UCLA

BIOSTATISTICS SEMINAR

SUMMER 2016

Friday, August 5, 2016

Computational Challenges with Big Environmental Data

Marc G. Genton, PhD

Professor, Applied Mathematics and Computational Science

King Abdullah University of Science and Technology (KAUST), Saudi Arabia

11:00am—CHS 33-105A

Coffee and Cookies served at 10:30AM in room 51-254 CHS

ABSTRACT: We discuss two types of computational challenges arising from big environmental data. The first type occurs with multivariate or spatial extremes. Indeed, inference for max-stable processes observed at a large collection of locations is among the most challenging problems in computational statistics, and current approaches typically rely on less expensive composite likelihoods constructed from small subsets of data. We explore the limits of modern state-of-the-art computational facilities to perform full likelihood inference and to efficiently evaluate high-order composite likelihoods. With extensive simulations, we assess the loss of information of composite likelihood estimators with respect to a full likelihood approach for some widely-used multivariate or spatial extreme models. The second type of challenges occurs with the emulation of climate model outputs. We consider fitting a statistical model to 1 billion global 3D spatio-temporal temperature data using a distributed computing approach. The statistical model exploits the gridded geometry of the data and parallelization across processors. It is therefore computationally convenient and allows to fit a non-trivial model to a data set with a covariance matrix comprising of 10^18 entries. We provide 3D visualization of the results. The talk is based on joint work with Stefano Castruccio and Raphael Huser.

****

A Stochastic Space-time Model for Intermittent Precipitation Occurrences

Ying Sun , PhD

Assistant Professor, Applied Mathematics and Computational Science

King Abdullah University of Science and Technology (KAUST), Saudi Arabia

3:30pm—CHS 33-105A

Refreshments served at 3:00PM in room 51-254 CHS

ABSTRACT: Modeling a precipitation field is challenging due to its intermittent and highly scale-dependent nature. Motivated by the features of high-frequency precipitation data from a network of rain gauges, we propose a threshold space-time t random field (tRF) model for 15-minute precipitation occurrences. This model is constructed through a space-time Gaussian random field (GRF) with random scaling varying along time or space and time. It can be viewed as a generalization of the purely spatial tRF, and has a hierarchical representation that allows for Bayesian interpretation. Developing appropriate tools for evaluating precipitation models is a crucial part of the model-building process, and we focus on evaluating whether models can produce the observed conditional dry and rain probabilities given that some set of neighboring sites all have rain or all have no rain. These conditional probabilities show that the proposed space-time model has noticeable improvements in some characteristics of joint rainfall occurrences for the data we have considered.