Reviewer: Aaron Christ
Alaska Department of Fish and Game

Mixed Effects Models and Extensions in Ecology with R

Alain F. Zuur, Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev, Graham M. Smith
Springer-Verlag, New York, 2009.
ISBN 978-0-387-87457-9. 574 pp. USD 84.95 (H).
http://www.highstat.com/

This book is written in a very approachable conversational style. The additional focus on the heuristics of the process rather than just a rote recital of theory and equations is commendable. This type of approach helps the reader get behind the 'why' of what's being done rather than blindly follow a simple list of rules. There are, however, such lists presented; generally outlines on the steps necessary to go from initial data inspection to final concluding remarks in the report. This structure is very useful to the ever increasing legions of statistically literate researchers who may stumble for lack of depth in their statistical experience. For more experienced readers, the text could become rather tedious, but it is still useful to see how proper procedures are applied specifically in R. Either way, the extra amount of explanation can be valuable.

The style of R coding is equally approachable. The authors go out of their way to make code more readable rather than using many shortcuts and densely abbreviated code. This helps to represent the linear thought process used during the analysis rather than just focus on getting output in the least number of keystrokes. Following the examples in the text is also made much easier through the availability of all code presented in the book from the publisher's website.

The content can be divided into three general sections. The first is an introduction to underlying concepts and some familiarity with the general concepts of linear regression and additive modeling is assumed. References listed in the introduction would be essential for those not familiar with basic theories of linear regression and additive modeling. (Appendix A gives a cursory review by way of example, and it is more useful for readers less familiar with R to see how such modeling exercises are carried out.) The inglorious realm of underlying assumptions and limitations of the data is addressed in Chapter 2. General concepts of data exploration along with illustrative examples of how data stray from assumptions help reiterate the importance of not just knowing the structure of your data, but also the need to be sure that the tools you want to use are appropriate. Throughout the book, there is also a marked effort to go beyond blindly reporting output by also looking at various model validation plots.

The next several chapters round out the introductions and background. Chapter 3 introduces the concept of additive modeling using the `gam` function in both **gam** and **mgcv** packages. Chapter 4 explores variance heterogeneity and how to recognize and treat it appropriately. Chapter 5 not only introduces mixed effect models for nested data, but also contains an interesting example that contrasts the step-by-step results from following two parallel model selections—one starting from a full model and the other not—and spends precious space following *wrong* steps for thoroughness of the comparison. Concepts of temporal and spatial correlation are introduced in Chapters 6 and 7 and finally this section closes with Chapter 8, a short review of the exponential family of distributions so the reader does not have to dig out his math stat book to review the basics.

The second section focuses on major models used for ecological data. GLM and GAM for count data, with nice examples contrasting Poisson, overdispersed Poisson, and negative binomial (Chapter 9, using `glm` in **stats**, `gam` in **mgcv** and `glm.nb` in **MASS**); GLM and GAM for absence-presence and proportional data (Chapter 10, also using `glm` and `gam`); zero-truncated and zero-inflated models for count data, with a nice discussion on how to decide which one of the many model types to use when excessive/lacking zeroes are an issue (Chapter 11, using `vglm` in **VGAM** and `zeroinfl`, `hurdle` in **pscl**); generalized estimation equations (Chapter 12 using `geepack` in **geeglm**); and GLMM and GAMM (Chapter 13, using `glmmPQL` in **MASS**, `lmer` in **lme4**, and `glmmML` in **glmmML**). As in the first section of the book, each chapter provides illustrative examples as the main vehicle for creating context before presenting background theory. Previously presented concepts and how they fit in to the new techniques are regularly cited, allowing the reader to see how the complexity of earlier models are built and expanded into the newly presented models.

The final section contains case studies providing details of how models previously presented are applied. Ten chapters, each focusing on a different data set and presenting a thorough description of the analyses conducted illustrate real applications along with the sometimes ugly decisions when dealing with less than perfect data. These case studies can be a great starting point for researchers to become familiar with various analyses before tackling their own problems. At the end of each is an interesting "What to Report in a Paper" section, which highlights some of the details and decisions that are important to relay when presenting one's work. The last case study even introduces, albeit briefly, Bayesian hierarchical modeling, potentially leading readers to consider another framework for analyses.

In short, this text is good for researchers with at least a little familiarity with the basic concepts of modeling and who want some solid step-by-step guidance with examples on how common ecological modeling tasks are accomplished using R.

Reviewer:

Aaron Christ
Alaska Department of Fish and Game
Division of Wildlife Conservation

333 Raspberry Rd
Anchorage, Alaska, 99518, United States of America
E-mail: aaron.christ@alaska.gov