Variable Selection in Ecological Habitat Modelling

M. Authier[a,c] and **C. Saraux**[b] and **C. Péron**[c]

[a]Observatoire PELAGIS UMS 3462
Université de La Rochelle
La Rochelle, France
authierm@gmail.com

[b] Laboratoire Ressources Halieutiques
IFREMER
Sète, France
Claire.Saraux@ifremer.fr

[c]Écologie Spatiale des Populations
Centre d'Écologie Fonctionnelle et Évolutive
Montpellier, France
Clara.Peron@cefe.cnrs.fr

Keywords: spatial ecology

Abstract: Habitat modelling in ecology is becoming increasingly relevant in population dynamics, biodiversity and conservation studies. A typical application of these models is to predict potential zones of specific conservation interest. The availability of many environmental covariates, measured *in situ* or with remote sensors, means that a large number of models can be investigated, which in turn often renders multi-model inference impratical. Model-averaging, however, is important to address model uncertainty and to prevent over-estimating effect size. Shrinkage regression deals with the identification and accurate estimation of effect size. In a Bayesian framework, we investigated the use of a recently proposed shrinkage prior, the Horseshoe prior (Carvalho *et al.* 2010) for variable selection in modelling the habitat of small pelagic fish in the Gulf of Lions, France. Using spatially-explicit generalized linear models on 5 different data sets from the `IFREMER` acoustic survey `PELMED 2011` and more than 10 environmental covariates, we compared the ability of a simple kriging model with no covariate, a saturated model with independent normal priors for regression coefficients and a saturated model with a Horseshoe prior for regression coefficients. Each model was calibrated on 80% of each data set, and the 20% remaining data was used for cross-validation. With respect to out-of-sample predictive ability the saturated model with a Horseshoe prior performed best, and the saturated model with independent normal priors worst. Overfitting and overestimated effect sizes were to blame: with an increasing number of covariates, extrapolation quickly became pervasive as we tried to predict from combinations of covariate values that were not in the calibration data. Shrinking regression coefficients toward zero with a Horseshoe prior avoided this problem and required only one model to be fitted to the data to obtain reasonable predictions.

References

Carvalho, C. and Polson, N. and Scott, J. (2010) The Horseshoe Estimator for Sparse Signals *Biometrika*, 97: 465-480.