

Comment on article by Gelfand et al.

Jay M. Ver Hoef*

I have enjoyed reading the paper by Gelfand et al. (2005). My congratulations go to the authors, as they have given us an important advance in the science of modeling species diversity. First, I would like to emphasize the importance of this topic. I fully agree with the authors that species diversity has been a central concept in ecology for many years, yet the mechanisms that determine species diversity are still enigmatic. How then has this paper helped us?

One of the first problems in assessing species diversity is to know where a species occurs. While this may seem simple, it is actually very difficult. The authors have a very fine data set that was systematically sampled in a very interesting, diverse part of the world, where high species diversity is compacted into a relatively small space. One of the questions that I want to ask is, “Can the methods of Gelfand et al. (2005) be used more generally?” That is, can I use them in Alaska? Alaska is a rather large state, but if we consider plants, being far to the north, it is not really very diverse. We know of only about 1600 different plant species in Alaska. Rhode Island has more plant species (2600). The methods of Gelfand et al. (2005) are fairly complex, but in principle it seems that they could be adapted for hundreds (perhaps thousands) of different plant species as computational power increases. However, for a more general application, there are problems with species presence data that do not occur for Gelfand et al. (2005). Sampling has not occurred uniformly over my state, or any large geographic area that I know of. For example, I’m pretty sure that if we added a covariate such as distance to the nearest university, there would be a highly significant, negative regression coefficient when modeling species presence or diversity. The reason is clear. For years, botany professors have been sending out legions of graduate students and classes to collect plants, and they stay relatively close to home. Thus, not all zeros are created equal. This is known as ascertainment bias in the epidemiology literature. Gelfand et al. (2005) have done an outstanding job in distinguishing other factors that do create zeros, such as transformed landscapes. This is an important step, but it is information that is relatively easy to gather as compared to effort. Eventually, it will be important to solve the effect of effort (ascertainment bias).

Now, what about prior information? Gelfand et al. (2005) use a hierarchical model with vague priors. This makes sense, given the complexity of the model. Eliciting priors from most plant collectors that I know would be very difficult. It would be hard for them to make sense of priors on parameters in a model with the complications of the potential and transformed surfaces, hidden random effects, etc. Still, these same plant collectors have a wealth of prior knowledge; they have spent years crawling through the bushes. Early in my career I collected plants as my job, and I lived by the maps drawn in Hulten’s (1968) *Flora of Alaska*. It was a big deal to extend any of the species ranges drawn in his book. Plant collectors, such as Hulten, simply used their experience and

*National Marine Mammal Laboratory, Seattle, WA, jay.verhoef@noaa.gov

knowledge of terrain, climate and the known collection locations for a species to draw a line on a map that formed the species range. How can we tap such information? One interesting approach has been taken by Lele and Das (2000), who did not adopt a Bayesian formulation. Their thesis is that we should elicit predictions, not priors, on parameters. I think this is the right idea, and it would be interesting to incorporate a Bayesian approach that uses elicited predictions into the models developed by Gelfand et al. (2005), and indeed many others.

The model that Gelfand et al. (2005) propose is very interesting; it is a major improvement on many other approaches. As noted earlier, it is fairly complex compared to almost all other approaches so far. Nevertheless, there is one part of the model that is perhaps too simple. In eq. (5), Gelfand et al. (2005) give us

$$\log \left(\frac{p_i^{(k)}}{1 - p_i^{(k)}} \right) = \mathbf{w}'_i \boldsymbol{\beta}_k + \psi_k + \rho_i.$$

This model allows each species to have its own intercept ψ_k and covariate response vector $\boldsymbol{\beta}_k$, but all species have a common spatial pattern ρ_i in the “residuals” – or that part of the model not explained by the fixed effects. A model such as

$$\log \left(\frac{p_i^{(k)}}{1 - p_i^{(k)}} \right) = \mathbf{w}'_i \boldsymbol{\beta}_k + \psi_k + \rho_i^{(k)}$$

has too many parameters, allowing a separate spatial pattern (in the residuals) for each species. Undoubtedly, some species are responding to similar spatial effects. As the authors point out, this residual spatial random effect accounts for (at least in part) unmeasured spatially-patterned covariates. Some species will respond in a similar manner to a particular unmeasured covariate, while other species will respond in a similar way to another unmeasured covariate. A more flexible approach that does not have too many more parameters would be to allow for just a few spatial patterns, and then assume that each species’ residuals are a linear combination of those spatial random effects:

$$\log \left(\frac{p_i^{(k)}}{1 - p_i^{(k)}} \right) = \mathbf{w}'_i \boldsymbol{\beta}_k + \psi_k + \sum_{m=1}^M \eta_k^{(m)} \rho_i^{(m)},$$

where M is, say, 1 to 5. Bayes factors, DIC, or reversible jump MCMC methods could be used to choose M .

None of this detracts from the fundamental contributions that Gelfand et al. (2005) have given us. I hope that both statisticians and ecologists take notice, and that they use and build upon the models and ideas that these authors have developed. The synergy of collaboration among statisticians and ecologists is apparent from this article.

Bibliography

Gelfand, A., Silander, J., Shanshan, W., Latimer, A., Lewis, P., Rebelo, A., and Holder, M. (2006). “Explaining species distribution patterns through hierarchical modeling.”

Bayesian Analysis, (in press).

Hulten, E. (1968). *Flora of Alaska and Neighboring Territories*, 1008. Stanford University Press.

Lele, S. and Das, A. (2000). “Elicited data and incorporation of expert opinion for statistical inference in spatial studies.” *Mathematical Geology*, 465–487.

