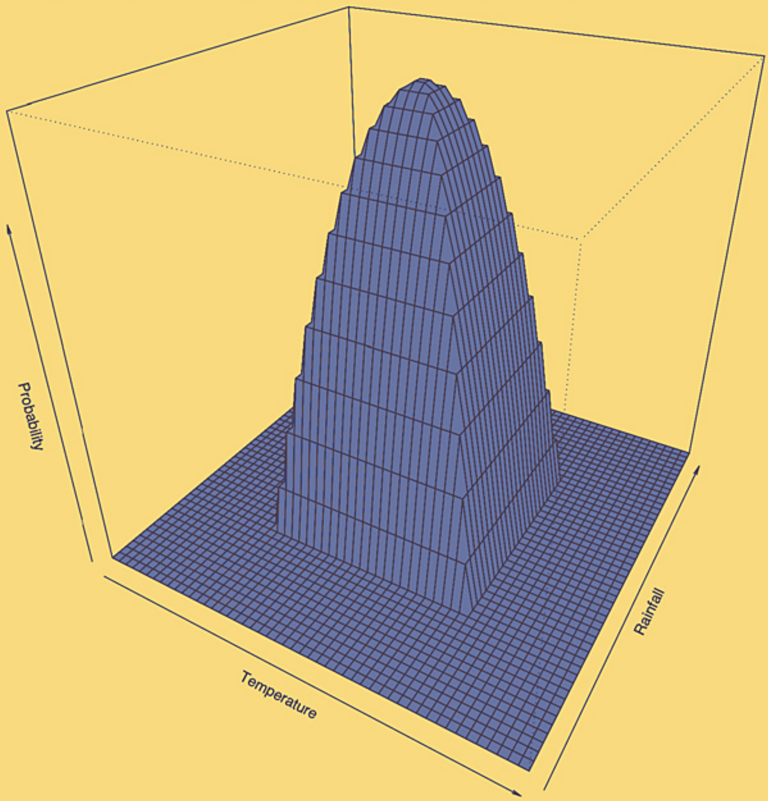
Chapman & Hall/CRC
Mathematical and Computational Biology Series

Niche Modeling

Predictions from Statistical Distributions



David Stockwell

Chapman & Hall/CRC
Taylor & Francis Group

Chapman & Hall/CRC Mathematical and Computational Biology Series

Niche Modeling

Predictions from Statistical Distributions

CHAPMAN & HALL/CRC
Mathematical and Computational Biology Series

Aims and scope:
This series aims to capture new developments and summarize what is known over the whole spectrum of mathematical and computational biology and medicine. It seeks to encourage the integration of mathematical, statistical and computational methods into biology by publishing a broad range of textbooks, reference works and handbooks. The titles included in the series are meant to appeal to students, researchers and professionals in the mathematical, statistical and computational sciences, fundamental biology and bioengineering, as well as interdisciplinary researchers involved in the field. The inclusion of concrete examples and applications, and programming techniques and examples, is highly encouraged.

Proposals for the series should be submitted to one of the series editors above or directly to:

CRC Press, Taylor & Francis Group
24-25 Blades Court
Deodar Road
London SW15 2NU
UK

Published Titles

Cancer Modelling and Simulation
Luigi Preziosi

Computational Biology: A Statistical Mechanics Perspective
Ralf Blossey

Computational Neuroscience: A Comprehensive Approach
Jianfeng Feng

Data Analysis Tools for DNA Microarrays
Sorin Draghici

Differential Equations and Mathematical Biology
D.S. Jones and B.D. Sleeman

Exactly Solvable Models of Biological Invasion
Sergei V. Petrovskii and Bai-Lian Li

Introduction to Bioinformatics
Anna Tramontano

An Introduction to Systems Biology: Design Principles of Biological Circuits
Uri Alon

Knowledge Discovery in Proteomics
Igor Jurisica and Dennis Wigle

Modeling and Simulation of Capsules and Biological Cells
C. Pozrikidis

Niche Modeling: Predictions from Statistical Distributions
David Stockwell

Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems
Qiang Cui and Ivet Bahar

Stochastic Modelling for Systems Biology
Darren J. Wilkinson

The Ten Most Wanted Solutions in Protein Bioinformatics
Anna Tramontano

Chapman & Hall/CRC Mathematical and Computational Biology Series

Niche Modeling

Predictions from Statistical Distributions

David Stockwell

Chapman & Hall/CRC
Taylor & Francis Group
Boca Raton London New York

Chapman & Hall/CRC is an imprint of the
Taylor & Francis Group, an informa business

Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 1-58488-494-0 (Hardcover)
International Standard Book Number-13: 978-1-58488-494-1 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Stockwell, David R. B. (David Russell Bancroft)
Ecological niche modeling : ecoinformatics in application to biodiversity / David R.B. Stockwell.
p. cm. -- (Mathematical and computational biology series)
Includes bibliographical references.
ISBN-13: 978-1-58488-494-1 (alk. paper)
ISBN-10: 1-58488-494-0 (alk. paper)
1. Niche (Ecology)--Mathematical models. 2. Niche (Ecology)--Computer simulation. I. Title. II. Series.

QH546.3.S76 2006
577.8'2--dc22 2006027353

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

List of Tables

List of Figures

0.1 Preface

Niche modeling is a relatively new field of research aimed at helping us to understand the response of species to their environment and predicting their distribution. The practice of niche modeling uses tools from mathematics and statistics, data management and geographic spatial analysis. The first six chapters are concerned with fundamentals, programming, theory and examples of niche modeling. When used in conjunction with more detailed and specific texts and manuals, students and researchers may successfully do niche modeling for the first time.

Successful niche modeling also requires an understanding of the limitations and potential pitfalls of prediction. Due to the importance of avoiding errors, the last six chapters are devoted to sources of errors. All are relatively novel topics in the field: autocorrelation, bias, long term persistence, non-linearity, circularity and fraud, and should be of interest to researchers.

While a statistical language like R or S-plus is not essential, it provides a way of describing these main concepts, showing someone how to use them, and hands on experience at solving problems through examples. It is assumed that readers have a basic knowledge of mathematics and programming.

Above all, successful niche modeling requires deep understanding of the process of creating and using probability distributions in multidimensional spatial and temporal application. Here simplified examples complement the rigor and completeness that can be found in the literature. The generality of the approach is illustrated by examples as diverse as invasive species dynamics, predicting house price increases, and detecting management of data or fraud.

I think there are many advantages in developing depth of intuition, such as capacity to develop novel approaches, and avoiding gross errors. Off-the-shelf statistical packages are tailored exactly to applications but can hide problematic complexity. Recipe book implementations fail to educate users in the details, assumptions and pitfalls of the analysis. As each situation is a little different, packages may not be able to adapt to the specific need of their study. Understanding of the basics, and the pitfalls, also creates confidence for communicating the results.

0.1.1 Summary of chapters

1. **Functions** This chapter summarizes major mathematical types, operations and relationships encountered both in the book and in niche modeling. This and the following two chapters could be treated as a tutorial in the R language. For example, the main functions for representing the

inverted U shape characteristic of a niche – step, Gaussian, quadratic and ramp functions – are illustrated both graphically and in R code. The chapter concludes with the ACF and lag plots, in one or two dimensions.

2. **Data** This chapter shows a simple biodiversity database using R. By using data frames as tables, it is possible to replicate the basic spreadsheet and relational database operations with R's powerful indexing functions, eliminating conversion problems as data is moved between systems while learning more about R.

3. **Spatial** R and image processing operations can perform many of the elementary spatial operations necessary for niche modeling. While these do not replace a GIS, it demonstrates generalization of arithmetic concepts to images and efficient implementation of simple spatial operations.

4. **Topology** Set theory helps to identify the basic assumptions underlying niche modeling, and the relationships and constraints between these assumptions. The chapter shows the standard definition of the niche as environmental envelopes around all ecologically relevant variables is equivalent to a box topology. A proof is offered that the Hutchinsonian environmental envelope definition of a niche when extended to large or infinite dimensions of environmental variables loses desirable topological properties. This argues for the necessity of careful selection of a small set of environmental variables.

5. **Environmental data collections** Management of data for niche modeling is poorly served by user-developed files stored in a local directory. A wide variety of data sets are currently available, and better quality niche modeling will result from using data in true archives – shared by many studies and trusted with the highest level of quality. A number of sources of data are described and access issues discussed.

6. **Examples** The three examples of niche models here were selected to contradict three main misconceptions of niche modeling. The house price increase example shows a niche that is bimodal and not an inverted U. The second example of the Brown Treesnake shows an asymptotic response with respect to precipitation. The third example of the zebra mussel shows how dynamic models of the spread of invasive species can be developed from the niche model, contrary to the view that niche models are restricted to equilibrium approaches.

7. **Bias** Here a simple theoretical model of range-shift is used to estimate the magnitude of potential bias in estimates of changes in range area due to climate change.

8. **Autocorrelation** This chapter shows the problem of validating models on autocorrelated data using internal or external validation. Holding

back data at random is shown to be inadequate to determine the skill of a model when the data are autocorrelated, particularly when using smoothed data.

9. **Nonlinearity** Procedures with linear assumptions are not reliable when the responses are non-linear. Here using simulations and a linear model for reconstructing past temperatures, niche model-like tree responses create artifacts including signal degradation, loss of variance, temporal shifts in peaks, and period doubling.

10. **Long Term Persistence** The natural world is more uncertain and more indeterministic than modeled using classical statistics. Here we show evidence that temporal and spatial natural series display LTP, or scale invariant distributions. These results provide no justification for models with preferred spatial or temporal scale, which greatly underestimate confidence limits.

11. **Circularity** A major source of error is due to conclusions encoded into the assumptions of the methodology, so allowing no other conclusion than the one obtained. Here we show a potential approach to the problem of quantifying circular reasoning. By feeding random data with the same noise and autocorrelation properties into a methodology, one obtains a null model with benchmarks for rejection regions, and expectations incorporating hidden model assumptions.

12. **Fraud** The accidental or fraudulent management of results can be detected using the distributional modeling methods of niche modeling. The second digit distribution postulated by Benford's Law allows detection of fabricated data in natural time series drawn from a single distribution. The approach is applied to a range of natural data.

I would like to express my thanks to providers of data used to illustrate issues in niche modeling. The Brown Treesnake point data were from a listing of the Australian Museum holdings provided by Gordon Rodda. Zebra Mussel occurrence data were provided by Amy J. Benson. Temperature reconstruction data were provided by Steve McIntyre. Thank you also to the San Diego Supercomputer Center, University of California San Diego, and to the National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, for providing financial support and office space, funded under a sabbatical research program by the United States National Science Foundation. The development and refinement of some of the sections of the book were assisted by exchanges via a weblog. Steve McIntyre, Demetris Koutsoyiannis, Martin Ringo, and anonymous correspondent TCO were particularly helpful. I would also like to express my deep appreciation for my wife Siriluck and two children, Lena and Victoria.