way we interact with ourselves and the planet.

Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA

*Correspondence: jnekola@unm.edu (J.C. Nekola).

References

1. Burger, J.R. et al. (2012) The macroecology of sustainability. PLoS Biol. 10, e1001345
2. Brown, J.H. et al. (2011) Energetic limits to economic growth. Bioscience 61, 19–26

Letter

Incorporating Imperfect Detection into Joint Models of Communities: A response to Warton et al.

Steven R. Beissinger,[1,*]
Kelly J. Iknayan,[1]
Gurutzeta Guillera-Arroita,[2]
Elise F. Zipkin,[3]
Robert M. Dorazio,[4]
J. Andrew Royle,[5] and
Marc Kéry[6]

Warton et al. [1] advance community ecology by describing a statistical framework that can jointly model abundances (or distributions) across many taxa to quantify how community properties respond to environmental variables. This framework specifies the effects of both measured and unmeasured (latent) variables on the abundance (or occurrence) of each species. Latent variables are random effects that capture the effects of both missing

environmental predictors and correlations in parameter values among different species. As presented in Warton et al., however, the joint modeling framework fails to account for the common problem of detection or measurement errors that always accompany field sampling of abundance or occupancy, and are well known to obscure species- and community-level inferences.

Detectability often differs among individuals within a species and among species within a community, and typically varies among observers, sampling sites, and survey methods [2]. These differences in detectability create biases in estimates of abundance, occupancy, and dynamics derived from raw counts of multispecies surveys, which are the basis for the joint modeling framework and the examples given in [1]. Undetected individuals result in underestimation of population size when species are common and in false absences when species are rare. As a result, inferences concerning the explanatory power of ecological covariates [3] or community patterns across gradients [4] can be seriously affected, with important effects being masked or spurious ones

detected when variation in detectability is not taken into account. The problem is ubiquitous across taxa, including both plants and animals [2]. Thus, imperfect detection is the rule rather than the exception. Fortunately, a class of models has been developed that specifically addresses this problem in the form of hierarchical, detection-based multispecies models, which treat species occurrence or abundance as an imperfectly observed (latent) state. For reviews with examples and the code to run these models, see Iknayan et al. [2], Royle and Dorazio (chapter 12 in [5]), and Kéry and Royle (chapter 11 in [6]).

The simple but powerful idea to model a community as a collection of single-species models linked by a mixture distribution was developed more than a decade ago by Dorazio and Royle [7] and Gelfand et al. [8]. These models typically include (i) an observation process that models detection, (ii) an ecological process related to abundance (or occupancy) and any covariates of interest, and (iii) a super-population process that models species as random effects from community-level distributions, using hyperparameters to model detection

---

Box 1. What is a Detection-based Joint Model for Abundance?

We extend the joint hierarchical model of species abundance in Box 1 of Warton et al. [1] by adding a layer to accommodate imperfect detection using measurements derived from repeated surveys over a period when the population is closed.

Let y_{ijk} be the number of individuals of species j detected for replicate k at site i. The only modification required is to treat true abundance N_{ij} as a latent state that is only partially observable and related to the observed counts (y_{ijk}):

$$y_{ijk} \sim \text{binomial } (N_{ij}, p_{ijk}) \qquad \text{\# Abundance measured with detection error}$$

Each individual of N_{ij} has a probability of being recorded or detected (p_{ijk}) in the count y_{ijk} [12]. The remainder is identical to models in Box 1 of [1], except we relabel their y_{ij} as N_{ij} to clarify that abundance is imperfectly observed and to distinguish true (N_{ij}) from measured (y_{ij}) abundance. For the latent variable model from [1] we have:

$$N_{ij} \mid \mathbf{z}_i \sim F(m_{ij}, \phi_j) \qquad \text{\# Model for latent abundance}$$
$$g(m_{ij}) = \alpha_i + \beta_{0j} + \mathbf{x}'_i\beta_j + \mathbf{z}'_i\lambda_j \qquad \text{\# Random effects, covariates and latent variables}$$
$$\mathbf{z}_i \sim N(\mathbf{0}, \mathbf{I}) \qquad \text{\# Model for latent variables}$$

To make the model identifiable, we need repeated abundance measurements (i.e., $k>1$) for at least some sites, and put some constraints on p_{ijk} (typically site-level covariates or random effects). Analogous models can be specified for occurrence instead of abundance [5–7].

---

and abundance (or occupancy) [2]. The hierarchical structure of these models allows the entire dataset collected during the sampling process to inform estimation of species-specific abundance or occupancy. This facilitates simultaneous modeling of data on all species, including rare ones, and even allows for inference on species not observed during sampling using data augmentation to estimate total species richness [7]. Multispecies models for occupancy have been broadly applied [2,6], while similar models for abundance are fewer but growing in number [9–11]. Both types are straightforward to implement in a Bayesian framework. Box 1 illustrates how to adapt the approach in Warton *et al.* [1] to create a detection-based joint model for abundance.

If a joint model approach using latent variables is the future of community ecology, as suggested by Warton *et al.* [1], then it will be critical to incorporate measurement error driven by detection processes directly into the modeling framework. Whether the objective is to examine biotic interactions, conduct ordinations, or predict species richness, models based on raw counts are likely to underperform at best and have the potential to result in seriously misleading conclusions at worst.

[1]Department of Environmental Science, Policy & Management, and Museum of Vertebrate Zoology, University of California, Berkeley, CA, USA
[2]School of BioSciences, University of Melbourne, Parkville, Australia
[3]Department of Integrative Biology and Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, East Lansing, MI, USA
[4]USGS, Wetland and Aquatic Research Center, Gainesville, FL, USA
[5]USGS Patuxent Wildlife Research Center, Laurel, MD, USA
[6]Swiss Ornithological Institute, Sempach, Switzerland

*Correspondence: beis@berkeley.edu (S.R. Beissinger).

References
1. Warton, D.I. *et al.* (2015) So many variables: joint modeling in community ecology. *Trends Ecol. Evol.* 30, 766–779
2. Iknayan, K.J. *et al.* (2014) Detecting diversity: emerging methods to estimate species diversity. *Trends Ecol. Evol.* 29, 97–106
3. Zipkin, E.F. *et al.* (2010) Multi-species occurrence models to evaluate the effects of conservation and management actions. *Biol. Conserv.* 143, 479–484
4. Tingley, M. and Beissinger, S.R. (2013) Cryptic loss of montane avian richness and high community turnover over 100 years. *Ecology* 94, 598–609
5. Royle, J.A. and Dorazio, R.M. (2008) *Hierarchical Modeling and Inference in Ecology,* Academic Press
6. Kéry, M. and Royle, J.A. (2016) In *Applied Hierarchical Modeling in Ecology* (1), Academic Press
7. Dorazio, R.M. and Royle, J.A. (2005) Estimating size and composition of biological communities by modeling the occurrence of species. *J. Am. Stat. Assoc.* 100, 389–398
8. Gelfand, A.E. *et al.* (2005) Modelling species diversity through species level hierarchical modelling. *J. R. Stat. Soc. Ser. C Appl. Stat.* 54, 1–20
9. Yamaura, Y. *et al.* (2012) Biodiversity of man-made open habitats in an underused country: a class of multispecies abundance models for count data. *Biodivers. Conserv.* 21, 1365–1380
10. Chandler, R.B. *et al.* (2013) A small-scale land-sparing approach to conserving biodiversity in tropical agricultural landscapes. *Conserv. Biol.* 27, 785–795
11. Dorazio, R.M. *et al.* (2015) Estimating the effects of habitat and biological interactions in an avian community. *PLoS ONE* 10, e0135987
12. Royle, J.A. (2004) N-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60, 108–115

Letter

Extending Joint Models in Community Ecology: A Response to Beissinger *et al.*

David I. Warton,[1,*]
F. Guillaume Blanchet,[2]
Robert O'Hara,[3]
Otso Ovaskainen,[4,5]
Sara Taskinen,[6]
Steven C. Walker,[2] and
Francis K.C. Hui[7]

The joint modelling of many variables in community ecology is a new and technically challenging area with many opportunities for future developments. The possibility of extending joint models to deal with imperfect detection has been highlighted by Beissinger *et al.* as an important problem worthy of further investigation [1]. We agree, and previously pointed to this potential extension as an outstanding question [2], alongside models that can estimate phylogenetic repulsion or attraction, nonlinearity in the response to latent variables, and spatial or temporal correlation, because further developments in all these directions are needed. An important advantage of the new model-based approach to multivariate analysis that we described [2] is that there is now the capacity to incorporate such important features into analyses. We make two key points here in response to Beissinger *et al.* First, we want to temper the enthusiasm expressed by those authors [1] concerning how often imperfect detection methods are in fact needed. Second, we clarify the extent to which imperfect detection methods are currently available for our motivating context, where there are many response variables [2].

It is undoubtedly true that there are plenty of situations where there is a need to account for imperfect detection, across many study species [3]. However, to suggest that this is the rule rather than the exception in ecology is an overstatement, in much the same way as it would be overreaching to suggest that all joint models need to account for phylogenetic correlation, nonlinearity, spatial correlation, or any other possible extension. The precise statistical model specified depends on both the question being asked and the properties of the data being modelled (which is in part informed by study design and sampling methods). The only plausible scenario in which all ecologists should be advised to use the same modelling approach is if all ecologists were asking the same question and collecting data in the same way, which would make for a rather uninteresting discipline! A constructive way forward would be to carefully identify the sorts of situations in which imperfect detection is needed (a function of the questions asked as well as the data collected), and what checks are needed (in study design and analysis) to verify whether such methods are needed. For example, one situation where models