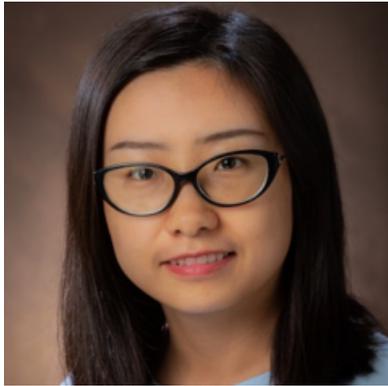# Special Colloquium

Spatial Heterogeneity Automatic Detection and Estimation

Dr. Xin Wang

Miami University, Oxford Ohio

Abstract: Studying the relationship between two surveys is important in improving the efficiency of the estimators in survey sampling. A linear regression model with common regression coefficients would be too simple to capture the regional differences and will lead to biases in the estimators. Oftentimes, the linear relationships vary across space, when some covariates have location-specific effects on the response. As a result, it is critical to detect the systematic variation in the model and identify which locations share common regression coefficients and which do not. Only a correct model structure can assure unbiased estimation and valid inferences. With this motivation, we propose a new procedure, called Spatial Heterogeneity Automatic Detection and Estimation (SHADE), for automatically and simultaneously subgrouping and estimating covariate effects for spatial regression models. The SHADE employs a class of spatially-weighted fusion type penalty on all pairs of observations, with location-specific weight constructed using spatial information, to cluster coefficients into subgroups. The SHADE is able to identify the true model structure with probability approaching one and estimate regression coefficients consistently. Furthermore, we develop an algorithm based on alternating direction method of multiplier algorithm (ADMM) to tackle the computation challenge. The numerical studies suggest that spatial information can enhance subgroup structure analysis in challenging situations when the spatial variation among regression coefficients is small or the number of repeated measures is small. Finally, the SHADE is applied to find the relationship between a natural resource survey and a land cover data layer to identify spatially interpretable groups.

Some extensions and potential future work are also discussed. The idea has been applied and extended to different models, such as linear mixed models in survey sampling, age and cohort models for obesity prevalence, functional data and recurrent event data.

Keywords: Areal data; Structure Selection; Penalization; Repeated measures; Spatial heterogeneity; Subgroup analysis

Tuesday, February 1, 12:05 – 12:50 pm