Invited abstracts of statGIS2007 Conference
24. September – 26. September 2007

Department of Statistics
University of Klagenfurt

Applied computational spatial statistics with R

Roger Bivand

Economic Geography Section, Department of Economics, Norwegian School of Economics and Business Administration, Bergen, Norway.
roger.bivand@nhh.no

Summary. The R programming environment and language for data analysis and statistics attracts many contributions because of its richness and flexibility. The wealth of contributions is however something of a challenge, because contributions sometimes do not work well together, and may be difficult to find. Admitting approaches that do not spring directly from Bayesian principles seems fruitful, because sampling from the posterior distribution of the parameters of a fitted model may also yield much insight. Indeed, sampling from analytically unwieldy problems has a long tradition in point pattern analysis, and both parametric and permutation bootstrap are occasionally used in areal data analysis. Examples will be given using data sets accompanying packages, and an extended example will use garbage data.

Interfacing R and OSGeo projects: status and perspectives

Roger Bivand

Economic Geography Section, Department of Economics, Norwegian School of Economics and Business Administration, Bergen, Norway.
roger.bivand@nhh.no

Summary. Over the four years since the previous StatGIS conference, the handling of spatial data in R has advanced substantially. This has involved the direct interfacing of OSGeo (open source geostatistical) software projects and contributed R packages, chiefly GDAL/OGR, and the implementation of spatial data classes in R (the sp package). A number of architectural choices are available for placing R and GIS software and data storage components. An interface has been written for GRASS 6, which also allows R to be used with the GRASS plugin in QGIS; R can also be used in QGIS via Rpy and Python. Since both GDAL/OGR and GRASS interface databases, spatial data can be moved to and from R in a client/server framework, in addition to desktop operation. In addition, the aRT R-Terralib link will be mentioned briefly.

Spatial Ecology: Putting the Myths into Perspective

Marie.-J. Fortin[1] and Mark R.T. Dale[2]

[1] Department of Ecology and Evolutionary Biology, University of Toronto, Canada
mjfortin@zoo.utoronto.ca
[2] Department of Biological Sciences, University of Alberta, Canada
mark.dale@ualberta.ca

Summary. A major aim of including the spatial component in ecological studies is to characterize the nature and intensity of spatial relationships between organisms and their environment. The growing awareness by ecologists of the importance of including spatial structure in ecological studies (for hypothesis development, experimental design, and statistical analyses) is beneficial because it promotes more effective research, but it has a negative aspect as we face some problems if there is spatial autocorrelation in our data or other problems even if there is none. Here, we put the conceptual and statistical issues concerning spatial autocorrelation into perspective by presenting them as a list of myths and challenges. We describe the problems for incorporating spatial considerations correctly in the analysis of spatially heterogeneous ecological systems but we also outline approaches to solutions to those problems.

Spatial Association Between Speciated Fine Particles and Mortality

Sujit K. Ghosh

Department of Statistics, North Carolina State University, USA
ghosh@stat.ncsu.edu

Summary. Particulate matter (PM) has been linked to a range of serious cardiovascular and respiratory health problems, including premature mortality. The main objective of this research work is to quantify uncertainties about the impacts of fine PM exposure on mortality. A multivariate spatial regression model is developed for the estimation of the risk of mortality associated to fine PM and its components across all counties the coterminous United States. Different sources of uncertainty in the data and model are explored using the spatial structure of the mortality data and the speciated fine PM. A flexible Bayesian hierarchical model is proposed for a space-time series of counts (mortality) by constructing a likelihood-based version of a generalized Poisson regression model that combines methods for point-level misaligned data and change of support regression. The results seem to suggest an increase by a factor of two in the risk of mortality due to fine particles with respect to coarse particles. This study also shows that in the Western United States, the nitrate and crustal components of the speciated fine PM seem to have more impact on mortality than the other components. On the other hand, in the Eastern United States, sulfate and ammonium explain most of the PM fine effect.

On the Importance of Being Spartan

Dionissios T. Hristopulos

Geostatistics Research Group, Department of Mineral Resources Engineering, Technical University of Crete, Greece
dionisi@mred.tuc.gr

Summary. Ancient Sparta was a Greek city state famous for its military prowess and the frugal lifestyle of its citizens. This presentation gives an overview of the current state of development of Spartan spatial random fields. The only connection between Sparta and Spartan spatial random fields (SSRFs) is the hope that the latter capture in their definition the spirit of frugality. The mathematical properties of the SSRFs that distinguish them from classical SRFs will be discussed. Then, the distinct possibilities offered by the Spartan framework for parameter estimation and spatial interpolation will be briefly presented and juxtaposed with the "classical" geostatistical approach, which rests on variogram estimation and kriging. Links to Markov random fields and statistical physics models (i.e., the Ginzburg-Landau model) will be discussed. Some technical issues, such as selection of a suitable step increment for the sample constraints will be highlighted. Case studies that involve real data will be presented. Finally the talk will touch upon open questions and perspectives for future development of the SSRFs.

Big Problems in the Spatial Epidemiology of Cancer: Human Mobility, Empirical Induction Period and Time

Geoffrey M. Jacquez

Department of Environmental Health Sciences, School of Public Health, University of Michigan, USA
jacquez@biomedware.com

Summary. We live in a dynamic and ever-changing world, characterized by daily, weekly and seasonal cycles (e.g. air quality differs from night to day; hospital admissions often peak on Mondays; vegetation changes dramatically from winter to spring), longer term trends (e.g. global warming) and episodic changes (e.g. Katrina). When considering cancer, exposures to potential carcinogens can change through time and vary from place to place, and people move through this exposure milieu on daily, weekly, annual and longer time scales. Until very recently analyses of cancer outcomes such as incidence and mortality have used place of residence at time of diagnosis or death to identify where and when the outcome took place. While this may be useful for assessing proximity to health services (e.g. where screening and treatment facilities should be placed), it can yield a poor and even misleading view of where and when potential causative environmental exposures may have occurred. This talk presents some of the salient obstacles in the spatial epidemiology of cancer, along with some possible solutions at the interface of Geographic Information Science, medical geography and space-time statistics.

Data Processing and Analysis of Uncertainty in Interpolation and Inverse Problems

Peter K. Kitanidis

Civil and Environmental Engineering, Environmental Fluid Mechanics and Hydrology, Stanford University, USA
peterk@stanford.edu

Summary. In interpolation and inverse problems, the objective is to identify a distributed parameter (e.g., one varying in two or three spatial dimensions) from sparse data. Although the true unknown is unique, the data are insufficient to identify a unique answer. Probability theory utilized by an appropriate inference method, such as Bayesian and geostatistical, assigns weights to solutions consistent with the data. In this talk, we discuss opportunities and challenges in stochastic methods for interpolation and inversing, such as: The need for adopting a logical, self-consistent, and honest system of inference; the need to stay "close to the data", by applying cross-validation and other tools that test the theory against the data; the need for practical and efficient computational methods that can solve actual problems without overly restrictive assumptions.

Developing Distance Metrics to Examine Ecological and Geographical Flows

David M. Theobald

Department of Human Dimensions of Natural Resources, Colorado State University, USA
davet@cnr.colostate.edu

Summary. Typically both spatial analysis in GIS and spatial statistics (and geostatistics) rely on simple distance within Euclidean space. Recently, in an effort to shift analysis from form to a concern for process that defines geospatial dynamics, researchers are calling for the computation of distance metrics that better capture the process of interest (e.g., economic distance or hydrologic flow/connectivity). In this talk, we provide examples of developing ecologically relevant measures of hydrologic flow to predict water quality and salmon habitat, and geographically relevant measures of human interaction for a demographic model. Some traditional geostatatistical approaches may be inappropriate because they rely on assumptions of symmetry (in distances) and non-negative, definite covariance matrix. We provide two options to avoid these difficulties.

Optimal Design, Management and Energy Conservation of Environmental Sensor Networks (ESNs)

T. Tsiligiridis

InfoLab, Agricultural University of Athens, Greece
tsili@aua.gr

Summary. Environmental Sensor Networks (ESNs) usually consist of a large number of low-cost, low-power, multifunctional sensor nodes that are small in size and communicate in short distances. One of the most important limitations in their operation is energy conservation. Most of the existing approaches to optimal network topologies for power conservation do not take into account the principles, characteristics and requirements of the specific application that sensor networks are used for. When these factors are considered, then the problem of optimal design and management of ESNs becomes much more complex. This is the reason why several heuristic algorithms, capable of finding good solutions in complex search spaces where conventional analytical techniques may fail, have been used in application-specific WSN designs. In the past two decades, different kinds of optimization algorithms have been designed and applied to solve real-parameter function optimization problems. Some of the popular approaches are real-parameter EAs, evolution strategies (ES), differential evolution (DE), particle swarm optimization (PSO), evolutionary programming (EP), classical methods such as quasi-Newton method (QN), hybrid evolutionary-classical methods, other non-evolutionary methods such as simulated annealing (SA), tabu search (TS) and others. However, in most of these approaches, either very limited network characteristics were considered, or several requirements of the application cases were not incorporated in the performance measure of the algorithm.
Genetic Algorithms (GAs) is one of the most powerful such heuristics. Their successful application in a sensor network design has led to the development

of several other GA-based application-specific approaches in ESN design. More specifically, network design can be investigated in terms of active sensors placement, clustering and signal range of sensors, while performance estimation includes, together with connectivity and energy-related characteristics, some application-specific properties of environmental measurements collection, like uniformity and spatial density of sensing points. A more integrated GA approach will be addressed, both in the direction of degrees of freedom of network characteristics and of application-specific requirements represented in the performance metric of the GA.