Chap. 8. In Chap. 9, Murphy and Evans follow with a second chapter addressing process-based issues, this time looking at the process of gene flow across landscapes. Huettmann and Gottschalk look at temporal and accuracy issues and engage in a debate on simplicity versus complexity in Chap. 10. Wiersma provides some perspective on various statistical issues in Chap. 11, and the implications of choosing one statistical model over another through a meta-analysis of statistical models applied in the literature over a 10-year period. Finally, in Chap. 12, Drew and Perera round out this section with a discussion on expert opinion and how it can be integrated with modeling to address knowledge and data issues.

The final section of the book, "*Designing Models for Increased Utility*," provides two examples. Roberts et al. (Chap. 13) show how models have been applied to the Gap Analysis Program (GAP) in Michigan to support management decision making, and Lawler et al. look at how models have been used in strategic conservation planning in Chap. 14. The book concludes with a chapter by the editors synthesizing key findings, and pointing to areas for future research. Although the fields of landscape ecology and predictive species/habitat modeling are rapidly evolving, it is our hope that many of the chapters in this volume will remain timeless, and help to set new standards. It is our wish that both the new and veteran modeler and landscape ecologist will find valuable insights in the reflections on the state of the practice of predictive modeling that we and our collaborators and colleagues offer in the pages of this book for advancing the science of landscape ecology world-wide.

References

Bissonette JA, Storch I (eds) (2003) Landscape ecology and resource management: linking theory with practice. Island Press, Washington DC.

Bolstad PV, Swank W, Vose J (1998) Predicting Southern Appalachian overstory vegetation with digital terrain data. Landsc Ecol 13:271–283.

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edition. Springer, New York.

Cushman S, Huettmann F (eds) (2010) Spatial complexity, informatics, and wildlife conservation. Springer, Tokyo.

Dale VH (2002) Ecological modeling for resource management. Springer, New York.

Franklin J (2009) Mapping species distributions: spatial inference and prediction. Cambridge University Press, New York.

Gustafson EJ, Crow TR (1996) Simulating the effects of alternative forest management strategies on landscape structure. J Environ Manage 46:77–94.

Keane RE, Arno SF, Brown JK, Tomback DF (1990) Modelling stand dynamics in whitebark pine (*Pinus albicaulis*) forests. Ecol Modell 51:73–95.

Manly BFJ, McDonald LL, Thomas DL, McDonald TL, Erickson WP (2002) Resource selection by animals: statistical design and analysis for field studies, 2nd edition. Kluwer Academic Publishers, Dordrecht.

Millspaugh JJ, Thompson FR Jr (2009) Models for planning wildlife conservation in large landscapes. Academic Press, Burlington MA.

Mladenoff DJ, Sickley TA, Haight RG, Wydeven AP (1995) A regional landscape analysis and prediction of favorable gray wolf habitat in the Northern Great Lakes. Conserv Biol 9:279–294.

Ostendorf B, Reynolds JF (1998) A model of arctic tundra vegeation derived from topographic gradients. Landsc Ecol 13:187–201.

Schamberger M (1982) Habitat: a rational approach to assessing impacts of land use changes on fish and wildlife. Env Pro 4:251–259.

Scott JM, Heglund PJ, Morrison MM, Haufler JB, Raphael MG, Wall WA, Samson FB (eds) (2002) Predicting species occurrences: issues of accuracy and scale. Island Press, Washington.

Urich DL, Graham JP (1983) Applying habitat evaluation procedures (HEP) to wildlife area planning in Missouri. Wildl Soc Bull 11:215–222.

Part I
Current State of Knowledge

Chapter 2
Integrating Theory and Predictive Modeling for Conservation Research

Jeremy T. Kerr, Manisha Kulkarni, and Adam Algar

2.1 Introduction

The need for effective techniques to predict how global changes will alter biological diversity has never been greater and continues to increase (Buckley and Roughgarden 2004; Thomas et al. 2004). Although accelerating climate and land use changes loom especially large, extinction rates have risen as a result of other types of threats as well – such as overkill and pollution. Individually, each of these perils is serious, but it is through their additive and sometimes synergistic interactions that the world is now in the midst of a sixth mass extinction (Wake and Vredenburg 2008).

Predicting responses of individual species or biological communities to the shifting environmental conditions with which they are increasingly confronted is the business of models (see Scott et al. 2002). Those predictions may focus on many different biological responses, including population fluctuations, shifting distributions, or collective changes in the species richness of communities (Woodward 1987; Root 1988; Parmesan 2005; White and Kerr 2006). Models may adopt spatial or temporal perspectives, or combine both. Regardless of the purpose, application, or spatio-temporal perspective intrinsic to models, the criteria for determining their utility can be narrowed down to two, alternative questions: did the model provide successful predictions and, if not, did it fail in interesting ways? In the former case, models are more likely to be helpful for policy-makers but much can be learned in the latter.

Although many modeling systems exist to build these predictions, few that incorporate extensive mechanistic justification are in wide use. There are at least two good reasons for this state of affairs. First, predictive models that lack strong mechanistic backup have proven surprisingly capable (Araújo and New 2007). Second, complex models with detailed biological underpinnings may include terms that are impractical to measure. So, higher model sophistication comes with a cost

---

J.T. Kerr (✉)
Canadian Facility for Ecoinformatics Research, Department of Biology,
University of Ottawa, 30 Marie Curie, Ottawa ON Canada, K1N 6N5
e-mail: jkerr@uottawa.ca

C.A. Drew et al. (eds.), *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*, DOI 10.1007/978-1-4419-7390-0_2,

in terms of ease of application but also can bring benefits, particularly in the form of greater reliability as environmental conditions change (Buckley 2008). Models built on such detailed theoretical underpinnings are not necessarily more effective, nor are models without such foundations more likely to be ineffective. For any model, from the theoretically derived to the purely phenomenological, utility is defined by the ability to test and the resultant tested predictions that may be used to inform policy (e.g., Dillon and Rigler 1974).

Here, we consider theoretical perspectives on modeling the distributions of species across landscapes and regions. Predicting spatial distributions of species and how those distributions change over time has been possible without using detailed mechanistic models; however, theoretical contributions to these models may improve their reliability as environmental conditions change. Different degrees of theoretical support for predictive models suggest an array of strategies to maximize model reliability for policy application (Kerr et al. 2007). We bring these together in a single framework that addresses the role of theory in model construction and the reliability of applications that result. Many examples demonstrate the importance of theoretical contributions to predictive modeling but that those models can often be successful even in the absence of such contributions.

2.2 Ecological Theory and a Framework for Predictive Modeling

2.2.1 Integrating Data, Testing, and Theory for Predictive Modeling

As the theoretical complexity of models increases, it becomes increasingly difficult to test their reliability across a range of environmental conditions. Collecting new data to test elaborate hypotheses about biological processes can quickly become impractical, whereas an able theoretician faces fewer constraints in pursuing increasingly sophisticated models able to detail mechanisms (McLean and May 2007). This does not suggest that very detailed models fail to illuminate potentially important biological processes, but if those models cannot be tested, their application to real world situations will be hampered. The point at which increasing model complexity, via the introduction of untestable mechanisms that may not actually be operating in the system in question, actually begins to *decrease* the reliability of real-world predictions could, perhaps, be viewed as the point beyond which modeling efforts serve a reduced practical purpose. Conservatively speaking, that point could be reached when data can no longer be collected to test additional model details.

The interaction between theoretical and empirical approaches to modeling is frequent and mutually indispensable (Fig. 2.1). In fact, most researchers work somewhere in the space between these extremes and use theory (or experimental observations) to inform data collection, and field observations to refine existing theory. The objective of this interplay is to produce reliably predictive models. It is possible to make progress on purely empirical grounds, or similarly on purely theoretical grounds, but rapid advances made based on such grounds may decrease the chances that scientific progress can reliably be translated into policy action.

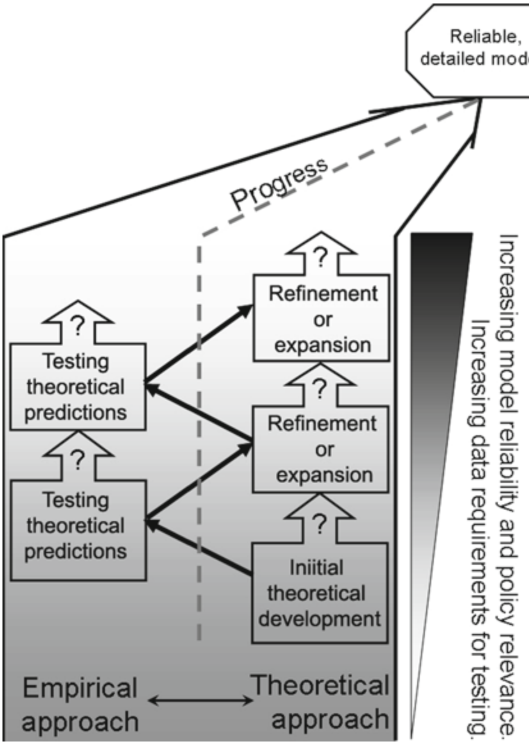

Fig. 2.1 One view of the interplay between theoretical and data-driven (empirical) approaches to predictive modeling (Box 1979; Austin 2002). In this framework, an initial theoretical model yields predictions that can be tested empirically. Results from those tests can then be used to refine (possibly refute) or expand models in terms of theoretical content, which may mean specification of additional biological processes. Scientific "progress" is expanded capacity to predict species responses to environmental change, across some spatial gradient (e.g., of habitat fragmentation) or through time. Discoveries do not require this interplay: progress can be driven entirely by empirical or theoretical perspectives, but such approaches may be less certain. With increased model detail and reliability, it may be possible to contribute to the development and implementation of policy, which are most likely to be implemented according to adaptive management principles, and not as a one-way process. Data requirements are likely to increase with increasing model detail, which can inhibit hypothesis testing. Hypothesis (model) testing can take many forms, including experimentation and simulations, but real-world use of models should include real-world tests wherever possible

2.2.2 Case Study: Butterfly Models Using Mechanistic Knowledge

In recent years, much work has focused on the impacts of observed climate changes on species distributions (Thuiller 2003; Peterson et al. 2004; Luoto et al. 2005; Guisan et al. 2008). Butterflies have proven to be a useful focal taxon in this kind of global change research (Dennis 1993; Wilson et al. 2005; Menendez et al. 2006). Because most butterflies, along with most other animals, are non-migratory, they must be able to tolerate a full year's range of environmental conditions or face local extinction. In northern environments, where climates are changing rapidly, butterflies cannot use behavior to avoid extreme winter conditions. Instead, they must be freeze tolerant. Keeping in mind that models predicting how changing climates will alter species distributions would be valuable, it would be helpful to know the mechanism limiting the northern distributions of these species. Possible mechanisms, on the ecology side, include limits imposed by host plant distributions or temperature-dependent foraging, whereas lab-based physiological research has demonstrated the molecular basis for freeze tolerance. If temperatures drop too far below the supercooling point of an individual, its antifreezes are no longer effective and its cells lyse as they freeze solid. This mechanism suggests an empirical step toward predicting climate change effects on butterfly distributions. Kukal et al. (1991) took a step toward quantifying this empirical step, and provided field experimental evidence – supported by molecular observation – that suggested the northern limits among at least two swallowtail butterflies (the Tiger Swallowtail, *Papilio machaon*, and Canadian Swallowtail, *P. canadensis*) were limited by minimum winter temperatures. If those temperatures drop sufficiently, mortality rates increase sharply. This work, in turn, supports the theoretical mechanism linking temperature to range limits for these butterflies, leading to further prediction on the empirical side of the equation: warming temperatures cause range expansion among butterflies. Species distribution models (Anderson et al. 2003; Phillips et al. 2006) can then be constructed. Consequently, range expansion among butterflies in Canada has been shown to track observed climate changes in the twentieth century (White and Kerr 2006; Kharouba et al. 2009). Species distribution models were validated with data that were spatially and temporally independent from data used to construct models, increasing their reliability for policy application.

From such examples, it is possible to generalize one approach to the modeling process beginning with a theoretically derived hypothesis at the outset, leading to predictions that are tested empirically, and an expansion or modification of the starting hypothesis which leads to further empirical evaluation, and so on (Fig. 2.1). The goal in this process is to construct reliable, predictive models for a biological phenomenon of interest, such as how climate changes will alter species' distributions. At each step in this process, the requirements for data increase, but so does the likelihood that modeling results up to that point could reasonably be used for policy decisions.

2.3 Models Missing Mechanisms

Predictive models of species distributions across landscapes and regions commonly exclude detailed mechanistic information, but their predictions have often proven quite accurate. Some of the most influential observations that species are actually shifting their ranges and phenological timing in the direction expected given climate change include relatively little environmental measurement (Parmesan 1996; Parmesan et al. 1999; Root et al. 2003). Although these efforts did not use predictive modeling, they provided critical evidence that biological responses to shifting environmental conditions were likely. Recent efforts link observed changes to environmental drivers (Rosenzweig et al. 2008).

2.3.1 Case Study: Madagascan Chameleons

Predictive models have had some surprising successe even when mechanisms are not directly considered. Raxworthy et al. (2003) used a common niche modeling technique (Genetic Algorithms for Rule-set Production; GARP) to estimate possible geographic ranges for chameleon species in Madagascar (Fig. 2.2). In this study, point observations for chameleon species were available in much of Madagascar, but GARP niche models consistently predicted some chameleon species could be present in a disjunct area from which no data had been collected. When that area was surveyed, several new chameleon species were discovered that were sister species to those that had been modeled. Missing from these GARP models were any considerations of the mechanisms that might be necessary for speciation – including isolation, the presence of dispersal barriers, mutation rates, or selection pressures. Predictive models can clearly succeed under some circumstances even without consideration of mechanism, although these models are strongly driven by expert knowledge of the niche requirements of chameleons.

Fig. 2.2 The modeled distribution (in *gray*) of *Furcifer pardalis*, a chameleon in Madagascar, included a large area in the eastern half of the country. This area (shown in *white*) was predicted to be suitable for the species and contained all historical observations. The model, developed using GARP, also predicted that this species should be present in a previously unexplored area (indicated by a *question mark* within a *black-shaded ellipse*) to the west of its known range (Raxworthy et al. 2003). When that unexplored area was surveyed, new chameleon species were discovered, including a sister species to *F. pardalis*. Although the GARP models linking environmental characteristics to chameleons' distributions in Madagascar did not specify biological processes or mechanisms that would generate the species' distribution, they provide a strong indication that such models can still generate surprising, useful, and correct predictions

It is even possible to predict the spatial distributions of species without any biological information at all. The simple reason this is possible is that environmental factors are spatially autocorrelated. This fact has long been recognized by ecologists, who have responded to the problems (and opportunities) that spatial autocorrelation presents with a series of increasingly sophisticated statistical tools, including adjustments to make probability tests more conservative, and the control of biases in coefficient estimation in regression models (see Legendre and Legendre 1998; Koenig 1999; Diniz-Filho and Hawkins, 2003). Commonly, spatial autocorrelation of environmental factors is considered after first investigating the potential, main effects of environmental variables thought to affect a biological response of interest, such as a species' distribution. Bahn and McGill (2007) take the reverse approach, simulating a series of species' ranges and modeling purely spatial predictors of those ranges, then testing whether adding environmental variables, after applying purely spatial variables, improved predictions of species ranges. Their result was striking: spatial structure did a better job of predicting simulated species ranges than mechanistic approaches. In other words, it is possible that spatial models attempting to predict the distribution of species may sometimes do so purely because of the underlying spatial structure of the species' range, not because the model successfully captures any biological mechanisms.

The view that prediction of a species' range requires no biological information is incorrect. Bahn and McGill (2007) argue that these models have little capacity to predict distributional shifts in changing environments. In fact, this is not strictly true: spatial models that do not depend on environmental conditions predict that distributions will remain relatively constant when those conditions change. Clearly, this proposition is refuted by many observations. Environmentally based models have made accurate predictions of distribution expansions and contractions, at least for some taxa (e.g., Kharouba et al. 2009). However, Bahn and McGill's warning should be heeded by those who predict distributional shifts. Observed correlations between environmental variables and species distributions may arise in whole or in part from spatial autocorrelation rather than mechanistic links. For example, Algar et al. (2009) have shown that accounting for potential confounding effects of spatial autocorrelation while training environment-diversity models results in more accurate predictions of diversity shifts in response to twentieth century climate change than similar models that ignore spatial autocorrelation.

2.4 Testing Spatial Models through Time

Of course, the search for mechanisms goes on, not least because of two problems created by ignoring them. First, if predictions succeed, the reason of "why" is a mystery in the absence of some mechanistic understanding. Second, it is impossible to distinguish correlation and causation. These problems reduce expectations that a model developed in one place may be applied in another geographic location or that model predictions will be continue to be effective as environmental conditions change. It is probably self-evident that strong theoretical thinking should support predictive modeling efforts, especially when models include a temporal component. Unfortunately, predictions of change through time often rely on models observing change across space.

2.4.1 The "Space-for-Time" Assumption

Ecologists recognize the perils of the "space-for-time" assumption; that is, using purely spatial models to predict changes through time. The reason many authors have been forced to make this risky assumption is the dearth of long-term data that can be used to build predictive models that include temporal components. In the case of species distribution modeling, the absence of historical species observations and/or environmental data often limits model construction to the purely spatial. Spatial environmental gradients within the species' range are then assumed to indicate how that species responds to environmental differences. Expectations of environmental change into the future can then be translated into predictions of biological impacts. Modeling techniques have proven highly capable for fitting spatial relationships between environmental data and observational data of species on the ground, and statistical assessments make them seem very accurate. When run through time, however, model behavior can be grossly unreliable. Using extensive datasets on the environmental factors that South African Protea are known to respond to, Pearson et al. (2006) developed highly accurate spatial models predicting each species' range using nine different modeling methods. When anticipated environmental conditions for 2030 were substituted into the different models, they returned wildly divergent predictions of biological responses for the same species, ranging from 92% loss of current modeled range to more than 300% expansions. Species-by-species modeling approaches might be more prone to return idiosyncratic results, but variability for individual species might diminish in relative importance when large species assemblages are considered. However, one study focusing on an assemblage of about 150 of the most thoroughly studied butterfly species in Canada found that the relationship between human population density and butterfly species richness was positive spatially, but negative when tested temporally (White and Kerr 2006). The message for models with only partial mechanistic support appears to be that temporal tests may be essential.

2.4.2 Global Change as a Pseudo-experiment

Fortunately, climate and land use changes during the twentieth century provide many pseudo-experimental opportunities to test otherwise spatial links between environmental factors and the distribution of species. Butterflies in Canada again provide very effective examples for a few reasons. First, long-term data are available for observations of many of these species, as are historical environmental data for factors that physiological ecology literature (e.g., Kukal et al. 1991) suggests should influence species ranges. Importantly, climate and land uses have changed in much of Canada during the twentieth century. In short, it is possible to test whether models

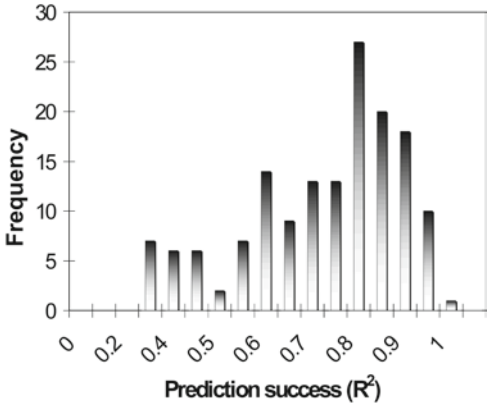

Fig. 2.3 Spatial models predicting species distributions may not work when used to predict distribution shifts through time (i.e., the "space-for-time" assumption may not hold), but a study of 139 butterfly species in Canada demonstrates that such predictions can be reliable (Kharouba et al. 2009). This histogram shows the distribution of conditional autoregressive R^2 values comparing predicted range shift to observed range shift during the twentieth century. For most species, models were sufficiently accurate to advise policy, but some predictions were erratic. Using large species assemblages reduces the impact of noisy predictions and creates a strong signal on which to base decisions. Observed, spatial correlations between environment and species distributions appear causative in this case

remain accurate through time using known climate and land-use changes as a kind of pseudo-experiment. That is, historical distributions can be estimated using models (in this case, Maximum Entropy). As environmental conditions change, the effects on species distributions can then be forecast using the same models, and actually tested using independent data collected from the present day. Models projected through time using this framework are generally accurate (Fig. 2.3), as would be expected if their spatial analogs specify environmental determinants of range correctly. Models developed for some species fail, on the other hand, suggesting that these techniques should only be applied to large species assemblages. Inferences around mechanisms are possible from purely spatial models, but adding a temporal component to them demonstrates whether projections of future conditions have any likelihood of being accurate. Macroecological research has few enough experimental opportunities: when they arise, it is essential to use them.

2.5 Theoretical Perspectives on Predictive Modeling

2.5.1 Niches, Neutrality and Predictive Models

Theoretical contributions to predictive modeling can include nearly any biological ingredient that affects the likelihood that a species will be present in a particular locality. Specific mechanisms may include physiological tolerances to the range of local environmental conditions (e.g., temperature, moisture, wind) and the capacity to survive negative biotic interactions (e.g., competition, parasitism, predation). Dispersal capabilities determine whether species can colonize suitable habitats (Thomas 2000) or rescue failing populations that are already present there (Brown and Kodric-Brown 1977). Species differ in all of these respects, but these differences may not always be essential to the development of models that successfully predict species distributions and responses to environmental change (Bell 2001). A conceptually simple modeling approach that could be applied to assemblages of species in a region that rely on similar trophic strategies is to ignore admitted differences among species entirely. Despite such a massive simplifying assumption, neutral theory is able to generate predictions for biological parameters that are essential for landscape ecologists – including extinction rates, abundance patterns, and species richness. Where neutral theory fails, additional complexity around biological mechanisms or species-specific data may be necessary, which is an area for hybrid theories that incorporate both neutral and niche-based perspectives. More sophisticated still are detailed mechanistic models that rely explicitly on taxon-specific physiology, dispersal, and trophic strategies to generate detailed predictions of the dynamics of species across landscapes. Other models for single species, which draw from metapopulation dynamics and habitat selection, are detailed elsewhere in this volume (*see* Chapter 6).

The niche concept, which lies at the core of ecological research, argues that each species has a unique, *n*-dimensional array of environmental tolerances and resource needs (Hutchinson 1957). Consequently, predictive models are strongly and justifiably influenced by the mechanisms that species' niches imply, such as differences in dispersal rates, physiological tolerances, and details of metabolism. Models often begin with substantial sophistication because such mechanisms certainly exist and are thought to determine species abundances and distributions. The distributions of individuals and species among habitat patches in fragmented landscapes, for instance, result from such niche-based predictions.

Neutral theory ignores differences in dispersal, birth, and death rates among all individuals within a community, a perspective imported from population genetics (Bell 2001). Beginning with the observation that biotic communities are saturated with individuals, neutral theory shows that differences in the properties of those communities through time can arise through stochastic processes of ecological drift rather than niche-based determinism (*see* Hubbell 2001 for a detailed discussion). Then, with information on the number of species and individuals within the regional meta-community, the spatiotemporal dynamics of that community can be predicted with remarkable accuracy. Neutral theory provides a theoretical framework predicting rank abundance distributions among biotic communities that range from log-series (highly inequitable, with nearly all individuals in the community drawn from a single species) to broken stick (relatively equitable distribution of individuals among species). It also provides a dispersal-based explanation for differences in the slope of species-area relationships observed in mainland areas relative to islands (i.e., species richness increases with area more slowly on mainlands because dispersal rates are higher in such areas, leading to lower beta diversity) and for the positive correlation between range-size and abundance. Long-term modeling can