

[TIES 2004 / ACCURACY 2004 Invited Program Abstracts \(pdf\)](#)

Physical Statistical Environmental Modeling

Mark Berliner
Department of Statistics
Ohio State University
1958 Neil Avenue
Columbus, OH 43210-1247 USA
mb@stat.ohio-state.edu

Abstract: I discuss the hierarchical Bayesian framework for analyzing environmental problems. This paradigm provides opportunities for the combination of physical reasoning and observational data in a coherent analysis framework, but in a fashion which manages the uncertainties in both information sources. A key to the hierarchical viewpoint is that separate statistical models are developed for the process variables studied and the observations conditional on those variables. Modeling process variables in this way enables incorporation of scientific models across a spectrum of levels of intensity ranging from qualitative use of physical reasoning to strong reliance on numerical models. Selected examples from this spectrum are reviewed.

Wildfire Chances and Probabilistic Risk Assessment

David R. Brillinger
Statistics Department
University of California
Berkeley, CA 94720 USA
brill@stat.berkeley.edu

Abstract. Forest fires are an important societal problem in many countries and regions. They cause extensive damage and substantial amounts are spent preparing for and fighting them. This talk will apply methods of probabilistic risk assessment to estimate chances of fires given a variety of explanatory variables. Updating methods and random effect models will be considered in particular. The work is collaborative with researchers at the U.S. Forest Service.

Efficient Posterior Inference and Prediction of Space-Time Processes Using Dynamic Process Convolutions

Catherine A. Calder
Department of Statistics
The Ohio State University
1958 Neil Ave.
Columbus, OH 43210
calder@stat.ohio-state.edu

Abstract. Bayesian dynamic process convolution models provide an appealing approach for modeling both univariate and multivariate spatial temporal data. Their structure can be exploited to significantly reduce the dimensionality of a complex spatial temporal process. This results in efficient Markov chain Monte Carlo (MCMC) algorithms required for full Bayesian inference. In addition, the dynamic process convolution framework readily handles both missing data and misaligned multivariate space-time data without the need for imputation. We review the dynamic process convolution framework and discuss these and other computational advantages of the approach. We present an application involving the modeling of air pollutants to demonstrate how this approach can be used to effectively model a space-time process and provide predictions along with corresponding uncertainty statements.

The Role of Precaution in Quantitative and Qualitative Analysis of Environmental Decisions

Alison Cullen
University of Washington
Seattle, Washington, USA
alison@u.washington.edu

Abstract. Environmental risks are often described quantitatively, but they are invariably interpreted in light of individual and social values, and receive attenuation (or not) according to both quantitative and qualitative factors. Legal, cultural, political, and practical expectations and constraints also influence individual and social tolerance of risks. The question remains, in the face of uncertain risks, what is the appropriate response of government? A range of responses are revealed across political and geographic boundaries as well as risk contexts, from strict precaution where the presence of uncertainty is argued to justify action, to an insistence on evidence of harm before action is taken.

Recent media coverage paints a picture contrasting precautionary European countries with a risk tolerant US, in matters of environmental health. However it is not difficult to identify risk decisions for any given country which fall along the entire precaution spectrum. In this paper we

undertake an international comparison of the driving factors characterizing several risk decisions about environmental health, such as those inherent in food safety, energy generation, waste management and climate policy. The role and limits of quantitative and qualitative analysis in this process are explored and the accompanying level of precaution revealed.

Hierarchical Bayesian spatio-temporal modeling for wind data

Li Chen, Montserrat Fuentes
Department of Statistics
and
Jerry M. Davis
Department of MEAS
North Carolina State University
Raleigh, NC 27695-8203
lchen4@unity.ncsu.edu

Abstract. Classical geostatistics and time series methods are powerful tools for stationary and separable space-time processes. However, it is well recognized that in real applications spatio-temporal processes are rarely stationary and separable. In this work, some new approaches to model and estimate nonstationarity and nonseparability are presented. We present new classes of nonseparable and nonstationary models for space-time processes. We also propose a test for separability to better understand the space-time dependence. We apply the above statistical methods to model spatio-temporal structures of wind fields and assess the performance of numerical models for wind prediction. Consequently improved wind field map can be obtained by combining observed wind data with numerical model output.

Keywords: stationary, separability, spatio-temporal, hierarchical, Bayesian

Data Quality and Uncertainty in Fine Particulate Monitoring

Alessandro Fassò and Orietta Nicolis
University of Bergamo - Dept. IGI
Via Marconi 5, 24044 Dalmine BG I, Italy
alessandro.fasso@unibg.it

Abstract: In order to assess compliance with air quality standards, (Italian) regulations prescribe to monitor concentrations of particulate matters. The accuracy varies with monitor type, temperature and pollution level often in a complex nonlinear manner. Consequently, comparisons, threshold exceedances interpretation and compliance assessment are often difficult.

For these reasons, in this paper we consider dynamical modelling of spatio-temporal and instrumental uncertainty. An application to north Italy air quality network allows to make some empirical conclusions.

Spatial-temporal modeling of the association between speciated fine particles and human health effects

M. Fuentes, H. R. Song, S. Ghosh, and D. Holland.
Patterson Hall 210 C
Box 8203 NCSU
Department of Statistics
North Carolina State University
Raleigh, NC 27695
fuentes@stat.ncsu.edu

Abstract. Particulate matter (PM) has been linked to a range of serious cardiovascular and respiratory health problems. Some of the recent epidemiologic studies suggest that exposures to PM may result in tens of thousands of excess deaths per year, and many more cases of illness among the US population. The main objective of our research is to quantify uncertainties about the impacts of fine PM exposure on mortality. We develop a multivariate spatial regression model for better estimation of the mortality effects from fine PM and its components across the coterminous US. Our approach adjusts for meteorology and other confounding influences, such as socioeconomic factors, age, gender and ethnicity, characterizes different sources of uncertainty of the data, and models the spatial structure of several components of fine PM. We consider a

flexible Bayesian hierarchical model for a space-time series of (mortality) counts by constructing a likelihood based version of a generalized Poisson regression model. The model has the advantage of incorporating both over and under dispersion in addition to correlations that occur in space and time. We apply these methods to daily mortality county counts, measurements of total and several components of fine PM from national monitoring networks in the US, and the output of deterministic air quality models.

Separable Approximations of Space-Time Covariances

Marc G. Genton
Department of Statistics
Box 8203
North Carolina State University
Raleigh, NC 27695-8203 USA
genton@stat.ncsu.edu

Abstract. Statistical modeling of space-time data has often been based on separable covariance functions, that is covariances that can be written as a product of a purely spatial covariance and a purely temporal covariance. The main reason is that the structure of separable covariances dramatically reduces the number of parameters in the covariance matrix and thus facilitates computational procedures for large space-time data sets. In this talk, we discuss separable approximations of space-time covariances. In particular, we describe the nearest (in the Froebinius norm) Kronecker product approximation of a space-time covariance matrix. The algorithm is simple to implement and preserves symmetry and positive definiteness of the solution. The separable approximation allows for fast kriging of large space-time data sets. We present several illustrative examples and an application to the Irish wind speed data.

Modeling Uncertainty about Pollutant Concentration and Human Exposure Using Geostatistics and a Space-time Information System: Application to Arsenic in Groundwater of Southeast Michigan

Pierre Goovaerts
Biomedware, Inc.
516 North State Street
Ann Arbor, MI 48104, USA
goovaerts@biomedware.com

Abstract. Assessment of the health risks associated with exposure to elevated levels of arsenic in drinking water has become the subject of considerable interest and some controversy in both regulatory and public health communities. The objective of this research is to explore the factors that have contributed to the observed geographic co-clustering in bladder cancer mortality and arsenic concentrations in drinking water in Michigan. The study requires: 1) the building of a probabilistic space-time model of arsenic concentrations, accounting for information collected at private residential wells and the hydrogeochemistry of the area, 2) the estimation of lifetime arsenic exposure, accounting for the impact of job location and occupational exposures to arsenic into the daily water ingestion habit.

Because of the small changes in concentration observed in time, the study has focused on the spatial variability of arsenic, which can be considerable over very short distances. Various geostatistical techniques, based either

on lognormal or indicator transforms of the data to accommodate the highly skewed distribution, have been compared using a cross validation procedure. The most promising approach involves a soft indicator coding of arsenic measurements, which allows one to account for data below the detection limit and the magnitude of measurement errors. Prior probabilities of exceeding various arsenic thresholds are also derived from secondary information, such as type of bedrock and surficial material, well casing depth, using logistic regression.

Computation of human exposure is achieved by modeling each study subject as a spatio-temporally referenced object that moves through space and time. Monte-Carlo simulation is used to propagate the uncertainty about arsenic concentration through the exposure model, leading to an individual model of uncertainty for arsenic exposure

Assessing Progress Towards Environmental Objectives

Anders Grimvall
Department of Mathematics
Linköpings universitet
58183 Linköping, Sweden
angri@mai.liu.se

Abstract. International and national bodies have established a large number of environmental objectives and interim targets. Furthermore, considerable efforts have been made to assess progress towards the established goals. In this paper, we review: (i) the reliability and validity of proposed indicators of progress, and (ii) the appropriateness of the currently used statistical procedures. In particular, we discuss and illustrate how currently used tools can be modified to facilitate the communication between scientists and decision-makers. Taking selected water and air quality objectives as a starting point, we show how the combined use of statistical tools and physics-based models can facilitate the interpretation of observed changes in the state of the environment. Also, we show how data reduction involving monotonic constraints can be employed to meteorologically normalize time series of environmental quality data and thereby clarify the presence of monotonic temporal and spatio-temporal trends.

Setting Environmental Standards: Some Case Studies and a Research Plan

Peter Guttorp
University of Washington
peter@stat.washington.edu

Abstract. The setting of environmental standards by government agencies, such as the US Environmental Protection Agency, are frequently done without taking into account measurement error as well as the statistical quality of the decision rule given. I will outline a statistical approach to the setting of standards, present some examples of such standards (and how they could/should be revised), and finally describe a research plan for how a scientist would go about protecting the health of the people from environmental insults.

Why aren't we making better use of uncertainty information in decision-making?

Kim Lowell
Centre de recherche en géomatique
Pavillon Casault, Université Laval
Québec, Québec G1K 7P4 CANADA
(418) 656-2131 ext. 7998
Fax : (418) 656-7411
Kim.Lowell@scg.ulaval.ca

Abstract. All decisions involve an assessment of potential risk relative to potential reward. In cases where the potential reward is clearly much greater than the potential risk or vice-versa, the course of action to pursue in any decision is very clear. However, if the potential reward and the potential risk are approximately the same, a better assessment of the two should be conducted. In fact, the two should be an integral part of the decision-making and not merely an afterthought.

Since the advent of statistical models, model users have been aware of potential errors in model estimates because of sampling issues associated with data. As the field of spatial uncertainty has developed over the last decade, users of spatial databases are similarly aware of potential errors due to cartographic methods and procedures. However, little work has been conducted to use such information as input into the decision-making process.

This paper will present 1) a discussion of the consequences of not considering model and data uncertainty in the decision-making process, 2) a conceptual model for doing this, and 3) a discussion of when it might not be necessary to do so. These topics will be addressed for spatial as well as aspatial models.

Estimating and modeling space-time variograms

Donald E. Myers
University of Arizona
myers@math.arizona.edu

Abstract. As with a spatial variogram or spatial covariance, a principal purpose of estimating and modeling a space-time variogram is to quantify the spatial temporal dependence reflected in a data set. The resulting model might then be used for spatial interpolation and/or temporal prediction which might take several forms, e.g. kriging and Radial Basis functions. There are significant problems to overcome in both the estimation and the modeling stages for space-time problems unlike the purely spatial application where estimation is the more difficult step. The key point is that a spatial-temporal variogram as a function must be conditionally negative definite (not just semi-definite) which can be a difficult condition to verify in specific cases. In the purely spatial context one relies on a known list of isotropic valid models, e.g., the Matern class as well as the exponential and gaussian models, as well as on positive linear combinations of known valid models. Bochner's Theorem (or the extension given for generalized covariances by Matheron) characterizes non-negative definite functions but does not easily distinguish the strictly positive definite functions.

Geometric anisotropies can be incorporated via an affine transformation and space-time might be viewed as simply a higher dimensional space but possibly with an anisotropy in the model. This approach implies that there is an appropriate and natural choice of a norm (or metric) on space-time analogous to the usual Euclidean

norm for space. The most obvious way to construct a model for space-time is to "separate" the dependence on space and on time. This is not new and in fact a similar problem can occur in spatial application, i.e., a zonal anisotropy. Early attempts used either the sum of two covariances or the sum of two variograms, in either case one component being defined on space and the other on time. It is easily shown that this leads to semi-definite models and hence if used in kriging equations, the result may be a non-invertible coefficient matrix. It is also easy to see that the product of two variograms (even on the same domain) can violate the growth condition. However it is well known that the product of two strictly positive definite functions is again strictly positive definite. In fact a gaussian covariance model might be viewed as product (of several gaussian models each defined on a lower dimensional space). Likewise one form of the exponential covariance, often used in hydrology applications, is also a product. When converted to variogram form, there is not only a product (with a negative sign) but also a sum. It turns out that the variogram form is more convenient in the estimation stage.

The simple product covariance is somewhat too limiting however, each component effectively must have the same "sill". An obvious extension is the product-sum model which when converted to variogram form is the same as for the product (but with different coefficients), This can be further generalized to an integrated product sum model.

In the estimation stage there are two separate problems, one is to determine the appropriate model type and the other is to estimate the model parameters. In a typical spatial application the list of possible models is usually kept small and hence the primary emphasis is on estimating the model parameters. In the spatial temporal context the list of possible models is likely to be much greater and model type selection more difficult. De Ioca, Myers and Posa have shown that the use of marginal variograms is one way to attack this problem and have given an example of extending to the integrated product sum model as well as to the multivariate case using a Linear Coregionalization Model.

The Assessment of the Biodiversity of Nature Reserves: Problems and

Opportunities

Michael W. Palmer
Botany Department
Oklahoma State University
104 LSE
Stillwater OK 74078 USA
carex@okstate.edu

Abstract. The science of ecology has moved away from the belief that natural systems are generally in a state of equilibrium, maintained by biotic interactions. The lack of a modern synthesis hampers our ability to articulate and implement conservation goals. However, given the loss of biodiversity worldwide, we are forced to make decisions based on an incomplete paradigm. Fortunately, some ecological principles seem to be relatively robust: communities are open, dynamic systems governed by spatial and temporal heterogeneity in the environment. The determinants of biodiversity are highly scale-dependent. Methods for assessing biodiversity are most limited when they rely on strict targets, and strongest when they appreciate the dynamic nature of natural communities. While cooperation and communication amongst conservation-related agencies are likely to be fruitful, it is likely that having independent agencies, each with their own targets and priorities, is a net benefit to biodiversity.

Wildfire Threats Count Analysis by Longitudinal Models

J. A. Quintanilha; L. L. Ho
Escola Politecnica – Universidade de São Paulo
Av. Prof. Almeida Prado trav2, n.83
São Paulo SP Brazil
05508-900
jaquinta@usp.br

Abstract. The current operational fire monitoring program conducted by IBAMA (Brazil) has collected data of wildfire threats counts and other explanatory of Amazon region. The aim of this paper is to present the results of statistical analysis of this dataset from 1999 to 2002. From original data, new variables were created. The sample unit was the municipality. The density of wildfire threats count (the ratio between the wildfire threats counts and the municipality area) was selected as a dependent variable. A longitudinal linear model was used and it identified as relevant explanatory variables: administrative limits, municipalities area, year, rain conditions, legal conditions of the areas, percentage of : deforestation, illegal human occupation, population growth index and agricultural area, as also it pointed out different structures of variance in the dependent variable for different type of the legal conditions of the areas. From residual analysis, most of standardized residuals (near 90%) are in the interval (-3, +3). However, some neighborhood municipalities must be considered differently since wildfire threats counts are not associated to any of explanatory variables used in this analysis.

Spatial Scale and Its Effects on Comparisons of Airborne and Ground-based Gamma-ray Spectrometry for Mapping Environmental Radioactivity.

E.M. Scott¹, D.C.W. Sanderson², A.J. Cresswell², J.J. Lang²
¹ Dept of Statistics, University of Glasgow, Glasgow G12 8QW, UK
² Scottish Universities Research and Reactor Centre, East Kilbride G75 0QF, UK
marian@stats.gla.ac.uk

Abstract. Airborne gamma ray spectrometry (AGS) has emerged as an important method for measuring environmental radioactivity (particularly radio caesium (Cs-137)) over wide areas. In early summer of 2002, a number of European AGS teams mapped three large common areas in a European inter-calibration exercise (RESUMÉ 2002). Part of the exercise also involved ground-based teams who conducted soil sampling, in-situ and dose rate measurements in the same areas.

The study design required airborne and ground-based measurements to be taken at three calibration sites (pre-characterised prior to the exercise), three common areas as well as a large composite area. The three calibration sites had 31 sampling points designed in an expanding hexagonal pattern with more than 500 laboratory measurements being made. The three common areas, X, Y and Z were measured by each of the AGS teams and a set of 42 ground based sites (control points) were defined within the common areas and investigated by in-situ gamma spectrometry, soil sampling and dose rate measurement. Analysis of the results focussed on an assessment of comparability of the AGS results and on the agreement between the AGS and ground based results.

The statistical issues in the analysis of the exercise results include the spatial resolution of the measurements made using the different measurement systems and the substantial natural variation compounded with the variation in measurement techniques (and the importance of the spatial (lateral) homogeneity of the source).

Bayesian Kriging and Bayesian Network Design

Richard L. Smith and Zhengyuan Zhu
Department of Statistics and Operations Research
University of North Carolina
Chapel Hill, NC 27599-3260, USA
rls@email.unc.edu

Abstract. One of the classical problems of spatial statistics is this: given a set of observations of a random field on a network of monitors, find a predictive distribution for the values at a location (or many locations) not on the network. The classical geostatistical solution is universal kriging, but the mathematical theory is only exact if the spatial covariance function is known up to a multiplicative constant. For the case where covariance parameters are estimated, various approximate solutions to the prediction problem are known, including Bayesian solutions by MCMC, but these are sometimes criticized on the grounds that their frequentist properties are unknown.

My objective in this talk is to explore these issues from a number of points of view. Laplace approximations are an alternative to MCMC methods for calculating Bayesian inferences; they lead to asymptotic approximations for the coverage probability and the expected length of a Bayesian predictive interval. The latter may be used as a criterion for network design, combining "estimative" and "predictive" criteria used in other recent approaches to network design. These concepts will be illustrated using the EPA PM2.5 network as a test example.

Issues of Spatial Scale in Rural Air Pollution Monitoring and Modelling

R. I. Smith and D. Fowler
Centre for Ecology and Hydrology Edinburgh
Bush Estate, Penicuik, Midlothian EH26 0QB, UK
ris@ceh.ac.uk
E. M. Scott and M. Giannitrapani
University of Glasgow
Glasgow, Scotland, United Kingdom

Abstract. Many applications of environmental policy require the prediction (from models) of levels of pollution at specified locations using current and future scenarios. Belief in the model outputs is largely supported, either explicitly or subjectively, by comparisons of model outputs with measurements. The classic problem is that the models provide output with some level of spatial averaging but measurements are usually from single point locations. Recent exercises in comparing deposition fluxes from a major European model show that the spatial scale component within these comparisons can be extremely important. The local patterns of sources and sinks for the pollutants can result in variability in concentration levels being as large at the local scale (<5km) as at the regional scale (300km), challenging both the interpretation of model output and the

strategy for sampling pollution concentration levels. The paper will focus on how far the measured data and the statistical techniques for comparing model output with measurements provide confidence in the predicted policy outcomes at different spatial scales.

Keywords: air pollution, transport models, sampling

Sensitivity Analysis of a Policy-Based Definition of the Wildland Urban Interface in the United States

Susan I. Stewart
North Central Research Station
USDA Forest Service
Evanston, Illinois, USA
sistewart@fs.fed.us
Roger B. Hammer and Volker C. Radeloff
University of Wisconsin
Madison, Wisconsin, USA

Abstract. Fire policy in the United States went through rapid changes following the fire season of 2000 and continues to evolve, with recent passage of the federal Healthy Forest Restoration Act which focuses hazardous fuels reduction treatments on lands in the wildland urban interface (WUI), where homes and vegetation coincide. In response to these policy developments, we undertook research to create a national map of the WUI, the zone where wildland fire is both difficult to fight and most likely to threaten houses and people. Despite its importance in fire fighting and US fire policy, there is no single, concise, operational definition of the WUI. Hence we began our research by creating a mappable definition of the WUI. Sensitivity analysis was conducted in three of the 48 states mapped; California (the state with most houses in WUI), North Carolina (most WUI area), and New Hampshire (highest percentage WUI area). We altered all parameters of our definition (housing density, vegetation definition and intermix vegetation density, and interface buffer distance). Both land area and number of homes classified as WUI proved most sensitive to the housing density threshold across all states, with a 50% increase (from minimum of 6 HU/km² to 12 HUs/km²) reducing land area classified as WUI by 29% (CA), 41% (NC), and 39% (NH). Results of most other parameter changes were more moderate with one exception: if wildland vegetation is defined only as forest land cover (no shrublands, etc.), the WUI shrinks dramatically in California, encompassing 4.4 million fewer housing units, a 723% drop. Variations in sensitivity to different definition parameters were also noted across states. Sensitivity analyses have implications for use of the definition we created, for alternative nationally-consistent WUI definitions, and for development of region-specific WUI criteria.

Ensemble Kalman Filtering for High-Dimensional Space-Time Models

Jonathan Stroud
The Wharton School
University of Pennsylvania
stroud@galton.uchicago.edu

Abstract. A major challenge in environmental statistics is combining numerical models, satellite observations and ground data. The Kalman filter provides a natural framework for combining all three, but it is difficult to

use in practice because of the high dimension of both the states and observations, and because it requires an accurate representation of model uncertainty. In this talk we develop a new Monte Carlo (ensemble) Kalman filter for state estimation in high-dimensional space-time models. We account for model uncertainty by generating ensembles of forcing variables and parameter values. We describe a hybrid Monte Carlo/variational approach and show how it provides samples from the posterior distribution of the states. A case study of Lake Michigan hydrodynamics is used to illustrate the approach..

Network Design for Prediction with Estimated Parameters

Zhengyuan Zhu
Department of Statistics and Operations Research
University of North Carolina
Chapel Hill, NC 27599-3260, USA
zhuz@email.unc.edu

Abstract. A common problem in spatial statistics is to observe a random process at a set of sample locations, and make inference about the process at unobserved locations. One type of network design problem is about how to choose the sample location network so that one can have the most accurate prediction of the process over the region of interest. We study network design for prediction of stationary isotropic Gaussian random fields using kriging, when the parameters of the covariance function have to be estimated from the same data that are used for prediction. The key issue is how to incorporate the parameter uncertainty into the design criterion, so that the designs which optimize such criterion will give good point predictions as well as accurate estimates of prediction variance. Several possible design criteria are discussed. An simulated annealing algorithm is developed to search for optimal designs of small sample size and a two-step algorithm is proposed for moderately large sample sizes. Simulation results are presented for the Matern class of covariance functions. A test example of optimally reducing an sulfur dioxide monitoring network in east north central US is used to illustrate the method.

Network Design for Semivariogram Estimation, Kriging, and Estimated Kriging

Dale L. Zimmerman
Department of Statistics and Actuarial Science
University of Iowa
Iowa City, IA 52242
dale-zimmerman@uiowa.edu

Abstract. Inferences drawn from spatial data are affected substantially by the spatial configuration of the network of sites where measurements are taken. In this talk, an approach to network design that emphasizes the utility of the network for estimating spatial dependence is contrasted with an approach that emphasizes prediction (kriging) of unobserved responses assuming known spatial dependence. It is shown, via contrived examples, that these design objectives are largely antithetical and thus lead to quite different "optimal" designs. Furthermore, a design approach that emphasizes prediction but also accounts for the sampling variation of spatial dependence parameter estimates is described and illustrated.

Invited Program Abstracts