INDIAN STATISTICAL INSTITUTE

P. C. Mahalanobis Memorial Lectures
January 9, 10, and 12, 2012

Speaker: Professor Alan E. Gelfand
Department of Statistical Science
Duke University
U.S.A.

Venue: Indian Statistical Institute
203, B.T. Road. Kolkata 700 108.

TITLES AND ABSTRACTS

∗ LECTURE 1 (Public Lecture) ∗

Venue: Auditorium, Platinum Jubilee Academic Building (First Floor)

Date & time: January 09, 2012; 4 p.m.

Space is the Place: Why Spatial Thinking Matters for Environmental Problems?

Spatial methods have become an increasingly used approach for analyzing data in many fields. In particular, it is now routine to collect data layers where there is some geographic referencing. This information should be used in order to enhance inference. From a statistical perspective, we think in terms of formal inference, utilizing probabilistic or stochastic modeling; we think beyond purely descriptive summaries. In this sense, we exceed the capabilities of Geographic Information Systems (GIS) software to investigate complex processes over space and time.

A particularly rich context for such investigation is environmental processes. Examples include analysis of weather/climate data, analysis of environmental exposure data, analysis of locations of disease occurrence, and analysis of distributions of species over a region. In this non-technical talk, I will describe the types of spatial (and, perhaps, spatio-temporal) data that we collect. I will discuss what we expect to see with regard to these types of data, i.e., what we mean by "spatial pattern." I will raise a variety of issues that arise in modeling such data - explanation of local behavior through spatially referenced explanatory variables, explanation of uncertainty through structured dependence. I will illustrate, with a variety of datasets involving the foregoing processes, hopefully to illuminate that statistical thinking does matter when we have inferential objectives such as explanation, interpolation, and prediction.

∗ LECTURE 2 ∗

Venue: NAB-1, New Academic Building (Ground Floor)
Date & time: January 10, 2012; 5 p.m.

Point Pattern Modeling for Degraded Presence-only Data over Large Regions

Explaining species distribution using local environmental features is a long standing ecological problem. Often, available data is collected as a set of presence locations only thus precluding the possibility of a presence-absence analysis. We propose that it is natural to view presence-only data for a region as a point pattern over that region and to use local environmental features to explain the intensity driving this point pattern. This suggests hierarchical modeling, treating the presence data as a realization of a spatial point process whose intensity is governed by environmental covariates. Spatial dependence in the intensity surface is modeled with random effects involving a zero mean Gaussian process. Highly variable and typically sparse sampling effort as well as land transformation degrades the point pattern so we augment the model to capture these effects. The Cape Floristic Region (CFR) in South Africa provides a rich class with such species data. The potential, i.e., nondegraded presence surfaces over the entire area are of interest from a conservation and policy perspective.

Our model assumes grid cell homogeneity of the intensity process where the region is divided into $\sim 37,000$ grid cells. To work with a Gaussian process over a very large number of cells we use predictive process approximation. Bias correction by adding a heteroscedastic error component is implemented. The model was run for a number of different species. Model selection was investigated with regard to choice of environmental covariates. Also, comparison is made with the now popular Maxent approach, though the latter is much more limited with regard to inference. In fact, inference such as investigation of species richness immediately follows from our modeling framework.

∗ LECTURE 3 ∗

Venue: CVPRU Seminar Room, S.N. Bose Bhavan (Eighth Floor)

Date & time: January 12, 2012; 12 noon

Analyzing Spatial Directional Data through the use of Gaussian Processes

Circular data arise in oceanography (wave directions) and meteorology (wind directions), and, more generally, with periodic measurements recorded in degrees or angles on a circle. In this talk we introduce a fully model-based approach to handle circular data in the case of measurements taken at spatial locations, anticipating structured dependence between these measurements. We formulate a wrapped Gaussian spatial process model for this setting, induced from a customary *inline* Gaussian process. We look at the properties of this process, including the induced correlation structure.

We build a hierarchical model to handle this situation and show how to fit this model straightforwardly using Markov chain Monte Carlo methods. Our approach enables spatial interpolation and can accommodate measurement error. We illustrate with a set of angular wave direction data from the Adriatic coast of Italy, generated through a complex computer model.

Then, we consider the projected normal spatial process built from a bivariate Gaussian process model. Such models are more flexible than usual wrapped or von Mises models and easily handle regression. However, they are more challenging to fit. We illustrate with a butterfly dataset.

∗ You are cordially invited to attend. ∗

Contact person: Dr. Sumitra Purkayastha, Applied Statistics Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata 700 108. ∗ **E-mail**: sumitra@isical.ac.in & s.purkayastha@gmail.com