Program

Ninth Annual Winter Workshop
Environmental and Environmental
Health Statistics

Department of Statistics
University of Florida
January 12-13, 2007

Contents

Sponsors

This year's symposium is funded by Institute of Food and Agricultural Science, Florida Fish and Wildlife Conservation Commission, Info Tech, Inc., University of Florida Graduate School, and the Department of Statistics.

Organizing Committee

Babette Brumback
Mary Christman
Michael Daniels
Yongsung Joo
Xueli Liu

Invited Speakers

Francesca Dominici, Johns Hopkins University
James Eisner, Florida State University
Alan Gelfand, Duke University
Marc Genton, Texas A&M University
Peter Guttorp, University of Washington
Tom Lumley, University of Washington
Doug Nychka, National Center for Atmospheric Research
Walt Piegorsch, University of Arizona
Richard Smith, University of North Carolina
Paul Switzer, Stanford University
Mark van der Laan, University of California, Berkeley

Acknowledgements

The organizers thank the Department of Statistics staff, Robyn Crawford, Tina Greenly, Allison Pipkin, Carol Rozear, Marilyn Saddler, and Juanita Williams for their tremendous efforts in helping to set up this meeting and make it run smoothly.

Conference Schedule

Thursday, January 11, 2007

6:00-8:00 PM Welcome Reception, Keene Faculty Center, Dauer Hall

Friday, January 12, 2007

8:00am Breakfast (J. Wayne Reitz Union, Room 282)

All Sessions Meet in the J. Wayne Reitz Union, Room 282

8:45am Welcome by Mary Christman, Conference Chair
Mark McLellan, Dean for Research, College of Agriculture and Life Sciences

9:00am **Session 1**

Speakers:

Doug Nychka, National Center for Atmospheric Research
Ensemble Kalman Filters: The movie

10:00-10:30am **Break and Conference photo at JWRU South side**

10:45-12:45pm **Session 2**

Speakers:
Richard Smith, University of North Carolina
Extreme Precipitation Trends over the Continental United States

James Eisner, Florida State University
Extreme Hurricane Winds in the United States

12:45-2:15pm **Lunch** (Gator Comer Dining Center) Map on page 17

2:15-4:15pm **Session 3**

Speakers:
Peter Guttorp, University of Washington
Some Statistical Aspects of Environmental Standards

Walt Piegorsch, University of Arizona
Statistical Methods for Benchmark Analysis in Environmental Risk Assessment

4:30-6:30pm **Poster Session:** J. Wayne Reitz Union, Rooms 346, 347,349

Saturday, January 13, 2007

8:00-8:30 Breakfast (J. Wayne Reitz Union, Room282)

All Sessions Meet in Reitz Union Lecture room 282

8:30-10:30am **Session 4**

Speakers:
Paul Switzer, Stanford University
Analysis of Monitoring Data for Time Trends Using Maximum Autocorrelation Factors

Alan Gelfand, Duke University
High Resolution Space-Time Ozone Modeling for Assessing Trends

10:30-10:45am **Break**

10:45-ll:45pm **Session 5**

Speakers:
Francesca Dominici, Johns Hopkins University
Health Effects of Chemical Composition and Size of Particulate Matter: New Statistical Challenges

ll:45-l:15pm **Lunch** - free time

l:15-2:15pm Tom Lumley, University of Washington
Air Pollution Source Apportionment and Health Effects Models

2:15-2:30pm **Break**

2:30-4:30pm **Session 6**

Speakers:
Marc Genton, Texas A&M University
Modeling and Testing Properties of Space-time Covariance Functions

Marc van der Laan, University of California, Berkeley
Targeted Maximum Likelihood Learning: Application to Air Pollution

5:00-8:30pm. **Barbecue:**
Hosted by Mary Christman
2219 NW 23rd Terrace, Gainesville, FL 32605
(352) 371-4093

Invited Talks

Ensemble Kalman Filters: The movie

Doug Nychka, National Center for Atmospheric Research

Combining a numerical model with observations to improve the estimate of the state of a system is at the heart of numerical weather prediction, air quality forecasts, target tracking and numerous other applications. This talk will explain how to do data assimilation even for complex problems using simple linear regression! This surprisingly simple, sequential algorithm is at the heart of the NCAR Data Assimilation Research Testbed and has been successful on a wide range of geophysical models. Data assimilation has roots as a Bayesian statistics problem where the prior distribution is the current guess of the state of the system, the likelihood is the distribution of the observations given the state of the system and the posterior is the updated estimate of the system given the observations. The Ensemble Kalman filter (EKF) is a Monte Carlo based algorithm for solving this Bayesian problem that has been successful in practice. Usually the EKF is presented using matrix expressions that make it difficult to understand its properties. As an alternative, we present the connection between sequentially updating the state with new observations and regression. This equivalent point of view shows how the EKF works and how to modify the EKF to handle outliers or non-Gaussian distributions. The use of ensembles of states to express uncertainty also has some interesting applications for the analysis of spatial data and is also a promising algorithm for estimating parameters for nonlinear models. Examples of these applications will be given in this talk.

Extreme Precipitation Trends over the Continental United States

Richard L. Smith, University of North Carolina, Chapel Hill

In recent years there has been much climatological literature devoted to the apparent increase in extreme precipitation events that may be a consequence of broader trends in climate due to increased greenhouse gases. However, most current papers (for example, Groisman et al, Journal of Climate, 2005) still use rather simple statistical methods to study this phenomenon. Here, we apply state of the art extreme value methods based on exceedances over high thresholds, including covariates to represent trend and seasonality, to estimate 25-year return levels, and trends in those return levels over 1970-1999. This is done separately for nearly 5,000 stations in the US climatological network, then the results are combined across stations using spatial statistics. The results provide a detailed picture of how trends in rainfall extremes have varied across the United States, as well as summary statistics computed for 19 regions. They indeed confirm an overall increase in extreme rainfall levels, but it is by no mean homogeneous across the whole country. Separate analysis based on model runs from NCAR's Community Climate System Model provide a first insight into how such changes may project into the future.

Extreme Hurricane Winds in the United States

James Elsner, Florida State University

Coastal hurricanes create tremendous environmental change and generate huge financial losses. The relative infrequency of severe coastal hurricanes implies that empirical probability estimates of the next big catastrophe will be unreliable. Here we model hurricane activity and resulting insured losses using extreme value theory and Bayesian models.

The occurrence of a hurricane above a specified threshold intensity is assumed to follow a Poisson distribution and the distribution of the maximum wind is assumed to follow a generalized Pareto distribution. The likelihood function is the product of the generalized Pareto probabilities for each wind speed estimate. Model parameters are first estimated using a maximum likelihood (ML) procedure. Results estimate the 100-year return level for the entire coast at 157 kt (+/-10 kt), but at 117 kt (+/- 4 kt) for the East coast region. Highest wind speed return levels are noted along the Gulf coast from Texas to Alabama.

We also examine how the extreme wind return levels change depending on climate conditions including the El Nino-Southern Oscillation, the Atlantic Multidecadal Oscillation, the North Atlantic Oscillation, and global temperature. The mean 5-year return level during La Nina (El Nino) conditions is 125 (116) kt, but is 140 (164) kt for the 100-year return level. This indicates that La Nina years are the most active for the occurrence of strong hurricanes, but that extreme hurricanes are more likely during El Nino years. Although El Nino inhibits hurricane formation in part through wind shear, the accompanying cooler lower stratosphere appears to increase the potential intensity of hurricanes that do form.

To take advantage of older, less reliable, data the models are reformulate using Bayesian methods. Gibbs sampling is used to integrate the prior over the likelihood to obtain the posterior distributions for the model parameters conditional on global temperature. Warmer temperatures are conditionally associated with a greater frequency of strong hurricanes and higher return levels for the strongest hurricanes. Results compare favorably with a ML approach as well as with recent modeling and observational studies. The maximum possible near-coastal wind speed is estimated to be 208 kt (183 kt) using the Bayesian (ML) approach.

Although wind speed is directly related to damage potential, the amount of damage depends on both storm intensity and storm size (spatial extent). Insured losses provide a direct measure of storm damage. As anticipated we find climate conditions prior to the hurricane season provide information about possible future insured hurricane losses. We exploit this information to model the distribution of likely annual losses and the distribution of a worst case catastrophic loss aggregated over the entire U.S. coast.

Some Statistical Aspects of Environmental Standards

Peter Guttorp, University of Washington

Environmental standards are set by government to protect health and property. In this talk we will first look at some different air quality standards. We will discuss where the design value of the standard came from, and how the government has decided to implement the protection. Hypothesis testing is one approach to implementing standards, but apparently not one frequently used among standard setters. Furthermore, most standards are evaluated by measurements at particular sites. We look at the issue of what exposure can happen away from the sites. Finally we discuss how uncertainty propagates through the system, and how one can go about treating this uncertainty properly to protect individuals within acceptable bounds of risk.

Statistical Methods for Benchmark Analysis in Environmental Risk Assessment

Walter W. Piegorsch, University of Arizona

Statistical models and methods are discussed for quantifying the risk associated with exposure to environmental hazards. Attention is directed at problems in dose-response modeling when low-dose risk estimation is a primary goal. Most common is the problem of estimating excess risk when data are in the form of proportions. Under this paradigm, excess risk measures are discussed which are used to construct benchmark doses" (BMDs); these BMDs represent the exposure/dose levels at which specified excess risk benchmarks are achieved. Simultaneous confidence bands are derived for making inferences on the excess risk. These are then inverted in order to construct confidence limits on the BMDs.

Analysis of Monitoring Data for Time Trends Using Maximum Autocorrelation Factors

Paul Switzer, Stanford University

Consider monitoring data from a regional monitoring network as a rectangular spatial-temporal array. The goal here is to re-express these monitoring data so as to extract dominant temporal components, filter out variations that are not part of these dominant time trends, and display the spatial variation of these dominant time trends. To this end regional weather or pollution data are commonly analyzed using methods closely related to principal components analysis, although such methods are not specifically optimized for time trend detection. Less common alternative analyses use variants of MAF, maximal autocorrelation factor analysis. I describe MAF optimality properties that are specific to time trend extraction for monitoring data and illustrate MAF's trend extraction possibilities in low signal-to-noise situations. A MAF analysis is applied to air pollution monitoring network data and a resampling approach is applied to assess the statistical stability of the MAF analysis. Finally, a method is described for examining the regional variation of the time trend components that are extracted by MAF. A second example applies MAF trend detection to climatic output of General Circulation Models.

High Resolution Space-Time Ozone Modeling for Assessing Trends

Alan E. Gelfand, ISDS, Duke University

The assessment of air pollution regulatory programs designed to improve ground level ozone concentrations is a topic of considerable interest to environmental managers. To aid in this assessment, it is useful to model the space-time behavior of ozone for predicting summaries of ozone across spatial domains of interest and for the detection of long-term trends at monitoring sites. These trends, possibly adjusted for the effects of meteorological variables, are needed for determining the effectiveness of pollution control programs in terms of their magnitude and uncertainties across space. We offer a space-time model for daily 8-hour maximum ozone levels. The model is applied to the analysis of data from the state of Ohio which contains a mix of urban, suburban, and rural ozone monitoring sites. The proposed space-time model is auto-regressive and incorporates the most important meteorological variables observed at a collection of ozone monitoring sites as well as at several weather stations where ozone levels have not been observed. This misalignment is handled through spatial modeling. In so doing we adopt a computationally convenient approach based on the successive daily increments in meteorological variables. The resulting hierarchical model is specified within a Bayesian framework and is fitted using MCMC techniques. Full inference with regard to model unknowns as well as for predictions in time and space, evaluation of annual summaries and assessment of trends are presented.

Health Effects of Chemical Composition and Size of Particulate Matter: New Statistical Challenges

Francesca Dominici, Johns Hopkins University

Optimal regulatory control of particulate matter (PM) is hindered by an insufficient understanding of the chemical characteristics of the ambient PM mixture that determine its toxicity. Ambient PM varies spatially and temporally in its characteristics, reflecting the blend of contributing sources and meteorological factors. With regard to size, $PM_{2.5}$ and $PM_{10-2.5}$ are distinct entities with fundamentally different sources and formation processes, chemical composition, and transport distances. With respect to chemical composition, approximately 50 correlated chemical components of $PM_{2.5}$ are monitored almost daily for several hundred monitoring stations in the US. A critical question is whether a small subset of these chemical constituents, alone or through their interactions are responsible for the estimated harmful effects of PM_{10} and $PM_{2.5}$.

To estimate the toxicity of the PM complex mixture and constituents we have assembled a national data set on health outcomes, PM chemical constituents, and all the important confounders. They include: 1) the National Medicare Cohort comprising individual-level data on diseases, age, gender, and race for virtually the entire population of US elderly; 2) concentrations of key air pollutants (gaseous pollutants, PM_{10}, $PM_{2.5}$, $PM_{10-2.5}$, and $PM_{2.5}$ chemical species) ; 4) weather for all monitoring stations across the country; and 5) US Census data. These data were collected for a variety of purposes, and are misaligned in time and space.

In this talk we will present Bayesian hierarchical regression models for analyses of these large and complex national datasets. Our modeling approaches will have the ultimate goal of estimating the health effects of PM size and chemical composition on a national scale accounting for several sources of uncertainty.

Air Pollution Source Apportionment and Health Effects Models

Thomas Lumley, University of Washington

Particulate air pollution comes from a variety of sources that have different physical, chemical, and political characteristics. Attributing health effects of pollution to specific sources would be very valuable. Unfortunately source contributions are not directly measurable and must be imputed. I will discuss the source apportionment (receptor) models used for imputation and the problems of inference about source-specific health effects.

Modeling and Testing Properties of Space-time Covariance Functions

Marc G. Genton, Texas A&M University

Modeling space-time data often relies on parametric covariance models and various assumptions such as full symmetry and separability. These assumptions are important because they simplify the structure of the model and its inference, and ease the possibly extensive computational burden associated with space-time data sets. We review various space-time covariance models and propose a unified framework for testing a variety of assumptions commonly made for covariance functions of stationary spatio-temporal random fields. The methodology is based on the asymptotic normality of space-time covariance estimators. We focus on tests for full symmetry and separability, but our framework naturally covers testing for isotropy, Taylor's hypothesis, and the structure of cross-covariances. The proposed test successfully detects the asymmetric and nonseparable features in two sets of wind speed data. We perform simulation experiments to evaluate our test and conclude that our method is reliable and powerful for assessing common assumptions on space-time covariance functions.

The talk is based on joint work with Bo Li (NCAR) and Michael Sherman (TAMU).

Targeted Maximum Likelihood Learning: Application to Airpollution

Mark van der Laan, University of California, Berkeley

We are concerned with estimating the causal effect of a decrease in ozone in the LA basin on hospital discharges due to asthma for ages birth to 19 years based on data consisting of quarterly measurements of ozone, asthma hospital discharges and other covariates from 1983 through 2000 inclusively over 195 geographical units. The immediate effect of ozone during a given quarter on asthma hospital discharges during that same quarter, adjusted for a user supplied set of covariates/confounders W, is modeled with a W-adjusted variable importance parameter defined in van der Laan (2005). By assuming a model only for the terms involving ozone and leaving the remainder of the regression unspecified we obtain a semi-parametric regression model in which the parametric component represents the adjusted effect of ozone of interest. We estimate the parametric model parameter with a new targeted maximum likelihood procedure which maps an initial maximum likelihood density estimator into a solution of the efficient influence curve estimating equation with an increased likelihood (and thereby less bias for the parametric component). This is done iteratively, by creating a hardest parametric sub-model with parameter epsilon through the given density estimator with score equal to the efficient influence curve at the density estimator of the parametric component of the semi-parametric regression model. This epsilon is then estimated with the maximum likelihood estimator, and finally a new density estimator is defined as the corresponding update of the original density estimator. This general targeted Maximum Likelihood Estimation procedure is an improvement over the maximum likelihood fit of the semi-parametric regression which suffers from bias for the parametric part due to over-smoothing/bias in the nonparametric part of the semi-parametric regression model. We apply this approach to obtain a W-adjusted variable importance estimate for user supplied subsets W (including singletons) with ozone, one at a time. We then select those subsets of covariates with significant associations using re-sampling based multiple testing adjustments.

Poster Abstracts

Spatial and Temporal Distribution of Basal Stem Rot Disease in Oil Palms (Elaeis guineensis) on a Former Coconut (Cocos nucífera) Plantation

Wong Wan Chew, University Malaysia Sabah

A study site with 9 hectares with 1124 oil palm *(Elaeis guineensis)* tress located in Sungai Balung Estate, Tawau region in Sabah, Malaysia. The selected study site was initially a field about 3.4% disease severity of basal stem rot disease caused by *Ganoderma boninense (suspected fungus).* It situated in a flat landscape with three different of soil type namely Inanam series, Lumisir and Paliau. First visual census assessment on BSR infection was conducted in October 2005 as a baseline guide for this study. Then visual census data two months intervals were conducted over 2006. The main objective was to identify the BSR disease pattern whether it occurs randomly or spatially clustered in a field. Subsequently, we attempt to gain understanding about the underlying mechanism that drives the proliferation of *Ganoderma* through these events of chronological infection pattern. For the purpose of the study, visual census data of BSR were analysed using Moran's "7" and Geary's "C" and kernel density analysis (ArcGIS9.0). Diseased trees increased gradually in early of 2006 with 43 palms and the end of same year with 61 palms. Global Moran's / and the relevant tests of significance (P< 0.001) demonstrated that results are not significant whereas the values of Geary's *C* coefficient in March, May, September and November in 2006 are greater than 1 indicating a negative spatial autocorrelation significantly and disease events distributed dissimilar (P< 0.001). The diseased palms distribution displayed spatially pattern within the whole area in that year. *Ganoderma* inocula were evenly distributed in the ground but influenced by different of soil type and immunity system of each individual palm for resistance of that fungus. Therefore, the key to reduce the severity of disease is to destroy the inoculum of the pathogen in infected stump and roots residues as soon as possible e.g. excavation to remove diseased stump before replanting as suggested by Idris *et al.* (2004), thereby inhibiting inoculum production.

Keywords: oil palm *(Elaeis guineensis),* basal stem rot, spatial statistics, Moran's I and Geary's C, kernel density, spatial autocorrelation

A New Bootstrap Method for Time Series Models is Proposed

Mihai Giurcanu, University of Florida

We propose a new method to bootstrap Z-estimators for time series in order to eliminate the need to recenter the estimating equations or to estimate the centering value in the bootstrap world. In the proposed method, bootstrap resamples are drawn from a weighted distribution on the blocks of consecutive observations, with the weights obtained by minimizing the Cressie-Read distance to the empirical distribution under the constraint that it satisfies the moment conditions at the Z-estimator. Using a recycling algorithm, we can iterate this biased bootstrap in order to increase the accuracy of the bootstrap approximation, preserving at the same time the computational effort required by the single level bootstrap. Consistency results, an application to optimal selection of the block size, and some simulations are provided.

A Composite Likelihood Cross-Validation Approach in Selecting Bandwith for the Estimation of the Pair Correlation Function

Yongtao Guan, Yale University

A useful tool while analyzing spatial point patterns is the pair correlation. In practice, this function is often estimated by some nonparametric procedure such as kernel smoothing, where the smoothing parameter (i.e., bandwidth) is often determined arbitrarily. In this article, a data-driven method for the selection of the bandwidth is proposed. The efficacy of the proposed approach is studied through both simulations and an application to a forest data example.

New Directional Dependence using Copula Function

Jong-Min Kim, University of Minnesota-Morris

We introduce a copula method for understanding multivariate distributions which has a relatively short history in statistics literature; most of the statistical applications related to Copula have arisen in the last fifteen years. Copula approach to Environmental data will be much useful and beneficial. Recently, Sungur (2005) proposed the concept of directional dependence in bivariate regression setting by using copulas. Using the copula directional dependence method proposed by Sungur (2005), we analyze and investigate environmental data such as Ozone data. Furthermore, we develop the copula directional dependence method in light of the financial perspective.

Calibrating OLS Confidence Interval in Long Memory Regression

Kyungduk Ko[1], Jaechoul Lee[1], and Robert B. Lund[2]
^oise State University, [2]Clemson University

Linear regression models with long memory errors have received considerable attention in many fields, such as climatology, engineering, medical imaging, and economics, because of their wide range of possible applications. Use of maximum likelihood or generalized least square estimates for inferential purpose is practically limited due to the complexity and computational demands in evaluating the estimates and their variances by the nature of long range dependency. This paper proposes a calibration method for the confidence interval of an ordinary least squares (OLS) estimator of the trend parameter for a fixed sample size in a linear regression model with long memory errors. The exact variances of the OLS estimators are derived in a closed form and an adjustment procedure of the degrees of freedom for Student's ¿-distribution is developed, which gives a prescribed coverage probability. Simulation results are reported together with a real application of the method to a global temperature series.

Kernel Estimation of Multivariate Cumulative Distribution Function

Rong Liu, Lij ian Yang, Michigan State University

A smooth kernel estimator is proposed for multivariate cumulative distribution function, extending the work on Yamato (1973) on univariate distribution function stimation. Under assumptions of strict stationarity and geometrically strong mixing, we establish that the proposed estimator follows the same pointwise asymptotically normal limiting distribution of the empirical cdf, while the new estimator is a smooth instead of a step function as the empirical cdf. We also show that under stronger assumptions the smooth kernel estimator converges to the true cdf uniformly almost surely at a rate of $(n^{-1/2} \log n)$. Simulated examples are provided to illustrate the theoretical properties. Using the smooth estimator, survival curves for US GDP growth are estimated conditional on the pollution growth rate to examine how GDP growth rate depends on the environmental policy.

Keywords: bandwidth; Berry-Esseen bound; GDP; kernel; pollution; rate of convergence; strongly mixing; survival function

Short Running Title. Kernel Distribution Estimation

Stochastic Precipitation Models and Simulations

Richard L. Smith, Brian J. Lopes & Lawrence Band, University of North Carolina

As recently as 2002 central North Carolina has experienced significant drought problems. The potential environmental and economic impacts of these low precipitation events have lead policy makers to evaluate their current approaches to water supply. In order to do so they require a stochastic model to evaluate the probability of extremal precipitation periods.

The purpose of this work is to expand on currently used methods for modelling precipitation events on a daily scale. With the proposed statistical techniques we can assess the probabilities of extreme precipitation events, and use simulated data as the input for hydrological models. Accounting for precipitation and the current progressive land use trend in North Carolina should give valuable insight into the major water supply watersheds of North Carolina.

The current work is an application of sampling-importance-resampling as discussed in Rubin (1987). We will exploit the analogy of this approach to the nearest-neighbor algorithms currently accepted in the hydrological community. The focus of our comparison will be the work discussed in Yates *et al.* (2003). By using this approach we expand the currently employed techniques for modelling daily precipitation to have more appropriate statistical implications behind the algorithm. A key component of this method is the use of climate models and reanalysis data in order to incorporate trends seen on the global climate scale when looking at the local events in North Carolina.

The final outcome will be daily precipitation for approximately 90 weather stations in North Carolina. Model validation will be done by comparing results with observed information for the 1990s. Then the algorithm will be used to get an ensemble of precipitation outcomes for the years 2000 through 2050.

References

Rubin, D.B. (1987), A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. *J. Amer. Statist. Assoc* **82,** 543-546.

Yates, D., Gangopadhyay, S., Rajagopalan, B. and Strzepek, K. (2003), A technique for generating regional climate scenarios using a nearest-neighbor algorithm. *Water Resources Research* **39,** No. 7, 1199, doi:10.1029/2002WR001769.

Maximum Likelihood For Spatially Correlated Discrete Data

Lisa Madsen, Oregon State University

Spatially correlated discrete data arise in such diverse disciplines as ecology, public health, and economics. A common inferential goal is to estimate the relationship between a spatially indexed discrete response and a set of explanatory variables. Techniques such as weighted least squares or maximum likelihood for regression of a spatially correlated response on a set of covariates are well-known when the response can be assumed Gaussian. When the response is discrete, satisfactory regression techniques have not been developed. Some authors have Liang and Zeger's (1986) generalized estimating equations (GEE) approach for non-normal spatial data.

This approach is unsatisfactory for three reasons. First, GEEs were developed for longitudinal data, where different subjects are assumed independent and correlation occurs only among observations made on the same subject. Spatial data do not naturally break down into independent groups. Secondly, these models assume that conditional on a spatially correlated latent variable, the responses are independent. The latent variable models place artificially low limits on the degree of correlation between observations. Thirdly, GEE estimators are not necessarily efficient.

I propose a copula-based maximum likelihood (ML) approach for spatially correlated discrete data. This approach not only allows modeling of correlations up to the theoretical limit, but will be fully efficient, given appropriate regularity conditions. The proposed model is not specific to spatially correlated discrete data but can be applied to any correlated discrete data including time series, space-time problems, and longitudinal data.

An application in insect ecology is discussed in which Japanese beetle grub counts are modeled as a function of soil organic matter. Specifically, if *Y(s)* and A(.v) are the number of grubs and percent soil organic matter, respectively, at 143 locations 5, then assume £[7(s)] = exp(/?$_0$ + ДX(s) + $\beta_2 X\{s)^2$ + Д,X(д·)^3). The proposed ML approach is used to obtain estimates of the β_k.

Performance of the ML estimator is illustrated with a simulation study in which the grub data are simulated. The estimator is shown to be approximately unbiased, and standard errors are close to the observed standard deviation. Nominal 95% confidence intervals for the β_k are shown to have coverage probabilities of over 93%.

Modeling Spatial Autocorrelation in Residuals from a River Monitoring Network using Shared Areas

Steve Patch, University of North Carolina at Asheville

Using coordinate data (latitude and longitude) does not adequately represent potential autocorrelation in river networks because a pair of sites located near each other may have little or no upstream watershed in common. Using the proportion of the larger watershed that is shared seems intuitively to be a better measure of closeness of two sites. For a river monitoring network the watershed flowing into each site may be determined using a digital elevation model. For every pair of monitoring sites, a measure of closeness of those sites can be defined to be the area those watersheds have in common divided by the area of the larger watershed. Multidimensional scaling can be applied to translate these pairwise closeness measures into a κ-dimensional coordinate system. Resulting site coordinates in that system can then be input to a spatial covariance model for the residuals from a statistical model considering the sites as random effects. An example is provided of how this methodology can be applied to an investigation of the effects of land use on nutrients in the Swannanoa River Watershed (North Carolina).

Extreme Precipitation: An Application Modeling N-Year Return Levels at the Station Level

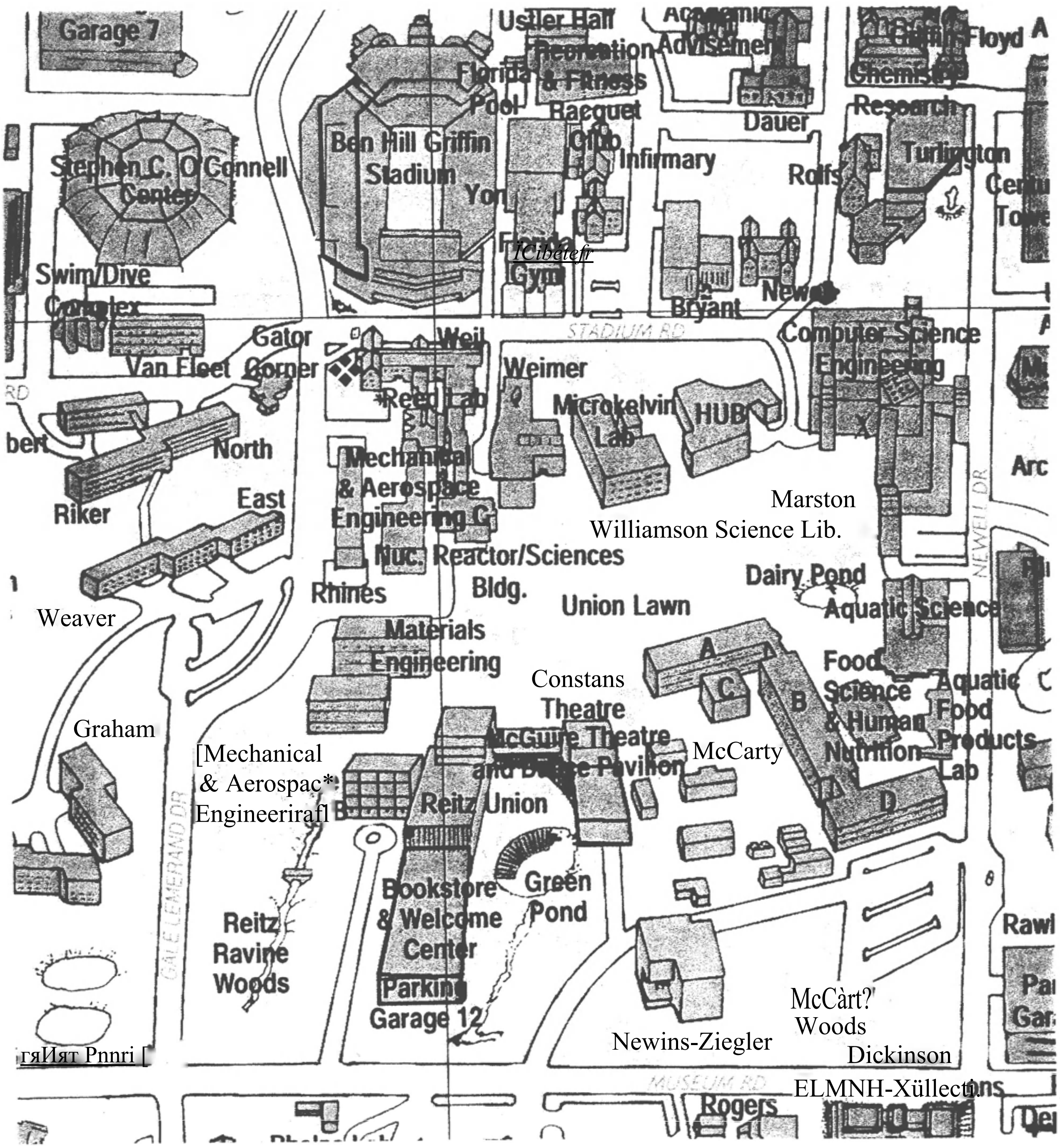
Elizabeth Shamseldin, Richard L. Smith, University of North Carolina at Chapel Hill
Stephan Sain, Doug Nychka, and Dan Cooley; National Center for Atmospheric Research

The question under investigation is whether regional climate model return level estimations can be used to obtain return level predictions at the station level. To begin, the relationship of grid cell data to the n-year return levels at point locations is explored. The tail of the Generalized Extreme Value distribution (GEV) is fit to the grid cell data above a given threshold. This is similar to the Peaks Over Threshold method. However, here the method used for parameter estimation is a Point Process approach, which leads directly to the GEV parameters. Different threshold values are tested for model stability. The parameter estimates are used to generate n-year return levels. The n-year returns at the point (station) locations are estimated the same way. Various models are explored to predict point location n-year return from grid cell n-year return. Current results are presented. Future work includes plans to test grid-point models on Regional Climate Model data and Community Climate System Model data, as well as investigation of seasonal and spatial effects in the model.

Workshop Participants

NAME	AFFILIATION	NAME	AFFILIATION
Tatiayana Apanasovich	Cornell University, SORIE	Brian Lopes	UNC at Chapel Hill
Alan Agresti	University of Florida	Lisa Madsen	Oregon State University
Babette Brumback	University of Florida	Melissa Martin	University of Florida
Cameron Carter	Florida Wildlife	Luke McEachron	Florida Wildlife
George Casella	University of Florida	Shabnam Mehra	University Of South Florida
Mary Christman	University of Florida	Curtis Miller	University of Florida
Elizabeth Dahlguist	University of South Florida	Gary Mohr	Florida Wildlife
Mike Daniels	University of Florida	Jensen Montambault	UF/Nat'l Resources
Kiranmoy Das	University of Florida	Trevor Park	University of Florida
Chip Deutsch	Florida Wildlife	Steve Patch	Environmental Quality Institute of UNCA
Bob Dorazio	University of Florida	Rachel Pawlitz	University of Florida-SNRE
Hani Doss	University of Florida	Ronald Randles	University of Florida
Claudio Fuentes	University of Florida	Elizabeth Shamseldin	University of North Carolina - Chapel Hill
Mihai Giurcanu	University of Florida	Katie Sieving	UF/Wildlife Ecology
Yongtao Guan	Yale University	Kagba Suaray	CSU Long Beach
James Hobert	University of Florida	Aixin Tan	University of Florida
Thomas Jagger	Florida State University	Zachary Thompson	University of South Florida
Bob Kawula	Florida Wildlife	Peter Toyinbo	University of South Florida
Richard Kiltie	Florida Fish & Wildlife	Alex Trindade	University of Florida
Paul Kubilis	Florida Wildlife	Melissa Tucker	Florida Wildlife
Christine Laufer	University of Florida	Joe Walsh	Florida Wildlife
Jaechoul Lee	Boise State University	Jing Wang	University of Illinois at Chicago
Yao Li	University of Florida	Peter Wludyka	University of Florida/University of North Florida
Zhen (Jane) Li	UNC Chapel Hill	Ping Xu	University of South Florida
Ramon Littell	University of Florida	Jie Yang	University of Florida
Rong Liu	Michigan State University	Linda Young	University of Florida
Ruitao Liu	University of Florida	Ricardo Zambrano	Florida Wildlife
Xueli Liu	University of Florida		

Map to Gator Corner Dining



Map

Mary Christman
2219 NW 23rd Terrace
Gainesville, FL 32605
352-371-4093