PUBLISHED AS:

Warton, D. I., & McGeoch, M. A. (2017). Technical advances at the interface between ecology and statistics: Improving the biodiversity knowledge generation workflow. *Methods in Ecology and Evolution, 8*(4), 396-397. doi:10.1111/2041-210X.12762

SPECIAL FEATURE

Technical advances at the interface between ecology and statistics – Improving the biodiversity knowledge generation workflow

David I. Warton and Melodie A. McGeoch

School of Mathematics and Statistics and Evolution & Ecology Research Centre, UNSW Sydney, NSW 2052, Australia

School of Biological Sciences, Monash University, Clayton 3800, Australia

Summary

1. Cross-disciplinary research between ecologists and statisticians has considerable potential for significant new advances, capitalising on recent advances in technology for collecting and analysing data.
2. We introduce a Special Feature of five papers showcasing interdisciplinary collaboration, centred around problems estimating biodiversity and how it changes over space and time.
3. Each paper arose around the Eco-Stats '15 conference at UNSW Sydney, held in December 2015, and more papers inspired by this conference are in preparation. This highlights the potential of interdisciplinary research, and of meetings designed to bring together researchers across disciplines, as a vehicle for scientific advances.

Key-words: biodiversity modelling, DNA barcoding, biodiversity turnover, detection probability, imperfect detection, joint species distribution model.

Some of the most exciting scientific breakthroughs arise from interdisciplinary research, through the translation of ideas and tools from one discipline to another. Research at the interface between ecology and statistics is a case in point, with many influential papers published in this journal arising through collaborations between ecologists and statisticians. The April 2015 Special Issue of this journal reported outcomes from a conference centred around this theme (Warton 2015). A particular opportunity comes from recent advances in technology, finding new ecological insights by exploiting the availability of new methods of collecting data, and new tools for analysing it.

This Special Issue reports on ideas generated by interdisciplinary teams around the Eco-Stats '15 conference held at UNSW Sydney in December 2015. Over 100 researchers attended, a mixture of statisticians, ecologists and researchers at the interface between the two. Six session themes were chosen, falling broadly under the umbrella of biodiversity modelling, and each involved two speakers known largely to different disciplines and with different skillsets, typically, a field or experimental ecologist was teamed up with a quantitative researcher. On the last afternoon of the meeting, breakout groups (organised by session theme) discussed potential opportunities for cross-disciplinary collaboration. Many of these formed a platform for subsequent research, and some of these are reported in this Special Issue.

The papers in the Special Issue relate to two key challenges in biodiversity science – accurately estimating diversity, as well as how it changes in space and time (McGill et al. 2015). Specifically, new techniques are presented which better estimate uncertainty associated with modelling occupancy and with assigning unique identities to taxonomic units, which simultaneously estimate how communities of species and their interactions respond to environmental variation, and which better capture the contribution of common species to biodiversity turnover in response to such variation.

Decision making for environmental management is seldom supported by unequivocal data (Cook et al. 2014). However, it is not the uncertainty *per se* that hinders evidence-based action, but rather not knowing the degree of uncertainty involved. DNA barcoding presents enormous opportunity for improving estimates of wholesale diversity. Not being able to accurately assign taxonomies to DNA sequences has hindered the full potential of high-throughput sequencing for DNA barcoding from being realised in biodiversity studies. Somervuo et al. (2017) test the performance of a probabilistic approach to assigning certainty of taxonomic placement (PROTAX) using case studies that represent a broad range of taxa and sources of potential error.

In addition to taxonomic uncertainty that may arise from biodiversity surveys, the relative abundance and species represented in a survey are well known to be affected by multiple factors ranging from the sampling technique used, to the behaviour and life history features of the species involved. One approach to accounting for imperfect species detection within and across sites is the use of occupancy-detection models (Guillera-Arroita, *in press*). Until now there have been limited techniques to diagnose occupancy detection models and where in the model any problems might be (*e.g.* in modelling detection probability or occupancy). Warton et al (2017) provide definitions of both detection residuals and occupancy residuals that can be used to tease the two apart, using conventional residual diagnostic plots.

Biodiversity data are typically available from surveys of different types, for example, it is not uncommon for data to be available in a given region as both presence-background data from opportunistic surveys, and presence-absence data from systematic field surveys. Koshkina et al (2017) propose a method of combining data from these different types, accounting for imperfect detection.

To quantify changes in biodiversity from one place or time to the next, pairwise measures of dissimilarity are widely used. Hui & McGeoch 2012 proposed zeta diversity, a measure that uses data from more than two sites, to offer broader insights. However many biodiversity analysis techniques still use pairwise metrics, such as generalised dissimilarity modelling (GDM, Ferrier et al 2007), to examine how quickly biodiversity changes along environmental gradients. Latombe et al (2017) extend GDM to handle zeta-diversity, and study what additional insights are provided by this extension.

Our inability to predict the consequences of species interactions for biodiversity responses to climate and other forms of environmental change has long been recognised as a major source of uncertainty in ecological forecasts (Kareiva et al. 1993). Joint species distribution modelling is a tool that has enabled progress in this area (Clark et al 2014, Pollock et al. 2014). To date available models have made the unrealistic assumption that species interactions are invariant across environmental gradients. Tikhonov et al (2017) provide a solution for estimating environment-dependent species associations that increase the information value that can be obtained from classic species by site biodiversity matrices.

A number of other projects are in progress that originated at Eco-Stats '15, but weren't completed in time to make this issue – including a project led by Wesley Brooks and Jay ver Hoef on methods for modelling presence-background data along stream networks, Paul Sunnucks and Rachel Fewster are modelling a rare wombat species using capture-recapture combined with modern genetic tools, Michael Kasumovic and Daniel Falster are applying game theoretical ideas to better understand male combat, and a multidisciplinary team (including Nigel Andrew, David Nipperess, Shinichi Nakagawa, Mariona Roige and Ezequial Marzinelli – with Melodie McGeoch, Guillaume Latombe and Cang Hui) are working on applying zeta diversity across scales and levels of organisation to better understand compositional turnover in biological systems. Interdisciplinary collaboration, and the opportunities offered by recent technological advances, have potential to lead to interesting and sometimes surprising findings, and will continue to be fertile ground for scientists in the forseeable future.

Acknowledgements

Thanks to the Statistical Society of Australia (Environmental Statistics Section) for sponsoring the Eco-Stats '15 conference, and Australian Research Council Discovery Project DP150103017 for some travel support.

Data accessibility

This manuscript does not use any data.

References

Clark, J. S., Gelfand, A. E., Woodall, C. W. and Zhu, K. (2014) More than the sum of the parts: forest climate response from joint species distribution models. *Ecological Applications*, **24** 990–999.

Cook, C.N., Inayatullah, S., Burgman, M.A., Sutherland, W.J. & Wintle, B.A. (2014) Strategic foresight: how planning for the unpredictable can improve environmental decision-making. *Trends in Ecology & Evolution,* **29,** 531-541.

Guillera-Arroita, G. (in press) Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography*, 40.

Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, **13**, 252–264.

Hui, C. & McGeoch, M.A. (2014) Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns. *The American Naturalist* **184**, 684-694.

Koshkina, V., Wang, Y., Gordon, A., Dorazio, R. & White, M. (2017) Integrated Species Distribution Models: Combining presence-background data and site-occupancy data with imperfect detection. *Methods in Ecology and Evolution* **8**, *ppp-ppp.*

Latombe, G., Hui, C. & McGeoch, M.A. (2017) Multi-site generalised dissimilarity modelling: using zeta-diversity to differentiate drivers of turnover in rare and widespread species. *Methods in Ecology and Evolution* **8**, *ppp-ppp.*

McGill, B.J., Dornelas, M., Gotelli, N.J. & Magurran, A.E. (2015) Fifteen forms of biodiversity trend in the Anthropocene. *Trends in Ecology & Evolution,* **30,** 104-113.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. (2014) Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution,* **5,** 397-406.

Somervuo, P., Yu, D.W., Xu, C.C.Y., Ji, Y., Hultman, J., Wirta, H. & Ovaskainen, O. (2017) Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. *Methods in Ecology and Evolution* **8**, *ppp-ppp.*

Tikhonov, G., Abrego, N., Dunson, D. & Ovaskainen, O. (2017) Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context. *Methods in Ecology and Evolution* **8**, *ppp-ppp.*

Warton, D.I. (2015) New opportunities at the interface between ecology and statistics. *Methods in Ecology and Evolution*, **6**, 363-365.

Warton, D.I., Stoklosa, J., Guillera-Arroita, G., MacKenzie, D.I. & Welsh, A.H. (2017) Graphical diagnostics for occupancy models with imperfect detection. *Methods in Ecology and Evolution* **8**, *ppp-ppp.*

Yu, D.W., Ji, Y., Emerson, BC., Wang, X., Ye, C., Yang, C. and Ding, Z. (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, **3**, 613–623.