Introduction to the Special Volume on "Ecology and Ecological Modelling in R"

Thomas Kneib
Ludwig-Maximilians-Universität München

Thomas Petzoldt
Technische Universität Dresden

Abstract

The third special volume in the "Foometrics in R" series of the Journal of Statistical Software collects a number of contributions describing statistical methodology and corresponding implementations related to ecology and ecological modelling. The scope of the papers ranges from theoretical ecology and ecological modelling to statistical methodology relevant for data analyses in ecological applications.

Keywords: statistical computation, ecological modelling, ecology, R.

1. Introduction

As one of the special volumes of the Journal of Statistical Software compiled under the heading "Foometrics in R", this special volume collects a number of contributions on statistical methodology and corresponding implementations related to ecology and ecological modelling. This is already the third special volume in this series, following its successful predecessors on "Spectroscopy and Chemometrics in R" and "Psychometrics in R", but the first contribution that files under Foometrics without actually having an explicit "metric" in the title (although the term ecolometrics is used from time to time to describe the field that is the scope of this special volume). Anyway, it fits well in the range of foometrics special volumes, intended to summarise data analysis techniques and mathematical models for different types of sciences "foo" (cf. Eastlake 3rd, Manros, and Raymond 2001, for an etymology of "foo") available in R (R Development Core Team 2007). As reflected by the background of the two co-editors, such a description requires a combination of knowledge of statistical methodology, R, and ecology, and one intention of this special volume is therefore also to promote the communication between statisticians and ecologists.

The scope of this special volume has been chosen to be as broad as possible, ranging from

theoretical ecology and ecological modelling to statistical computation and data analysis in ecology. Both aspects are well-represented in the selection of published papers. In the concluding section, some references on additional work that may be of interest to our readership are given.

The linkage between the contributions to this special volume from the diverse fields of ecology is their focus on R as a software framework that enables the rapid yet reliable development and distribution of software for ecologists. As a positive side-effect, R can also serve as an interface that bridges the gap between different ecological schools. Two very nice examples in this special volume are the contributions by Dray and Dufour (2007) and Calenge (2007) which make methods developed in the French school of *analyse des données* (a special branch of multivariate statistics) available to the non-French speaking ecological community. In particular, the **ade4** package also forms a connection to the previous special volume of JSS on Psychometrics, where it is mentioned as one possibility to perform correspondence analyses (de Leeuw and Mair 2007).

2. Outline of the contributions

The volume starts with an application of R to analyse subfossil remains of organisms preserved in aquatic sediments. The package **analogue** (Simpson 2007) contains functions to perform modern analogue technique (MAT) transfer functions, which can be used in palaeoecology to reconstruct past changes in the environment, such as lake-water pH or climate change. In addition to this, analogue matching (AM) is concerned with identifying modern sites that are floristically and faunistically similar to fossil samples and can therefore be used to define reference conditions in conservation biology. Taking a related direction, the second paper by Yuan (2007) describes the package **bio.infer** that implements functionality for inferring environmental conditions from assemblage composition using maximum likelihood predictions.

A group of papers arranged around the French school of *analyse des données* follows, that embraces various features of multivariate statistics, such as principal components or correspondence analysis. Although the corresponding statistical theory has been available for quite a while, it has been largely overlooked by the (non-French) ecological community due to the lack of non-French references. Dray and Dufour (2007) provide both a review of the methodology, making it available to a larger community, and an introduction to the **ade4** package. On top of **ade4**, a graphical user-interface is available in package **ade4TkGUI** making it even easier to access **ade4** features for first-time users and R novices (Thioulouse and Dray 2007). Finally, as a third **ade4** related contribution, Calenge (2007) present the package **adehabitat** as a further add-on. Besides demonstrating the exploration of habitat selection data using **ade4**-functionality, the paper discusses further **adehabitat** features such as home range estimation or visualisation tools.

The contribution by Goslee and Urban (2007) introduces the **ecodist** package for investigating ecological data based on dissimilarities. Various variants of the Mantel test for inspecting the relationship between dissimilarity measures are introduced, in particular partial tests and tests that allow for nonlinear spatial structures.

Jachner, van den Boogart, and Petzoldt (2007) focus on a theme of general relevance when comparing results from ecological simulation models with real data: The assessment of similarity between time series. The paper discusses several deviance measures that allow to ignore

location, scale or distance of the measured values as well as exact time and speed or inequality constraints and time continuity. Jachner *et al.* (2007) provide a systematic overview over dissimilarity measures and the corresponding implementations are made available in the package **qualV**.

Moving from statistically-oriented papers to papers with focus on ecological modelling, Petzoldt and Rinke (2007) describe an object-oriented approach that allows one to implement and simulate ecological models in a fairly general framework. The relevant infrastructure is included in the package **simecol** and demonstrated through a series of examples ranging from cellular automata over differential equation models up to individual-based simulations of age-structured populations.

The in-depth analysis of age-structured populations and population models is the focus of the following two contributions. In the first of them Jones (2007) presents his package **demogR** which includes tools for the construction and analysis of matrix population models. In addition to the standard analyses commonly used in evolutionary demography and conservation biology, **demogR** contains a variety of tools from classical demography. This includes the construction of period life tables, and the generation of model mortality and fertility schedules for human populations. The tools in **demogR** are generally applicable to age-structured populations but are particularly useful for analyzing problems in human ecology.

The related package **popbio** from Stubben and Milligan (2007) is intended to do both, the construction and analysis of projection matrix models from a demography study of marked individuals classified by age or stage. The package covers methods described in Matrix Population Models by Caswell (2001) and Quantitative Conservation Biology by Morris and Doak (2002). The package also includes methods to estimate vital rates and construct projection matrix models from census data typically collected in plant demography studies.

The volume is completed by the very first paper submitted to the special volume, the **untb** package of Hankin (2007a), that can be used to analyze ecosystem data in the context of Hubbell's striking "neutral theory of biodiversity" (Hubbell 2001). His theory states that observed population dynamics may be explained on the assumption of per capita equivalence amongst individuals. The package provides a comprehensive set of R routines for numerical simulation of neutral ecosystem dynamics, analysis of field data, and visualization of datasets. Moreover, it is noticeable that the **untb** package circumvents numerical problems with large numbers by using either the logarithmic representation employed by the **Brobdingnag** package provided by the same author (Hankin 2007b) or alternatively by employing the **PARI/GP** computer algebra system (Batut, Belabas, Bernardi, Cohen, and Olivier 2006).

3. Conclusion and outlook

The range of contributions compiled in this special volume gives an impression about the vast potential of the R framework and, even more important, the R community and its open source platform in bringing together high level data analysis and scientific knowledge across the borders of disciplines. We agree that "many biologists do not want to spend the rest of their lives as programmers"[1], but this is not necessarily required to benefit from the increasing availability of relevant R packages. Typically it is sufficient to be an attentive reader and to have a thorough understanding of the methodological background to use R packages on a

---

[1] https://listserv.umd.edu/cgi-bin/wa?A2=ind0505d&L=ecolog-l&D=0&P=5326

specific problem.

As already mentioned in the introduction, this special volume is also intended to promote interaction between ecologists and statisticians. This is particularly important since statistical methods are evolving continuously, and according to Hobbs, Twombly, and Schimel (2006) "reliance on traditional methods presents a major barrier to understanding fundamental ecological patterns and processes". For example, recent developments supplement traditional H_0-based tests by likelihood-based model-selection techniques (Hobbs and Hilborn 2006; Johnson and Omland 2004) and, as demonstrated in this volume, special procedures are tailored explicitly to address ecological questions.

Of course, this special volume is not intended to be the exclusive source for employing R in ecology and there are already numerous books and online resources available demonstrating both power and practical use of R for ecological applications. There are new upcoming books (e.g. Bolker 2007), new manuals and online-tutorials, e.g. about the **vegan** package for multivariate statistics (Oksanen 2007) or the use of R in plant pathology (Sparks, Esker, and Garrett 2007), and new software packages for microscopic image analysis built on top of R (Grosjean and Denis 2007), to mention only a few. Almost surely, the future will see several additional exciting new developments and possibilities for utilising R in ecology.

Much thought has already been spent on the question of why R is so successful (Narasimhan 2004; Tierney 2005; Mullen and van Stokkum 2007) and de Leeuw (2005) argued that the reasons are not solely technical. But why is the user base so large and why do we have more than 1140 contributed packages today? We are certain that one of the main non-technical reasons is the culture applied for the development of scientific open source software: an active community interested primarily in research results, publication credits and exchange of ideas. Everything is transparent and everything can be reproduced without unreasonable financial and technical demands. Documentation, reproducibility and hands-on experience are key factors to understand the increasingly complex ecological data analysis ourselves. Rigorous testing, user feedback and continuous improvements are required to establish credibility and, finally, transparency and reliability should help to convince environmental decision makers.

Acknowledgments

References

Batut C, Belabas K, Bernardi D, Cohen H, Olivier M (2006). "User's Guide to **PARI**/**GP**, Version 2.3.0." Technical Reference Manual. URL http://pari.math.u-bordeaux.fr/.

Bolker B (2007). *Ecological Models and Data in R*. Princeton University Press, Princeton. In press.

Calenge C (2007). "Exploring Habitat Selection by Wildlife with **adehabitat**." *Journal of Statistical Software*, **22**(5). URL http://www.jstatsoft.org/v22/i05/.

Caswell H (2001). *Matrix Population Models: Construction, Analysis, and Interpretation.* Sinauer, Sunderland, MA, 2nd edition.

de Leeuw J (2005). "On Abandoning XLISP-STAT." *Journal of Statistical Software*, **13**(7), 1–5. URL http://www.jstatsoft.org/v13/i07/.

de Leeuw J, Mair P (2007). "An Introduction to the Special Volume on 'Psychometrics in R'." *Journal of Statistical Software*, **20**(1). URL http://www.jstatsoft.org/v20/i01/.

Dray S, Dufour AB (2007). "The **ade4** Package: Implementing the Duality Diagram for Ecologists." *Journal of Statistical Software*, **22**(4). URL http://www.jstatsoft.org/v22/i04/.

Eastlake 3rd D, Manros C, Raymond E (2001). "Etymology of 'Foo'." RFC 3092 (Informational). URL http://www.ietf.org/rfc/rfc3092.txt.

Goslee SC, Urban DL (2007). "The **ecodist** Package for Dissimilarity-Based Analysis of Ecological Data." *Journal of Statistical Software*, **22**(7). URL http://www.jstatsoft.org/v22/i07/.

Grosjean P, Denis K (2007). **Zoo**/'pkgPhytoImage Version 1.2-0. Computer-Assisted Plankton Image Analysis.* User's Manual, URL http://www.sciviews.org/zooimage/.

Hankin RKS (2007a). "Introducing **untb**, an R Package For Simulating Ecological Drift Under the Unified Neutral Theory of Biodiversity." *Journal of Statistical Software*, **22**(12). URL http://www.jstatsoft.org/v22/i12/.

Hankin RKS (2007b). "Very Large Numbers in R: Introducing Package **Brobdingnag**." *R News*, **7**(2), 62–64.

Hobbs NT, Hilborn R (2006). "Alternatives to Statistical Hypothesis Testing in Ecology: A Guide to Self Teaching." *Ecological Applications*, **16**(1), 5–19.

Hobbs NT, Twombly S, Schimel DS (2006). "Deepening Ecological Insights Using Contemporary Statistics." *Ecological Applications*, **16**(1), 3–4.

Hubbell SP (2001). *The Unified Neutral Theory of Biodiversity and Biogeography.* Princeton University Press.

Jachner S, van den Boogart KG, Petzoldt T (2007). "Statistical Methods for the Qualitative Assessment of Dynamic Models with Time Delay (R Package **qualV**)." *Journal of Statistical Software*, **22**(8). URL http://www.jstatsoft.org/v22/i08/.

Johnson Jerald G, Omland KS (2004). "Model Selection in Ecology and Evolution." *Trends in Ecology and Evolution*, **19**(2), 101–108.

Jones JH (2007). "**demogR**: A Package for Evolutionary Demographic Analysis in R." *Journal of Statistical Software*, **22**(10). URL http://www.jstatsoft.org/v22/i10/.

Morris WF, Doak DF (2002). *Quantitative Conservation Biology.* Sinauer, Sunderland, MA.

Mullen KM, van Stokkum IHM (2007). "An Introduction to the 'Special Volume Spectroscopy and Chemometrics in R'." *Journal of Statistical Software*, **18**(1), 1–5. URL http://www.jstatsoft.org/v18/i01/.

Narasimhan B (2004). "Lisp-Stat to Java to R." *Journal of Statistical Software*, **13**(4), 1–20. URL http://www.jstatsoft.org/v13/i04/.

Oksanen J (2007). *Multivariate Analysis of Ecological Communities in R:* **vegan** *Tutorial.* URL http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf.

Petzoldt T, Rinke K (2007). "**simecol**: An Object-Oriented Framework for Ecological Modeling in R." *Journal of Statistical Software*, **22**(9). URL http://www.jstatsoft.org/v22/i09/.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Simpson GL (2007). "Analogue Methods in Palaeoecology: Using the **analogue** Package." *Journal of Statistical Software*, **22**(2). URL http://www.jstatsoft.org/v22/i02/.

Sparks A, Esker P, Garrett K (2007). "Web-based Teaching Modules for Plant Pathology Applications in the R Programming Environment." URL http://www.k-state.edu/pdecology/RModules.html.

Stubben CJ, Milligan BG (2007). "Estimating and Analyzing Demographic Models Using the **popbio** Package in R." *Journal of Statistical Software*, **22**(11). URL http://www.jstatsoft.org/v22/i11/.

Thioulouse J, Dray S (2007). "Interactive Multivariate Data Analysis in R with the **ade4** and **ade4TkGUI** Packages." *Journal of Statistical Software*, **22**(6). URL http://www.jstatsoft.org/v22/i06/.

Tierney L (2005). "Some Notes on the Past and Future of Lisp-Stat." *Journal of Statistical Software*, **13**(9), 1–15. URL http://www.jstatsoft.org/v13/i09/.

Yuan LL (2007). "Maximum Likelihood Method for Predicting Environmental Conditions from Assemblage Composition: The R Package **bio.infer**." *Journal of Statistical Software*, **22**(3). URL http://www.jstatsoft.org/v22/i03/.

Affiliation:

Thomas Kneib
Institut für Statistik
Ludwig-Maxmimilans-Universität München
D-80539 München, Germany
E-mail: thomas.kneib@stat.uni-muenchen.de
URL: http://www.stat.uni-muenchen.de/~kneib/

Thomas Petzoldt
Institut für Hydrobiologie
Technische Universität Dresden
D-01062 Dresden, Germany
E-mail: thomas.petzoldt@tu-dresden.de
URL: http://tu-dresden.de/Members/thomas.petzoldt/