PREFACE

This book is a gentle introduction to applied Bayesian modeling for ecologists using the highly acclaimed, free WinBUGS software, as run from program R. The bulk of the book is formed by a very detailed, and yet, I hope, enjoyable tutorial consisting of commented example analyses. These form a progression from the trivially simple to the moderately complex and cover linear, generalized linear (GLM), mixed and generalized linear mixed models (GLMM). Along the way, a comprehensive and largely non-mathematical overview is given of these important model classes, which represent the core of modern applied statistics and are those which ecologists use most in their work. I provide complete R and WinBUGS code for all analyses; this allows you to follow them step by step and in the desired pace. Being an ecologist myself and having collaborated with many ecologist colleagues, I am convinced that the large majority of us best understands more complex statistical methods by first executing worked examples step by step and then by modifying these template analyses to fit their own data.

All analyses with WinBUGS are directly compared with analyses of the same data using standard R functions such as `lm()`, `glm()` and `lmer()`. Hence, I would hope that this book will appeal to most ecologists regardless of whether they ultimately choose a Bayesian or a classical mode of inference for their analyses. In addition, the comparison of classical and Bayesian analyses should help to demystify the Bayesian approach to statistical modeling. A key feature of this book is that all data sets are simulated (= “assembled”) before analysis (= “disassembly”) and that fully commented R code is provided for both. Data simulation, along with the powerful, yet intuitive model specification language in WinBUGS, represents a unique way to truly understand that core of applied statistics in much of ecology and other quantitative sciences; GLMs and mixed models.

This book traces my own journey as a quantitative ecologist towards an understanding of WinBUGS for Bayesian statistical modeling, and of GLMs and mixed models. Both the simulation of data sets and model fitting in WinBUGS have been crucial for my own advance in these respects. The book grew out of the documentation for a 1-week course that I teach at the graduate school for life sciences at the University of Zürich, Switzerland, and elsewhere to similar audiences. Therefore, the typical readership would be expected to be advanced undergraduate and graduate students and researchers in ecology and other quantitative sciences. To maximize your benefits, you should have some basic knowledge in R computing as well as statistics at the level of the linear model (i.e., ANOVA and regression).

After three introductory chapters, normal linear models (LM) are dealt with in chapters 4–11. In chapter 9 and especially 12, they are generalized to contain more than a single stochastic process, i.e., to the (normal) linear mixed model (LMM). Chapter 13 introduces the generalized linear model (GLM), i.e., the extension of the normal linear model to allow error distributions other than the normal. Chapters 13–15 feature Poisson GLMs and chapters 17–18 binomial GLMs. Finally, the GLM, too, is generalized to contain additional sources of random variation to become a generalized linear mixed model (GLMM) in chapter 16 for a Poisson example and in chapter 19 for a binomial example. I strongly believe that this step-up approach, where the simplest of all linear models, that “of the mean” (chapter 4), is made progressively more complex until we have a GLMM, helps you to get a synthetic understanding of these model classes, which have such a huge importance for applied statistics in ecology and elsewhere.

The final two main chapters go one step further and showcase two fairly novel and non-standard versions of a GLMM. The first is the site-occupancy model for species distributions (chapter 20; MacKenzie *et al.* 2002; 2003; 2006) and the second is the binomial (or N-) mixture model for estimation and modeling of abundance (chapter 21; Royle 2004a). These models allow one to make inference about two pivotal quantities in ecology: distribution and abundance of a species (Krebs 2001). Importantly, these models fully account for the imperfect detection of occupied sites and individuals, respectively. Arguably, imperfect detection is a hallmark of all ecological field studies. Hence, these models are extremely useful for ecologists, but owing to their relative novelty are not yet widely known. Also, they are not usually described within the GLM framework, but I believe that recognizing how they fit into the larger picture of linear models is illuminating. The Bayesian analysis of these two models offers clear benefits over that by maximum likelihood, for instance in the ease with which finite-sample inference is obtained (Royle and Kéry 2007), but also just heuristically, since these models are easier to understand when fit in WinBUGS.

Owing to its gentle tutorial style, this book should be excellent to teach yourself. I hope that you can learn much about Bayesian analysis using WinBUGS and about linear statistical models and their generalizations by simply reading it. However, the most effective way to do this obviously is by sitting at a computer and working through all examples, as well as by solving the exercises. Fairly often, I just give the code required to produce a certain output but do not show the actual result; so to fully grasp what is happening it is best to execute all code.

If the book is used in a classroom setting and plenty of time is given to the solving of exercises, then up to two weeks might be required to cover all material. Alternatively, some chapters may be skipped or left for the students to go through for themselves. Chapters 1–5, inclusive, contain key material. If you already have experience with Bayesian inference you may skip chapters 1–2. If you understand well (generalized) linear models, you may also skip chapter 6 and just skim chapters 7–11 to see whether you can easily follow. Chapters 9 and 12 are the key chapters for your understanding of mixed models, whether LMM or GLMM, and should not be skipped. The same goes for chapter 13, which introduces GLMs. The next chapters, 14–19, are examples of (mixed) GLMs and may be sampled selectively as desired. There is some redundancy in content, e.g., between the following pairs of chapters, which illustrate the same kind of model for a Poisson and a binomial response: 13/17, 15/18, and 16/19. Finally, chapters 20 and 21 are somewhat more specialized and may not have the same importance for all readers (though I find them to be the most fascinating models in the whole book).

As much as I believe in the great benefits of data simulation for your understanding of a model, data assembly at the start of each chapter may be skipped. You can download all data sets from the book website or simply execute the R code to generate your own data sets, and only go to the line-by-line mode of study where the analysis begins. Similarly, comparison of the Bayesian solutions with the maximum likelihood estimates can be dropped by simply fitting the models in WinBUGS and not in R also.

All R and WinBUGS code in this book can be downloaded from the book website at http://www.mbr-pwrc.usgs.gov/software/kerybook/ maintained by Jim Hines at the Patuxent Wildlife Research Center. The website also contains some bonus material, a list of WinBUGS tricks, an Errata page, solutions to exercises, a text file containing all the code shown in the book as well as the actual data sets that were used to produce the output shown in the book. It also contains a real data set (the Swiss hare data) we deal with extensively in the exercises. The Swiss hare data contain replicated counts of Brown hares (*Lepus europaeus*: see chapter 13) conducted over 17 years (1992–2008) at 56 sites in 8 regions of Switzerland. Replicated means that each year two counts were conducted during a two-week period. Sites vary in area, elevation, and belong to two types of habitat (arable and grassland): hence, there are both continuous and discrete explanatory variables. Unbounded counts may be modeled as Poisson random variables with log(area) as an offset, but we can also treat the observed density (i.e., the ratio of a count to area) as a normal or the incidence of a density exceeding some threshold as a binomial random variable. Hence, you can practice with all models shown in this book and meet features of genuine data sets such as missing values and other nuisances of real life.