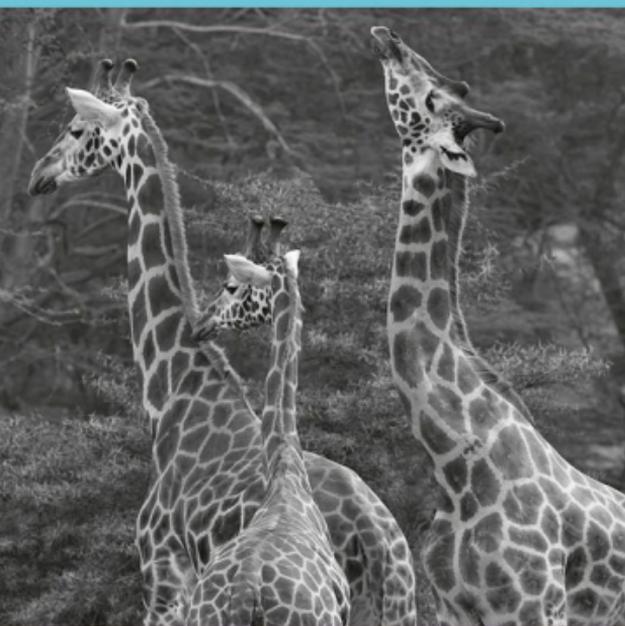


Quantitative Ecology and Evolutionary Biology

Integrating models with data

Otso Ovaskainen
Henrik Johan de Knegt
Maria del Mar Delgado



OSEE

Oxford Series in Ecology and Evolution

Oxford Series in Ecology and Evolution

Edited by Robert M. May, H. Charles J. Godfray, Jennifer A. Dunne, and Ben Sheldon

The Comparative Method in Evolutionary Biology

Paul H. Harvey and Mark D. Pagel

The Cause of Molecular Evolution

John H. Gillespie

Dunnock Behaviour and Social Evolution

N. B. Davies

Natural Selection: Domains, Levels, and Challenges

George C. Williams

Behaviour and Social Evolution of Wasps: The Communal Aggregation Hypothesis

Yosiaki Itô

Life History Invariants: Some Explorations of Symmetry in Evolutionary Ecology

Eric L. Charnov

Quantitative Ecology and the Brown Trout

J. M. Elliott

Sexual Selection and the Barn Swallow

Anders Pape Møller

Ecology and Evolution in Anoxic Worlds

Tom Fenchel and Bland J. Finlay

Anolis Lizards of the Caribbean: Ecology, Evolution, and Plate Tectonics

Jonathan Roughgarden

From Individual Behaviour to Population Ecology

William J. Sutherland

Evolution of Social Insect Colonies: Sex Allocation and Kin Selection

Ross H. Crozier and Pekka Pamilo

Biological Invasions: Theory and Practice

Nanako Shigesada and Kohkichi Kawasaki

Cooperation Among Animals: An Evolutionary Perspective

Lee Alan Dugatkin

Natural Hybridization and Evolution

Michael L. Arnold

The Evolution of Sibling Rivalry

Douglas W. Mock and Geoffrey A. Parker

Asymmetry, Developmental Stability, and Evolution

Anders Pape Møller and John P. Swaddle

Metapopulation Ecology

Ilkka Hanski

Dynamic State Variable Models in Ecology: Methods and Applications

Colin W. Clark and Marc Mangel

The Origin, Expansion, and Demise of Plant Species

Donald A. Levin

The Spatial and Temporal Dynamics of Host-Parasitoid Interactions

Michael P. Hassell

The Ecology of Adaptive Radiation
Dolph Schluter

Parasites and the Behavior of Animals
Janice Moore

Evolutionary Ecology of Birds
Peter Bennett and Ian Owens

The Role of Chromosomal Change in Plant Evolution
Donald A. Levin

Living in Groups
Jens Krause and Graeme D. Ruxton

Stochastic Population Dynamics in Ecology and Conservation
Russell Lande, Steiner Engen, and Bernt-Erik Sæther

The Structure and Dynamics of Geographic Ranges
Kevin J. Gaston

Animal Signals
John Maynard Smith and David Harper

Evolutionary Ecology: The Trinidadian Guppy
Anne E. Magurran

Infectious Diseases in Primates: Behavior, Ecology, and Evolution
Charles L. Nunn and Sonia Altizer

Computational Molecular Evolution
Ziheng Yang

The Evolution and Emergence of RNA Viruses
Edward C. Holmes

Aboveground–Belowground Linkages: Biotic Interactions, Ecosystem Processes, and Global Change
Richard D. Bardgett and David A. Wardle

Principles of Social Evolution
Andrew F. G. Bourke

Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics
John Harte

Ecological Speciation
Patrik Nosil

Energetic Food Webs: An Analysis of Real and Model Ecosystems
John C. Moore and Peter C. de Ruiter

Evolutionary Biomechanics: Selection, Phylogeny, and Constraint
Graham K. Taylor and Adrian L. R. Thomas

Quantitative Ecology and Evolutionary Biology: Integrating Models with Data
Otso Ovaskainen, Henrik Johan de Knegt, and Maria del Mar Delgado

Quantitative Ecology and Evolutionary Biology

Integrating Models with Data

OTSO OVASKAINEN

Department of Biosciences, University of Helsinki, Finland

HENRIK JOHAN DE KNEGT

Department of Environmental Sciences, Wageningen University, The Netherlands

MARIA DEL MAR DELGADO

Research Unit of Biodiversity (UMIB, UO-CSIC-PA), Oviedo University – Campus Mieres, Mieres, Spain

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© Otso Ovaskainen, Henrik Johan de Knecht & Maria del Mar Delgado 2016

The moral rights of the authors have been asserted

First Edition published in 2016

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2015957822

ISBN 978-0-19-871486-6 (hbk.)

ISBN 978-0-19-871487-3 (pbk.)

Printed and bound by
CPI Litho (UK) Ltd, Croydon, CR0 4YY

Preface

We have written this book primarily for two kinds of target audiences. The first one consists of those ecologists and evolutionary biologists who may not be very experienced with mathematical and statistical modelling yet, but who are interested in learning more about modelling approaches in order to integrate them with data. The second audience consists of those mathematicians, statisticians, and computer scientists who are interested in ecological and evolutionary questions, and thus wish to see how the mathematical, statistical, and computational tools that they already master can be applied in these fields of sciences. We have intended the book for researchers and graduate students, as well as advanced master students. We also hope it will be useful as a textbook for graduate-level courses in ecological modelling, statistical ecology, theoretical ecology, and spatial ecology.

After an introductory chapter, we devote one chapter to movement ecology, one to population ecology, one to community ecology, and one to genetics and evolutionary ecology. We have followed the same structure in each chapter. We start with a conceptual section, which provides the necessary biological background and motivates the modelling work. The next three sections present mathematical modelling approaches, followed by one section devoted to statistical approaches. We end each chapter with a perspectives section, where we summarize some of the key messages of the chapter, discuss some limitations of the approaches we have considered, and mention some alternative approaches.

We assume that the reader is familiar with the basics of calculus and probability. These are covered in good depth in the context of ecological modelling, e.g., in the books of Otto and Day (2007), Mangel (2006), and Hastings (1996). From the statistical side, we expect the reader to have basic knowledge of linear models (e.g., ANOVA and regression), covered, e.g., by the book by Grafen and Hails (2002) or that of Zuur et al. (2007). Some parts of this book utilize rather advanced mathematical and statistical methods compared with the background typical for graduate students in ecology and evolution. Therefore, we finish the book with two appendices, one related to mathematical and the other to statistical methods. In these appendices, we summarize a wide range of methods that are utilized in this book and that we expect to be useful for the reader more widely.

In a single book, it is not possible to present a comprehensive review of all the mathematical and statistical approaches developed for movement ecology, population ecology, community ecology, population genetics, and evolutionary biology.

Therefore, the selection of approaches and models we present is necessarily limited, and biased by our own research interests. The reason why we cover such a broad range of topics rather than focusing on a more limited scope is that one key aim of this book is to link modelling approaches used in different fields of ecology and evolutionary biology to each other. With this aim, we have constructed the models in a stepwise manner, starting from simple baseline models, and then adding gradually new components. As one example, in Chapter 2 we start from a random walk model in homogeneous space. We then extend the model to heterogeneous space by considering how moving individuals may respond to spatial variation in their environments, after which we model movements in a highly fragmented landscape as a particular case study. In Chapter 3, we add to the movement model the births and deaths of individuals, resulting in a model that we will call the butterfly metapopulation model. As another example, in Chapter 2 we link dispersal kernels to movement models. In Chapter 3, we utilize dispersal kernels to construct a model for the ecological dynamics of sessile organisms, which we call the plant population model. In Chapter 4, we add to the plant population model competition among heterospecifics, resulting in what we call the plant community model. Finally, in Chapter 5 we extend the plant population model into an eco-evolutionary model by letting the individuals carry genes that influence the dispersal of their propagules.

While our aim is to integrate models with data, we discuss real data only in the context of few motivating examples. All the data analysed in the statistical sections have been generated by the mathematical models of the same chapter. We have decided to do so because with simulated data the reader has the full knowledge of the process that has generated the data, which would be never the case with real data. Therefore, knowing the underlying process enables one to pinpoint the potentials and pitfalls of different kinds of statistical methods more effectively than would be the case with real data.

As our starting point is movement ecology, much (though not all) of our emphasis is on the spatial aspects of ecology and evolution, especially on the interplay between environmental heterogeneity and ecological processes. In particular, each chapter has one section that applies the mathematical modelling approaches to study the consequences of habitat loss and fragmentation. In addition to presenting perspectives to ecological modelling, we thus hope to provide some insights on how habitat loss and fragmentation influence the movements of individuals, the dynamics of populations and communities, and evolutionary dynamics.

Writing this book would not have been possible without the support we have obtained from many people. First of all, working with the publisher has been as smooth as one could hope for. We would therefore like to thank Charles Godfray for taking the initiative of inviting us to write this book, Ian Sherman for guiding us with the initial planning, and Jennifer Dunne for providing feedback. Lucy Nash has worked as an excellent editor; her help has been invaluable through the writing process. For permission to reproduce copyrighted material, we would like to thank the National Academy of Sciences USA (Figures 2.2A–C and 4.2A–C), the Ecological

Society of America (Figure 2.2D), Oxford University Press (Figure 3.2CD), Nature Publishing Group (Figure 3.2E), and John Wiley and Sons (Fig. 5.2A,B,C).

Most importantly, the book has benefited greatly from comments that we have received from several colleagues. We would thus like to thank Nerea Abrego, Florian Hartig, Arild Husby, Jussi Jousimo, Etsuko Nonaka, Iñaki Odriozola, Bart Peeters, Johannes Signer, Tord Snäll and Eugenia Soroka for providing a large number of suggestions and corrections that greatly improved the contents. We thank Kaisa Torppa for helping us to compile the index, Sami Ojanen for producing the maps of Figures 2.2 and 3.2., and Juha Merilä for permission to use the photo of Fig. 5.2D. Our warmest thanks go to Ilkka Hanski, the leader of the Metapopulation Research Centre, where this book has been written. This is not only for the feedback and encouragement that he has provided during the writing process, but also and especially for sharing with us his passion for science.

The research that resulted in the development of the modelling approaches presented in this book, as well as the writing of this book, has been supported by grants from the European Research Council (Grant 205905), Academy of Finland (Grants 12424, 129636, and 25044), the Kone Foundation (Grant 44-6977), The Netherlands Organisation for Scientific Research (Rubicon Grant 825.11.009), and the Research Council of Norway (CoE Programme Grant 223257).

Contents

| | | |
|----------|---|-----------|
| 1 | Approaches to ecological modelling | 1 |
| 1.1 | Forward and inverse approaches | 2 |
| 1.2 | The interplay between models and data | 3 |
| 1.3 | The many choices with mathematical and statistical models and methods | 6 |
| 1.4 | What a biologist should learn about modelling | 8 |
| 2 | Movement ecology | 10 |
| 2.1 | Why, where, when, and how do individual organisms move | 10 |
| 2.1.1 | Internal state: why to move | 11 |
| 2.1.2 | Motion capacity: how to move | 12 |
| 2.1.3 | Navigation capacity: when and where to move | 13 |
| 2.1.4 | Different types of movement | 13 |
| 2.1.5 | Approaches to movement research | 14 |
| 2.1.6 | Outline of this chapter | 15 |
| 2.2 | Movement models in homogeneous environments | 17 |
| 2.2.1 | The Lagrangian approach | 18 |
| 2.2.2 | Translating the Lagrangian model into an Eulerian model | 20 |
| 2.2.3 | Dispersal kernels | 21 |
| 2.2.4 | Adding directional persistence: correlated random walk models | 23 |
| 2.2.5 | Adding directional bias: home-range models | 25 |
| 2.3 | Movement models in heterogeneous environments | 27 |
| 2.3.1 | Random walk simulations in heterogeneous space | 27 |
| 2.3.2 | Diffusion models with continuous spatial variation in movement parameters | 29 |
| 2.3.3 | Diffusion models with discrete spatial variation in movement parameters | 34 |
| 2.3.4 | Using movement models to define and predict functional connectivity | 35 |
| 2.3.5 | The influence of a movement corridor | 40 |
| 2.4 | Movements in a highly fragmented landscape | 43 |
| 2.4.1 | The case of a single habitat patch | 44 |
| 2.4.2 | The case of a patch network | 47 |

| | | |
|----------|---|-----------|
| 2.5 | Statistical approaches to analysing movement data | 50 |
| 2.5.1 | Exploratory data analysis of GPS data | 51 |
| 2.5.2 | Fitting a diffusion model to capture-mark-recapture data | 60 |
| 2.6 | Perspectives | 63 |
| 2.6.1 | Limitations and extensions of random walk and diffusion models | 64 |
| 2.6.2 | The many approaches of analysing movement data | 66 |
| 3 | Population ecology | 68 |
| 3.1 | Scaling up from the individual level to population dynamics | 68 |
| 3.1.1 | Factors influencing population growth through birth and death rates | 69 |
| 3.1.2 | How movements influence population dynamics | 70 |
| 3.1.3 | How population structure influences population dynamics | 71 |
| 3.1.4 | The outline of this chapter | 72 |
| 3.2 | Population models in homogeneous environments | 73 |
| 3.2.1 | Individual-based stochastic and spatial model | 76 |
| 3.2.2 | Simplifying the model: stochasticity without space | 77 |
| 3.2.3 | Simplifying the model further: without stochasticity and space | 80 |
| 3.2.4 | Another way of simplifying the model: space without stochasticity | 81 |
| 3.3 | Population models in heterogeneous environments | 83 |
| 3.3.1 | Environmental stochasticity | 84 |
| 3.3.2 | Spatial heterogeneity in continuous space: the plant population model | 86 |
| 3.3.3 | Spatial heterogeneity in discrete space: the butterfly metapopulation model | 89 |
| 3.3.4 | The Levins metapopulation model and its spatially realistic versions | 93 |
| 3.4 | The persistence of populations under habitat loss and fragmentation | 97 |
| 3.4.1 | Habitat loss and fragmentation in the plant population model | 98 |
| 3.4.2 | Habitat loss and fragmentation in the butterfly metapopulation model | 100 |
| 3.4.3 | Habitat loss and fragmentation in the Levins metapopulation model | 100 |
| 3.5 | Statistical approaches to analysing population ecological data | 102 |
| 3.5.1 | Time-series analyses of population abundance | 102 |
| 3.5.2 | Fitting Bayesian state-space models to time-series data | 105 |
| 3.5.3 | Species distribution models | 112 |
| 3.5.4 | Metapopulation models | 115 |
| 3.6 | Perspectives | 117 |
| 3.6.1 | The invisible choices made during a modelling process | 118 |
| 3.6.2 | Some key insights derived from population models | 119 |
| 3.6.3 | The many approaches to analysing population data | 120 |

| | | |
|----------|---|------------|
| 4 | Community ecology | 122 |
| 4.1 | Community assembly shaped by environmental filtering and biotic interactions | 122 |
| 4.1.1 | Ecological interactions | 124 |
| 4.1.2 | Fundamental and realized niches and environmental filtering | 125 |
| 4.1.3 | Organizational frameworks for metacommunity ecology | 126 |
| 4.1.4 | The outline of this chapter | 127 |
| 4.2 | Community models in homogeneous environments | 129 |
| 4.2.1 | Competitive interactions | 129 |
| 4.2.2 | Resource–consumer interactions | 135 |
| 4.2.3 | Predator–prey interactions | 138 |
| 4.3 | Community models in heterogeneous environments | 140 |
| 4.3.1 | The case of two competing species | 141 |
| 4.3.2 | The case of many competing species | 142 |
| 4.4 | The response of communities to habitat loss and fragmentation | 145 |
| 4.4.1 | Endemics-area and species-area relationships generated by the plant community model | 145 |
| 4.5 | Statistical approaches to analysing species communities | 150 |
| 4.5.1 | Time-series analyses of population size in species communities | 150 |
| 4.5.2 | Joint species distribution models | 155 |
| 4.5.3 | Ordination methods | 161 |
| 4.5.4 | Point-pattern analyses of distribution of individuals | 162 |
| 4.6 | Perspectives | 164 |
| 4.6.1 | Back to the metacommunity paradigms | 164 |
| 4.6.2 | Some insights derived from community models | 165 |
| 4.6.3 | The many approaches to modelling community data | 167 |
| 5 | Genetics and evolutionary ecology | 168 |
| 5.1 | Inheritance mechanisms and evolutionary processes | 168 |
| 5.1.1 | Genetic building blocks and heritability | 168 |
| 5.1.2 | Selection, drift, mutation, and gene flow | 170 |
| 5.1.3 | Connections between ecological and evolutionary dynamics | 172 |
| 5.1.4 | The outline of this chapter | 173 |
| 5.2 | The evolution of quantitative traits under neutrality | 175 |
| 5.2.1 | An additive model for the map from genotype to phenotype | 175 |
| 5.2.2 | Coancestry and the additive genetic relationship matrix | 177 |
| 5.2.3 | Why related individuals resemble each other? | 179 |
| 5.2.4 | The animal model | 180 |
| 5.2.5 | Why related populations resemble each other? | 181 |
| 5.3 | The evolution of quantitative traits under selection | 184 |
| 5.3.1 | Evolution by drift, selection, mutation, recombination, and gene flow | 185 |
| 5.3.2 | Selection differential and the breeder's equation | 186 |
| 5.3.3 | Population divergence due to drift and selection | 190 |

| | | |
|---|--|------------|
| 5.4 | Evolutionary dynamics under habitat loss and fragmentation | 194 |
| 5.4.1 | Evolution of dispersal in the Hamilton–May model under adaptive dynamics | 194 |
| 5.4.2 | Evolution of dispersal in the plant population model under quantitative genetics | 198 |
| 5.5 | Statistical approaches to genetics and evolutionary ecology | 201 |
| 5.5.1 | Inferring population structure from neutral markers | 201 |
| 5.5.2 | Estimating additive genetic variance and heritability | 203 |
| 5.5.3 | Using association analysis to detect loci behind quantitative traits | 204 |
| 5.5.4 | Detecting loci under selection from genotypic data | 206 |
| 5.5.5 | Detecting traits under selection from genotypic and phenotypic data | 207 |
| 5.6 | Perspectives | 209 |
| 5.6.1 | Mathematical approaches to modelling genetics and evolution | 209 |
| 5.6.2 | Some insights derived from evolutionary models on dispersal evolution | 211 |
| 5.6.3 | The many uses of genetic data | 212 |
| Appendix A: Mathematical methods | | 215 |
| A.1 | A very brief tutorial to linear algebra | 215 |
| A.2 | A very brief tutorial to calculus | 217 |
| A.2.1 | Derivatives, integrals, and convolutions | 217 |
| A.2.2 | Differential equations | 218 |
| A.2.3 | Systems of differential equations | 220 |
| A.2.4 | Partial differential equations | 221 |
| A.2.5 | Difference equations | 222 |
| A.3 | A very brief tutorial to random variables | 222 |
| A.3.1 | Discrete valued random variables | 222 |
| A.3.2 | Continuous valued random variables | 223 |
| A.3.3 | Joint distribution of two or more random variables | 225 |
| A.3.4 | Sums of random variables | 226 |
| A.3.5 | An application of random variables to quantitative genetics | 228 |
| A.4 | A very brief tutorial to stochastic processes | 229 |
| A.4.1 | Markov chains | 229 |
| A.4.2 | Markov processes | 231 |
| Appendix B: Statistical methods | | 233 |
| B.1 | Generalized linear mixed models | 233 |
| B.1.1 | Linear models | 233 |
| B.1.2 | Link functions and error distributions | 234 |
| B.1.3 | Relaxing the assumption of independent residuals | 236 |
| B.1.4 | Random effects | 238 |

| | |
|--|-----|
| B.1.5 Multivariate models | 241 |
| B.1.6 Hierarchical models | 244 |
| B.2 Model fitting with Bayesian inference | 245 |
| B.2.1 The concepts of likelihood, maximum likelihood, and parameter uncertainty | 246 |
| B.2.2 Prior and posterior distributions, and the Bayes theorem | 246 |
| B.2.3 Methods for sampling the posterior distribution | 249 |
| References | 253 |
| Index | 277 |

I

Approaches to ecological modelling

The motivation for conducting any kind of research in ecology and evolutionary biology, including modelling, is to advance knowledge about ecological and evolutionary processes. This happens mostly with small steps that refine the existing views by bringing new pieces of knowledge, but sometimes through revolutionary ideas that challenge existing paradigms. For an excellent introduction to the philosophy of science in the context of ecological research, we refer to the book by Hilborn and Mangel (1997). Briefly, the process of doing science involves the interplay between theories, hypotheses, models, and data. Theories summarize our present understanding on how we consider nature to work. They are the highest form of scientific knowledge, as they have resisted repeated testing and scrutiny through experiments and observations. Hypotheses are proposed explanations of how nature might work, requiring new data to test the validity of the proposed explanation. Theories and hypotheses can be expressed verbally, but they can also be made more specific by expressing them with the help of mathematical models. Different models of the same phenomena can be viewed as alternative hypotheses, and alternative parameter values of those models can be seen as even more specific hypotheses. Data are essential for examining the extent to which hypotheses are supported. This can be done by examining whether a single hypothesis is supported by data, or by quantifying the degree of support for alternative hypotheses simultaneously. Observing a contradiction between a hypothesis and data is a key step for making scientific progress, as it makes one think why nature does not work as was hypothesized. Such thinking can lead to new hypotheses, which can refine the existing theories or even eventually replace them entirely by new ones.

The word ‘modelling’ is used in ecological research in many different ways, ranging from mathematical models used to develop theoretical ecology to statistical models used to analyse empirical data. Mathematical modelling helps to understand the causal pathways from ecological mechanisms to the resulting phenomena, whereas statistical models are needed to confront theories and models with data. As a simple example, a theoretically minded ecologist might model resource–consumer dynamics with a system of differential equations (as we will do in Chapter 4). Insights from the model may be derived using just pen and paper, e.g. to show that a consumer population will persist in the long term if and only if the rate of resource production exceeds some threshold value, or to understand more generally how the dynamics of the consumer depends on the underlying ecological processes and their

parameter values. An empirically minded researcher could study the same system by acquiring experimental data on consumer abundance for different levels of resource availability. Statistical models may then be applied to show that in the experiment the growth rate of the consumer increases with the availability of resources, as suggested by the mathematical models constructed by the theoretical ecologist.

1.1 Forward and inverse approaches

One way of classifying different approaches to ecological modelling is that of ‘forward’ and ‘inverse’ approaches. Forward approaches are those where assumptions are made about the underlying mechanisms, and mathematical or simulation tools are used to study the consequences of those assumptions. With an inverse approach, the aim is to uncover the mechanisms that might have generated the patterns observable in empirical data. Both the forward (moving from mechanisms to patterns) and the inverse (moving from patterns to mechanisms) approaches involve their own challenges. To illustrate, let us consider a researcher who is interested in the movement ecology of butterflies, specifically on how the characteristics of the species and the structure of the environment influence the displacements that the individuals make over their lifetimes (as we will do in Chapter 2). If taking the forward approach, the researcher could track the detailed movement tracks of individuals to learn about the mechanisms that influence lifetime displacements, such as flight speed and directional persistence under different environmental conditions. Such data could be used to construct a movement model, which could be used to simulate lifetime movements in heterogeneous landscapes. While this can be relatively straightforward to do, the critical question is how much the model can be trusted. This is because the researcher is likely to obtain detailed movement data only for a limited number of individuals, only for a limited set of environmental conditions, and only over small temporal and spatial scales. This can be problematic, as the key features of movement behaviour may differ among individuals, environmental conditions, and temporal and spatial scales. More generally, a major challenge in the forward approach is that of extrapolation.

Conversely, if taking the inverse approach, the researcher could start by acquiring capture-mark-recapture data, i.e. marking and releasing butterflies when encountering them, and then recapturing some of the marked individuals later at another location. Such data may be acquired at the scale of lifetime displacements, and thus they would avoid the problem of extrapolation. However, it can be highly challenging to use such data for inference about the mechanisms that might have generated the observed displacements. This is because the data are simultaneously influenced by a myriad of processes, such as the intrinsic behaviour of the species, the spatial distribution of habitat characteristic, the prevailing weather conditions, and where and when the captures were attempted. Inverse problems are generally ill posed in the mathematical terminology, meaning that they do not have a unique solution. For example, assume that only few butterflies were recaptured. This may be because the

death rate was high or alternatively because the emigration rate out of the study area was high, or because the capture rate was low.

Much of the current research in ecology and evolutionary biology uses both forward and inverse approaches in an integrated manner. This is partly due to the increasing interest in conducting ecological research at the interface between theory and data, and partly due to developments in hardware and software that allow for computationally demanding analyses. In this book, our emphasis is to combine the two. To do so, we will use mathematical models to generate various kinds of data sets. We will then apply statistical methods to those data, with the aim of asking how much of the assumed mechanisms we can recover.

1.2 The interplay between models and data

Figure 1.1 illustrates the interplay between models and data in the context of the forward approach. Often one wishes to understand how some phenomenon at a higher level of biological organization emerges from the processes at a lower level of biological organization. In the example of Figure 1.1, we are interested in how the dynamics of a population emerge from the deaths and births of individuals. In the process of model construction, knowledge about the ecological context is required. This could involve e.g. whether the organism has discrete or overlapping generations, and whether it inhabits a seasonal or unseasonal environment. These assumptions translate into the model structure. In addition to deciding about the model structure, one needs to decide about the parameter values. In scenario simulations, the parameter values are assumed rather than estimated, which is a perfectly valid approach if the aim is to ask how the parameter values influence the model's behaviour. To confront the model with data, the parameter values need to be estimated. In the most pure form of the forward approach, the model parameters are simply measured one by one. For example, one might monitor the lifetimes of individuals at different population densities to measure the mean death rate and its dependency on population density. Once the parameters have been measured, the model can be used to generate predictions. To validate the model, the predictions are to be contrasted with independent data. For example, one may compare the dynamics generated by the model with data acquired directly at the population level. If the match between model prediction and data is good, one may conclude that the model is compatible with the data. This is, of course, encouraging, yet it does not mean that the model is correct, as different kinds of models may end up with the same prediction. To refine the model or to select between alternative types of models, new data, ideally from a different context, are needed. Testing whether the model can predict such new data is called cross-validation. If the model does not match well with the data, that is actually not such a bad result. In this case, the conclusion is that the hypothesis that corresponds to the model is not compatible with the data. Thus, something is wrong either with the structural model assumptions, or with the parameter values. Through analysing the mismatch between the model's prediction and the data in more detail, one may

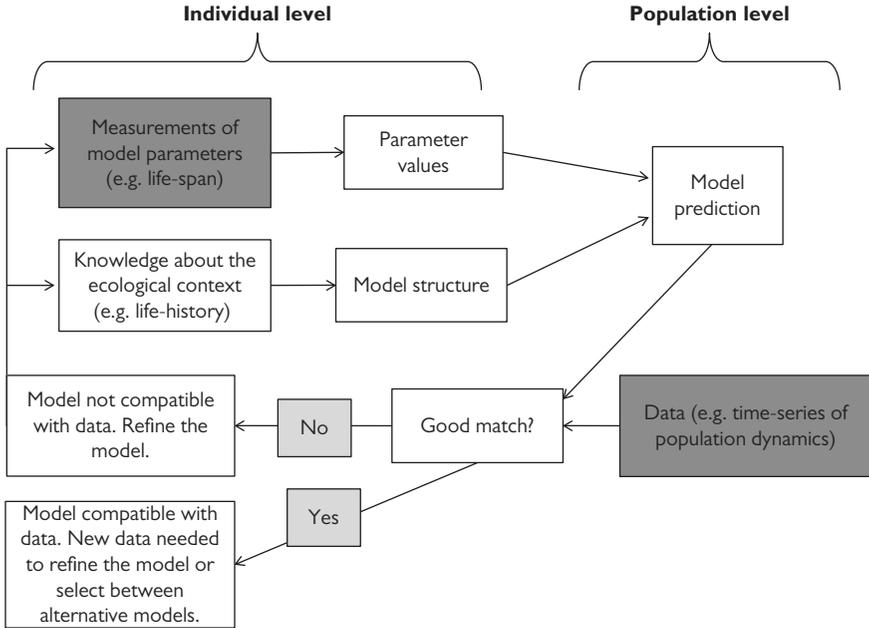


Figure 1.1 A flow diagram illustrating the interplay between models and data by the forward approach. We have assumed here that the model is constructed at the individual level, with the aim of predicting dynamics at the population level.

be able to infer reasons for the failed predictions, and thus learn more about the ecological context or the parameter values.

Applying the pure forward approach of Figure 1.1 can be challenging due to three problems. The first problem is that ecological processes are complex and influenced by manifold factors; thus, it is very difficult to derive the structural model assumptions from first principles. The second problem is that it is usually very difficult to perform direct measurements of all the model parameters. Even if this would be possible, the third problem is that the parameter estimates usually involve large amounts of uncertainty, and when a large number of uncertain estimates are put together through uncertain and nonlinear model structures, the amount of uncertainty becomes amplified.

Figure 1.2 presents an overview of the processes of model construction, parameterization, and inference, now performed in a manner that combines elements from both the forward and the inverse approaches. Like with the forward approach, the researcher uses prior knowledge about the ecological context to formulate a model, or a family of models, that describe the structural assumptions about the assumed underlying biology. While in the pure forward approach we assumed that all model parameters were measured directly, here we assume that direct information is possibly available only for some of the parameters, and that this information

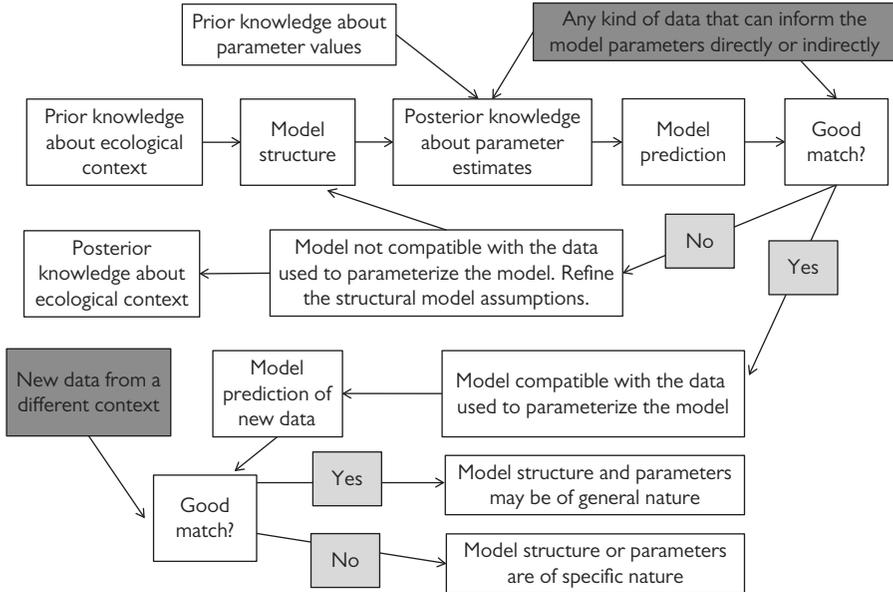


Figure 1.2 A flow diagram illustrating the interplay between models and data by combining elements from the forward and inverse approaches.

may be uncertain. We use in Figure 1.2 the terminology of Bayesian inference (see Appendix B.2); hence we refer to this information as prior knowledge about parameter values. In the Bayesian context, parameter estimation corresponds to the computation of the posterior distribution of the parameters. A key difference to the forward approach is that in the inverse approach any kind of data can be used to inform the model parameters. For example, even if the model parameters are formulated at the individual level, also population level data can be informative about the parameter values.

Parameter estimation should always be followed by model validation. To start with, it is always possible to compare model predictions to the data used to parameterize the model. If the model is not able to reproduce the data used to parameterize it, the problem is not in the parameter values but in the structural model assumptions, and thus in the assumptions made when describing the ecological context. If the model is able to reproduce the data used to parameterize it, it is encouraging, but does not necessarily mean that the mechanisms incorporated in the model actually operate in reality. To move one step forward, one may submit the model to a harder test, which is to make a prediction in a context that is different from the one used for parameter estimation. If the model is successful in predicting the new data as well, there is a stronger support that the assumptions behind the model reflect reality. If this is not the case, one may think about how to refine the model so that it simultaneously fits both the original data and the new data.