

Rejoinder

Alan E. Gelfand, John A. Silander Jr., Shanshan Wu, Andrew Latimer,
Paul O. Lewis, Anthony G. Rebelo and Mark Holder

We very much appreciate the positive comments of both Jay Ver Hoef and Jennifer Hoeting. We were particularly delighted by their appreciation of our “synergy between statisticians and ecologists” and our demonstration “that interdisciplinary work can advance several disciplines simultaneously.” We briefly address the key points they have brought up.

Regarding Jennifer’s criticism of the “independence” assumption in expression (2) of the paper, we agree that this is surely not true though it may be roughly true. However, the “correlation” calculation in (2) is a bit more complicated than it initially appears in that the calculation is with respect to a uniform distribution over say, the locations in unit i . In fact, the randomness arises from the fact that the objects being integrated are random functions, rather than from randomness in the choice of locations. So, in fact, the integral in (2) is a stochastic integral and the assumption demands that the resulting random variable factor almost surely into the product of the random variables $1 - U_i$ and $p_i^{(k)}$. In the absence of this assumption, we would have little choice but to model $P(V_i^{(k)} = 1)$ analogously to (5) in the paper and, as a consequence, we would sacrifice the ability to consider both potential and transformed species distributions.

Jay raises the important issue of sampling bias in the data collection. We recognize that this occurs in most presence-absence species sampling and, indeed, it does for our data as well. The expert botanist on our team (Rebelo) was *sure* that for large areas within the CFR, no protea would be found and thus that there was no need to sample in these areas. We did not take this information as “data,” e.g., in the form of null sites; rather, we counted upon the second stage spatial modelling to provide smoothing for the random effects associated with unsampled grid cells. In this regard, Jennifer also comments upon sampling concerns, particularly with, say if one were working with museum data where there are no nulls. We note that there seems to be a component of the ecology community that is comfortable with developing species ranges in this setting (see, for instance, Engler et al., 2004). We are troubled by such inference for handling presence-only data (as most statisticians would be) and, in a manuscript in preparation, will attempt to illuminate more clearly the flaws in this work along with possible remedies.

Jay has noted the limitations in expression (5) of the additive form in species random effects and spatial random effects. We completely agree and have looked at for example, an additional multiplicative term of the form $\alpha\psi_k\rho_i$ as well as other possibilities. However, in forthcoming work (Latimer, et al., 2005) we have focused on species level spatial random effects, $\rho_i^{(k)}$. Assuming these to be independent across species enables us to fit our model one species at a time. This allows simple parallelization of the computation and is permitting us to make our way through range prediction for the more than 300 protea species in the CFR. This approach also fits in nicely with Jennifer’s suggestion

that we investigate model selection at the species level. Indeed, in the Latimer et al paper we carry out a modest version of this. We note that in other new work we consider the issue of spatial scale, working with 1 min \times 1 min grids as well as 4 min \times 4 min and 16 min \times 16 min grids. The differences in predicted species range are expected and noteworthy.

Finally, two issues were raised which we have not investigated. First, Jennifer suggests that an uncertainty weighted sum of the β 's would be more appropriate to work with. The challenge in this case would be to specify these weights so that the resulting sum is still a parametric function in order that we can examine its posterior. In our current formulation the $\beta_l^{(k)}$ are i.i.d. for each l making it unclear where the weights would come from? Second, Jay suggests that informative prior specification could be elicited in the form of predictions. This is an attractive idea and has some history in the Bayesian community. See, for instance the work of Ibrahim and colleagues, e.g., Ibrahim (1997).

In summary, we thank Jay and Jennifer for their thoughtful remarks and expect that they, along with us, will continue to collaboratively address challenging model problems that arise when studying complex ecological processes. There is much opportunity for statisticians to contribute in such enterprise.

Bibliography

- Engler, R., Guisan, A., and Rechsteiner, L. (2004). "An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data." *Journal of Applied Ecology*, 41:263–274.
- Ibrahim, J. (1997). "On properties of predictive priors in linear models." *The American Statistician*, 51:333–337.
- Latimer, A., Wu, S., Gelfand, A., and Silander, J. J. (2005). "Building statistical models to analyze species distributions." *Ecological Modelling*. (forthcoming).