Geophysical Statistics Project
External Advisory Panel Meeting
27 April 2006
Research Talks

Dan Cooley

Modeling Extreme Precipitation for Colorado's Front Range

A common tool used by flood planners to assess a location's potential for extreme precipitation are return-levels maps. The n-year return level is the (precipitation) amount which is exceeded on average once every n years.

We have developed a Bayesian methodology to produce a precipitation return level maps for the Front Range region of Colorado. To develop the map, we have relied on the theory of extreme values, specifically, using the generalized Pareto distribution (GPD) to model precipitation above a threshold. We use a hierarchical strategy to model the spatial variation of the stations' GPD parameters. Our spatial analysis is done in a unique space in which a location's coordinates are given by climatological measures rather than longitude and latitude.

We are currently working on the question of how to model precipitation events different duration (e.g., 1 hour and 6 hour) periods. In the past, different durations have been modeled separately which can lead to inconsistent estimates.

Joint work with Richard Smith

Dorin Drignei

Parameter Estimation for Computationally Intensive Nonlinear Regression with an Application to Climate Modeling

Nonlinear regression is a useful statistical tool, relating observed data and a nonlinear function of unknown parameters and covariates. When the parameter-dependent nonlinear function is computationally intensive a straightforward regression analysis by maximum likelihood is not feasible. The method presented in this talk proposes to construct a faster running surrogate for such a computationally intensive nonlinear function, and to use it in a related nonlinear statistical model that accounts for the uncertainty associated with this surrogate. The statistical method is applied to a model calibration problem where the computationally intensive nonlinear function is the MIT intermediate size climate model.

Joint work with Chris Forest (MIT)

Anders Malmberg

A Hierarchical Bayesian Spatio-Temporal Model for Tropospheric Carbon Monoxide

Tropospheric carbon monoxide (CO) is an important trace gas and air quality indicator and is one of the few atmospheric species that can be remotely sensed from space. The Measurement of Pollution in the Troposphere (MOPITT) instrument, on board the Terra satellite, provides CO mixing ratios at seven pressure levels at a global scale and can be used to better understand the transport and transformation of CO. To sensibly validate MOPITT data using in-situ observations and for use in assimilation contexts we here propose a hierarchical Bayesian spatio-temporal model based on a advection/diffusion idea. These models and techniques for estimating parameters also have a general use for a variety of environmental problems that include remotely sensed measurements. We present some results, discuss the properties of our model and quantify the impact sparse data have on our model estimates.

Joint work with David Edwards (NCAR/ACD)

Mikyoung Jun

Statistical Analysis of IPCC Climate Model Biases

There are extensive efforts to develop climate models to study the climate change. As a study problem we consider the suite of 20 climate model experiments initiated by the Intergovernmental Panel on Climate Change (IPCC). The previous work with climate model output usually assumes that the output are random samples from a symmetric distribution centered around the true climate. One of the most interesting problems to climate scientists and modelers is to verify the bias of climate models and understand how the biases of different models are correlated. We propose the following solution: model the climate model outputs as a spatio-temporal processes on sphere *x* time and focus on the spatial and temporal covariance structure. We consider covariance models not only for each climate model outputs but also cross-covariance models for pairs of climate model outputs. This facilitates drawing inferences about model biases and also classifying climate models with common biases.

Joint work with Reto Knutti (NCAR/CGD)

Elizabeth Shamseldin

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

The question under investigation is whether regional climate model return levels can be used to obtain return level predictions at the station level. As a first step, we explore the relationship of grid cell data to the n-year return levels at point locations. The tail of the Generalized Extreme

Value distribution (GEV) is fit to the grid cell data above a given threshold and 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 and the parameter estimates are used to generate n-year return levels. The n-year returns at the point (station) locations are estimated in the same way. Various models are explored to predict point location n-year return from grid cell n-year return and some results are presented that relate daily station precipitation to the NCEP reanalysis for the conterminous U.S. Future work includes plans to test grid-point models on CCSM model output.

Joint work with Richard Smith, Steve Sain (UCD), Dan Cooley, Doug Nychka

---

Shree Khare

An Ensemble Smoother for Geophysical Data Assimilation

A smoother is an algorithm which yields a probabilistic estimate of a system state conditioned on both future and past observations. The goal of our current research is to develop a robust smoother tool which can be applied to a wide variety of geophysical assimilation problems of interest. Recently, we have incorporated a Monte Carlo implementation of a smoother into the Data Assimilation Research Testbed (DART). In DART, the smoother can be applied to a hierarchy of geophysical prediction models. Results will be shown from both a low-order dynamical system and an intermediate size atmospheric general circulation model. Critical issues pertaining to the implementation of an ensemble smoother which have been investigated will be discussed.

Joint work with Jeff Anderson (IMAGe)

Eva Furrer

A Generalized Linear Modeling Approach to Stochastic Weather Generators

This work is part of a much larger and broader project on “Climate, Agriculture, and Complexity in the Argentine Pampas”, which is funded by NSF through “Biocomplexity in the Environment: Coupled Natural-Human Systems.” The project is centered at the University of Miami but also includes collaborators at several institutions in the US and Argentina. Our contribution is the generation of daily series of site-specific climatic variables such as precipitation and temperature that are consistent with seasonal climate forecasts (the ENSO index in our case). One possibility to generate such series is the classical (parametric) weather generator, for which we developed a generalization allowing to incorporate seasonal climate forecasts as covariates and extends the work of Chandler and Wheater (2002). The objective of generating weather series within the project is to use them as input for crop models and economic models, which simulate crop yields and economic returns. In view of this objective our main focus is to characterize the uncertainty inherent in the generated weather series.

The Kriging Estimator as a Local Smoother

The objective of this mathematical research project is to apply well-known but probably under-used local properties of smoothing splines to a new, but actually not so different situation: the Kriging estimator. The equivalence between smoothing splines and kriging is well established and it is straightforward - with a few tricks - to show that the kriging estimator is a smoothing spline estimator on a Hilbert space with reproducing kernel based on the covariance function of the underlying Gaussian process. We characterize the kriging solution via a variational problem, express it as the sum of the observations weighted by the kriging weight function and use this representation to suggest a first order approximation is accurate and inherents the same kind of bias and variance properties associated with spline and kernel estimators.

Joint work with Rick Katz and Doug Nychka

Tomoko Matsuo

Nonstationary Covariance Modeling for Incomplete Data

Multi-resolution basis can provide a useful representation of nonstationary processes that are typically encountered in the geosciences. The main advantage over existing nonstationary covariance models is its scalability to large spatial problems. Because of the localized support of the basis functions in a wavelet basis the sparsity of the covariance is naturally introduced as part of the model, and therefore it is amenable to efficient statistical computations. The blocky and diagonal semi-parametric representation of the covariance of the basis coefficients is shown to give a good approximation to the Matérn family of covariances. We also outline some theory that justifies the sparsity properties, which are central to the development of the estimation procedure. Since efficient algorithms exist for a wavelet transform on regular grids, we take advantage of this computational feature through a Monte Carlo Expectation Maximization (EM) algorithm when faced with irregularly distributed incomplete observations. This approach allows one to work with a likelihood (or a pseudo-likelihood) function that is dominated by sparse matrices, and therefore the evaluation for large spatial problems can be facilitated. This feature is further exploited by a nonparametric method using a method of moments in order to efficiently compute an estimate of a nonstationary covariance. Finally, we apply this method to surface ozone data from an environmental monitoring network, and demonstrate how nonstationary structures are captured by our model. Some issues of model selection are also discussed in the context of this application.

Joint work with Debashis Paul (UC-Davis) and Doug Nychka
