Melissa Anthony, Saint Louis University, St. Louis, MO, USA

Daryl Trumbo, Washington State University, Pullman, WA, USA

Jason Knouft, Saint Louis University, St. Louis, MO, USA

Spatial Autocorrelation in Species Distribution Models: Simultaneous Incorporation of Multiple Scales of Influence Using a Bayesian Framework

Spatial autocorrelation (SAC), defined as the positive association between sample similarity and spatial proximity, is pervasive in ecological data. Several methods exist for the incorporation of SAC into statistical models, but not until recently have these methods been applied to species distribution data. When incorporated into species distribution models, SAC has been found to significantly alter model coefficients, subsequently changing the statistical inference of the model. Models without a SAC component run the risk of overestimating the relationship between environmental variables and the presence or abundance of a species, potentially resulting in poor predictive ability of the model. Numerous studies have demonstrated the improvement in species distribution model performance after the incorporation of SAC suggesting inclusion of this spatial component is essential for developing statistical models that accurately predict species distributions in novel environments.

All methods of incorporating SAC into species distribution models operate by scaling the probability that a species is present in one location by the probability that the species is present in nearby locations. Many of these methods only quantify SAC on a single scale resulting in models that lack the ability to take into account how SAC may affect the parameter estimates of environmental variables across different spatial scales. For example, climate varies on a larger geographic scale than variables such as soil or vegetation type, and the effect of SAC when determining the relative ability of these variables to explain a species' distribution should be scaled accordingly. This study attempts to address this scaling issue by using Bayesian additive models which incorporate SAC in the form of a spatial random effects parameter specified by a Gaussian conditional autoregressive model. The use of additive models allows the inclusion of a separate spatial random effects parameter for each scale of SAC present in the model as dictated by the scales of the environmental variables. For example, the spatial random effects coefficient when modeling the association between climate and species presence will be estimated using samples over a larger geographic distribution than more local measurements of the environment, such as vegetation type. By correctly partitioning the variation in observed species presence between the environment and the appropriate scale of influence, species distribution models are expected to exhibit higher performance and predictive power.

Another weakness of current methods of incorporating SAC into species distribution models is the frequentist approach to model parameterization, which does not appropriately quantify model uncertainty. Species distribution models that are built using data from a relatively small geographic range are often extrapolated to make predictions over larger areas. Accurate quantification of the uncertainly of model predictions is crucial in conservation planning and failure to do so can hinder the decision making process. This problem is addressed in the current study by using Bayesian methods of model parameterization that incorporate the uncertainly of model parameters in making predictions, unlike frequentist methods which make predictions based solely on a single value for each parameter.

To investigate how this new strategy performs, data are used from a survey of 103 ponds sampled for six amphibian species along with corresponding abiotic data that represents the environment at multiple scales. Model performance will be assessed and compared to other commonly used methods to quantify the improvement in predicting species distributions.