THE GRADUATE COLLEGE OF THE
UNIVERSITY OF OKLAHOMA HEALTH SCIENCES CENTER

ANNOUNCES THE FINAL EXAMINATION OF

Claude Larry Hill, Jr.

FOR THE DEFENSE OF THE DOCTOR OF PHILOSOPHY DEGREE
GRADUATE COLLEGE
DEPARTMENT OF BIOSTATISTICS AND EPIDEMIOLOGY

Wednesday, August 10, 2016, 3:00 p.m.
Room 204, College of Health Building, OUHSC

MODELING NONSTATIONARY ANISOTROPIC GEOSTATISTICAL DATA PROCESSES

COMMITTEE IN CHARGE: David M. Thompson, Ph.D., Chair, Julie A. Stoner, Ph.D., Michael P. Anderson, Ph.D., Sheryl L. Magzamen, Ph.D., Jennifer Koch, Ph.D.

ABSTRACT: Introduction: Spatial prediction models are based on the field of geostatistics. Assumptions of these models include spatial autocorrelation, stationarity, and isotropy. Relaxing of the assumptions may provide for models that better predict the data generating process; however, when the assumptions are relaxed, the sample size used to make the predictions are reduced. To relax the assumptions of stationarity and isotropy while maintaining the sample size, we develop a technique using Conditional Kriging.

Methods: The theoretical model for Conditional Kriging is developed. Using the parameters assigned to a partitioned conditioned on the location of the point to be predicted, we are able to maintain the properties of the semivariogram. We create a set of 400 simulated domains. We then compare three common techniques with the Conditional Kriging model. Additionally, we compare the four methods using fine particulate matter ($PM_{2.5}$) air pollution from the period of January 2014 until June 2015. The performance measures are used to compare the models.

Results: In the simulated settings, the Partitioned Domain model performs well when considering prediction error and interval, but does very poorly in the other measures. Conditional Kriging and the Anisotropic Model perform best overall. In the $PM_{2.5}$ analyses, the Partitioned Domain model performed well.

Discussion: When the sample size is small, the more complex models start to have difficulty particularly with the prediction interval and coverage measures. With a large enough sample size, the Partitioned Domain model does best. The Conditional Kriging and Anisotropic Models performed good in all sample sizes. The real world analysis was based on a large sample size, thus the Partitioned Domain was appropriate.

Conclusion: More complex models generally outperformed the baseline model. However, with small sample sizes, these models may not be appropriate. For mid-size samples (200 to 500), the Conditional Kriging and Anistropic Models were best at predicting exposure levels.