STATISTICS SEMINAR

Brook Russell
Colorado State University

Thursday, January 29, 2015
Dickens Hall, Room 207, 4:00-5:00 pm
Refreshments: Dickens 108, 3:30-4:00 pm

Data Mining for Extreme Behavior with Application to Ground Level Ozone

Abstract:

This project aims to increase understanding of the meteorological conditions which lead to extreme ground level ozone conditions. Our approach focuses only on the tail behavior by utilizing the framework of regular variation. Our approach has two parts. The first is an optimization problem: given a set of meteorological covariates, we aim to find the linear combination of these covariates which has the highest degree of tail dependence with ozone. The second is a data mining problem: given a long list of possible meteorological covariates, we seek to find the ones which are linked to extreme ozone.

We use a constrained optimization procedure which maximizes a measure of tail dependence and whose constraint enforces a requirement on the marginal distribution. Our optimization procedure requires that we consider tail dependence estimators with a smooth threshold, rather than the hard threshold typical of extremes. Data mining is performed within the model selection context, and because the model space cannot be explored completely, we employ an automated model search procedure. We present a simulation study which shows that the method can detect complicated conditions leading to extreme responses. We apply the method to ozone data for Atlanta and Charlotte and find similar meteorological drivers for these two Southeastern US cities. We identify several covariates which may help to differentiate the meteorological conditions which lead to extreme ozone levels from those which lead to merely high levels. Our current work includes a spatial extension of our modeling procedure.