Something occupancy models

Lauren C. Ponisio[1,2], Nicholas Michaud[1], Perry de Valpine[1]

1. Department of Environmental Science, Policy, and Management
   University of California, Berkeley
   130 Mulford Hall
   Berkeley, California, USA
   94720

2. Department of Entomology
   University of California, Riverside
   417 Entomology Bldg.
   Riverside, California, USA
   92521

Abstract

1. occupancy models are everywhere, but model fitting and assessment are extremely computationally intensive
2. because models are so computationally intensive, users often forgo model assessment (determining if a model provides an adequate fit to a particular dataset) because if involves simulating from and refitting the model many times.
3. Using the NIMBLE package for R, we develop combined computational approaches including user defined and automatic blocking of parameters for MCMC, filtering over latent states, and customized MCMC samplers for specific parameters to improve efficiency. We test these approaches using three representative occupancy models of varying levels of complexity including a single species model with spatial auto-correlation, a single species dynamic (multi-season) model, and a multi-species model. We also develop and inplement methods for calculating calibrated predistive posterior p-values to assess model fit within the open source modeling software, NIMBLE.
4. These computation approaches lead to an improvement in MCMC sampling efficiency over, particularly with models including random effects.
5. Ours results highlight the need for more customizable approaches to MCMC to fit and assess hierarchical models in order to ensure occupancy models are accessible to practitioners. By implementing MCMC procedures and model assessment techniques in open source software, we have made progress toward this aim.
6. *Implications:*

NIMBLE, Markov chain Monte Carlo, latent states, block sampling, dynamic occupancy, mutli species occupancy, spatial occupancy, JAGS

Introduction

Estimating the proportion of sites occupied by a species is common challenge for many sub-disciplines ecology and evolution including meta-population, endangered species and invasion biology. Greater acceptance of the baises introduced by imperfect detection has lead to the development and proliferation of occupancy models — models where the occurrence of a species at a site as a latent state layered underneath a detection process (e.g., MacKenzie *et al.*, 2006; Royle & Kéry, 2007). Now only a little over a decade after occupancy models were introduced to ecology, they are being used to model the occurrence of everything from bees (M'Gonigle *et al.*, 2015) to tigers (Hines *et al.*, 2010) in an endless variety of complexity.

Occupancy models are part of a larger class of models known as Hidden Markov Models. For discrete Hidden Markov Models like occupancy models where a species is either present or absent from a site, likelihood calculation involves summing over the distribution of latent states. Because estimating the effect of explanatory variables on site occupancy or shared variation of in occupancy across species is often of greatest interest to ecologists (e.g., Iknayan *et al.*, 2014), the Hidden Markov Models are embedded within a hierarchical model. In such cases, practitioners often rely on Markov chain Monte Carlo (MCMC) to perform a Bayesian analysis. Standard MCMC software will including the lantent state varaibles in MCMC sampling (e.g., **?**)

Such models are computationally intensive, and large models requiring hundreds or thousands of dimensions which require MCMC can be intractable.

In addition, fitting these models is such a challenge that users often forgo adding any additional computation to asses model fit. A common idea behind evaluating whether a model provides an adequate fit to a particular dataset is that if data is simulated from the model, the simulated data should resemble the observed data. This is the basis of posterior predictive P-values, which compare the distribution of summary statistics calculated from simulated datasets to the observed statistic. Posterior predictive P-values alone, however, often fail to reject poor-fitting models (Bayarri & Berger, 2000; Robins *et al.*, 2000; Hjort *et al.*, 2006). Methods for correcting posterior predictive P-values for better performance have been proposed (e.g., calibrated posterior predictive P-values, Hjort *et al.*, 2006), but refitting the model via MCMC iterativly.

To ensure occupancy models are accessible to practitioners, more efficient methods for fitting and assessing these models are necessary.

Materials & Methods

A common idea behind evaluating whether a model provides an adequate fit to a particular dataset is that if data is simulated from the model, the simulated data should resemble the observed data. This is the basis of posterior predictive P-values, which compare the distribution of summary statistics calculated from simulated datasets to the observed statistic. Posterior predictive P-values alone, however, often fail to reject poor-fitting models (Bayarri & Berger, 2000; Robins *et al.*, 2000; Hjort *et al.*, 2006). Methods for correcting posterior predictive P-values for better performance have been proposed, but previously were too computationally intensive to be feasible (e.g., calibrated posterior predictive P-values, Hjort *et al.*, 2006).

Results

Discussion

Acknowledgments

References

Bayarri, M. & Berger, J. (2000) P values for composite null models. *Journal of the American Statistical Association*, **95**, 1127–1142.

Hines, J., Nichols, J., Royle, J., MacKenzie, D., Gopalaswamy, A., Kumar, N. & Karanth, K. (2010) Tigers on trails: occupancy modeling for cluster sampling. *Ecological Applications*, **20**, 1456–1466.

Hjort, N.L., Dahl, F.A. & Hognadottir, G. (2006) Post-processing posterior predictive p values. *Journal of the American Statistical Association*, **101**, 1157–1174.

Iknayan, K.J., Tingley, M.W., Furnas, B.J. & Beissinger, S.R. (2014) Detecting diversity: emerging methods to estimate species diversity. *Trends in ecology & evolution*, **29**, 97–106.

MacKenzie, D., Nichols, J., Royle, J., Pollock, K., Bailey, L. & Hines, J. (2006) *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Elsevier, Burlington, Massachusetts, USA.

M'Gonigle, L., Ponisio, L., Cutler, K. & Kremen, C. (2015) Habitat restoration promotes pollinator persistence and colonization in intensively-managed agriculture. *Ecol Appl*, **25**, 1557–1565.

Robins, J., van der Vaart, A. & Ventura, V. (2000) Asymptotic distribution of p values in composite null models. *Journal of the American Statistical Association*, **95**, 1143–1156.

Royle, A. & Kéry, M. (2007) A bayesian state-space formulation of dynamic occupancy models. *Ecology*, **88**, 1813–1823.