

# International Graduate Institute on Modeling Environmental Space – Time Processes



University of Washington, July 9 - 13, 2007

The International Graduate Summer School on Modeling Environmental Space - Time Processes will provide a program of lectures and labs designed for graduate students or Postdoctoral fellows in the statistical sciences as well as others with a solid background in statistics and probability.

## Instructors:

### Lectures:

- Peter Guttorp, University of Washington
  - <http://www.stat.washington.edu/peter>
- Paul Sampson, University of Washington
  - <http://www.stat.washington.edu/pds>
- Nhu Le, British Columbia Cancer Research Agency
  - [http://www.bccrc.ca/ccr/people\\_nle.html](http://www.bccrc.ca/ccr/people_nle.html)
- Douw Steyn, University of British Columbia
  - <http://www.eos.ubc.ca/public/people/faculty/D.Steyn.html>
- Jim Zidek, University of British Columbia
  - <http://hajek.stat.ubc.ca/~jim/homepage.html>

### Labs:

- Yiping Dou, University of British Columbia
  - <http://www.stat.ubc.ca/People/Home/index.php?person=ydou>
- Zhong Liu, University of British Columbia
  - <http://www.stat.ubc.ca/People/Home/index.php?person=zliu>

## DESCRIPTION:

This course will provide an overview of statistical methods in environmental statistics. Basic concepts on measuring environmental fields, including sampling techniques spatial sampling, data quality and measurement error and its potential effects will be covered. The primary focus will be on the hierarchical modeling of space-time processes and the corresponding spatial interpolation.

Much recent research in air quality modeling involves spatial or spatiotemporal estimation. We will describe a variety of tools, originating in geostatistics but having found their own direction in environmetrics, which are suitable for such things as estimating the air quality at unmonitored locations with realistic estimates of uncertainty. In particular, we adapt the geostatistical kriging method to nonstationary spatial fields and to nonseparable space-time processes, and describe how one can even include

deterministic model output in the estimation. The methods are illustrated with examples out of the lecturers' experience, using freely available software tools.

Some specialized R-based software will be given to participants on a CDROM along with the necessary compilers for its installation on Windows XP and its use illustrated through several examples involving air pollution fields. That same CDROM will contain the course lecture notes as well as supplementary material for the lectures and labs. Much of this software will be based on material given in the reference for the course.

### **LEARNING OUTCOMES:**

The course focuses on giving participants an understanding of various tools that have recently been developed in spatial statistics for use in environmental applications. We do not intend to go into great technical detail, but rather to use examples and illustrations to give participants an understanding of the types of questions that can be addressed using these tools. At the same time, we intend to be relatively comprehensive (although of course it is not possible to cover all approaches in the literature). While participants do not need specific background in spatial statistics or geostatistics, familiarity with methods of regression analysis and maximum likelihood estimation at the level of an M.S. in statistics will be assumed. At the end of the course we expect participants to have an understanding of

- (a) the various simplifying assumptions that have commonly been made in geostatistics
- (b) ways to overcome some of these assumptions
- (c) knowledge of both classical and Bayesian approaches to spatial and spatiotemporal data
- (d) some insight into how one can combine deterministic numerical models with data and stochastic models
- (e) approaches that have been taken to designing networks for monitoring environmental space – time fields and how these may be implemented in practice

## TIMETABLE FOR LECTURES AND LABS:

**Lectures:** HUB 309

**Labs:** Mary Gates Hall 030

**Coffee:** In the lecture room.

**Lunch:** Anywhere

**Monday night mixer:** McMahon Hall Patio

		<b>Mon</b>	<b>Tue</b>	<b>Wed</b>	<b>Thu</b>	<b>Fri</b>
Lectures	0900-1015	1.1 Environmental risk	2.1 Stationary processes	3.1 Space-time modeling	4.1 Mapping mean fields	5.1 Network design
	1015-1045	Coffee				
	1045-1200	1.2 Geostatistics	2.2 Non – stationary covariances	3.2 Spatial prediction	4.2 Model assessment	5.2 Case study
	1200-1300	Lunch				
Labs	1300-1700	1 geoR & geostatistics	2 Modeling non-stationary covariances	3 Using DLM & other methods	4 Melding & and model assessment	5 Student presentations
	1800-2000	Mixer				

## CONTENTS:

### 1.1 The nature of environmental risk

We take participants through the minefield of difficulties facing the environmental risk analyst including the ecological effect and other problems caused by aggregating data, measurement error, collinearity, as well as the transfer of causality. A case study will be presented to demonstrate the use of space-time models in reconstructing individual histories of exposure to air pollution in a study of its association with cancer, a disease with a long latency period.

### 1.2 Geostatistics

An outline of the geostatistical kriging tools, and their relationship to regression, is provided. We discuss optimality properties of kriging for known covariance function, and the consequences (particularly for the prediction variance) of having to estimate the covariance. A Bayesian approach to kriging is also described.

### 2.1 Stationary processes

We describe some classes of valid spatial covariance functions and variograms, and their estimation. The assumptions of isotropy and stationarity are introduced and discussed. Bayesian estimation tools are also provided. An introduction to spectral theory and Fourier analysis of stationary spatial processes will be included.

## **2.2 Nonstationary covariance functions**

The need for nonstationary methods is illustrated. We describe four such approaches: Haas' moving window kriging approach, a method for kernel averaging of locally stationary processes due to Fuentes, a spatially varying moving average ("process convolution") approach due to Higdon and Swall and the Sampson-Guttorp deformation approach.. Fourier approaches to the estimation enable testing for nonstationarity in some cases.

The Sampson-Guttorp approach is illustrated in a variety of examples. Classical as well as Bayesian approaches to the estimation, and assessment of the uncertainty of the estimates, are discussed. Some of the advantages and disadvantages of the different approaches are outlined.

## **3.1 Approaches to modeling space time processes**

An overview is given of the many approaches that have been taken to modeling space-time processes, including those in current research. The use of deterministic models is also discussed. Some basic probability theory needed in subsequent lectures will be included along with discussion of aspects of uncertainty in spatial prediction.

## **3.2 Spatial prediction**

An approach to spatial prediction for multivariate Gaussian fields will be presented. The need for a multivariate theory will be argued, and how it can be applied even when a single environmental response is of interest. Methods for the empirical assessment of predictors will be described. The concept of separability of time and space is introduced, and nonseparable processes are illustrated with examples.

## **4.1 Modeling mean fields**

It is common in the space-time modeling literature to decompose observations into the sum of a systematic trend component and residuals. In geophysics and meteorology, variants of principal components called empirical orthogonal functions (EOFs) have long been used to describe leading modes of variability in space-time processes. Here we illustrate the use of smoothed EOFs to model the spatio-temporal mean of a random field viewed as spatially varying systematic temporal trends.

## **4.2 Assessment of deterministic models and data assimilation**

Recent research in space-time processes has started to bridge the gap between deterministic and stochastic modeling. In particular, we illustrate two methods for assessing air quality models: a geostatistical approach that is suitable for data-rich situations, and a variant of the Bayesian melding due to Fuentes and Raftery. The latter is also useful in combining data with numerical models

## **5.1 Design of monitoring networks**

Societal concerns, which change over time with the accumulation of knowledge, lead to the need to monitor (measure) various environmental processes over time and

space. But where should the monitors be placed, how many should there be and how often should measurements be taken? Above all, what principles should be used to establish such networks, given that their purposes may well change over time because of changing knowledge, concerns and values. This lectures focuses on such issues as well as the practical ones designers face. Various strategies for design will be presented.

## **5.2 A case study**

The Greater Vancouver Regional District recently embarked on a review of its network for monitoring air pollution. This case study will describe the network, how it was established and why a review at this time was thought necessary. Some of the individuals involved in the review will describe the outcomes to the extent that they are known.

### **COURSE CREDIT:**

Two semester credits for this course may be obtained through either STAT 593 (4 quarter credits) at the University of Washington or STAT 547 (2 Credits) at the University of British Columbia. That means in particular that students at PIMS universities should be able to obtain credit through existing inter-university credit sharing arrangements.

**Evaluation:** To obtain such credit on a pass–fail basis, a student will need to submit to one of the instructors, a report no longer than 5 pages in length (not counting figures and tables) no later than July 30, 2007. That report should model or analyze a space-time process, either one encountered in one of the labs or drawn from the participant’s experience.

### **Reference:**

Le, N.D & Zidek, J.V. (2006). Statistical analysis of environmental space – time processes. New York: Springer