

STATISTICS SEMINAR

Narmadha M. Mohankumar

PhD Defense
Monday, April 5th, 2022
K-State Student Union, Big 12 Room, 12 Noon
(Central Time)



Data fusion and spatio-temporal approaches to model species distribution

Species distribution models (SDMs) are increasingly used in ecology, biogeography, and wildlife management to learn about the distribution of species across space and time. Determining the species-habitat relationships and the distributional pattern of a species is important to increase scientific knowledge, inform management decisions, and conserve biodiversity. I propose approaches to address some of the most pressing issues encountered in studies of species distributions and contribute towards improving predictions and inferences from SDMs.

First, I present a modeling framework to model occupancy data that accounts for both traditional and nontraditional spatial dependence as well as false absences. Occupancy data are used to estimate and map the true presence of a species, which may depend on biotic and abiotic factors as well as spatial autocorrelation. Traditionally, spatial autocorrelation is accounted for by using a correlated normally distributed site-level random effect, which might be incapable of modeling nontraditional spatial dependence such as discontinuities and abrupt transitions. Machine learning approaches have the potential to model nontraditional spatial dependence, but these approaches do not account for observer errors such as false absences. I combine the flexibility of Bayesian hierarchical modeling and machine learning approaches and present a modeling framework to account for both traditional and nontraditional spatial dependence and false absences. I illustrate the framework using six synthetic data sets containing traditional and nontraditional spatial dependence and then apply the approach to understand the spatial distribution of Thomson's gazelle (*Eudorcas thomsonii*) in Tanzania and sugar gliders (*Petaurus breviceps*) in Tasmania.

Second, I develop a model-based approach for data fusion of distance sampling (DS) and capture-recapture (CR) data. DS and CR are two widely collected data types to learn about species-habitat relationships and abundance; still, they are seldomly used in SDMs due to the lack of spatial coverage. However, data fusion of the sources of data can increase spatial coverage, which can reduce parameter uncertainty and make predictions more accurate, and therefore, can be used for species distribution modeling. My modeling approach accounts for two common missing data issues: 1) missing individuals that are missing not at random (MNAR) and 2) partially missing location information. Using a simulation experiment, I evaluated the performance of the modeling approach and compared it to existing approaches that use ad-hoc methods to account for missing data issues. I demonstrated my approach using data collected for Grasshopper Sparrows (*Ammodramus savannarum*) in north-eastern Kansas, USA.

Third, I extend my data fusion approach to a spatio-temporal modeling framework to investigate the influence of temporal support in spatio-temporal point process models to model species distribution. Temporal dynamics of ecological processes are inherently complex, and their influence on species-habitat relationships and abundance operate in multiple spatio-temporal scales. Spatio-temporal point process models are widely used to model species-habitat relationships and estimate abundance across multiple spatio-temporal scales; however,

the robustness of the models to changing temporal scales is rarely studied. Understanding the temporal dynamics of ecological processes across the entirety of spatio-temporal scales is key to learning about species' distribution. Therefore, investigating the influence of temporal support on the robustness of spatio-temporal point processes to model species distributions is needed. In my approach, I combine DS and CR data in a spatio-temporal point process modeling framework and investigate the robustness of the model to changing temporal scales. My fused data spatio-temporal model alleviates constraints in individual data sources such as lack of spatio-temporal coverage and enables the study of complex phenomena on multiple-scale species-habitat relationships and abundance. To demonstrate the modeling framework and illustrate the impact of the temporal support to model species-habitat relationships and abundance, I use four synthetic data examples and data on Grasshopper Sparrows (*Ammodramus savannarum*) in north-eastern Kansas, USA.