Spatially Evaluating Resource Selection Functions using Conditional Randomization

Mary Smulders[1], Trisalyn Nelson[1], Scott Nielsen[2]

[1]Spatial Pattern Analysis and Research (SPAR) Laboratory,
Geography Department, University of Victoria
[2]Department of Renewable Resources, University of Alberta

Abstract

Resource selection function (RSF) models are increasingly being used to predict maps of the relative probability of wildlife occurrence. Current methods evaluating a RSF's accuracy are reported as singular values, representing the overall ability of the model to correctly determine species occurrence. These methods do not indicate the spatial location or variation in accuracy. The spatial dependence in error values may relate to ecological processes unaccounted for in the original RSF. The purpose of this research is to explore spatial methods of evaluating RSF models using a conditional randomization approach. A case study on adult female grizzly bears (*Ursus arctos*) is used to demonstrate our approach. Local test statistics computed from bear telemetry locations are used to identify RSF scores that are statistically lower than expected. Through examining landscape characteristics associated with significant areas, factors that may contribute to the unexpected RSF values can be identified.

Background and Relevance

A resource selection function (RSF) is a model that determines the probability of use of a particular resource unit (Manly et al. 2002). They are often used as a quantification tool in ecology since resource location determines the distribution and abundance of organisms (Boyce and McDonald, 1999). RSF models statistically correlate field observations to a set of habitat variables with the intention of reflecting essential elements of the organism's ecological requirements, such as climate, land-cover, and geology.

The predictive capacity of a RSF is commonly reported as a singular accuracy value (Fielding and Bell, 1997) that measures the overall ability of the model to predict species occurrence. Common metrics used are: sensitivity, specificity, the Kappa statistic, receiver operating characteristics (ROC) and area under the curve (AUC) (Fielding and Bell, 1997; Raes and ter Steege, 2007). By reporting only a single value, differences in the spatial distribution of model errors are ignored. The spatial distribution of model errors may indicate ecological processes that have not been accounted for in the RSF model (Fielding and Bell, 1997). By characterizing the spatial distribution of model errors, areas that under perform can be identified and supplementary data evaluated to determine source of errors and, when included in the model, the level of improved predictive success.

Methods and Data

Randomization tests assess statistical significance based on empirical distributions generated from the observed sample (Nelson and Boots, 2005). Randomizations are appropriate for ecological data since many traditional methods of statistical analysis are based on probability and distribution theories that may not be known (Fortin and Jacquez, 2000). A reference distribution through randomization, by contrast, is derived from the observed data by randomization and is used to determine the significance of a statistic calculated from the actual observed data (Fortin and Jacquez, 2000). The common randomization tests, based on the complete spatial randomization (CSR) of observations, are not appropriate for ecological data (Fortin and Jacquez, 2000) since they are inherently spatially autocorrelated (Legendre, 1993) and therefore data independence is violated. To avoid such dependence and to take into account spatial structure in data, some researchers have employed a restricted randomization process (e.g., Fortin et al. 1996).

Our goal is to show how conditional randomization methods can be used to quantify the spatial variability in the predictive success of a RSF. We outline methods and demonstrate them with a case study on patch-level selection (Johnson, 1980) for grizzly bears (*Ursus arctos*) in the Yellowhead Ecosystem in the Northeastern slopes of the Canadian Rockies. We use presence/available data to fit a RSF model. The RSF was developed as a 30 m by 30 m grid for each grizzly bear foraging season based on food availability and plant phenology (Munro et al., 2006; Nielsen et al., 2003).

The pattern of observed adult female grizzly bear radio telemetry data is quantified using spatially local statistics generated for 900 m by 900 m quadrats. The conditional randomization is based on the expected number of grizzly bear telemetry validation points that should fall within each RSF score bin (Johnson et al., 2006). Reference distributions are generated by applying the same local statistics to 99 permutations randomized conditionally on the RSF model (Edgington, 1995). Comparison of the observed data to reference distributions allows us to identify quadrats where the spatial pattern of habitat selection is unexpected given conditioned random use of the RSF model. A map showing unexpected locations is produced and compared with environmental data, such as elevation and land cover, in order to explore how the RSF model can be improved.

Results and Conclusions

We expect most of the bear telemetry locations will coincide with higher RSF values. Telemetry points that consistently coincide with lower RSF values indicate areas where the RSF is poorly predicting grizzly bear habitat selection. By quantifying the spatial distribution of RSF model errors we can identify regions where the RSF under performs. These areas can be linked to supplementary data to evaluate variables that may be contributing to errors in the RSF model.

References

Boyce, M. S., and L. L. McDonald. 1999. Relating populations to habitats using resource selection functions. TRENDS in Ecology and Evolution **14**:268-272.

Edgington, E. S. 1995. Randomization tests. Marcel Dekker Inc, New York, New York.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of predictions errors in conservation presence/absence models. Environmental Conservation **24**:38-49

Fortin, M.-J., P. Drapeau, and G. Jacquez. 1996. Quantification of the spatial co-occurrences of ecological boundaries. Oikos **71**:51-60.

Fortin, M. J., and G. Jacquez. 2000. Randomization tests and spatially autocorrelated data. Bulletin of the Ecological Society of America **81**:201-205.

Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. Ecology **61**:65-71.

Johnson, C. J., S. E. Nielsen, E. H. Merrill, T. L. McDonald, and M. S. Boyce. 2006. Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. Journal of Wildlife Management **70**:347-357.

Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? Ecology **74**:1659-1673.

Manly, B. F. J., L. L. McDonald, D. L. Thomas, T. L. McDonald, and W. P. Erickson. 2002. Resource selection by animals, 2 edition. Kluwer Academic Publishers, Dordrecht.

Munro, R. H. M., S. E. Nielsen, M. H. Price, B. Stenhouse, and M. S. Boyce. 2006. Seasonal and diel patterns of grizzly bear diet and activity in West-Central Alberta. Journal of Mammalogy **87**:1112-1121.

Nelson, T., and B. Boots. 2005. Identifying insect infestation hot spots: an approach using conditional spatial randomization. Journal of Geographic Systems **7**:291-311.

Nielsen, S. E., M. S. Boyce, G. B. Stenhouse, and R. H. M. Munro. 2003. Development and testing of phenologically driven grizzly bear habitat models. Ecoscience **10**:1-10.

Raes, N., ter Steege, H. 2007. A null-model for significance testing of presence-only species distribution models. Ecography. **30**: 727-736