Preface

Statistical models are developed and used to make inferences from observational data in many areas of ecology, such as conservation biology, fisheries and wildlife management, and more theoretical areas, which include macroecology, metapopulation biology, and landscape ecology. The range of applications spans large, continental-scale monitoring programs that often have several competing objectives, and small-scale studies designed to address well-defined, and often more narrowly focused, scientific hypotheses. Regardless of context, statistical modeling is needed to compute inferences about the form or function of ecological systems, to predict the effects of ecological processes, and, in applied problems, to predict the consequences of management actions.

In this book we attempt to provide a modern view of statistical modeling and inference as applied in the analysis of populations, metapopulations, and communities. We rely exclusively on hierarchical models for this purpose. The hierarchical modeling framework that we advocate makes a clear distinction between an observation component of the model - which typically describes nuisance variation in the data – and the process component of the model - which is usually fundamental to the object of inference. Historically, these two components have not usually been distinguished in model formulation. Hierarchical models are sometimes referred to as state-space models. More recently, the term *hierarchical modeling* has been widely adopted and has become synonymous with Bayesian analysis. However, we believe that hierarchical modeling is a conceptual and philosophical approach to doing science, and it has tangible benefits independent of the framework that is adopted for formal inference, i.e., regardless of whether one is "Bayesian" or not. This distinction is not subtle, nor is it a matter of semantics. Hierarchical modeling is a conceptual framework that emphasizes model construction, whereas Bayesian analysis is concerned with technical aspects of inference. Many of our colleagues, when confronted with a complicated modeling problem, declare the need for a Bayesian, hierarchical model, usually with an emphasis on the Bayesian part. However, the Bayesian mode of inference is often selected merely as a computational expedient (owing to the availability of software for fitting these models), not because of any inferential benefits that might be gained from a Bayesian analysis of the problem. Such models can just as well be fitted using classical methods (i.e., maximizing an integrated likelihood). Thus, our book is not about Bayesian analysis per se; instead, we emphasize construction of hierarchical models. That said, we do apply the Bayesian inference paradigm to many problems, especially those whose solution by classical inference methods is incomplete.

A second motivation for this book is that our impression is that the statistical education of most ecologists does not include sufficient exposure to "modeling" – that is, the construction of probability models to describe ecological systems. For example, in modeling occurrence of species almost everyone knows that logistic regression is the preferred modeling framework. Unfortunately, such regression models are usually viewed as *procedures* for inference or prediction with little consideration given to the underlying probability model of species occurrence. In this way ecologists often fail to appreciate or understand what motivates the distributional assumption (independent Bernoulli outcomes) used in logistic regression. In general, we find that many ecologists are not adequately trained in the formulation of probability models for solving ecological inference problems. Our book is intended to help ecologists develop these skills.

We do not attempt comprehensive coverage of any particular topic in this book; however, we think we have achieved relatively broad coverage both in terms of ecological scale (populations, metapopulations, communities) and in the traditional methodological context by which ecological inference problems are classified and rendered into named procedures (e.g., site-occupancy models, capture-recapture, etc.). Moreover, we have tried to cover all of these problems in a cohesive way as opposed to viewing them as distinct entities requiring unique ("black box") solutions. We have tried to describe these problems in a relatively broad context that is unified by their basic formulation and a few common concepts. Our effort is incomplete to be sure, and many interesting and important classes of problems, new or alternative statistical formulations, and novel or distinctive applications remain.

0.1 SOFTWARE, IMPLEMENTATION AND WEBSITE

Another theme of our book is the development of fairly advanced implementation skills. It is our belief that if ecologists don't achieve these implementation skills, they can only do what someone else thinks they should do (say, using some specialized software package). In that regard, we adopt an implementation framework based on the flexible, comprehensive, and freely available software packages **R** and **WinBUGS**. We use these packages to illustrate calculations involved in data analysis and to plot results. If ecologists master these packages, they can do what they want to do. Together these packages provide a flexible computational platform with which any ecologist or data analyst can grow.

R (R Development Core Team, 2004) is an implementation of the S computing language (Becker et al., 1988; Chambers, 1998) and is freely available at the following web site:

`http://www.R-project.org.`

WinBUGS is an implementation of the BUGS language (Gilks et al., 1994) for specifying models and doing Bayesian analyses. **WinBUGS** is freely available at the following web site:

`http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`

and can be used alone or accessed remotely from other software packages. We prefer the latter method and use the **R** library, **R2WinBUGS** (Sturtz et al., 2005), to access the **WinBUGS** computational algorithms remotely while working in **R**.

There are easier ways to do some of the analyses described in our book. For example, we ask the reader to write **R** functions to specify and fit logistic regression models, even though such models are implemented in most statistical software packages. (Even **R** has its own generic functions for doing this). However, learning to write such functions will pay dividends over the long term. By learning to do the easy things the hard way, the hard things will become easy.

Complete, functional **R** and **WinBUGS** implementations for most analyses in the book can be accessed from the Web Supplement for the book, which can be found at the URL: `http://www.mbr-pwrc.usgs.gov/HMbook/`.

0.2 ORGANIZATION OF THIS BOOK

One logical way to organize the book is according to ecological scale of organization: models relevant to populations, models relevant to metapopulations, and models relevant to communities. However, the approach we take is organized more along the lines of technical relationships among the models. For example, after some introductory material (Chapters 1 and 2) we introduce occupancy models (Chapter 3), a metapopulation concept. We then introduce closed population models in their simplest form (Chapter 5). This progression allows us to establish a technical equivalence between the two models via "data augmentation" (Royle et al., 2007a). Though of marginal interest in this particular pair of problems, data augmentation is exploited repeatedly throughout the book as a means of developing useful generalizations or extensions of certain models. For example, in Chapter 6 we use it to develop models of heterogeneity and individual covariates for closed populations. The formulation of these models as "occupancy" models (using data augmentation) yields a generic and flexible solution to these respective problems. In another topical progression, we discuss site-occupancy models for open populations (Chapter 9) and then consider models of population dynamics, including both classical "Jolly-Seber" models (Chapter 10) and models of animal survival (classical "Cormack-Jolly-Seber" models, (Chapter 11)). This progression allows us to show that these two classes of models are, in terms of probability structure, effectively equivalent to a multi-season, site-occupancy model (admitting certain constraints). Thus, an important point of our book is to show that hierarchical formulations of these models establish a formal duality between models of occurrence (closed and open models of site occupancy) and models of individual encounter histories (closed and open models of populations). Such interrelationships would be difficult to emphasize in a different organization of the book.