Bayesian Models

A Statistical Primer for Ecologists

N. Thompson Hobbs and
Mevin B. Hooten

Bayesian Models

Bayesian Models

A Statistical Primer for Ecologists

N. Thompson Hobbs and Mevin B. Hooten

PRINCETON UNIVERSITY PRESS

PRINCETON AND OXFORD

Published by Princeton University Press, 41 William Street, Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press, 6 Oxford Street, Woodstock,
Oxfordshire OX20 1TW

press.princeton.edu

Jacket art © Irena Peziene/Shutterstock

Library of Congress Cataloging-in-Publication Data

Hobbs, N. Thompson.
Bayesian models : a statistical primer for ecologists / N. Thompson Hobbs and Mevin B. Hooten.
pages cm
Includes bibliographical references and index.
ISBN 978-0-691-15928-7 (hardcover : alk. paper) 1. Ecology–Statistical methods.
2. Bayesian statistical decision theory. I. Hooten, Mevin B., 1976– II. Title.
QH541.15.S72H63 2015
577.01′5195–dc23
2015000021

British Library Cataloging-in-Publication Data is available

This book has been composed in Times LT Std and Futura Std

Printed on acid-free paper. ∞

Typeset by S R Nova Pvt Ltd, Bangalore, India
Printed in the United States of America

1 3 5 7 9 10 8 6 4 2

Contents

Preface

Why This Book?

This book is about the process of gaining new knowledge about ecology using models and data. We wrote it for several reasons. An overarching motivation is that our satisfaction with scientific work has been palpably enhanced by understanding this process from start to finish, with no gaps where faith must fill in for understanding. We write this book because what we write about is easily the most intellectually satisfying material we have learned in our careers. It has made the work we do every day more enjoyable. We are confident that your research in ecology will be accomplished with greater satisfaction and reward if you master the concepts we describe here.

There are more specific motivations as well. We teach graduate students ecological modeling using Bayesian methods. Colleagues whose students have taken our classes often ask us what they should read to get a big picture understanding of the Bayesian approach. They seek a book explaining statistical principles of Bayesian modeling written in language accessible to nonstatisticians. They may never write a line of computer code implementing Bayesian methods, but they realize the importance of understanding them. Our colleagues want to be able to appreciate contemporary scientific literature, to review papers and proposals, and to mentor students who *do* know how to write code. We decided to write the book they asked for.

Of course, there are now many excellent texts on Bayesian modeling in ecology (e.g., Clark, 2007; McCarthy, 2007; Royle and Dorazio, 2008; Link and Barker, 2010; Kéry, 2010; Kéry and Schaub, 2012), texts that we use all the time. For the most part, these books emphasize computational methods over concise explanation of basic principles. Many of these are difficult to appreciate without a background in mathematical statistics—training that most ecologists lack. Our book will complement the existing crop by providing the basic understanding of mathematical and statistical principles needed to use more advanced texts effectively.

A third motivation also comes from our personal experience. Students and colleagues often come to us with a modeling problem they are working

on, a problem that uses Bayesian methods. They can write code in one of the popular implementations of Gibbs samplers, WinBUGS (Lunn et al., 2000) or JAGS (Plummer, 2003), and often they bring a thick stack of code with them. Although they can create computer programs that give them answers, they cannot write the mathematical expression for the model that underpins their work. They are unsure about their starting point and hence are not entirely confident about where they have ended up. They have difficulty writing a mathematical expression that clearly communicates their analysis in manuscripts and proposals. As you will see as this book unfolds, we believe that reliable analysis must begin with a model written in mathematical symbols and operators, not in a computer language. Writing models is the foundation of good science.

The ability to write models leads to our next motivation. There is a diminishing set of important questions in ecology that can be answered by a single investigator working alone, even a very good one. Instead, what remains are problems solvable only by the application of intersecting sets of talents, skills, and knowledge. This book offers hold-in-your-hand evidence of the value of collaboration between a statistician and an ecologist, but there are many other examples (Gross et al., 2002, 2005; Clark, 2003b, 2005; Latimer et al., 2006; Farnsworth et al., 2006; Cressie et al., 2009; Rotella et al., 2009; Webb et al., 2010; Eaton and Link, 2011; Wilson et al., 2011; Fiechter et al., 2013; Peterson et al., 2013). These collaborations require a mutual understanding of basic principles and a shared vocabulary.

Our final reason for writing this book relates to the first one. Most ecologists working today, save perhaps some of the youngest ones, received training in statistics emphasizing procedures over principles. We emerged from this training understanding how to do "data analysis" using a suite of recipes—t-tests, analysis of variance, regression, general linear models, and so on. The diligent and ambitious among us might have added work in sampling and multivariate statistics. Mathematical statistics was reserved for statistics majors.

The outcome of this approach to training is seen in a revealing way in the frontispiece of a once widely used text (Sokal and Rohlf, 1995), which displays a table of analyses resembling a dichotomous key in taxonomy. If our data are like x, then we should use analysis y; otherwise, we should use analysis z. Those of us trained this way had precious little understanding of *why* we should use analysis y over z beyond the authority provided by that table. Moreover, there was a limited range of kinds of data for which this taxonomic approach would serve, and we were stymied if the observations we worked hard to obtain were not found somewhere in the table. Sometimes we would heroically bend those observations to make them fit in one of the table's narrowly defined cells. A narrow range of

approaches to analysis constrains the questions that ecologists are willing to ask—we are uncomfortable posing questions for which we see no analytical route to insight. If the only tool in our locker is analysis of variance, then the world we study must be composed of randomized plots. Some of the current Bayesian books are organized in a similar way—around procedures rather than principles.

All research problems that ecologists seek to solve have aspects in common and aspects that are unique. The unique features of these problems argue for a principled approach to insight. A lean set of modeling principles can substitute for volumes of statistical facts. Understanding these principles enables us to design routes to insight uniquely suited to each of the diverse problems we confront over the course of a research career. Providing this understanding is the main reason for writing this book.

Goals

The overarching goal of this book is to train ecologists in the basic statistical principles needed to use and interpret Bayesian models. An essential part of meeting that goal is to teach how to write out accurate mathematical expressions for Bayesian models linking observations to ideas about how ecological systems work. These expressions form the foundation for inference. It is our ultimate aim to increase the intellectual satisfaction of ecologists with their teaching, research, and peer review by providing a solid, intuitive understanding of how we learn from data and models in the Bayesian approach. Finally, we aim to enhance the quality of collaboration between ecologists and statisticians by training ecologists in the mathematical concepts and language of statistics.

Approach

Organization

The book is organized to provide the understanding needed to support a general process for model building in ecology, a process that applies to virtually all research problems. We outline a flow of tasks in building models that we have found helpful explaining the process (figure 0.0.1). The sequence here is not immutable, but it offers a useful schematic of the steps needed to build a revealing Bayesian model. We will return to this diagram throughout the book to show how specific topics fit into the larger task of model building.

We have organized the book in three parts. The aim of part I is to provide a basic understanding of the principles underpinning Bayesian analysis (fig. 0.0.1). We begin part I with a preview of the entire book to motivate what will follow. We then spend some time encouraging thinking about deterministic models and how they have been traditionally used in ecology (fig. 0.0.1 A). Next, we cover basic principles of probability and probability distributions. Think of this as a crash course in mathematical statistics, a prospect that might not be thrilling on the face of it, but we urge you to read this material carefully if it is not familiar to you. Gaining familiarity will allow you to understand the powerful ideas that follow and, ultimately, to understand what you are doing when applying Bayesian methods to your own research.

Knowledge of statistical distributions provides the foundation for understanding maximum likelihood approaches to parameter estimation. Likelihood is a central element of Bayesian models. We explain the link between likelihood and Bayesian approaches to inference by explaining the theory underpinning the Bayesian approach. We develop the concept of Bayes as "likelihood reweighted." We then dissect Bayes' theorem piece by piece to explain its components: the posterior distribution, the likelihood, the prior, and the marginal distribution of the data. We introduce some uniquely Bayesian concepts likely to be unfamiliar to most ecologists, for example conjugate relationships between likelihoods and priors, which turn out to be critical in the second part of the book. We finish part I by applying the basic statistical concepts we have developed to specific problems in ecological research, illustrating how mathematical expressions are built to link models and data (fig. 0.0.1 B). We show how models of complex phenomena, models with many parameters and latent quantities, can be broken into manageable chunks using hierarchical modeling.

Part II lays out the nuts and bolts of how we use the principles developed in part I to learn about parameters, unobservable states, and derived quantities. We clearly explain Markov chain Monte Carlo and Gibbs sampling, the numerical algorithms that have revolutionized the ability to gain insight from hierarchical models (fig. 0.0.1 C). Part II closes by describing how we check models to assure their fidelity to statistical assumptions and how we make inference from a single model or from multiple models (fig. 0.0.1 D).

Part III includes a series of problems and worked solutions drawn from several subdisciplines of ecology. These problems require application of the concepts and principles we developed in part I and II, emphasizing model specification (fig. 0.0.1 B), a skill that we believe is not emphasized in other texts. Our intention in part III is to encourage active model building and to show how the same approach to model specification can be fruitfully applied to a broad range of problems in ecology.

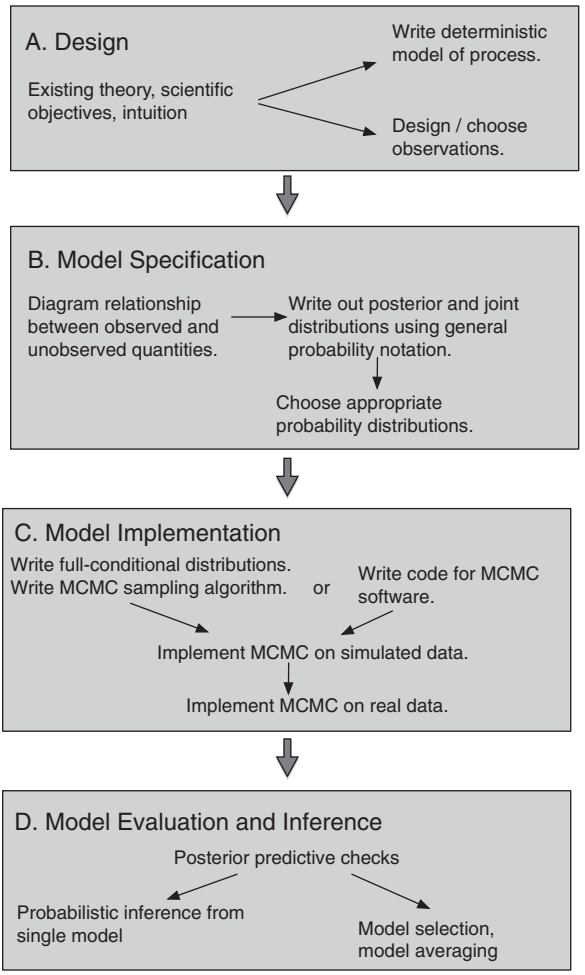

Figure 0.0.1. Gaining insight from Bayesian models involves the same sequence of steps for virtually all research problems, steps that fall into four broad groups. The sequence of steps is indicated by the long and short arrows. This book is organized to explain these steps in a logical way. (**A**) Design is not uniquely Bayesian, but we include it here because we want to encourage the thoughtful development of mathematical models of ecological processes as a starting point for analysis (chapter 2). (**B**) The premise of this book is that mathematical models must be combined with data to allow us to learn about how ecological systems operate. Chapters 3, 4, 5, and 6 show how we specify models to include data. (**C**) A key idea is that a properly specified model provides all we need to know to implement the enormously powerful algorithm Markov chain Monte Carlo (MCMC). We provide a principled understanding of how and why MCMC works in chapter 7. (**D**) We then cover how to use output from MCMC as a basis for inference from single models (chapter 8) and from multiple ones (chapter 9). Finally, we return to the key process of model specification (**B**) in chapter 11 by providing a series of problems challenging you to formulate Bayesian models.

Crosscutting Themes

We err on the side of excessive explanation of notation and equations. We believe that a formidable impediment to understanding statistics by ecologists is that ecologists, for the most part, don't get up in the morning every day and write statistical models. The authors of most statistical textbooks *do*. Consequently, notation that is compact and efficient in the eyes of the practiced is murky for the rest of us. We promise to use a consistent notation in expressions that are fully explained, believing that clarity trumps elegance. We cannot avoid equations, but we can strive to make them understandable.

We use model diagrams[1] throughout the book. These sketches portray stochastic relationships among data, unobserved states, and parameters. We show how these diagrams, properly composed, provide a blueprint for writing out Bayesian models, and ultimately, for designing the samplers that allow us to obtain probability distributions of the quantities we seek to understand.

Unlike many existing books on Bayesian methods, ours will not emphasize computer code written in any specific language. We will teach algorithms but not coding. We choose this approach because there are an increasing number of software packages that implement algorithms for Bayesian analysis (e.g., Lunn et al., 2000; Plummer, 2003; INLA Development Team, 2014; Stan Development Team, 2014). The diversity of this software is likely to expand in the future. Including today's favorite flavor of code in our book assures that it will become obsolete as a new favorite emerges. Moreover, writing your own algorithms in the programming language of your choice is not all that difficult and offers the added benefit of forcing you to think through what you are doing. We make a single exception to this "no-code" rule in part II by including a script to illustrate the relationship between mathematical expressions and their implementation.

[1] Formally known as directed acyclic graphs or, more succinctly, Bayesian networks.

I Fundamentals

1 Preview

Art is the lie that tells us the truth.
—*Pablo Picasso*

All models are wrong but some are useful.
—*George E. P. Box*

Pablo Picasso was a contemporary of George Box's, a statistician who had an enormous impact on his field, writing influential papers well into his 90s. Both men sought to express truth about nature, but they used tools that were dramatically different—Picasso brushing strokes on canvas, and Box writing equations on paper. Given the different ways they worked, it is remarkable that Box and Picasso made such similar statements about the central importance of abstraction to insight. Abstraction plays a role in all creative human enterprise—in art, music, literature, engineering, and science. We create abstractions because they allow us to focus on the most important elements of a problem, those relevant to the objectives of our work, without being distracted by elements that are not relevant.

Scientific models are, above all else, abstractions. They are statements about the operation of nature that purposefully omit many details and thus achieve insight that would otherwise be discursively obscured. They provide unambiguous statements of what we believe is important. A key principle in modeling and statistics—in science for that matter—is the need to reduce the dimensions of a problem. A data set may contain a thousand observations. By reducing its dimensions to a model with a few parameters, we are able to gain understanding.

However, because models are abstractions and reduce the dimensions of a problem, we must deal with the elements we choose to omit. These elements create uncertainty in the predictions of models, so it follows that assessing uncertainty is fundamental to science. Scientists, journalists, logicians, and attorneys alike can rightly claim to make statements based on evidence, but only scientific statements include evidence tempered by uncertainty quantified. We know what is certain only to the extent that we can say, with confidence, what is uncertain. Sharpening our thinking about uncertainty and learning how to estimate it properly is a main theme of this book.

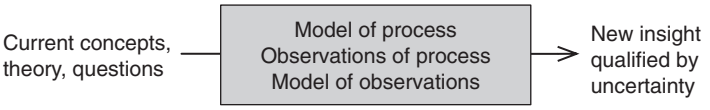

Figure 1.1.1. The fundamental challenge in ecological research is to establish a credible line of inference extending from concepts and theory to new insight tempered by uncertainty.

Your science will have impact to the extent that you are able to ask important questions and provide compelling answers to them. Doing so depends on establishing a line of inference that extends from current thinking, theory, and questions to new insight qualified by uncertainty (fig. 1.1.1). This book offers a highly general, flexible approach to establishing this line of inference. We cannot help you pose novel, interesting questions, but we can teach an approach to inference applicable to an enormous range of research problems, an approach that can be understood from first principles and that can be unambiguously communicated to other scientists, managers, and policy makers. We emphasize that understanding the principles of this framework allows you to customize your analyses to accommodate the inevitable idiosyncrasies of specific problems in research.

We sketch that framework in this chapter to give a general sense of where this book is headed, a preview we use to motivate the development of concepts and principles in the chapters that follow. There should be details of our approach that are unfamiliar, otherwise you probably don't need this book. Soon enough, we will explain those details fully. For now, we offer a somewhat abstract overview followed by a concrete example as an enticement to read on. It will be rewarding to return to this section after you have worked through the book. We hope you will be pleasantly surprised by your increased understanding of our small preview. The only part of this chapter that is essential for the remainder of the book is understanding our notation, which we describe in section 1.1.1.

1.1 A Line of Inference for Ecology

Virtually all research problems in ecology share a set of features. We want to understand how the state of an ecological system changes over time, across space, or among individuals. We seek to understand why those changes occur. Our understanding usually depends on a sample drawn from all possible instances of the state because we want to make statements about a system that is too large to observe fully. The observations in that sample

are often related imperfectly to the true state. In the subsections that follow, we lay out an approach first described by Berliner (1996)[1] for modeling the imperfect observations that arise from a process we want to understand. It does not apply to all research problems, but it is sufficiently general and flexible that it applies to most.

1.1.1 Some Notation

Before we proceed, we must introduce some notation. Boldface lowercase letters will indicate vectors (e.g., $\boldsymbol{\theta}$, $\mathbf{a}$), and lightface lowercase letters, scalars (θ, a).[2] Bold capital letters will be used for matrices (e.g., $\mathbf{A}$). The symbol $\boldsymbol{\theta}$ will indicate a vector of parameters, and, of course, θ will indicate a single parameter.[3] The letter $\mathbf{y}$ will indicate a vector of data, $\mathbf{Y}$ a matrix, and y or y_i a single observation. Corresponding notation using $\mathbf{x}$, x, and x_i will be used for predictor variables, also called covariates. The notation $[a|b, c]$ will be used for the probability distribution of the random variable a conditional on the parameters b and c.[4] Deterministic models will be denoted by $g()$ with arguments necessary for the model within the parentheses. Notation will be added as needed, in context.

1.1.2 Process Models

Process models include a mathematical statement that depicts a process and a way to account for uncertainty about the process. To compose a process model we start by thinking about the true state (z) of an ecological system. That state could be the size of a population, the flux of nitrogen from the soils of a grassland, the number of invasive plants in a community, or the area of landscape annually disturbed by fire. We seek to understand influences on that state, the things that cause it to change. We write an equation,[5] a deterministic model that represents our ideas about the

[1]For elaboration, see Wikle (2003); Clark (2003b); Cressie et al. (2009), and Wikle et al. (2013).

[2]It might be useful to review scalars, vectors, and matrices. Most ecologists are familiar with matrices—rows and columns of numbers. Vectors are "one dimension" of a matrix, that is, a row or a column. Scalars are a single element. So, a matrix might be $\mathbf{A} = \begin{pmatrix} c & d \\ d & f \end{pmatrix}$; a vector, $\mathbf{a} = (c, d)'$; and a scalar, c. We will use the notation $()'$ to list the elements of a vector.

[3]We illustrate with the familiar example $y = \beta_0 + \beta_1 x$. The vector of parameters from the model is $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. When discussing parameters generically, we will use $\boldsymbol{\theta}$.

[4]The symbol "|" reads conditional on.

[5]Or equations. To keep things simple, we focus on a single equation here, but we might build a system of equations making multiple predictions of states.

behavior of the state of interest and the quantities that influence it.[6] When we say the model is deterministic, we mean that for a given set of parameters and inputs, it will make precisely the same predictions. We use the notation $g(\boldsymbol{\theta}_p, \mathbf{x})$ to represent the deterministic part of a process model, where $g()$ is any mathematical function, $\boldsymbol{\theta}_p$ is a vector of parameters in the model, and $\mathbf{x}$ is one or more explanatory variables that we hypothesize influence the true state.

Our deterministic model is an abstraction, so it follows that we have omitted influences on the true state from the model, and we must deal with our omissions. If we model aboveground net primary production of a grassland as a function of growing season precipitation, we have brushed aside the influence of grazing intensity and precipitation that occurs during the dormant season. If we model reproductive success of individuals as a function of age and genotype, we have ignored variation contributed by their nutritional status. A model of harvest from a fishery based on observations of stock size and sea temperature omits the effect of variation in the food web. We recognize that these neglected influences shape the behavior of the true state by treating them stochastically, by estimating a parameter, σ_p^2, that subsumes all the unmodeled influences on the true state. Including this stochastic component allows us to estimate a statistical distribution (fig. 1.1.2A) for the true state,

$$\underbrace{\left[z|g\left(\boldsymbol{\theta}_p, \mathbf{x}\right), \sigma_p^2\right]}_{\text{process model}}, \tag{1.1.1}$$

where the bracket notation $[z|g(\boldsymbol{\theta}_p, \mathbf{x}), \sigma_p^2]$ means the distribution of z conditional on $g(\boldsymbol{\theta}_p, \mathbf{x})$ and σ_p^2.[7] If the notation is somewhat unfamiliar at this point, don't worry; equation 1.1.1 simply says that if we know the functional form $g()$ and the values of $\boldsymbol{\theta}_p$, $\mathbf{x}$, and σ_p^2, we can specify the probability distribution of the true state, z (fig. 1.1.2 A).

We want to determine the probability distribution of the true state as well as the probability distributions of the parameters in our model. Doing so

[6] As a tangible example, we might model the influence of phytoplankton biomass (x) on zooplankton biomass (z) with a linear model, $z = \theta_0 + \theta_1 x$. In this case $z = g(\boldsymbol{\theta}, x) = \theta_0 + \theta_1 x$.

[7] This notation was first introduced by Gelfand and Smith (1990) as a way to reduce clutter. It has become widely used by statisticians and ecologists. There is a small caveat needed here. We are using the arguments for probability distributions as if they were the mean ($g(\boldsymbol{\theta}_p, x)$) and variance ($\sigma_p^2$) of the distributions. The arguments for most of the statistical distributions are parameters that are functions of the mean and variance, a topic we treat in detail in section 3.4.4. For now, we take the liberty of treating the mean and variance as the needed parameters because they are familiar to ecologists.

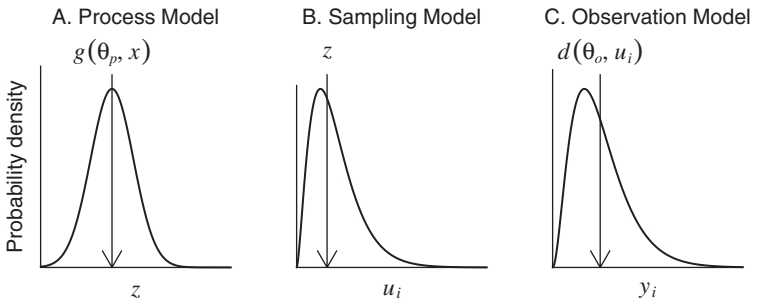

Figure 1.1.2. An observation on the operation of an ecological process (y_i) is linked to ideas about how the process works via three linked probability distributions. Hierarchical models consist of linked distributions. As we move from panel **A** to **C**, notice how the random variables (on the x-axis) becomes the the mean (or other central tendency) of the distribution in the next panel. **(A)** The deterministic model, $g(\boldsymbol{\theta}_p, \mathbf{x})$ predicts a true state, z, as a function of parameters $\boldsymbol{\theta}_p$ and predictor variables $\mathbf{x}$. There is uncertainty in the predictions, σ_p^2, that arises because there are influences on z that are not represented in the model. Thus, the distribution in **A** is $[z|g(\boldsymbol{\theta}_p, \mathbf{x}), \sigma_p^2]$. If the deterministic model predicts z well, then the distribution in **A** shrinks toward the $g(\boldsymbol{\theta}_p, \mathbf{x})$ arrow. **(B)** There are almost always more instances of z in nature than we can hope to observe. Observations of individual instances of z define a sampling distribution, $[u_i|z, \sigma_s^2]$, the breadth of which depends on σ_s^2. As variation among observations declines, the distribution in **B** shrinks toward the z arrow. Note that the mean of the distribution shown by the arrow is not at the peak of the distribution, because the distribution is skewed. (**C**) Observations (y_i) of instances of the true state u_i are often biased, such that $y_i \neq u_i$. A deterministic observation model, $d(\boldsymbol{\theta}_o, u_i)$, corrects for this bias. Uncertainty in this correction (σ_o^2) leads to the observation distribution $[y_i|d(\boldsymbol{\theta}_o, u_i), \sigma_o^2]$. As uncertainty in the observation model declines, the distribution in **C** shrinks toward the $d(\boldsymbol{\theta}_o, u_i)$ arrow.

requires evaluating the predictions of the process model against data. The data can be obtained in experiments or observational studies; they can be measurements we plan to collect or have already collected. This linkage between process models and observations is discussed next.

1.1.3 Sampling Models

We can rarely observe all instances of the true state in the system we study. Instead, we take a sample of $i = 1, \ldots, n$ observations of the true state and we notate the ith observation as u_i. This sample might be biomass from plots on a grassland landscape where we seek to understand the true state, aboveground productivity. It might be presence or absence of an exotic fungus on trees in a stand where we want to understand

infestation of the stand. It might be classifications of zooplankton in aliquots from a stream where we want to estimate the stream's species richness. Uncertainty arises because our sample assuredly will not represent the true state perfectly. Again, we represent this uncertainty stochastically using a probability distribution relating the true state to an observation

$$\underbrace{\left[u_i|z, \sigma_s^2\right]}_{\text{sampling model}}, \tag{1.1.2}$$

where σ_s^2 represents sampling variation (fig. 1.1.2 B). Expression 1.1.2 implicitly assumes that we can observe instances of the true state without bias, which simply says that if we collect many observations, then the average (i.e., also called the *expected value* of the observations, $\mathrm{E}(u)$) of the observations equals the mean of the distribution of the true state, $\mathrm{E}(u) = z$. (Expectation will be treated in detail in chapter 3.) We realize samples of the true state is a nuanced concept—soldier on, things will become clear in the next section.

1.1.4 Observation Models

The assumption that we can observe the true state perfectly may not be reasonable. When we count animals, some are overlooked. When we use Lidar[8] to estimate the heights of 10,000 trees, we do not measure the height of each tree using a ladder and a meter tape (thankfully) but instead observe backscatter from a laser beam. When we measure nitrogen mineralization, we do not follow the fate of individual nitrogen atoms but measure the net change in the extractable soil ammonium pool over time. The mismatch between what we observe and the true state requires a model of the observations, which we notate as $d(\boldsymbol{\theta}_o, u_i)$, where $\boldsymbol{\theta}_o$ are parameters. It is important to understand that u_i is the quantity we would observe if we could *perfectly* observe the instance of the true state in a draw from all the instances, without any bias injected by our observation process. We use y_i to notate the actual measurements we have in hand, including error resulting from the way we observe the u_i. The observation model serves to eliminate the bias found in our observations, y_i relative to an instance (u_i) of the true state drawn from the distribution of z. The probability distribution of the observations (fig. 1.1.2 C) arising from the

[8]Lidar, light detection and ranging, a technique used in remote sensing.

observed instances of the true state is

$$\underbrace{\left[y_i | d(\boldsymbol{\theta}_o, u_i), \sigma_o^2\right]}_{\text{observation model}}, \tag{1.1.3}$$

where the σ_o^2 represents all the influences on the y_i that are not represented in $d(\boldsymbol{\theta}_o, u_i)$. As a simple example, the σ_o^2 could be the variance of the predictions of a regression model used to calibrate observations against true values. We emphasize that model $d(\boldsymbol{\theta}_o, u_i)$ is needed to offset bias in the y_i. If our observations are unbiased, then there is no need for an observation model, (i.e., $y_i = u_i$) and the uncertainty in our data arises solely from sampling variation (i.e., eq. 1.1.3). For simplicity, we have ignored the sampling variation and observation errors that might influence the $\mathbf{x}$, but these could be handled in the same as we have done for the $\mathbf{y}$ (i.e., we would use probability distributions like eqs. 1.1.2 and 1.1.3).

1.1.5 Parameter Models

Because the approach we sketch is Bayesian, we also require models of the parameters expressing what we knew about the parameters when we began our investigation, that is, our *prior* knowledge. This knowledge is expressed in probability distributions, one for each parameter we seek to estimate[9]

$$\underbrace{\left[\boldsymbol{\theta}_p\right]\left[\boldsymbol{\theta}_o\right]\left[\sigma_p^2\right]\left[\sigma_s^2\right]\left[\sigma_o^2\right]}_{\text{parameter models}}. \tag{1.1.4}$$

These distributions must have numeric arguments that specify our current knowledge of the probability distribution of the parameters. The arguments can be chosen to make the distributions informative or vague, but as you will see, we will encourage you to make priors as informative as knowledge and scholarship allows. We might know a lot about a parameter or we might know very little.

1.1.6 The full Model

We are now equipped to write a mathematical expression representing our ideas about the operation of an ecological process linked to data in a way

[9]Equation 1.1.4 requires the assumption that the parameters $\boldsymbol{\theta}_p, \boldsymbol{\theta}_d, \sigma_p^2, \sigma_s^2$, and σ_o^2 are statistically independent. We will explain this idea in greater detail later.

that includes all sources of uncertainty—in the process, in our sample of the process, and in the way we observe it (fig. 1.1.2)

$$\underbrace{\left[y_i | d(\boldsymbol{\theta}_o, u_i), \sigma_o^2\right]}_{\text{observation model}} \underbrace{\left[u_i | z, \sigma_s^2\right]}_{\text{sampling model}} \underbrace{\left[z | g\left(\boldsymbol{\theta}_p, \mathbf{x}\right), \sigma_p^2\right]}_{\text{process model}} \underbrace{\left[\boldsymbol{\theta}_p\right]\left[\boldsymbol{\theta}_o\right]\left[\sigma_p^2\right]\left[\sigma_s^2\right]\left[\sigma_o^2\right]}_{\text{parameter models}} . \tag{1.1.5}$$

Equation 1.1.5 is Bayesian and hierarchical.[10] It is *Bayesian* because it treats the unobserved quantities as random variables. This treatment allows us to make statements about the probability distributions of all of the unobserved quantities based on the observed ones. It is *hierarchical* because the z and u_i are found on both sides of a conditioning symbol " | ", illustrating a powerful tool for simplifying problems that we will discuss more fully soon. The observation model includes our knowledge of the relationship between the true state and our observations of it and the uncertainty that occurs because that relationship is imperfect. The sampling model includes the uncertainty that comes from observing a subset of instances of the true state. The process model represents our hypotheses about the ecological process by specifying a probability distribution defining our knowledge and our uncertainty about the true state and the factors that control its behavior. The parameter models allow us to exploit previous estimates of parameters that we have made ourselves or that have been made by others. Together these models provide a line of inference extending from concepts to insight for a broad range of research problems (fig. 1.1.1). We can use equation 1.1.5 to obtain estimates of unobserved states, parameters, and quantities of interest derived from parameters and states. All of these estimates are properly tempered by uncertainty in a statistically coherent way.

In the remainder of this book, we develop the principles needed to understand equation 1.1.5 and to apply it to research problems in ecology. We tailor it to match the needs of the particular problem at hand. But first, we provide an example of its use.

[10]The symbol $\propto$ means "is proportional to." It may be confusing at this point that the observed $\mathbf{x}$ does not show up in the list of observed quantities on the left-hand side of the proportionality. Bear with us. We will make that clear as we proceed.

1.2 An Example Hierarchical Model

We are now going to apply the general framework we have previewed to a specific problem. Remember that we don't expect this section to be familiar to you. Although the application here focuses on learning about population dynamics of a large mammal from a time series of observations in East Africa, it could be about any topic, any research design, and any location. We urge you to draw analogies to your own work as the example develops.

The Serengeti wildebeest (*Connochaetes taurinus*) population migrates across the grasslands of Tanzania and Kenya in an annual cycle driven by availability of green plant biomass (Boone et al., 2006). During the late 1800s, the viral disease rinderpest created a panzootic, decimating the wildebeest and other wild and domestic ruminants. Numbers of these animals remained low until the 1950s when a campaign to vaccinate cattle in the pastoral lands surrounding the Serengeti eliminated the virus in wildlife (Plowright, 1982). Survivorship of wildebeest, particularly juveniles, doubled after rinderpest was eradicated, and the population grew rapidly until the mid 1970s when density-dependent mortality produced a quasi-equilibrium. This steady state appears to have been caused by intraspecific competition for green plant biomass during the dry season. This biomass varies in approximate proportion to annually variable rainfall (Mduma et al., 1999).

Developing informative models depends on a clearly stated question. An unambiguous question is vital because it guides the abstraction that we formulate; it informs our decisions about which variables and parameters we need to include in our model. In this particular example we ask the question: How does variation in weather modify feedbacks between population density and population growth rate in a population of large herbivores occupying a landscape where precipitation is variable in time? Answering this question requires a model that portrays density dependence, effects of precipitation, and their interaction.

1.2.1 Process Model

We pose the deterministic model portraying growth of the wildebeest population as

$$N_t = g(\boldsymbol{\beta}, N_{t-1}, x_t, \Delta t) = N_{t-1} e^{(\beta_0 + \beta_1 N_{t-1} + \beta_2 x_t + \beta_3 N_{t-1} x_t)\Delta t}, \qquad (1.2.1)$$

where N_t is the unobserved, true abundance of wildebeest in year t (corresponding to z in eq. 1.1.5); x_t is a measure of the annual rainfall

that influences population growth during $t - 1$ to t; and $\Delta t = 1$ year.[11] The meaning of the coefficients $\boldsymbol{\beta}$ requires some thought. A critical part of gaining insight from models and data is careful thinking about the biological meaning of the parameters in our models and their relationship to existing concepts and theory, the need for which is illustrated here.

We first consider the intercept, β_0, a logical place to start. If the units of precipitation, x_t, are centimeters per year, ranging from 0 to a large number, then algebra dictates that $e^{\beta_0 \Delta t}$ is the proportional change in the population during the period t to $t + 1$ that occurs when abundance is zero and rainfall is zero. The abundance equals zero part might seem a bit odd at first glance, because there is no population to grow when $N_t = 0$, but the parameter nonetheless serves to define the upper limit on population growth rate as numbers approach zero. There is no problem here. But the part about zero rainfall is troubling because it makes β_0 difficult to interpret biologically or at least makes it somewhat unuseful in its biological interpretation. Why would we want to estimate a parameter determining population growth when the population is low and rainfall is low? Alternatively, if we define x_t in terms of divergence from the long-term average by subtracting the mean rainfall from each year's observation, then the definition of β_0 becomes more sensible. In this case, $e^{\beta_0 \Delta t}$ is the proportional change in population size that occurs when the population is zero and rainfall is average. This definition allows us to relate our model to well-established theory; β_0 is analogous to the intrinsic rate of increase in the logistic equation.[12]

Once we have defined the intercept in a sensible way, the slope terms are easily interpreted. The parameter β_1 represents the strength of density dependence, that is, the change in the per capita population growth rate that occurs for each animal added to the population.[13] Again relating our model to classical theory, $\beta_1 = -(r_{max}/K)$, where K is the carrying capacity, that

[11] The identical form $\log(N_t) = \log(N_{t-1}) + (\beta_0 + \beta_1 N_{t-1} + \beta_2 x_t + \beta_3 N_{t-1} x_t)\Delta t$ might be more familiar to ecologists accustomed to a general linear modeling framework.

[12] This might require a bit of explanation. Equation 1.2.1 can be rearranged to $\log(N_t/N_{t-1}) = (\beta_0 + \beta_1 N_{t-1} + \beta_2 x_t + \beta_3 N_{t-1} x_t)\Delta t$. If we drop the $\beta_2 x_t$ and $\beta_3 x_t N_{t-1}$ terms and assume Δt is small, the resulting expression approximates $1/N \cdot dN/dt = r - rN/K$, where N is the population size, r is the intrinsic rate of increase, and K is the carrying capacity—the long-term average population size at which $1/N \cdot dN/dt = 0$. Thus, the portion of our model representing density dependence is a version of the logistic equation in discrete time, also known as the *Ricker equation*, where $\beta_0 = r$, and $\beta_1 = r/K$.

[13] You might be tempted to make the sign for β_1 negative, because density dependence reduces the per capita population growth rate as population size increases. There is a good reason to use addition. When the model is fit to data, we want to estimate the sign and the value of β_1 without needing to "back transform" the sign. Always use addition in additive models that you will fit to data.

is, the population size at which the long-term average population growth rate equals zero. The parameter β_2 expresses the strength of the effect of variation in rainfall, that is, the change in the per capita population growth rate per unit deviation from the long-term mean. Finally, the parameter β_3 determines the magnitude of the effect of rainfall on the effect of density.

Estimating the β's will inform the question we posed, but we don't pretend that they capture all the influences on wildebeest population dynamics. We recognize that there are other processes, for example, predation, poaching, and disease, that shape wildebeest numbers over time. We can acknowledge these other processes exist without portraying them explicitly. Instead, we lump these unmodeled effects into a single parameter, σ_p^2. The model of the process for a population at time t including deterministic and stochastic components is

$$\left[N_t | g\left(\boldsymbol{\beta}, N_{t-1}, x_t, \Delta t\right), \sigma_p^2\right]. \tag{1.2.2}$$

1.2.2 Sampling Model

The wildebeest population was estimated[14] on 20 occasions during 1961–2008 using spatially replicated counts of animals on georectified aerial photographs arrayed along transects, with each photograph covering a known area (Norton-Griffiths, 1973). For simplicity, we ignore the aspect of sampling contributed by the transects and treat the photographs as if they were a random sample from the area used by the population. Thus, we assume that each photograph provided a statistically independent estimate of population density calculated as the observed count divided by the area covered by the photograph. The stochastic sampling model representing the relationship between these observations and the true population size is

$$\left[y_{tj} \,\middle|\, \frac{N_t}{a}, \sigma_s^2\right], \tag{1.2.3}$$

where y_{tj} is the density of animals at year t on photograph j, and a is the total area from which the sample of photographs was drawn—the area used by the population we are modeling—which we assume to be a known constant. Note that equation 1.2.3 implies that there is no bias in the estimate of animal density on a given photograph, which is the assumption made

[14]Portions of this time series have been published in Hilborn and Mangel (1997) and Mduma et al. (1999). The most recent data were graciously provided by A.R.E. Sinclair, Ray Hilborn, and Grant Hopcraft. We use these data later to illustrate making inference from a single model.

by the researchers who collected the data. Although this assumption is reasonable for large animals counted in open habitat, we could obviate the need for this assumption by modeling the animals that were present and not observed.

As a purely hypothetical example, imagine that we fitted a sample of animals with high-resolution telemetry instruments that would allow us to know whether each animal was within a photograph. Imagine also that we marked them with highly visible neckbands to allow us to distinguish between animals that were fitted with instruments and those that were not. We could use these observations to estimate the probability (ψ) that an animal truly present was counted. In this case our combined observation and sampling model would be

$$\underbrace{\left[y_{tj} | \psi n_{tj}, \sigma_o^2\right]}_{\text{observation model}} \underbrace{\left[n_{tj} \,\middle|\, \frac{N_t}{a}, \sigma_s^2\right]}_{\text{sampling model}}, \tag{1.2.4}$$

where n_{tj} are animals that are truly present on photograph j during year t, and σ_o^2 is the uncertainty associated with the estimate of ψ. We could also deal explicitly with the likely cases in which it was not certain whether an animal was on a sampled photograph.

To simplify the example and to exploit published data, we make the heroic assumption that rainfall at time t, x_t, is measured without error; we are treating it as known, but we are not obliged to do so. We could develop sampling models and observation models for the rainfall data in the same way we did for the count data (i.e., eqs. 1.2.3 and 1.2.4).

1.2.3 Full Model

Combining the sampling model (eq. 1.2.3) and process model (eq. 1.2.2) with models for the parameters, we obtain the foundation for a full Bayesian analysis capable of estimating the parameters and the unobserved, true population size (fig. 1.2.1); however, work remains to be done. The full model written thus far (fig. 1.2.1) applies to a single year of observations and a single photograph, but, of course, we want to use all the years and all the observations within a year. We have not yet chosen specific probability distributions for the stochastic components, and we must do so in a sensible way. We may need to deal with correlation among photographs or among annual estimates in the time series of data, and with errors in the estimation of rainfall. We need to lay out a method for numerically estimating the parameters and unobserved states. However, although tasks remain, all

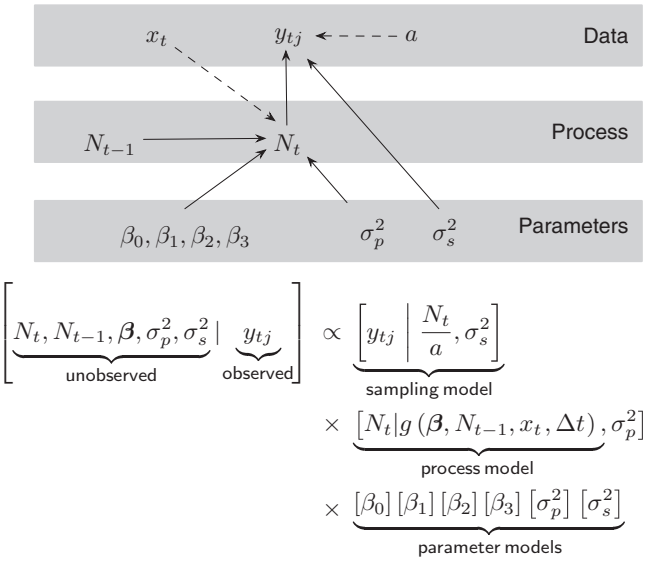

Figure 1.2.1. Hierarchical Bayesian model of the dynamics of the Serengeti wildebeest population. The true population size at time t is modeled using the deterministic model $g(\boldsymbol{\beta}, N_{t-1}, x_t, \Delta t) = N_{t-1}e^{(\beta_0+\beta_1 N_{t-1}+\beta_2 x_t+\beta_3 N_{t-1} x_t)\Delta t}$, which represents the effects of the true, unobserved population size (N_t); the dry season rainfall (x_t); and their interaction on population growth. The biological interpretation of the parameters $\boldsymbol{\beta}$ are given in the text. The parameter σ_p^2 represents all the effects on wildebeest abundance not represented by the $\boldsymbol{\beta}$. The y_{tj} is a single observation of population density obtained from one of $j = 1, \ldots, n_t$ photographs chosen from area = a km^2. The parameter σ_s^2 represents sampling error. The arrow diagram is a Bayesian network, also called a *directed acyclic graph*, that can be used as a visual guide for properly writing the full model. The solid lines show stochastic relationships. The dashed lines show deterministic relationships, implying that the quantities at the tails of the arrows are known without error. Instructions for drawing diagrams like this one will be developed in subsequent chapters.

proceed in a logical way from the foundation we have built (fig. 1.2.1). All are manageable. Completing them allows us to reliably estimate unobserved states, parameters, and derived quantities of interest.

1.3 What Lies Ahead?

One of the aims of this book is to enable ecologists to write equations for models that allow data to speak informatively. Our goal is to provide

an understanding of principles needed to accomplish this vital task. We will show how an enormous range of research problems in ecology can be decomposed into a set of sensible parts, as we have illustrated. The specific example we offered no doubt raised more questions than it answered, which was our intention. Notable among these questions might be the following:

1. Why does the Bayesian approach to analysis work? What are the probabilistic foundations of inference from models like this one?
2. How does a Bayesian analysis relate to analyses based on maximum likelihood?
3. How do we choose the appropriate statistical distributions for the stochastic components of the model (fig. 1.2.1)?
4. How can we incorporate multiple sources of data in estimates of parameters and states?
5. How do numerical methods allow us to implement the model and obtain results? How do these methods work?
6. How can we evaluate the model to assure it adequately represents the data?
7. What can we conclude about the operation of the ecological process based on estimates of the parameters and states? What do we do about derived quantities; for example, how do we make inferences about the population size where growth rate is maximum?
8. What if we want to evaluate the strength of evidence for this model (equation 1.2.1) relative to competing ones, for example, a model with nonlinear density dependence?
9. How do we predict future states with honest estimates of uncertainty?

Answering these questions in a clear and accessible way for a broad range of research problems in ecology is the goal of remainder of this book.

2 Deterministic Models

We write models to express our ideas about how ecological processes work (figure 0.0.1 A). We then use data to decide if those ideas are any good. Combining data with models provides insight about ecology by answering a fundamental question: What is the probability that I will observe the data if the model faithfully represents the processes that give rise to the data? Our starting point for answering this question will always be consideration of an ecological process. We formalize our thoughts by writing a deterministic equation or equations making predictions that can be compared with observations (fig. 0.0.1 A). There is no uncertainty in our predictions because the model is deterministic. In this chapter we talk about deterministic models as expressions of ecological hypotheses. In the next chapter we lay out the principles of probability that allow us to combine these models with data in a way that provides new knowledge qualified by uncertainty.

2.1 Modeling Styles in Ecology

We wrote this book to be broadly useful across all the subdisciplines of ecology, offering a way of working with models and data that is recognizably valuable, whether you study genes or ecosystems. To that end, we briefly discuss the different ways that deterministic models have been used in ecology. We want to ensure that ecologists who work on different kinds of problems recognize a familiar style of modeling, and have confidence that what lies ahead is relevant to their own research.

Modeling has formed a productive, diverse activity throughout the history of ecology (Kingsland, 1985; Golley, 1993). Subdisciplines of