Challenges in Developing an Information Management System for Lake Erie—a case study
Neela M. Akhouri, Jignesh M. Patel and Stephen L. Goldman
Lake Erie Research Center, University of Toledo, Oregon, Ohio, USA
Neela.Akhouri@Utoledo.edu

Lake Erie, one of the largest fresh water lakes in the world was moribund in the early 1970’s. With the enactment of the Clean Water Act of 1972 and the Great Lakes Water Quality Agreement between Canada and the United States, water quality management agencies have been successful to a large extent in controlling nutrient and contaminant levels in the ecosystem and a resurgence of fish and wildlife has been evident. The present state of Lake Erie reported by US Environment Protection Agency (2002 LaMP, USEPA) indicates land-use, nutrient enrichment, habitat degradation and loss, pathogen and toxic chemical contamination, and human health still continue to be priority management issues. To understand the various factors influencing sustainability of the lake ecosystem at a larger spatial and temporal scale (e.g. entire length of the river or creek) and to mitigate any detrimental changes, management goals and objectives should be based on the current state of its ecosystem components. To establish these management goals, it is necessary to form a baseline understanding of change over time in the watershed ecosystems. To develop baseline indices; analyze trends, changes, adaptations, and to establish relationships between factors influencing ecological changes, massive sets of data and a tool to “re-purpose” previously-existing data are required. However, historically, ecological data have been collected by a single or small group of investigators in plots of ≤ 1 m² over relatively short periods of time (Kareiva and Anderson, 1988; Brown and Roughgarden, 1990). Therefore, it is not possible to accrue data at large spatial and temporal scales by any one project or program.

For the past several decades, many disjointed programs and policies have existed at the federal, state, and local levels to record and analyze the processes occurring at different spatial and temporal scales in Lake Erie and its drainage basin. Even though the need for availability of extensive data sets (Meyer et al., 1999), and integrated analysis framework (Hardy, 1998) has been established, the daunting task of integration and synthesis of voluminous and complex data (Great Lakes Commission, 2003) has been stymied due to its (1) spatial dispersion, (2) structural and semantic heterogeneity, and (3) lack of detailed information regarding available data and availability of data. Details of these
problems and our approach to finding a solution using technology tools will be presented in this paper.

**Climate and the multi-decadal properties of rainfall and river flow**

W.J.R. Alexander  
Department of Civil and Biosystems Engineering, University of Pretoria, South Africa.  
alexwjr@iafrica.com

A comprehensive database consisting of 11 804 years of South African annual hydrometeorological data from 183 sites was assembled and analysed. The principal conclusion was that there has been a beneficial increase in the mean annual rainfall over South Africa by at least 9% during the 78-year period of record. There was some indication of an acceleration of the increase during the past two decades. It was also established that there was a corresponding increase in open water surface evaporation during this period. It is postulated that the concurrent increase in these two processes is due to naturally occurring global warming. There is also a statistically significant 18 to 22-year periodicity in many South African rainfall and river flow records. It is characterised by sudden reversals from periods of below average conditions to conditions of high rainfall and floods. It is probably related to solar activity. The analyses produced no evidence to support the views that climate change will result in a meaningful reduction in rainfall, or an increase in the frequency or severity of floods and droughts within the next 50 years. All evidence was to the contrary.

**Statistical analysis of extreme floods**

W.J.R. Alexander  
Department of Civil and Biosystems Engineering, University of Pretoria, South Africa  
alexwjr@iafrica.com

In this presentation it is demonstrated that widespread, severe floods are caused by infrequent, but not rare, meteorological phenomena, including tropical cyclones and cut-off low pressure systems. The magnitude of the severe floods relative to the series of annual maximum floods at any one site can be readily determined, but all direct statistical analysis methods seriously under-estimate their frequency of occurrence. This is demonstrated by the statistical analyses of wide area, severe flood-producing rainfall. The statistical analyses confirm that the widely used log Pearson type 3 distribution using conventional moment estimators, remains the preferred method for the statistical analysis of hydrological and meteorological data, but that no direct statistical analysis methods can be used with confidence for return periods exceeding 50 years. It is also demonstrated that there is no justification for making any allowances for future climate change in design flood estimation procedures, as any increases, should they occur, will be undistinguishable against the background of the large natural variability.
Stochastic rainfall generation for rainfall-runoff modelling
Enrica Bellone
University College London
enrica@stts.ucl.ac.uk

Rainfall and evaporation are important inputs to the rainfall-runoff models used for flood design. We overview the stochastic simulation methods being studied at IC/UCL. For single-site rainfall, a wide range of Poisson-based models have been investigated. Simulations from these models may be combined with synthetic evaporation series, derived using a Generalized Linear Model (GLM)-based approach, to obtain the necessary fine scale inputs for runoff generation. A spatial extension of the Poisson-based models (fitted to UK radar data) provides a continuous space - continuous time approach to the rainfall simulation problem. Alternatively, GLMs have been applied to daily raingauge records, where radar data are not available or the underlying hypotheses of the point process models are not satisfied. While the Poisson-based models require both spatial and temporal stationarity, the GLM approach does not involve a stationarity assumption. Combining the two approaches, for regions where both radar and gauge data are available, may be a way to allow spatial and temporal non-stationarities, including climate change scenarios, in continuous simulations.

A Statistician's Take on the Benthic Index of Biotic Integrity
Grace Chiu
Pacific Institute for the Mathematical Sciences (PIMS) and Department of Statistics, University of Washington

Stream health is often measured by the multimetric benthic index of biotic integrity* (B-IBI) developed by Kerans and Karr (1994). B-IBI metrics quantify the well-being of benthic inhabitants of the stream. For the Puget Sound Lowland streams, such metrics include the total number of taxa, the proportion of predatory taxa, etc. Each metric is converted into a score of 1, 3, or 5, where a higher value indicates a healthier stream with respect to the metric. Summing the metric scores yields the B-IBI.

Conventional scoring requires subjective and space/time-dependent input on the cutoff points for the metric-to-score conversion. To a statistician, simple standardization (centering and division by the standard deviation) may be a more natural, non-study-specific conversion. Our more statistically oriented B-IBI (SOBIBI) is the sum of all scores computed this way. Currently, we are comparing the performance of our SOBIBI to the reported properties of the B-IBI via plots and bootstrap simulations.

In this talk, I will introduce the background of the project, discuss our current progress, and invite ideas from the audience.

A Quantitative Decision Support Instrument for Water Quality Management
Marius Claassen, Alta de Waal, Xolani Hadebe and Thabo Sekonyela
The promulgation of the National Water Act (NWA), 1998 (Act No. 36 of 1998) and various other acts, as well as the publication of various policies and White Papers (such as Water Supply and Sanitation, 1994, etc.) have given a new direction to water resources management, and specifically Water Quality Management (WQM), in South Africa. The purpose of the NWA is to ensure that the South Africa’s water resources are protected, used, developed, conserved, managed and controlled in an equitable, efficient and sustainable manner. This necessitates a change in the approach of WQM to an integrated source, resource and remediation focused management approach.

Currently no quantitative technique exists to assist licence officers in making decisions regarding the allocation of water licences. Licence officers are frustrated because of the lack of structure and support that currently exists to alleviate their tasks. To make this integrated approach operational, management instruments are needed for regional offices to include resource-directed water quality issues into licence allocations. This calls for a decision support tool that will guide regional offices in making important licence allocation decisions. The decision support tool must be able to facilitate a decision despite incomplete, imprecise, and highly variable data. Also, the decision needs to be based on multiple criteria such as socio-economic factors, race and gender considerations, alignment with catchment strategy, etc.

We suggest that Bayesian networks provide a possible solution to this problem. Bayesian networks have proven to be an extremely powerful technique for reasoning under uncertainty and the graphical structure explicitly represents cause-and-effect assumptions between variables. In this article we explain the role of Bayesian networks in licence allocation decision making: The procedure consists of identifying the indicators and considerations relevant to WQM. The Bayesian network links the indicators and considerations and enables us to calculate a value (within some probability bounds) for each consideration, given the values of indicators. These values are then used to calculate the probability of allocating the licence (or not). This methodology allows decision support, even under circumstances of severe uncertainty and lack of information.

**Pesticides risk assessment of Goulburn-Murray Water's (G-Mw) irrigation supply channels, Victoria, Australia**

Ray Correll\(^1\), Mary Barnes\(^1\), Rai Kookana\(^1\), Golam Kibria\(^2\)*, Peter Butcher\(^2\)

CSIRO Contaminant Interactions & Remediation, Urrbrae, Adelaide, Australia\(^1\)

Goulburn-Murray Rural Water Authority, Tatura, Victoria, Australia\(^2\)

Goulburn Murray Rural Water Authority (G-Mw) is a major rural water authority in Australia. It supplies water for irrigation, domestic and stock drinking and aquaculture. A preliminary survey (2001) found than more than 70 pesticides were used as herbicides, insecticides, and fungicides. G-MW suspected that
water supply channels contaminated with pesticides can be unfit for human consumption, stock drinking, irrigation, food processing and aquaculture and could have an adverse impact on aquatic biota (including native fish) living in natural waterways.

A first tier (1\textsuperscript{st}) assessment was made of the risk from pesticides to water quality and through water quality to human, stock & domestic supply, food industry, pastures, aquaculture and aquatic ecosystems. Risk was measured using various exposure pathways to channels (drift, accidental spills, unlawful acts, and drainage discharge) to different receptors (human, stock, domestic use, food processing industry, aquaculture, aquatic ecosystem, pastures, and crops).

Various recommendations were made to reduce risks to water supply channels. The recommendations include monitoring of pesticides in fruit, vegetables and leguminous pastures and aquaculture production sites, the use of best-management practices, and a new survey to obtain pesticides uses data in the new and expanding farming zones (not covered in the 1\textsuperscript{st} tier assessment) and obtaining of further ecotoxicological information of harmful pesticides.

### A methodology for seismic risk assessment with an application to the insurance industry

Nicholas Davies and Andrzej Kijko
Hannover Re Africa Ltd, Johannesburg and Council for Geoscience, Private Bag X112, Pretoria 0001, South Africa
Kijko@geoscience.org.za

The study concentrates on a methodology for probabilistic seismic hazard and seismic risk assessment. It starts with an introduction and a historical perspective on the estimation of seismic damage to buildings. The method for the estimation of expected damage from a probabilistic point of view is then presented. The work closes with an application of the described methodology to several sites around South Africa.

### Using the Worsley likelihood ratio test to define changes in longitudinal profiles of rivers

ESJ Dollar\textsuperscript{1} and LH Dollar\textsuperscript{2}
CSIR Environmentek, Pretoria\textsuperscript{1} and School of Civil and Environmental Engineering, University of the Witwatersrand\textsuperscript{2}

The questions of ‘which’ and ‘how many’ rivers need to be conserved in South Africa in order to meet the objectives of sustainability and use, and in order to protect ‘river biodiversity’ remains unanswered. The South African National Water Act (No. 36 of 1998) makes provision for the classification of water resources in order to achieve a balance between protection and use. Currently, a multi-disciplinary project aiming to answer the aforementioned questions is underway. One of the objectives of this project is to divide the longitudinal profiles of South African rivers in ‘macro-reaches’. Macro-reaches are defined as stretches of river of variable length within which inherited controls (such as lithology, structure, tectonics and so on) are sufficiently similar to result in a
uniform channel type. This has been achieved for a small number of South African rivers, as this process is time consuming and data intensive. In this paper we explore the possibility of using a method (Worsley, 1979) for identifying significant changes in river longitudinal profiles that can be applied to all rivers in South Africa.

Worsley’s method (1979) defines a single change point in a set of data. This was applied sequentially to all data points known along a river profile in order to identify multiple change points. Change point positions derived from the statistical method were compared to longitudinal profiles that had previously been sub-divided using survey data, geological maps, aerial photographs and video footage. An assumption was that if the statistical method could identify the same change points, then the method can be applied with reasonable certainty to rivers for which detailed assessments have not been performed. An example of the Crocodile River is presented. It was found that the usefulness of the method depended on assumptions made to apply the Worsley likelihood ratio. Various questions are raised to determine whether there are more appropriate methods of finding change points and how the data set should be sub-divided to provide statistically valid answers that are also useful to decision makers.

**Modified likelihood function with environmental applications**

Abdel H El-Shaarawi

National Water Research Institute, Burlington, Ontario, Canada and Department of Chemical Engineering, University of Genoa, Italy

Abdel.el-shaarawi@cciw.ca

Inferences about a vector of parameters $\theta$ using the likelihood function $L(\theta, x)$ based on independent observations $x$, remain unchanged under permutations of the elements of $x$. This fact leads to a simple approximation $L'(\theta, x)$ of $L(\theta, x)$ using order statistics. $L'(\theta, x)$ produces a closed form estimates for the parameters of the location and scale family of distributions and a single non-linear equation for estimating the shape parameter. $L'(\theta, x)$ is then used to fit the generalized extreme value distribution to several environmental data sets. Simulation results will also be presented to study the performance of the method.

**Modeling the accumulation of contaminants in the aquatic environment**

Abdel H. El-Shaarawi and Abdalla ElbergaI

National Water Research Institute, Burlington, Ontario, Canada

abdel.el-shaarawi@cciw.ca

The main dynamic features of the growth of an organism or a community are usually modeled as a set of deterministic differential equations. Models of this type do not account for natural environmental and or biological variability. This paper discusses how to account for the variability by the including stochastic components in such models and how to use the Fokker-Planck or diffusion equation to compute the moments of the associated random process. Expressions for the first and second order moments are required for the
application of quasi likelihood method to make inferences about the process parameters. The results of using this approach to model the accumulation of contaminants in the tissue of fish from the Great Lakes will be presented and discussed.

**Modeling particulate matter data in sparse monitoring networks**

A. Fasso and O. Nicolis  
University of Bergamo, Italy  
fasso@unibg.it

In many countries, particulate matters are being monitored only recently. Since these monitoring networks are often sparsely distributed over the land, it is of interest to assess spatio-temporal correlation between particulate matters and other quantities. In this talk, some models describing the relation between PM10 and Oxide Nitrate are developed and compared using data from North Italy.

**Modeling waterborne infectious outbreaks: when, where, and how bad will they be?**  
Nina H. Fefferman and Elena N. Naumova  
Tufts University

We offer mathematically rigorous definitions of previously ambiguous epidemiological concepts of waterborne infections. These definitions include notions of environmental exposure, diseases temporal patterns, and outbreak signatures. We define an outbreak as an occurrence of either of the following: the probability of exposure is more than one standard deviation higher than the norm in at least one location, or else the size of the population open to possible exposure to infection at a given time is more than one standard deviation higher than the norm. By using the specific properties of the disease as parameters to generate a disease temporal pattern, we portray an outbreak as a composition of these patterns that form a unique outbreak signature. Based on the proposed definitions, we derive a sequential combinatorial model of symptomatic manifestation of disease outbreaks. By considering the observed number of new cases per unit time as a composite function of the probabilities of exposure, infection, and the timing of the onset of symptoms, we are able to differentiate between spatial and temporal spreads of sustained exposure. Finally, we use the demographic distribution of affected populations in order to further decompose an outbreak signature. We demonstrate our model performance using a simulated hypothetical outbreak. Using ten years of Massachusetts’s surveillance data on laboratory-confirmed cryptosporidiosis we compile a set of parameters essential for modeling. Then, using a proposed model we attempt to describe a suspected outbreak of cryptosporidiosis occurred in Worcester, MA in 1995. The proposed modeling procedure allow us (1) to decompose an outbreak signature and differentiate between different disease spread scenarios, (2) to predict endemic Poisson-like fluctuations prior to an outbreak, and (3) to examine
the likely reporting errors associated with all phases of an outbreak manifestation.

Towards a more integrated approach to the management of land, water and external inputs at a catchment scale

W Enkerlin, A Fajgelj, IG Ferris, L Gourcy, K Gross, LK Heng, L Foglund, O Perera, J Turner, G Voigt, F Zapata

Food and Environmental Protection Section, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture; Soil and Water Management and Crop Nutrition Section, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture; and Soil Science Unit, Agency’s Laboratories, International Atomic Energy Agency

I.Ferris@iaea.org, F.Zapata@iaea.org and L.K.Heng@iaea.org

One of the key challenges for sustainable intensification of agricultural production is to ensure sufficient, safe and nutritious food production for an ever growing population on limited land and water resources while promoting natural resource conservation (Lal, 2000; Walling, 2001). The recently held World Summit on Sustainable Development (WSSD), September 2002 reaffirmed land degradation as one of the major global environment and sustainable development challenges of the 21st Century. Agricultural lands have been more severely affected by degradation processes. About one third of agricultural land has been degraded during the last 50 years. Land degradation affects soil quality through several processes including water or wind erosion, waterlogging, salinization, acidification, soil compaction, nutrient mining and soil organic matter (organic carbon) depletion. The main causes of land degradation are inappropriate land use, unsustainable farming practices, deforestation, and overgrazing. Another serious challenge, albeit less visible, is environmental pollution that degrades soil and water quality (Lal, 2000; UNEP, 2000). Pesticides are often perceived as a cause of groundwater and surface water pollution. Both challenges require an integrated approach at the landscape level and linkages between production and monitoring activities (OECD, 2001). These integrated studies must be planned and implemented by multi-disciplinary teams with a range of knowledge and expertise that often is found in several institutions (Chalk et al., 2002). To change current practices it is imperative that all stakeholders, including the decision and policy makers (farmers and regulatory officials) are involved at the onset (OECD, 2001).

The South African Rainfall Enhancement Programmes: An overview

Lizelle Fletcher
Department of Statistics, University of South Africa

The Water Research Commission (WRC) came into being in 1971 to address the management and utilisation of South Africa’s meagre water resources in a sustainable and responsible manner. Cloud seeding research in this country,
focusing on rainfall enhancement, was thus driven by the WRC and the SA Weather Bureau (SAWB) from the early 1980s onwards. The Bethlehem Precipitation Research Project (BPRP), conducted from 1984 to 1989, was an exploratory experiment initiated with the primary objective to investigate natural and artificially modified precipitation processes within summertime convective clouds. Simultaneously, a similar cloud-seeding experiment, the Program for Atmospheric Water Supply (PAWS), was independently conducted by a private company CloudQuest at Carolina in the eastern part of the country. The National Precipitation Research Programme (NPRP) followed on the BPRP, and was the result of the amalgamation in 1990 of the above two projects, combining the efforts of the Bethlehem meteorologists and the Carolina researchers. This project came to an abrupt end in April 1997, at the height of progress and success, when the then Minister of Water Affairs and Forestry withdrew the WRC’s funding for rainfall enhancement research in South Africa. It was during this period that a novel hygroscopic seeding technique was developed which was successfully implemented not only in South Africa, but also in Mexico, under the auspices of the US National Center for Atmospheric Research (NCAR) as well as in Thailand. Early in 1995 the Northern Province government approached the WRC and the SAWB to employ cloud seeding at very short notice as an emergency response to drought in the province. This operational project ran until 1997. The South African Rainfall Enhancement Program (SAREP) proposal was drafted in the second half of 1997. This semi-operational project was conducted from December 1997, building on the previous work of the NPRP. Funding for this project was unfortunately unstable, leading to its demise in April 2001.

**Estimating space-time trends combining observations with output of numerical models**

Montserrat Fuentes  
Statistics Department  
North Carolina State University

Estimating spatial temporal trends of air pollution levels is vital for air quality management, and presents statistical problems typical of many environmental and spatial applications. Ideally, such trends would be based on a dense network of monitoring stations, but this does not always exist. Instead, there are generally two main sources of information about pollution levels: one is pollution measurements at a sparse set of monitoring stations and the other is the output of the regional scale air quality models. Here we develop formal methods for combining sources of information with different spatial resolutions for space-time trend estimation. We formulate this problem using a Hierarchical model, in which scientific information and output of numerical models is introduced to improve the trend estimation. We also offer a review of the current literature and approaches that use output of numerical models as a prior for trend estimation.
We present applications to air quality and wind fields trend estimation using output of numerical air quality models and weather prediction numerical models.

**Modeling Several Levels of Uncertainty in Hearing Threshold Data**
Byron J. Gajewski
University of Kansas
bgajewski@kumc.edu

According to the world health organization (WHO) and many other organizations, noise pollution is a very important environmental problem. Additionally, the WHO also states that there is not enough knowledge of the effect of the “dose-response” relationship of noise pollution control on humans. To better understand human hearing loss, for example, good statistical methods for analyzing this type of data are necessary. Over the past few years we have been involved in analyzing data sets using hearing threshold as the primary response variable. In general, the distribution of hearing thresholds \( y \) tend to have heavy tails and be more peaked than normal or log-normal distributions. A typical remedy to this problem is to transform the hearing threshold with the natural logarithm of the response plus a parameter, or \( \ln(y + \alpha) \), where alpha is an unknown parameter. In the past, some researchers have set alpha to 20, a priori, while others estimate alpha with maximum likelihood estimation. We use a Bayesian approach to estimate \( \alpha \).

Additionally, we wish to obtain the probability of a 15 decibel (db) drop in subjects’ hearing threshold over time. Furthermore, as a nuisance, the data typically suffers from right censoring, grouping and missing observations. Therefore, we desire a method that simultaneously accounts for the uncertainty in: (1) usual parameters associated with a parametric model; (2) an added parameter, alpha (3) the transformation to a probability measure and (4) the nuisance tendencies of the data. To account for the list of uncertainties, we formulate a Bayesian method. We then demonstrate the method with a case study. One can extend the method for use on many types of hearing threshold studies.

**Isotonic regression for the normalisation of environmental quality data**
Anders Grimvall
Department of Mathematics, Linköping University, SE-58183 Linköping, Sweden
angri@mai.liu.se

Trends in environmental quality can emerge more clearly if the collected data are normalised by removing meteorologically driven fluctuations and other forms of natural variation. The most widely used normalisation procedures are based on linear or non-linear regression models. In this paper, we discuss how we can incorporate prior knowledge about monotone response to one or more explanatory variables. In particular, we propose a technique that facilitates estimation of monotone temporal trends in the presence of seasonal variation and one or more covariates. The basic idea is simple. First, the seasonal pattern
is decomposed into increasing and decreasing components prevailing under different parts of the year, and then an algorithm for isotonic regression is employed to extract a monotone temporal trend. Because existing algorithms for least squares isotonic regression can be used only for problems involving a small number of explanatory variables we propose a computationally simpler two-step procedure in which a smooth function is first fitted to data and then made monotone. More precisely, we suggest that a kernel smoother is used to produce a response surface for a grid of values of the explanatory variables and that a simple averaging algorithm developed by Mukerjee and Stern (1994) and improved by Strand (2003) is then used to produce an increasing or decreasing response surface. The performance of our method is tested on water and air quality data collected over several seasons.

Using wavelet tools to estimate and assess trends in atmospheric data
Peter Guttorp
University of Washington, USA

Recently, wavelet methods have been applied in a variety of geophysical time series. I will describe some of these tools (the modified wavelet transform, multiscale decomposition, wavelet variance, etc) and show how they can be used to estimate long-term temporal structure with attendant standard errors. The methodology will be illustrated in two contexts: estimation of hemispherical mean temperature, and an experiment measuring atmospheric turbulence. Possible extensions to space-time data will also be discussed.

Developing ecosystem health monitoring programs for rivers and streams
Bronwyn D. Harch, Maree O’Sullivan, Ross Sparks
CSIRO Mathematical and Information Sciences
CMIS, Qld Biobscience Precinct, 306 Carmody Road, St Lucia 4067, Australia
Bronwyn.Harch@csiro.au

During the last decade, the long-term ecological health of Australian rivers and streams has emerged as one of the biggest national natural resource management objectives. The National Water Quality Management Strategy forms a national program in Australia to achieve ecologically sustainable use of water resources by protecting and enhancing their quality, while maintaining economic and social development.

The main focus for monitoring the ‘health’ of rivers and streams has progressed from only measuring water chemistry attributes to measuring aspects of ecosystem health – chemical, physical and biological. Ecosystem health for freshwaters is now increasingly being diagnosed using aquatic macro-invertebrates (water bugs), fish, frogs, nutrients (forms of N and P), algae, macrophytes, the production and consumption of organic carbon, sediment bacteria and the more traditional water chemistry parameters. Examples of our involvement with the development of ecosystem health programs in South-East Qld, South-West Qld and the Sydney area will be presented with an emphasis on the contributions made by statistics. Aspects of
Kernel Estimates of Hazard Functions and Application
Ivana Horova, Jiri Zelinka and Marie Budikova

In recent years, considerable attention has been paid to methods for analyzing
data on events over time and to the study of factors associated with recurrence
rates for these events.
In summarizing survival data, there are two functions of central interest, namely
the survival and the hazard functions. The well-known product-limit estimator of
the survival function was proposed by Kaplan and Meier in the year 1958. A
single sample of survival data may also be summarized through the hazard
function. We focus on nonparametric estimates of the hazard function and its
derivatives under random censoring based on kernel estimate of the Nelson
estimator of the cumulative hazard function. Methods of kernel estimates
represent one of the most effective nonparametric methods. An automatic
procedure for the simultaneous choice of the parameters of kernel estimates is
applied to the estimate of hazard functions for two carcinoma data sets kindly
provided by the Masaryk Memorial Cancer Institute in Brno. The attention is also
paid to the points of the most rapid change of these hazard functions.

A Class of Linear Space-Time Models Applied to Estimation of Fish
Populations
Gudmund Høst
Norwegian Computing Centre, Norway
gudmund.host@nr.no

Joint modeling of population dynamics in space and time may be important when
estimating population abundance. In estimating abundance of marine
populations, both the data collection and the population may have space-time
characteristics. We will present a class of models for population density derived
from a linear partial differential equation with random forcing. We show that the
model has an equivalent and more familiar representation as smoothing of
uncorrelated noise.
We estimate parameters by Markov Chain Monte Carlo (MCMC) and apply the
model to a herring population in Norway.

The power of the test in analysis of variance of Poisson distributed
variables
Zuzana Hrdlièkova
Masaryk University Brno,Czech Republic
Department of Applied Mathematics
zuzka@math.muni.cz
The power of the test for the linear hypothesis testing in the classical ANOVA model with the normal error can be found by means of the noncentral F distribution. The knowledge of the power of the test is very useful for planning biometrical experiments.

If the assumptions of the classical linear model are disturbed then the generalized linear models are often used. The tests are based on the Pearson's chi^2 statistics and the deviance or scaled deviance. In such situation it is in general very complicated to derive exact power of the test. In some situation the asymptotic power can be used. Therefore in the paper we focus on the ANOVA models in situation when error terms have Poisson distribution. We aim at calculating the power of the test for particular values of parameters by means of simulation. Obtained results are applied to environmental data.

**Diagnosing Hydroclimatic Variations and Change based on a Quantile Regression framework**

Shaleen Jain  
NOAA-CIRES Climate Diagnostics Center and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80305-3328  
Shaleen.Jain@noaa.gov

Decadal and longer-term variations in regional hydroclimate directly impact the sustainability of water-dependent natural systems (e.g., stream ecosystems), as well as the planning and operations of water-related infrastructure (dams, reservoirs, levees, and water supply systems). Diagnosis of the variations and change from observational data is often made difficult due to the presence of nonstationarities, outliers, and departures from normality. Quantile regression methodologies, employed on full and moving-window subsets of the data largely overcome these problems, and allow a robust characterization of the sensitivity the full empirical probability distribution. Results are presented for an analysis framework to diagnose long-term variations across various quantiles for the major river systems along the North American west coast. Preliminary results are also presented for extending the diagnosis framework to a multivariate setting. Finally, we discuss the relevance of the framework toward the characterization of climate-related vulnerability of water resources.

**Seismological Outliers: L1 or Adaptive Lp-Norm Application**

Andrzej Kijko  
Council for Geoscience, Private Bag X112, Pretoria 0001, South Africa  
Kijko@geoscience.org.za

In the early 1930's Jeffreys recognized that the errors in determination of arrival times of seismic waves do not follow the Gaussian distribution required by least squares ($L_2$-norm) location techniques. Large errors in the arrival times can displace the least-squares estimate of the earthquake hypocenter far away from the true location. As an alternative to the least-squares procedure, the minimization of the sum of absolute residuals ($L_1$-norm), as protection against the
effects of outliers, is often used. In this note a modification of inversion procedures is proposed which incorporates an adaptive algorithm for $L_p$-norm estimation. A detailed study by numerical simulation demonstrates that in the presence of outlying observations, an $L_p$-norm procedure can select the proper value of $p$, which may not necessarily equal 1. The advantage of this procedure is that no apriori decision to use $L_1$-norm is to be made, but instead the data prescribes the most appropriate value of $p$.

In on-line seismic networks where the system is fully automated, and where there is no chance for manual intervention, it is suggested that the adaptive $L_p$-norm offers the most reliable method for data processing.

Keywords: Outliers, adaptive $L_p$ norm

**Probability Modeling of the Water-Quality: Assessment in Mid-Atlantic Region**

M. Liu$^1$, N. K. Neerchal$^1$, E. A. Greene$^2$, and A. E. LaMotte$^2$

Department of Mathematics and Statistics, UMBC, Baltimore, MD 21250, and United States Geological Survey, Baltimore, MD 21237

Nitrate concentration level is an important indicator of water quality. Thus it is very useful to relate the levels of nitrate concentrations in a watershed to various indicators of stress, such as fertilizer applications, percentage of agricultural land in the watershed. In this project, existing nitrate concentration data from the U. S. Geological Survey (USGS) National Water Quality Assessment (NAWQA) were used in conjunction with geographic data to develop logistic regression equations to predict the presence of elevated levels of nitrate concentration. The resulting logistic-regression equations were transformed to determine the probability of nitrate concentration exceeding a threshold relating to a specific designated use. Generalized Estimating Equations methodology is used to estimate models that account for intra-watershed correlation. Statistical tests are also performed to investigate possible overdispersion resulting from intra-watershed correlation. Even though this model was developed for nitrate contamination, the method can be used for other water-quality parameters that may affect human or environmental health.

**Functional data analysis methods in the environmental sciences**

Wendy Meiring

University of California, Santa Barbara

Functional data are observations from curves or surfaces. Many environmental data sets may be considered as functional data. For example a balloon-based ozonesonde measures the ozone partial pressure profile as the sonde ascends through the atmosphere. Each launch time provides observations from one ozone partial pressure profile as a function of altitude. A sequence of sonde launches provides observations from a time series of ozone profile "curves", with each curve a function of altitude. The shape of these curves evolves over time in response to complex dynamical and chemical processes. We present functional data analysis methodology to estimate altitude-dependent non-linear time trends.
and other space-time modes of variability. We illustrate throughout with a study of altitude-dependent variation in stratospheric ozone partial pressures over Hohenpeissenberg in Germany, based on a time series of ozonesonde flights. We illustrate the value of functional data analysis methods for studying altitude dependent non-linear time trends and ozone variation related to the Quasi-Biennial Oscillation. Due to the large number of observations, our analysis combines dimension reducing functional basis approximations, with flexible additive models on the low-dimensional basis function coefficient scale. We fit additive coefficient models composed of cubic and periodic splines, as special cases of smoothing spline anova models (SSANOVA). We provide preliminary estimates of uncertainty of altitude dependent QBO effects and non-linear time trends, building on SSANOVA Bayesian confidence intervals. We discuss the importance of non-linear time trends to study ozone recovery as stratospheric chlorine levels decrease. We indicate continued improvements of the uncertainty measures as well as model extensions for additive and interactive effects of additional atmospheric explanatory variables.

Analysis of Multidimensional Discrete data with Rasch type Models. Application to the WHO Housing and Health Survey.

Mounir Mesbah
Université de Bretagne-Sud, Vannes, France.
mounir.mesbah@univ-ubs.fr

Most general population surveys, nowadays, include jointly with measurement of various manifest observed variables, a set of questions giving categorical ordinal responses. Each set is inserted to measure a specific latent trait, i.e. an unobserved causal variable. Rasch type models relate an unobserved latent continuous variable (ability) to manifest observed ordinal unidimensional items. This model, belonging to the item response theory family, was first developed in the psychometry fields to deal with measurement of univariate trait of student in educational testing. In our work, we will use graphical and Rasch type models, to show how one can construct valid multivariate distribution to deal with such complex situations. Then, we present data from a large Pan-European Survey, the “Housing and Health Survey” including various set of questions measuring Quality of Life, Immediate Environment, Health, Housing conditions and other related indicators. Finally, using this large data set, we will show how to build and validate a complex causal model predicting Health Related Quality of Life by housing condition variables.

References:
J.B. Hardouin and M.Mesbah. “Clustering binary variables in subscales using an extended Rasch model and Akaike information criterion” Submitted.
Statistical analysis of data with negative binomial distribution
Jaroslav Michalek,
Masaryk University Brno, Czech Republic

The data from the population with negative binomial distribution (NBD) are very frequent in environmental applications. The NBD can be considered as a mixture of Poisson distributions with a Gamma distribution of the means. The distribution has two parameters, $\mu$ and $\kappa$, here $\mu$ is an expectation and the variance of the distribution is given by $\mu + \mu^2/\kappa$. Thus the parameter $\kappa$ is related to the so called over-dispersion of the Poisson distribution.

If $\kappa$ is known, the standard GLM technique can be used. If $\kappa$ is unknown, it can be estimated. The data analysis is then often performed in the same way as in the case where $\kappa$ is known. The situation is more complicated if $\kappa$ is unknown and the estimate of $\mu$ depends on covariates. In that case, the general method of maximum likelihood can be applied and asymptotic results can be used only if the sample size is large enough.

In this contribution simulated data will be used to demonstrate different approaches of estimating and testing parameters of the NBD. The robustness of the applied methods will be compared. The results of a statistical analysis of ecological data that used with NBD will be presented in the second part of the paper. The data which we use were obtained from an ecological study oriented to verifying the hypothesis that the distribution of large herbivores in a forests depends on the presence of a shrub layer.

Estimating rainfall in Zimbabwe
Kingstone Mutsonziwa
School of Statistics and Actuarial Science, University of the Witwatersrand

Annual data from 1953-1993 in Zimbabwe were used to extrapolate rain gauge readings at each of the stations considered to estimate the amount of rainfall received in areas where no readings were made using Kriging. Rainfall in Zimbabwe is correlated with latitude and longitude with rainfall being approximately constant in the NW-SE direction. The areas in the NE direction receive substantial amounts of rainfall in comparison to the rest of the areas in Zimbabwe. Some limitations in the analysis will be discussed and some recommendations are suggested.

Evaluation of effects of environmental temperature on seasonality of infections
Elena N. Naumova, PhD and Jyotsna Jagai, MS
Tufts University School of Medicine
To explore long-term, mid-term, and short-term effects of environmental stressors, and extreme events in particular, on incidence of infectious diseases, analytical tools have to be adapted to the specific properties of a health outcome. Infectious diseases are typically manifested via periods with low incidence alternated by periods of outbreak clusters, which form a unique seasonal pattern. Each infectious disease has specific clinical properties, such as incubation time, and dominant routes of exposure/transmission, that can be influenced by environmental factors. While climatic conditions typically define or restrict a habitat area of a pathogen, meteorological factors affect timing and intensity of infectious outbreaks. Extreme weather events may have pronounced immediate effects, induce a shift in seasonal pattern, and have long-term consequences.

The objective of this study was to develop methodology for assessing long-term, mid-term, and short-term effects of ambient temperature on a temporal pattern of enteric infections (EI). We discussed the notion of seasonality and approaches to parametric modeling of a seasonal pattern. We suggest a two-stage hierarchical regression modeling procedure. The first stage provided the estimates for seasonal characteristics, which are used in the second stage to estimate the degree of associations among those characteristics.

Using 45,816 records of reported cases of six enteric infections collected over a ten-year period (1992-2001) in Massachusetts, we described diseases temporal patterns. We linked these temporal patterns to daily measures of ambient temperature abstracted from databases of the National Oceanic and Atmospheric Administration. All EI, except one, exhibited well-defined seasonal patterns. Two diseases were almost mimicking the temperature seasonal curve; and in three EI we observed significant delays in peak timing relative to a peak in ambient temperature. Extreme temperature predicted the intensity of a seasonal increase and the timing of seasonal peak in waterborne infections. Foodborne infection incidence was the most sensitive to the short-term temperature fluctuations, significantly rising on 2, 8 and 15 days after an increase in temperature. Our findings offer insights to the possible dominant routes of transmission for these diseases.

We suggest that weather forecast can be considered for forecasting of EI, so public health measures to prevent these diseases can be better targeted and focused.

**MFWIS: A Surveillance System For Waterborne Infections**

Elena N. Naumova and Ian B. MacNeill

Tufts University School of Medicine, Boston, MA 02111 and Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada N6A 5B7

elena.naumova@tufts.edu and ibmacneill@hotmail.com

MFWIS is a system for monitoring case count data on infections with a view to early detection of outbreaks and to forecasting the extent of detected outbreaks. Historical data are smoothed using a loess-type smoother. Upon receipt of a new datum, the smoothing is updated and estimates are made of the first two
derivatives of the smooth curve and these are used for near-term forecasting. Recent data and the near-term forecasts are used to compute a warning index. The algorithm for computing the warning index and the interpretation of the index have been designed to effect a balance between type I errors (false prediction of an epidemic) and type II errors (failure to correctly predict an epidemic). If the warning index signals a sufficiently high probability of an epidemic, then a forecast of the possible size of the outbreak is made. This longer term forecast is made by fitting a “signature” curve to the available data. The effectiveness of the forecast depends upon the extent to which the signature curve captures the shape of outbreaks of the infection under consideration.

Crossing Problems
Jiri Neubauer
University of Ostrava
neubauer@kmi.vvs-pv.cz

Observing some random process it seems to be interesting to know the probability that the process is crossing any level during a particular time interval. Another thing which could be useful to know is time the process spends above a given level. This contribution is dedicated to the crossing problem, to the question of mean number of points the process crosses the given level and to the question of time spent above the level. It was necessary to use the spectral analysis of time series, particularly the estimation of a spectral density function, to reach the solution of these issues. Results will be presented on simulated time series and on time series of some air pollutants, which were measured in the town of Vyskov in Czech Republic.

Power of non-parametric trend tests – a semi-parametric approach
Anders Nordgaard
Linköping University, Sweden
annor@mai.liu.se

Non-parametric tests for monotone trends in environmental quality data have become widespread during the past decades. The basic ideas were formulated by Mann (1945) and Kendall (1975). More specific features, such as the handling of serial correlation and adjustments involving covariates, have been discussed by Hirsch and Slack (1984), Libiseller and Grimvall (2002), and several other authors. However, despite the popularity of non-parametric trend tests, there is no generally accepted procedure to calculate the power of such tests. In this paper, we develop a semi-parametric method for such calculations. The method is based on simulations involving resampling from measured data. It is non-parametric in the sense that it does not require any specific model of the random variation in collected data. However, it is parametric in the sense the trend scenarios are expressed as parametric functions. The method is applied on measured concentrations of nutrients in Swedish rivers.
Keywords: Time series, Mann-Kendall test, power function, resampling, bootstrap

References

Practical issues and tools for modeling temporal and spatio-temporal trends in atmospheric pollutant monitoring data
Paul D. Sampson
University of Washington

There is a substantial literature on methods for modeling trends in space-time monitoring data of atmospheric pollutants. The choice of approach will reasonably depend on the spatio-temporal scales of the monitoring data as well as the scientific aims of the analysis, such as testing of long-term regional trends, computing various metrics of long-term exposure in chronic health effect models, or utilizing spatio-temporal trends in the computation of spatial estimates of exposure, for both long-term and acute health effects analyses. Any such modeling and analysis must consider the common issues of temporal and spatial correlation. We discuss our experience in assessing long-term trends in tropospheric ozone and in modeling spatially varying seasonal structure for purposes of spatial estimation. In the latter we introduce an illustrate a simple but flexible approach to modeling spatio-temporally varying seasonality and long-term trend in terms of basis functions derived from a singular value decomposition of the space x time data matrix.

Air pollution and short-term health effect studies in Kuwait
Lukman Thalib 1, Rashid Al Rashood 2 and Philip W Moody 1.

1Department of Community Medicine, Faculty of Medicine, Kuwait and 2EPA, Kuwait

A number of recent studies have found that relatively low levels of air pollution, levels below current air pollution guidelines are associated with adverse health effects including daily death counts and daily hospital admissions. Although, the risk associated with air pollution, however, small can no longer be considered trivial, detecting the health risks associated with air pollution requires careful statistical modeling. We have previously presented the air pollution and health effects from Australia and in this study we present for the first time the association from Kuwait.
The data presented here come from Kuwait, an Arabian Gulf country with desert climate. The data used in this study were collected during 1990 –2000 on the daily air pollution levels, cause specific daily hospital admissions and cause
specific daily mortality data, to explore the association between air pollution and short-term health effects. Daily weather data were also used to adjust for any confounding effects of changing weather on the mortality and morbidity. Statistical modeling techniques used here were based on the protocol developed by the APHEA (Air Pollution on Health: European Approach).

Some pollutant levels were found to have significant associations with total daily morbidity. Results also indicated that significant positive association between pollution levels and admissions for asthma and respiratory diseases, in particular. This is marked in some age groups.

There are significant associations between some components of air pollution and hospital admission. Kuwait data analyses are not completed as yet and initial analyses indicate that increasing levels of respiratory disorders like asthma and cardiac disorders could be attributed to the higher levels of air pollution that is experienced in the country. These findings are very similar to the findings from overseas, particularly compared to senior author's previous studies from Australia.

Acknowledgment: This study was supported by Kuwait University Research Grant MXS001 (EPA Funded).

Environmental Chemical Databases Ask for Environmetrics, Chemometrics and Information Metrics
Kristina Voigt¹, Rainer Brueggemann², and Stefan Pudenz³
GSF - National Research Center for Environment and Health, Institute of Biomathematics and Biometry, Department of Biostatistics, 85764 Neuherberg¹,
Leibniz-Institute of Freshwater Ecology and Inland Fisheries, 12587 Berlin²
Criterion Evaluation and Information Management, Berlin, Germany³
kvoigt@gsf.de

Achieving sustainable development in the chemicals sector is the main objective of the European Commission's White Paper on a future chemicals strategy. Initiative was taken, and later reinforced by the White Paper, to collect data on chemicals for their risk assessment leading to, where necessary, risk reduction. The available information on existing substances should be thoroughly examined and best use made of it in order to waive testing, wherever appropriate. An important role play the chemical databases which can be regarded as an information turnover. In this presentation the emphasis lies on the evaluation of 12 numerical databases available on the free Internet which focus on environmental fate and ecotoxicity. The following 12 databases are chosen: Biocatalysis/Biodegradation Database, Chemicals Information System for Consumer-relevant Substances (CIVS), ECOTOX, Envirofacts, Environmental Fate Database, Environmental Health Criteria Monographs (EHCs), EXTOXNET, HSDB (Hazardous Substances Databank), NRA Chemical Review Program, Pesticide Database, Japan, SIDS (OECD-Screening Information Data Set), UmweltInfo.

The question to answer is whether data is available on 12 chosen well-known high-production volume chemicals and 15 environmental parameters
(ecotoxicological and environmental fate and pathways). Hence we have to cope with a 12x27 data-matrix. Different mathematical multi-criteria evaluation and decision support instruments will be applied. The fundamentals of these multi-criteria evaluation methods will be presented as well as their advantages and disadvantages demonstrated on the data-matrix 12x27. The methods applied are the Hasse diagram technique (HDT), a method derived from discrete mathematics, the Method of Evaluation by Order Theory (METEOR), as well as other multi-criteria evaluation methods, e.g. a simple approach of the utility function theory, PROMETHEE and AHP (analytical hierarchy process). It is shown that the degree of participation of stakeholders are the main characteristics. Whereas the Hasse diagram technique is directed to the scientifically given data-matrix, and therefore yields an objective and transparent evaluation and data analysis, METEOR, concordance analysis utility function theory, PROMETHEE and AHP have their preferences as participants or stakeholders have an active part in the decision process. Significant data gaps even on the chosen well-known high production volume chemicals as well as on ecotoxicity and environmental fate parameters are identified by all chosen methods. This indicates an alarming signal concerning the new existing chemicals policy of the EEC. For future steps concerning the data availability of chemicals several steps involving chemometrical and environmetrical approaches should be considered, like e.g. the fostering of several other (also not well-known databases), estimating of data by well-established methods (QSAR) and filling up data gaps in established databases.

The influence of large scale climate variability on winter precipitation extremes over the North America
Xuebin Zhang1, Pavel Ya. Groisman2, and Francis W. Zwiers3

Climate Research Branch, Meteorological Service of Canada, Downview, Ontario, Canada1 National Climate Data Center, Asheville, NC, USA2 Canadian Centre for Climate Modelling and Analysis, Victoria, BC, Canada3

Extremes are important aspects of climate. Changes in the magnitude and frequency of climate extremes will have environmental and socio-economical consequences. Understanding the mechanism of changes in extremes is therefore very important. The North America climate is strongly influenced by large scale climate variability, such as El Nino/Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), and the North Atlantic Oscillation (NAO). The connections between seasonal precipitation totals over the region and those variabilities are relatively well understood. However, there are not many studies addressing the impacts of large-scale climate variability on precipitation extremes. This study attempts to identify the statistical relationship between maximum daily precipitation in winter and large scale climate variability represented by ENSO, PDO, and NAO, by conducting extreme value modelling with the ENSO, PDO, and NAO indices being co-variates. It was found that the ENSO, PDO, and NAO indeed have profound impacts on winter extreme
precipitation. For example, enhanced extreme precipitation occurs in the Pacific and Atlantic coasts of the North America in the El Nino years.

A Web-based Rainfall Atlas for Southern Africa
Walter Zucchini, Oleg Nenadic and Guenther Kratz
University of Goettingen, Germany
onenadi@uni-goettingen.de

We describe the development of a web-based rainfall atlas for Southern Africa, a decision support system for the management of water resources. The construction of the atlas (accessible at http://134.76.173.220/rainfall/index.html) encompasses four major phases. In the first phase, a 16 parameter model was fitted to rainfall data from 2050 sites and later updated and extended to 5070 sites. The interpolation of the parameters (second phase) was performed using Kriging with external drift. This way, the model was extended to cover the entire region of Southern Africa (South Africa, Lesotho and Swaziland). In the third phase, the interpolated parameters were used to generate artificial rainfall sequences at a spatial resolution of 1 degree square. Based on the simulated sequences, rainfall-related statistics, such as percentiles of annual and monthly rainfall distributions, probabilities associated with droughts, and additional measures regarding the timing and intensity of rainfall were calculated. The information transfer is covered by the fourth phase - the construction of the atlas website. It offers online access to the rainfall-related statistics (1.3 GB of numerical data) and to some 5000 maps of the data. To facilitate the computation of statistics that are not available in the database, users can generate artificial sequences based on the above mentioned model online. The simulated sequences can then be imported into any statistical software for further processing.