Preface

'*Which test should I apply?*' During the many years of working with ecologists, biologists and other environmental scientists, this is probably the question that the authors of this book hear the most often. The answer is always the same and along the lines of 'What are your underlying questions?', 'What do you want to show?'. The answers to these questions provide the starting point for a detailed discussion on the ecological background and purpose of the study. This then gives the basis for deciding on the most appropriate analytical approach. Therefore, a better starting point for an ecologist is to avoid the phrase 'test' and think in terms of 'analysis'. A test refers to something simple and unified that gives a clear answer in the form of a p-value: something rarely appropriate for ecological data. In practice, one has to apply a data exploration, check assumptions, validate the models, perhaps apply a series of methods, and most importantly, interpret the results in terms of the underlying ecology and the ecological questions being investigated.

Ecology is a quantitative science trying to answer difficult questions about the complex world we live in. Most ecologists are aware of these complexities, but few are fully equipped with the statistical sophistication and understanding to deal with them.

Even data gathered from apparently simple ecological research can require a level of statistical awareness rarely taught at the undergraduate or even the postgraduate level. There is little enough time to teach the essentials of ecology, let alone finding the time to teach 'advanced' statistics. Hopefully, for post graduates moving into academia there will be some advanced statistical support available, but many ecologist end up working in government, a voluntary organisation or consultancy where statistical support is minimal.

Although, the authors of this book believe that a quantitative approach is at the core of being a good ecologist, they also appreciate how challenging many ecologists find statistics. This book is therefore aimed at three levels of reader.

At one level it is aimed at making ecologists aware of how important it is to design scientifically robust ecological experiments or monitoring programmes, and the importance of selecting the best analytical technique. For these readers we hope the book, in particular the case studies, will encourage them to develop their personal statistical skills, or convince them they need statistical support.

On the next level it is aimed at the statistically literate ecologist, who may not be fully aware of the techniques we discuss, or when to use them. Hopefully, we have explained things well enough for these readers to feel confident enough to use some of the techniques we describe. Often these techniques are presented in a

fairly impenetrable manner, even for the statistically aware ecologist, and we have tried to make our presentation as 'ecologist friendly' as possible.

Finally, we hope the book will be of value to statisticians, whether they have a background in ecology or statistics. Ecological data can be particularly challenging to analyse, and we hope that providing an insight into our approach, together with the detailed case studies, will be of value to statistician readers, regardless of their background and expertise.

Overall, however, we hope this book will contribute in some small way to improving the collection and analysis of ecological data and improve the quality of environmental decision making.

After reading this book, you should be able to apply the following process: 'These are my questions', 'This is my statistical approach', 'Here is proof that I did it all correct (model validation)', 'This is what the data show' and 'Here is the ecological interpretation'.

Acknowledgement

A large part of the material in this book has been used by the first two authors as course material for MSc and PhD students, post-docs, scientists, both as academic and non-academic courses. We are greatly indepted to all 1200–1500 course participants who helped improve the material between 2000 and 2005 by asking questions and commenting on the material.

We would also like to thank a series of persons who commented on parts of this book: Ian Jolliffe, Anatoly Saveliev, Barry O'Neill, Neil Campbell, Graham Pierce, Ian Tuck, Alex Douglas, Pam Sikkink, Toby Marthews, Adrian Bowman, and six anonymous reviewers and the copy-editor. Their criticisms, comments, help and suggestions have greatly improved this book.

The first author would like to thank Rob Fryer and FRS Marine Laboratory for providing the flexibility to start the foundation of this book.

We would also like to thank the people and organizations who donated data for the theory chapters. The acknowledgement for the unpublished squid data (donated by Graham Pierce, University of Aberdeen) used in Chapters 4 and 7 is as follows. Data collection was financed by the European Commission under the following projects: FAR MA 1.146, AIR1-CT92-0573, FAIR CT 1520, Study Project 96/081, Study project 97/107, Study Project 99/063, and Q5CA-2002-00962. We would like to thank Roy Mendelssohn (NOAA/NMFS) for giving us a copy of the data used in Mendelssohn and Schwing (2002). The raw data are summaries calculated from the COADS dataset. The COADS references are Slutz et al. (1985) and Woodruff et al. (1987). We thank Jaap van der Meer (NIOZ) for allowing us to use the Balgzand data, The Bahamas National Trust and Greenforce Andros Island Marine Study for providing the Bahamas fisheries dataset, Chris Elphick (University of Connecticut) for the sparrow data, and Hrafnkell Eiríksson (Marine Research Institute, Reykjavik) for the Icelandic Nephrops time series. The public domain SRTM data used in Chapter 19 were taken from the U.S. Geological Survey, EROS Data Center, Sioux Falls, SD. We thank Steve Hare (University of Washington) for allowing us to use the 100 biological and physical time series

from the North Pacific Ocean in Chapter 17. A small part of Chapter 13 is based on Zuur (1999, unpublished PhD thesis), which was partly financed by the EU project DYNAMO (FAIR-CT95-0710).

A big 'thank you' is also due to the large number of folks who wrote R (www.r-project.org) and its many libraries. We made a lot of use of the lattice, regression, GLM, GAM (mgcv) and mixed modelling libraries (nlme). This thank you is probably also on behalf of the readers of this book as everything we did can be done in R.

Finally, we would like to thank John Kimmel for giving us the opportunity to write this book, and his support during the entire process. On to the next book.

Alain F. Zuur
Elena N. Ieno
Graham M. Smith

February 2007

http://www.springer.com/978-0-387-45967-7

Analyzing Ecological Data
Zuur, A.; Ieno, E.N.; Smith, G.M.
2007, XXVI, 672 p., Hardcover
ISBN: 978-0-387-45967-7