

Annual Report 1999-2000

National Research Center for Statistics and the Environment



NRCSE

The NRCSE was established in 1997 through a cooperative agreement with the United States Environmental Protection Agency which provides the Center's primary funding.





Annual Report 1999-2000 National Research Center for Statistics and the Environment

1. SUMMARY	3
2. OUTREACH ACTIVITIES	3
2.1 Seminars	3
2.2 Web site	3
2.3 Workshops	4
2.3.1 EPA Las Vegas	4
2.3.2 Slovakia workshop	4
2.3.3 Large Data Sets	5
2.3.4 Mini-workshops	5
2.3.5 Collaborative working group	6
2.3.6 EPA site visit	6
2.4 Conference presentations	6
2.5 Educational activities	7
2.6 Professional service and recognition	7
2.7 Other	8
3. RESEARCH ACTIVITIES	8
3.1 Internal funding	8
3.1.1 Ranked set sampling: costs and applications	8
3.1.2 Assessment of deterministic models	9
3.1.3 Statistical modeling of censored data	9
3.1.4 Statistical adjustment of ozone for meteorological variables	9
3.1.5 Stochastic analysis of precipitation	10
3.1.6 Statistical aspects of setting and implementing environmental standards	11
3.1.7 Monitoring network design	11
3.1.8 Nonhomogeneous global covariance estimation	12
3.1.9 Trend estimation using wavelets	12
3.1.10 Semiparametric trend estimation and model selection	13
3.1.11 Bayesian methods for assessment of environmental fate and transport models	13
3.1.12 Receptor modeling for air quality data in space and/or time	14
3.1.13 ORCA: A visualization toolkit for high-dimensional data	15
3.1.14 Bayesian estimation of nonstationary spatial covariance structure	15
3.1.15 Agricultural modeling for watershed management	16
3.1.16 Composite sampling	16
3.2 External funding	17
3.2.1 Visiting researchers	17
3.2.2 Particulate matter air pollution	18
3.2.3 Particulate matter field study in Slovakia	19
3.2.4 Ecological assessment of riverine systems by combining information from multiple sources	20



3.3 Submitted and published research papers	21
3.4 Graduating students	26
PhD degrees	26
MSc degrees	26
Foreign degrees	26
4. ADMINISTRATION	27
4.1 Executive and advisory committees	27
4.1.1 Executive committee	27
4.1.2 Advisory committee	27
4.2 Members	27
4.3 Hiring	27
4.4 Space	27
4.5 Relations to other statistical research groups	28
4.5.1 NCAR (National Center for Atmospheric Research)	28
4.5.2 IMPACT	28
4.5.3 Other research groups	28
5. A VIEW TOWARDS THE FUTURE	28
APPENDIX A. SEMINARS	30
APPENDIX B. TECHNICAL REPORTS 1999-2000	32
APPENDIX C. CONFERENCE PRESENTATIONS	35
APPENDIX D. WORKSHOP AGENDAS	38
D1. EPA Las Vegas	38
D2. Exposure assessment in environmental and occupational health	40
D3. Large Data Sets	41
D4. Internal workshop	43
D5. Teaching Environmental Statistics at the UW	45
Who is the audience for Environmental Statistics courses?	45
I. Offerings at the UW	45
II. Areas of Need	46
III. Problems in Teaching Statistics	46
IV. Suggestions for improvements	46
V. Proposed Program Level Changes	47
E. Course description for EPA Region X Risk Assessment Course	48



1. Summary

The fourth year of operation of NRCSE (Oct. 1, 1999 – Sep. 30, 2000) has been extremely productive. Center operations are functioning smoothly, and there is a steady stream of visitors, both in person ([section 3.2.1](#)) and on the worldwide web ([section 2.2](#)). Members are continually presenting work at scientific meetings ([Appendix C](#)), and technical reports ([Appendix B](#)) and papers ([section 3.3](#)) are produced in a variety of areas.

Three major workshops ([section 2.3](#)) were produced off-site this year: at [EPA Las Vegas](#), in [Slovakia](#), and at [NCAR](#) in Colorado. [Smaller workshops](#) complemented the Center [seminar series](#) during the year.

The Center [web site](#) finally has a functioning [software repository](#), and software can be downloaded for free in areas of spatial statistics, model assessment, visualization, and regression with censored data.

2. Outreach activities

2.1 Seminars

The University has several weekly statistical seminar series. Consequently, there has been relatively sparse attendance at some NRCSE seminars. We therefore limited the NRCSE seminars to about three per quarter, and made them joint with other campus departments and groups. The list of seminars is in [Appendix A](#). In addition to the seminars, the Center has been holding afternoon workshops to cover in more depth areas of interest to Center members. These workshops are described below in [section 2.3.4](#).

2.2 Web site

The Center web pages, located at <http://www.nrcse.washington.edu>, are the main source of information about the Center. During the time period of this report there were 207,893 successful requests for pages from over 14, 000 different hosts. This is an increase by about 25% in the requests over the previous year. On average, about 14 Mb of data were transferred per day. About 26% of the visitors came from .edu domains, and 10% from each of .com, .net and .gov/.us domains. Italy, Australia, France and Canada were the leading non-US domains. About 16% of the hosts could not be identified. The most visited directory was the research directory, where project descriptions and research reports are stored. As during the previous year, the highest demand was for Technical Report 15: “Meteorological Adjustment of Western Washington and Northwest Oregon Surface Ozone Observations with Investigation of Trends” by Joel H. Reynolds, Barnali Das, Paul D. Sampson and Peter Guttorp.



The software page at <http://www.nrcse.washington.edu/software/software.asp> of the Center web service was activated. It currently contains four links: the Orca visualization software, Doug Nychka's Funfit package implemented for SPlus, the POMAC package for Pareto optimal process model development and assessment, and SPlus code for maximum likelihood estimation in linear regression with interval or left censored data.

2.3 Workshops

2.3.1 EPA Las Vegas

The Center has pursued a series of workshops at various EPA locations intended to give EPA researchers a feel for the kind of research being conducted and to initiate new research contacts. The second of these workshops took place at the EPA laboratory in Las Vegas, Nevada, on December 13-14, 1999. Eight Center members and graduate students participated in the workshop, giving eight talks to a fairly large audience. The titles of the talks are found in [Appendix D.1](#).

The program was developed jointly with EPA scientists. It consisted of topics ranging from Tom Lewandowski's work on toxicodynamic/toxicokinetic modeling to Adrian Raftery's methods for incorporating expert opinion in deterministic models. These presentations alternated with group discussions of statistical issues arising in the EPA Lab's work. For example, there was a lively discussion of the best ways to sample to assess pollution levels in industrial sites, and the best ways to handle possibly inadequate sets of measurements. There have been follow-up contacts between the lab and the Center related to this visit, and we hope that this and future workshops will help bring EPA and NRCSE researchers closer together.

2.3.2 Slovakia workshop

Alison Cullen was involved in the planning of a NRCSE sponsored workshop in Dovaly, Slovakia on October 24-26, 1999. The purpose of the workshop was to enhance capabilities to identify, assess and manage high priority environmental and/or occupational health issues. The workshop had approximately 40 Slovak and Czech participants. The audience included decision-makers, who must deal with contemporary environmental and occupational health problems, and scientific staff who support the decision-making. The workshop was designed to follow a case where an environmental / occupational issue has been identified through planning, implementation, analysis and communication of a data collection program in order to support risk management decision making. Further information about this project is in [section 3.2.3](#), and details about the workshop are given in [Appendix D.2](#).



2.3.3 Large Data Sets

In July 2000, a workshop on large data sets was held at the National Center for Atmospheric Research in Boulder, CO. This workshop, sponsored by NRCSE, the Geophysical Statistics Project, and NCAR, acquainted statisticians with substantive scientific problems that hinge on the analysis of large data sets, these can be either observational or generated as the output of numerical models and presented recent statistical advances for large problems. Topics included visualization strategies, computational algorithms and new methods, including techniques from data mining. Although the statistical methodology is relevant to wide range of problems, the focus was on continuous variables and multivariate or spatial-temporal contexts. Funding was available to support attendance with special emphasis given to graduate students and other young researchers. About 50 researchers participated in the workshop, including 10 from government laboratories or industry. The detailed program can be found in [Appendix D.3](#).

2.3.4 Mini-workshops

Internal planning workshop

In order to plan for the expected request for proposals for renewal funding of NRCSE, the Center held an internal planning workshop to familiarize Center members with the range of projects being conducted, and discuss what has worked and what has been less successful in our organizational structure and communication. Oral presentations as well as poster presentations were given, and three group discussions were planned. The program can be found in [Appendix D.4](#).

Statistical downscaling of precipitation

On May 24, an NRCSE-sponsored workshop on statistical downscaling of precipitation was held in Padelford Hall. Attendees heard presentations by Jim Hughes (UW Biostat and NRCSE), Bryson Bates (CSIRO, Australia), Dennis Cox (Rice University and NCAR) and Claudia Tebaldi (NCAR). Jim Hughes opened the workshop by giving a review of the downscaling problem (predicting local atmospheric measurements, such as precipitation, from broad scale measurements such as sea level pressure) and summarizing the various approaches that have been used to solve it. Bryson Bates discussed applications of downscaling in southwestern Australia and future research plans in that area. Dennis Cox discussed his work on rainfall modeling at five rain gauge stations in the Southeast U.S. and the development of methods for model goodness of fit assessment. Finally, Claudia Tebaldi discussed her work using eight rain gauge stations in the southeast U.S., which has focused on the development of atmospheric summary measures at temporal scales that are useful for downscaling. In addition to these four talks there was extensive discussion on the utility, strengths and limitations of downscaling.



Environmental Statistics Teaching at UW

A workshop on teaching environmental statistics was organized by Alison Cullen on May 26, 2000. 14 participants from 10 departments discussed current offering of environmental statistics, and identified some areas of need. In particular, undergraduate courses on correlated data, multivariate analysis, and risk analysis/decision making are lacking on campus. Among the ideas for improvement were: development of a web site as clearing house for environmental statistics courses; development of a data set or case repository; development of a Speakers' Bureau to bring in researchers to talk about their use of statistics in their work. A list of participants and details of the discussion can be found in [Appendix D5](#).

2.3.5 Collaborative working group

Following the Joint Statistical Meetings in Indianapolis (see details in [Appendix C](#)), Paul Sampson organized a small but internationally diverse collaborative working group the week of August 21 on spatial deformation methods for nonstationary spatial covariance modeling. Two of the speakers at his session at the JSM, Alexandra Schmidt (Brazil), a student at the University of Sheffield, and Olivier Perrin (France), who is now at the University of Toulouse, came to Seattle to present their work and collaborate with our former student Wendy Meiring (South Africa), now at the University of California, Santa Barbara, and current students Doris Damian (Israel) and Sinjini Mitra (India). Comparisons of the different deformation methods were initiated, and data sets to pursue such comparisons were exchanged.

2.3.6 EPA site visit

On July 19, 2000, Peter Preuss, Jack Puzak and Chris Saint from the EPA National Center of Environmental Research (the EPA office in charge of our funding) conducted a site visit at NRCSE. Mary Lou Thompson, David Ford, Paul Sampson, June Morita, Thomas Lumley and Peter Guttorp presented a variety of aspects of NRCSE work. The EPA response was very positive.

2.4 Conference presentations

A number of Center members and graduate students have given presentations and organized sessions at national and international meetings of various scientific organizations. These include the Society for Risk Analysis annual meeting, the Fourth Bernoulli World Congress, the Joint Statistical Meetings, the joint SPRUCE/Environmetrics conference, and the Fourth International Chemometrics/Environmetrics Conference. In particular, NRCSE has continued the pattern at the Joint Statistical Meetings of organizing one session as an invited organization, and one through the ASA Section on Statistics and the Environment. A detailed list of presentations is given in [Appendix C](#).



2.5 Educational activities

The middle school curriculum on the scientific method, entitled “The Truth About Science,” has been adopted as a recommended curriculum for the Seattle School District. Plans are underway for the materials to be published by the National Science Teacher's Association Press. In all, about 25 local teachers have gone through training in using the curriculum. A web page at <http://www.nrcse.washington.edu/truth> contains downloadable resources, lesson plans, and general information about the product.

Elaine Faustman, Scott Bartell and Bill Griffith presented a course on Risk analysis to personnel at EPA Region X. The course outline can be found in [Appendix E](#). It was attended by over a dozen Region X scientists. As an outgrowth of the original course we have also been presenting monthly statistics lectures at EPA (Oct, Nov, Dec) and plan on continuing. These are each about two hours in length and involve topics of interest to the EPA participants. There are about 25 people attending these lectures, including a few state agency people.

2.6 Professional service and recognition

Jon Wakefield was awarded the annual Guy Medal in Bronze for 2000 by the Royal Statistical Society for his recent work in research on the development of statistical methods, particularly for spatial epidemiology and population pharmacokinetic modeling.

Loveday Conquest served as the Chair of the ASA Section on Statistics and the Environment for 1999.

Peter Guttorp is the President-Elect of the International Environmetrics Society. His term goes for two years. In 2002 he will be President of the society, also for two years.

As institutional members of the International Environmetric Society (TIES), the Center receives two full memberships, which the executive committee has decided to award to outstanding research assistants. The 2000 award went to Nicolle Mode and Marianne Turley.

Ashley Steel won the student methods paper competition at the North American Benthological Association annual meeting 2000 with a paper on horizontal Secchi disks for measuring water clarity.

Loveday Conquest received a *Women Who Make a Difference* Award “for outstanding achievements and contributions to the science, engineering, and technology industries,” Women of Color Technology Awards, August 2000.



Peter Guttorp is Section Editor for the Spatial/temporal section of Wiley's *Encyclopedia of Environmetrics*, to appear in autumn of 2001. Several NRCSE member have contributed articles to the *Encyclopedia*.

Peter Guttorp participated in the development of the EPA PM Criteria Document.

2.7 Other

NRCSE is co-sponsoring the Student Paper Awards of the American Statistical Association Section on Statistics and the Environment. This co-sponsorship was motivated by a desire to ensure that awardees would be able to attend and participate in the Joint Statistical Meetings where the award is presented, something that had not previously been assured. The first recipient was Deepak K. Agarwal from University of Connecticut.

The Center publishes a newsletter about four times a year, with the latest developments, publications, and other items of potential interest to the membership. The newsletters are available at <http://www.nrcse.washington.edu/newsletter/newsletter.asp>.

3. Research activities

3.1 Internal funding

The internal NRCSE research funds are allocated annually by the Executive Committee after a competitive application process. Criteria for awards include

- * Scientific merit.
- * Relevance to the Center's agenda.
- * Evidence of involvement by EPA scientists.
- * Feasibility of project and likelihood of substantial products.
- * Results and EPA contacts of previous NRCSE support

3.1.1 Ranked set sampling: costs and applications

Center member: Loveday L. Conquest
EPA collaborator: Barry Nussbaum
W ESTAT collaborator: David Marker
Research assistant: Nicolle Mode

Ranked set sampling (RSS, McIntyre 1952) is a two-phase sampling procedure that reduces the number of samples required from a more exacting and expensive measurement by using expert knowledge or other frugal measurement to select the sampled values. Current work is focusing on [1] generalizing the cost analyses to unbalanced designs (i.e. unequal set sizes where each order statistic appears an unequal number of times), and [2]



other sampling schemes where the initial data are not just ranked, but also placed in strata based upon expected ranges (double sampling with cutpoints).

The paper submitted last year to *Environmetrics* (Ranked Set Sampling for Ecological Research: Accounting for the Total Costs of Sampling, by Mode, Conquest, and Marker) has been published (Vol. 10, pp. 179-194). A second manuscript, "Incorporating Prior Knowledge in Ecological Sampling: Ranked Set Sampling and Other Double Sampling Procedures" has been submitted to *Environmetrics* for publication. L. Conquest also presented this paper at the Fourth International Chemometrics and Environmetrics Conference, Las Vegas, Nevada, September 17-20, 2000. Loveday Conquest also attended the International Environmetrics Society Conference, Sheffield, U.K., where she participated in poster discussions and sessions on sampling techniques for environmental and ecological studies

3.1.2 Assessment of deterministic models

Center members: David Ford and Joel Reynolds
Research assistant: Marianne Turley

Software for the POMAC process assessment tools was made available at the NRCSE software web page.

3.1.3 Statistical modeling of censored data

Center member: Mary Lou Thompson
Research assistant: Kerrie Nelson

We have developed a maximum likelihood approach to point and interval estimation for multiple linear regression in the presence of Type I interval and left censoring. We have evaluated and compared the characteristics of the ML estimates to those obtained from simple midpoint substitution under different assumptions as to the degree of censoring, strength of correlation and sample size. The methodology has been implemented in Splus and a program for general implementation is available from the NRCSE website <http://www.nrcse.washington.edu/research/projects/software/smmcd.txt>. We are working on a paper detailing these results, which we expect to submit for publication soon.

3.1.4 Statistical adjustment of ozone for meteorological variables

Center members: Mary Lou Thompson, Joel Reynolds, Peter Guttorp, and Paul Sampson
EPA collaborator: Larry Cox
Research assistants: Barnali Das and David Caccia



A review paper on meteorological adjustment of ozone (NRCSE TRS 26) is to appear in *Atmospheric Environment* (2001, **35** (3): 617–630). In conjunction with this work, several of the approaches suggested in the literature have been applied to Chicago ozone data from the AIRS database for the period 1981-1991. The work highlights the need for development of techniques for extreme value analysis of space-time processes, as well as for analysis of networks designed to measure extreme values of a random field. Work with the **IMPACT** consortium of European researchers ([section 4.5.2](#)) continues, focusing on meteorological adjustment for air quality data from Paris and vicinities.

3.1.5 Stochastic analysis of precipitation

Center members: Jim Hughes, Peter Guttorp and Paul Sampson
Research assistants: Enrica Bellone, Tamre Cardoso

We consider the use of stochastic models of precipitation in assessing climate variability and climate change and in downscaling (doing sub-grid scale simulation) of general circulation models of global climate. Technically, our method involves fitting non-stationary hidden Markov models to sequences of multi-station precipitation data. The states of the model are identified as weather states, and daily observations from atmospheric fields drive the transitions between weather states.

Current work focuses on developing a model for precipitation amounts in Washington State, using a small network of 10 stations. Issues of sensitivity to measurement error, particularly for small precipitation amounts, are important and difficult. The choice of the number of weather states is another technically challenging question. The work this year resulted in a dissertation by Enrica Bellone entitled *Nonhomogeneous hidden Markov models for downscaling synoptic atmospheric patterns to precipitation amounts*, accepted for the PhD degree in Statistics.

Bryson Bates from CSIRO in Perth, Australia, visited NRCSE in May, and worked with Jim Hughes on development of the methodology, with particular emphasis on how to summarize atmospheric variables for use in the model, and on seasonal fits.

A hierarchic Bayesian approach to estimating precipitation rate using data from different sources, such as rain gauges, weather radar, and distrometers, is being developed by Tamre Cardoso. Traditionally, rain gauge data has been regarded as “ground truth” for calibration purposes, although gauges have known biases, particularly in windy conditions. This modeling project will enable researchers to improve radar-gauge calibration exercises, and will eventually be used to improve precipitation observation networks and satellite calibration.



3.1.6 Statistical aspects of setting and implementing environmental standards

Center members: Peter Guttorp, Paul Sampson, and Mary Lou Thompson

EPA collaborator: Larry Cox

Other collaborators: Ronit Nirel, Israel, and Bruno Sanso, Venezuela

Graduate research assistants: David Caccia and San San Ou

Undergraduate research assistant: Anthony Nguyen

The typical environmental standard is what may be called an ideal standard. Based on various health effects studies, a target value not to be exceeded is determined, and the standard may be that this value not be exceeded, or only be exceeded with a certain probability, or a certain number of times per year.

Our work in this project focused on analyzing the statistical properties of the present US ozone standards, and developing alternative, statistically more appropriate, implementations of ideal versions of the present standards. Two technical reports (NRCSE TRS 48 and TRS 59) have been produced, and these papers have been submitted for publication.

With Center visitors Nirel and Sanso, we have started developing theoretical tools needed for defining areal standards, and estimating the associated quantities. This work will also enable the use of process model output (such as Models-3) as prior information regarding the mean field of an air pollution field. Additional funding for this project has been received from EPA.

3.1.7 Monitoring network design

Center member: Paul D. Sampson

Undergraduate research assistant: Brooke Hoem

NRCSE interest in monitoring network design led to the sponsorship and organization of a special invited session on Air Quality Monitoring Network Design at the Joint Statistical Meetings to take place in Atlanta, Aug 2001. At this session we will present results of current research deriving from a proposal to consider a Pareto optimality approach to the spatial design problem with multiple objective functions. An undergraduate student, Brooke Hoem, is currently developing software for the evaluation of numerical design objective criteria, beginning with the conventional kriging-based average predictive error, to be interfaced with the Pareto Optimality algorithms implemented in the NRCSE "Pareto Evolve" software system developed primarily for purposes of multiple criteria assessment of environmental or ecological process models.



3.1.8 Nonhomogeneous global covariance estimation

Center members: Peter Guttorp, Paul Sampson, and Tilmann Gneiting

NCAR collaborator: Doug Nychka

Research assistant: Barnali Das

We have developed computationally rather heavy tools for analyzing meteorological time series on a global scale, taking into account spatial heterogeneity and the fact that data are collected on a globe (an oriented sphere). A flexible class of parametric nonstationary global covariance functions has been developed, and applied to global temperature data with likelihood tools that enable use of incomplete monitoring data without requiring imputation. The methodology enables realistic estimates of prediction variance for regional and global averages, and allows comparison of gridded model output data to suitably processed observational data. This work would not have been possible without the generous cooperation of the Geophysical Statistics Projects at NCAR in Boulder, CO.

The main result of this year's efforts was the successfully defended Ph.D. dissertation by Barnali Das, entitled *Global covariance modeling: a deformation approach to anisotropy*. We expect several papers to result during the coming year.

3.1.9 Trend estimation using wavelets

Center members: Don Percival, Peter Guttorp

Research assistant: Peter Craigmile

A common problem in the analysis of environmental time series is how to deal with a possible trend component, which is usually thought of as large scale (or low frequency) variations or patterns in the series that might be best modeled separately from the rest of the series. Trend is often confounded with low frequency stochastic fluctuations, particularly in the case of models that can account for long memory dependence (slowly decaying auto-correlation) and non-stationary processes exhibiting quite significant low frequency components.

We have developed both an approach to estimating trend at a given temporal scale and procedures for testing the presence of a trend, valid for a large range of assumptions. This work is described briefly in Section 9.4 of the book "Wavelet Methods for Time Series Analysis," D. B. Percival and A. T. Walden, Cambridge, UK: Cambridge University Press, 2000, and two NRCSE Technical Reports (NRCSE TRS 47 and 49). This work forms the basis for the central part of Peter Craigmile's doctoral dissertation *Wavelet*



Based Estimation for Trend Contaminated Long Memory Processes, which will be completed in December 2000.

The trend estimation procedure has also been used in health effects studies for particulate matter air pollution (see [section 3.2.2](#)).

3.1.10 Semiparametric trend estimation and model selection

Center member: Peter Guttorp

Research assistant: Florentina Bunea

The partially linear regression model is a semiparametric extension of the linear regression model, in which the mean of the observations are the sum of a linear function of some covariates and an arbitrary nonlinear function of another set of covariates. In the trend estimation framework, this second set of covariates would be time and/or space. This work deals with optimal estimation of the nonlinear function in the presence of model selection for the linear part. A method has been devised allowing for adaptive estimation of the nonlinear function and simultaneous selection of variables. The method has been applied to an analysis of the ozone level at Chicago O'Hare airport, yielding results that are quite comparable to other studies of the same data (cf. [section 3.1.4](#)). For 1981–1985, very few meteorological variables are needed (temperature and possibly relative humidity) to explain the ozone variation. Also, based on this 5 year period of observation, we could not detect a trend. This work also resulted in a Ph.D. dissertation by F. Bunea entitled *A Model Selection Approach to Partially Linear Regression*.

3.1.11 Bayesian methods for assessment of environmental fate and transport models

Center members: Alison C. Cullen, Adrian Raftery

Research assistant: Samantha Bates

The aim of this project is to develop Bayesian methods for assessing uncertainty and variability in risk assessment models, building on the Bayesian melding approach of Poole and Raftery (2000). The work this year has focused on the development and implementation of this Bayesian approach in the context of risk assessment. There have been four main foci of our work. The first is the development and application of the sampling-importance-resampling (SIR) algorithm for making inference about the parameters of the deterministic simulation models involved given all available evidence and uncertainty. This has been investigated in the context of three main examples: a one compartment air-to-soil model developed originally by Alison Cullen for PCBs in the New Bedford Harbor area, a model for the population dynamics of whales, and a simulated model designed to investigate higher-dimensional situations. The second focus has been the extension of



these methods to multiple-compartment models, and we have focused on the air-to-soil-to-plant extension of Cullen's air-to-soil model.

The third focus has developed from the observation that the SIR algorithm is inefficient in high-dimensional models with the ridge-like posteriors characteristic of these models, and we have been developing an MCMC method as an alternative to the SIR algorithm. Standard MCMC methods do not work well in this context, and we have developed an entirely new MCMC method that does perform well for these applications, the nearest-neighbor MCMC method. Our fourth focus has been the development of model validation methods based on the Bayesian melding approach.

This year we have published one paper in the *Journal of the American Statistical Association* (Poole and Raftery 2000), and a second paper will appear in the *Proceedings of the American Statistical Association* (Bates, Raftery and Cullen 2000). This latter paper has also been issued as NRCSE TRS 58. We anticipate submitting three papers for publication in FY01, the last year of the project, on inference from such models, on our new nearest neighbor MCMC method, and on model validation. Two invited lectures have been given at conferences (the Interface meeting and the Joint Statistical Meetings).

3.1.12 Receptor modeling for air quality data in space and/or time

NRCSE member: Peter Guttorp

NRCSE postdoc: Eun Sug Park

An important problem in environmental science is to identify where pollution comes from given air pollution data. Multivariate receptor modeling aims to achieve this goal by decomposing ambient concentrations of pollutants into components associated with source emissions. This is a difficult problem in its most general form and typically restrictive assumptions are required. One assumption is that the observations are temporally independent, which is inappropriate for most of hourly measurements. We have developed a multivariate receptor model for temporally correlated data, which can incorporate extra sources of variability due to dependence in estimation of model parameters and uncertainty. A paper, *Multivariate receptor modeling for temporally correlated data by using MCMC*, has been written and submitted to *Journal of the American Statistical Association*.

Assumptions on the number of pollution sources and identifiability conditions are the main source of model uncertainty in multivariate receptor models, which is often overlooked. A Bayesian approach based on the marginal likelihood for assessing model uncertainty in multivariate receptor models has been developed. The work has resulted in a paper, *Multivariate Receptor Models and Model Uncertainty*, presented at Fourth Inter-



national Conference on Environmetrics and Chemometrics in Las Vegas, and submitted to *Chemometrics and Intelligent Laboratory Systems*.

We are currently focusing on extending receptor models to spatially correlated data obtained from multiple monitoring sites. Two cases, measurements on a single species from multiple monitoring sites, and measurements on multiple species from multiple monitoring sites, are being investigated. The first type of data can be used to locate the major pollution sources by estimating their spatial profiles, while the second type of data is ideal for characterizing spatial structure of source contributions and errors. Two manuscripts are expected from this research. They will be presented as invited papers at the Joint Statistical Meetings meeting, August 2001, Atlanta, Georgia and at the International Statistical Institute meeting, August 2001, Seoul, Korea.

3.1.13 ORCA: A visualization toolkit for high-dimensional data

Center member: Thomas Lumley

Other investigators: Dianne Cook (Iowa State), Peter Sutherland, and Tony Rossini

A paper describing the structure and development of ORCA has been produced and submitted for publication (NRCSE TRS 46). The web page for the project has been moved to <http://software.biostat.washington.edu/orca>. The link is available from the Center software page.

3.1.14 Bayesian estimation of nonstationary spatial covariance structure

Center members: Paul D. Sampson, Peter Guttorp

Research assistant: Doris Damian

The approach to modeling nonstationary (or non-homogeneous) spatial covariance structure through a deformation of the geographic coordinate system, as implemented first by Sampson and Guttorp and then by Meiring, has left the calculation of uncertainty in the estimated structure exceedingly difficult using bootstrap methods. We have long recognized the appeal of a formal Bayesian estimation of this spatial covariance model assuming a Gaussian model for the space-time process, and have now completed the specification of a Bayesian estimation paradigm for the spatial deformation model and its implementation using MCMC methods. In the process of this investigation, a number of results concerning likelihood-related estimation of variograms and spatial deformations have been revealed. The first manuscript on this methodology has been published in *Environmetrics*. It was also the basis of Doris Damian's recently presented general exam and invited plenary talks at two recent international meetings. The first, this past September in Sheffield, England, was the joint meeting of The International Environmetric Society (TIES) and the organization called Statistics in Public Resources and Utilities, and in



Care of the Environment (SPRUCE). The second is the Third European Conference on Geostatistics for Environmental Applications (geoEnv 2000) in Avignon this November.

The computations are still highly demanding, but with attractive benefits for a fully Bayesian spatio-temporal estimation (kriging) methodology now being developed. A number of issues, both theoretical and practical, are still under study. These include careful study and interpretation of prior distributions for the spatial deformation and accounting for spatial trend in the mean and variance of the spatio-temporal process.

3.1.15 Agricultural modeling for watershed management

Center member: Alison Cullen

Graduate student: Valerie Lertyaovarit

EPA Region X collaborators: Chris Feise and Karl Arne

The purpose of this project is twofold:

- (i) to build a web site for EPA programs to address agriculture issues;
- (ii) to identify the interrelationships related to agriculture among the EPA programs and state agencies.

A web site was needed to fully understand the relationships different departments within EPA have with agriculture issues. It was believed that many of the departments shared the same agricultural issues and were not collaborating with each other to find further information. The project was to create and design a site for staff within EPA to find out which EPA departments and state agencies had common agricultural interests.

The web site has been designed and is near completion. It is still a work in progress, because of the lack of EPA staff to maintain the site using Lotus Notes software. Access to this web site is restricted since EPA has not yet decided whether it should be made available to the public or kept internal.

3.1.16 Composite sampling

Center member: Gerald van Belle

WESTAT collaborator: David Marker

The Department of Housing and Urban Development and the National Institute of Environmental Health Sciences are sponsoring a national survey of dust hazards in housing. Westat developed the survey and was to conduct the data collection between June and October 1998. The survey assessed children's potential household exposure to lead and



allergens by estimating the levels of lead in dust, soil, and paint, the prevalence of hazardous levels of lead, and levels and patterns of allergens in dust in homes. The survey is an area probability sample of 1,000 homes representing the entire U.S. housing stock. The survey collected multiple floor dust samples from every house, all of which were to be measured individually. The dust samples were sent to analytical laboratories for chemical analysis for lead and selected allergens. NRCSE funded an add-on to generate empirical data on matched individual samples and composites for lead. The analysis is not yet completed.

A paper on *Composite Sampling* has been accepted in *Environmental and Ecological Statistics*.

3.2 External funding

NRCSE has considerable funds available for researchers at other institutions. These are used in three different categories: travel funds for short-term visitors, setting up future projects or working on intensive research collaboration, funds for visiting researchers, who spend substantial time at the Center, and funds for subcontracts, where the bulk of the work is done at the researcher’s home institution.

3.2.1 Visiting researchers

	Organization	Arrival	Departure	Collaborators
Nussbaum, Barry	EPA	990929	991003	Conquest
Sinha, Bimal	U. Maryland	990929	991003	Conquest
Pollak, Moshe	Hebrew Univ	991010	991013	NRCSE, Statistics
Goovaerts, Pierre	Univ of Michigan	991202	991204	Guttorp, Sampson
Barring, Hans	Univ of Amsterdam	991201	991207	Guttorp, Sampson
Van Storch, Hans	GKSS Germany	000106	000107	Guttorp, Wallace
El-Shaarawi, Abdel	Canadian Water Res. Inst.	000120	000120	Internal workshop
Cox, Larry	EPA	000120	000120	Internal workshop
Switzer, Paul	Stanford Univ.	000120	000120	Internal workshop
Urquhart, N. Scott	Oregon State Univ.	000207	000210	Conquest, Courbois
Tebaldi, Claudia	NCAR	000409	000531	Guttorp, Bellone
Bates, Bryson	CSIRO Land & Water	000430	000528	Hughes
Stein, Alfred	Wageningen Univ.	000517	000519	Besag, NRCSE
Henry, Ronald	USC	000531	000602	Guttorp, Park



Nirel, Ronit	Hebrew Univ., Israel	000701	000815	Sampson, Guttorp, Sanso
Sanso, Bruno	Univ. Simon Bolivar	000708	000810	Guttorp, Sampson, Nirel
Meiring, Wendy	UC Santa Barbara	000820	000824	Spatial Cov. Group
Perrin, Olivier	Univ. of Toulouse, Fr.	000817	000901	Spatial Cov. Group
Schmidt, Alex	Univ. of Sheffield, Eng.	000817	000825	Spatial Cov. Group
Kim, Ho	Seoul Nat. Univ. Korea	000822	000825	Park
Harte, David	Statistical Associates, NZ	000823	000828	Craigmile, Mathsoft
Vere-Jones, David	Victoria University, NZ	000823	000828	Guttorp, Mathsoft

Names of long-term visitors are listed in **boldface**. There were no sabbatical visitors during this year. During next year Paul Murtaugh from Oregon State University and Caren Marzban from NOAA's National Severe Storms Laboratory will be visiting for the entire year, and several shorter visits are also planned.

3.2.2 Particulate matter air pollution

Under separate EPA funding, a project on particulate matter air pollution was initiated in 1998. Participants include researchers from Iowa State, Ohio State, NISS, University of British Columbia, Duke, Chapman, and NRCSE.

The main work during this period was production of a special issue of *Environmetrics*, with Peter Guttorp as guest editor, which will appear as the December issue of the journal.

An editorial by Guttorp, Alison Cullen, NRCSE, and Richard Smith from NISS/UNC discussed scientific aspects of the recent court rejection of revised National Ambient Air Quality Standards for particulate matter and ozone. The scientific papers covered a broad range of analyses of particulate matter air pollution:

Lawrence H. Cox

Statistical Issues in the Study of Air Pollution Involving Airborne Particulate Matter

Michael Phelan

Timing and Scope of Emission Reductions for Airborne Particulate Matter: A Simplified Model

Anep Dewanji and Suresh Moolgavkar:

A Poisson Process Approach for Recurrent Event Data with Environmental Covariates

Li Sun, Nhu Le, James V. Zidek and Haluk Özkaynak:

Interpolating Vancouver's Daily Ambient PM10 Field

Lianne Sheppard and Doris Damian:

Estimating Short-Term PM Effects Accounting For Surrogate Exposure Measurements From Ambient Monitors.

Thomas Lumley and Drew Levy:

Bias in the Case--Crossover Design: Implications for Studies of Air Pollution



Thomas Lumley and Lianne Sheppard:

Assessing Seasonal Confounding and Model Selection Bias in Air Pollution Epidemiology Using Positive and Negative Control Analyses

Richard A. Smith, Jerry Davis, Jerome Sacks, Paul Speckman and Patricia Styer:

Air Pollution and Daily Mortality in Birmingham, Alabama: A Reappraisal

Merlise Clyde:

Model Uncertainty and Health Effect Studies for Particulate Matter

Merlise Clyde, Duke University, worked with NRCSE member Peter Guttorp and research assistant Erin Sullivan, on Bayesian model assessment approaches to analyzing Phoenix air pollution data. This resulted in a paper presented at the PM 2000 conference in South Carolina in January (NRCSE TRS 40). In addition, Sullivan defended a Master's thesis entitled *Estimating the Association Between Ambient Particulate Matter and Elderly Mortality in Phoenix and Seattle Using Bayesian Model Averaging* (NRCSE TRS 54), which used the ideas developed in the work with Clyde together with the wavelet-based trend estimation approach of Craigmile ([section 3.1.9](#)) to analyze Phoenix and Seattle data.

Several other participants in this project presented papers at the PM 2000 conference in South Carolina in January of 2000.

3.2.3 Particulate matter field study in Slovakia

Center member: Alison Cullen

EPA collaborator: John Vandenberg

Other collaborators: Michael Brauer, UBC, Canada; Eleonóra Fabiánová, Eva Mikhaliková, Peter Miskovic, Frantiska Hrubá, SUHE, Slovakia.

Recent interest in the levels of and health effects associated with airborne particulate matter exposure have sparked studies in the US and worldwide. Working with local scientists we have examined ambient exposure to inhalable particulate matter (PM₁₀ and PM_{2.5}), nitrogen oxides, sulfates, and nicotine in occupational settings, both industrial and office type, and in the home. 49 subjects were selected from those residing in either the Banska Bystrica city center or the Sasova residential area, because earlier studies in both areas suggested that ambient levels of particulate matter were significantly lower in the residential area than the city center.

Results indicate that central site monitors underpredict actual human exposures to PM₁₀ and PM_{2.5}. Personal exposure to sulfates was found to be predicted by outdoor sulfate levels, location of receptor residence and time activity information. From these results we



concluded that personal exposure measurements and precise daily activity data are crucial for accurate evaluation of exposure.

A workshop in Slovakia took place in October 1999 ([Appendix D2](#)). The paper “Personal Exposure to particles in Banska Bystrica Slovakia” (authors: M. Brauer, F. Hrubá, E. Mihaliková, E. Fabianová, P. Misković, A. Plzíkova, M. Lendacká, J. Vandenberg and A. Cullen) which was presented at the PM 2000 conference in South Carolina in January, 2000, has appeared in *Exposure Analysis and Environmental Epidemiology*.

The EPA has approved additional funding for this project.

3.2.4 Ecological assessment of riverine systems by combining information from multiple sources

This is a collaborative project with co-investigators Mark Handcock, UW, Joe Sedransk, Case Western, and Tony Olsen, EPA Corvallis. It originated from the NRCSE workshop on combining information from multiple sources in 1997, and was supported by NRCSE in 1998-99. Based on the preliminary development the project was awarded an NSF grant under the EPA/NSF Partnership for Environmental Research program for the 2001-2003 period.

The objective of the project is to improve understanding of the biological integrity of stream and river systems in the United States Mid-Atlantic Region by combining information from separate monitoring surveys, available contextual information on hydrologic units and remote sensing information. We now have developed the heart of the research program: to complement the mapping presented in the Atlas with new hierarchical spatial statistical models for environmental indicators on the streams and rivers that capture the spatial variation in the measures. These models have been used to estimate the indicators through the riverine system based on the information from multiple sources and aggregate scales. We quantify the uncertainty based on the information from multiple sources and aggregate scales, quantify the uncertainty in the estimates, and develop methods to visualize the resulting estimates and uncertainties.

We have developed a general framework for comparative distributional analysis of environmental variables. The methods are based on the “relative spatial distribution.” The spatial models developed are used to predict spatial distributions and relative spatial distributions. These methods are then used to combine county-level social science data with the different sources of environmental data. This makes it possible to investigate questions of environmental justice in a systematic and rigorous way.



Preliminary results of the project were presented at the Joint Statistical Meetings in Indianapolis, Indiana.

3.3 Submitted and published research papers

- R. A. Arnold, I. Diamond, and J. C. Wakefield (2000): Population denominator data. In *Spatial Epidemiology*. Elliott, P., Wakefield, J.C., Best, N.G. and Briggs, D. (editors), Oxford University Press.
- P. Aylin, R. Maheswaran, J. Wakefield, S. Cockings, L. Jarup, R. Arnold, G. Wheeler, P. Elliott (1999): A national facility for small area disease mapping and rapid initial assessment of apparent disease clusters around a point source: the UK Small Area Health Statistics Unit. *Journal of Public Health Medicine* **21**: 289–98.
- N. G. Best and J. C. Wakefield (1999): Accounting for inaccuracies in population counts and case registration in cancer mapping studies. *Journal of the Royal Statistical Society, Series A* **162**: 363–382.
- D. Billheimer (2001): Compositional Receptor Modeling. To appear, *Environmetrics*.
- D. Billheimer (2001): Space-time modeling of compositional data. To appear, A. El-Shaarawi and W. Piegorsch (eds.): *Encyclopedia of Environmetrics*. London: Wiley.
- D. R. Brillinger, P. Guttorp and R. P. Schoenberg (2001): Point process, temporal. To appear, A. El-Shaarawi and W. Piegorsch (eds.): *Encyclopedia of Environmetrics*. London: Wiley.
- M. Brauer, F. Hruba, E. Mihalikova, E. Fabianova, P. Miskovic, A. Plzikova, M. Lendacka, J. Vandenberg and A. Cullen (2000): Personal exposure to particles in Banska Bystrica, Slovakia. *Journal of Exposure Analysis and Environmental Epidemiology* **10**: 478–487.
- F. Bunea and J. Besag (2000): MCMC for contingency tables. In N. Madras (ed.): *Monte Carlo Methods*: 25–36. *Fields Institute Communications*. Providence, RI: American Mathematics Society.
- S. P. Charles, B. C. Bates and J. P. Hughes (2000): Statistical Downscaling from Numerical Climate Models for Southwest Australia. Proc. 3rd International Conference on Water Research and Environmental Research, the Institution of Engineers, Australia.



- M. Clyde, P. Guttorp and E. Sullivan (2000): Effects of ambient fine and coarse particles on mortality in Phoenix, Arizona. Submitted to *Journal of Exposure and Environmental Epidemiology*.
- P. F. Craigmile, D. B. Percival, and P. Guttorp (2000): Wavelet-based parameter estimation for trend contaminated fractionally differenced processes. Submitted to *Journal of Time Series Analysis*.
- P. F. Craigmile and D. B. Percival (2001): Wavelet-based trend detection and estimation. To appear, A. El-Shaarawi and W. Piegorsch (eds.): *Encyclopedia of Environmetrics*. London: Wiley.
- A. C. Cullen, P. Guttorp and R. L. Smith (2000): EDITORIAL: Special issue on statistical analysis of particulate matter air pollution data. *Environmetrics* **11**: 609–610.
- D. Damian, P. D. Sampson and P. Guttorp (2001): Bayesian Estimation of Non-Stationary Semi-Parametric Spatial Covariance Structures. To appear, *Environmetrics*.
- P. J. Diggle, S. E. Morris, and J. C. Wakefield (2000): The analysis of matched case-control studies in spatial epidemiology. *Biostatistics* 1: 89–105.
- C. P. Doberstein, J. R. Karr, L. L. Conquest (2000): The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. *Freshwater Biology* **44**: 1–17.
- P. Elliott, R. Arnold, S. Cockings, N. Eaton, L. Jarup, J. Jones, M. Quinn, M. Rosato, I. Thornton, M. Toledano, E. Tristan and J. Wakefield (2000): Risk of mortality, cancer incidence and stroke in a population potentially exposed to cadmium. *Occupational and Environmental Medicine* **57**: 94–97.
- P. Elliott, J. C. Wakefield, N. G. Best, and D. Briggs (2000): Spatial Epidemiology: methods and applications. In *Spatial Epidemiology*. Elliott, P., Wakefield, J.C., Best, N.G. and Briggs, D. (editors), Oxford University Press.
- P. Elliott and J. C. Wakefield (2000): Bias and confounding in small-area studies. In *Spatial Epidemiology*. Elliott, P., Wakefield, J.C., Best, N.G. and Briggs, D. (editors), Oxford University Press.



- E. M. Faustman, S. M. Silbernagel, R. A. Ponce, T. Burbacher and R. Fenske (2000): Mechanisms underlying children's susceptibility to environmental toxicants. *Environmental Health Perspectives* 108: 13–21.
- E. D. Ford (2000): *Scientific Method for Ecological Research*. Cambridge, U.K.: Cambridge University Press.
- N. Gertler and A. C. Cullen (2000): Effects of a Transient Cancer Scare on Property Values: Implications for Risk Valuation and the Value of Life. *Human and Ecological Risk Assessment*. **6**: 731–745.
- T. Gneiting and M. Schlather (2001): Space-time covariance models. To appear, A. El-Shaarawi and W. Piegorsch (eds.): *Encyclopedia of Environmetrics*. London: Wiley.
- T. Gneiting (2000): Criteria of Pólya type for radial positive-definite functions. To appear in *Proceedings of the American Mathematical Society*.
- T. Gneiting (2000): Power-law correlations, related models for long-range dependence, and their simulation. To appear in *Journal of Applied Probability*.
- T. Gneiting (2000): Addendum to “Isotropic correlation functions on d -dimensional ball”. To appear in *Advances in Applied Probability*.
- T. Gneiting (2000): Nonseparable, stationary covariance functions for space-time data. Under revision for *Journal of the American Statistical Association*.
- P. Guttorp, D. R. Brillinger and R. P. Schoenberg (2001): Point process, spatial. To appear, A. El-Shaarawi and W. Piegorsch (eds.): *Encyclopedia of Environmetrics*. London: Wiley.
- S.H. Kang and E.S. Park (2000): The actual size of the chi-squared and the likelihood ratio test of independence in a contingency table. Submitted to *Journal of Statistical Computation and Simulation*.
- J. E. Kelsall, S. E. Morris and J. C. Wakefield (2000): Disease surveillance and cluster detection. In *Spatial Epidemiology*. Elliott, P., Wakefield, J.C., Best, N.G. and Briggs, D. (editors), Oxford University Press.
- D. Levy, T. Lumley, L. Sheppard, J. Kaufman, H. Checkoway (2001): Referent selection in case-crossover analyses of health effects of air pollution. To appear, *Epidemiology*. 12:000-000, 2001. (in press)



- R. Maheswaran, S. E. Morris, S. Falconer, A. Grossinho, J. C. Wakefield and P. Elliott, (1999). Magnesium in drinking water supplies and mortality from acute myocardial infarction in North West England. *Heart* 82: 455–460.
- S. E. Morris, R. Sale, J. C. Wakefield, S. Falconer, P. Elliott and B. J. Boucher (2000): Hospital admissions for asthma and chronic obstructive airways disease in east London hospitals and proximity to major roads. *Journal of Epidemiology and Community Health* 54: 75–76.
- E. S. Park, C. H. Spiegelman, and R. C. Henry (2000): Estimating the number of factors to include in a multivariate bilinear model. *Communications in Statistics*, B **29**, Issue. 3.
- E. S. Park, C. H. Spiegelman, and R. C. Henry (2001), Bilinear estimation of pollution source profiles and amounts by using multivariate receptor models. To appear, *Environmetrics*.
- E. S. Park, P. Guttorp, and R. Henry (2000): Multivariate receptor modeling for temporally correlated data by using MCMC. Submitted to *Journal of the American Statistical Association*.
- E. S. Park, Man-Suk Oh and P. Guttorp (2000), Multivariate Receptor Models and Model Uncertainty. Submitted to *Chemometrics and Intelligent Laboratory Systems*.
- C. Pascutto, J. Wakefield, N. Best, L. Bernardinelli, P. Elliott, S. Richardson and A. Staines, (2000). Statistical issues in the analysis of disease mapping data. To appear, *Statistics in Medicine*.
- D. B. Percival and A. T. Walden (2000): *Wavelet Methods for Time Series Analysis*. Cambridge, U.K.: Cambridge University Press.
- D. B. Percival (2001): Wavelet methods. To appear, A. El-Shaarawi and W. Piegorsch (eds.): *Encyclopedia of Environmetrics*. London: Wiley.
- D. J. Poole and A. E. Raftery (2000): Inference for Deterministic Simulation Models: The Bayesian Melding Approach. *Journal of the American Statistical Association* **95**: 1244–1255.
- P. D. Sampson (2001): Nonstationary spatial covariance modeling. To appear, A. El-Shaarawi and W. Piegorsch (eds.): *Encyclopedia of Environmetrics*. London: Wiley.



- R. P. Schoenberg, D. R. Brillinger and P. Guttorp (2001): Point process, spatial-temporal. To appear, A. El-Shaarawi and W. Piegorsch (eds.): *Encyclopedia of Environmetrics*. London: Wiley.
- L. Sheppard, T. Lumley: Comments on Combining evidence on air pollution and daily mortality from the 20 largest U.S. cities: a hierarchical modeling strategy by Francesca Dominici, Jonathan M. Samet and Scott L. Zeger. *JRSS B* **163**:297.
- L. Sheppard, D. Levy, H. Checkoway (2001): Correcting for the effects of location and atmospheric conditions on air pollution exposure analysis in a case-crossover study. To appear, *J. of Exp. Anal. and Environ. Epi*.
- L. A. Sheppard (2001): Ecological study design. To appear, A. El-Shaarawi and W. Piegorsch (eds.): *Encyclopedia of Environmetrics*. London: Wiley.
- E. A. Steel, P. Guttorp, J. J. Anderson, and D. C. Caccia. Modeling juvenile salmon migration using a simple Markov chain. In press, *Journal of Agricultural, Biological, and Environmental Statistics*.
- M. L. Thompson (2001): Meteorological adjustment of air quality data. To appear, A. El-Shaarawi and W. Piegorsch (eds.): *Encyclopedia of Environmetrics*. London: Wiley.
- G. Van Belle (2000): Composite sampling. To appear, *Environmental and Ecological Statistics*.
- J. C. Wakefield and P. Elliott (1999). Issues in the statistical analysis of small-area health data. *Statistics in Medicine* **18**: 2377–2399.
- J. C. Wakefield and S. E. Morris (1999). An application of spatial errors-in-variables modelling: investigating the relationship between ischaemic heart disease and water constituents. In *Bayesian Statistics 6*; Proceedings of the Sixth Valencia International Meeting, Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. (editors), p. 657–684, Oxford University Press.
- J. C. Wakefield, N. G. Best and L. A. Waller (2000): Bayesian approaches to disease mapping. In *Spatial Epidemiology*. Elliott, P., Wakefield, J.C., Best, N.G. and Briggs, D. (editors), Oxford University Press.
- B. Whitcher, P. Guttorp and D. B. Percival (2000): Multiscale detection and location of multiple variance changes in the presence of long memory. *Journal of Statistical Computing and Simulation* **68**: 65–88.



3.4 Graduating students

The following students, graduating during the time period of this report, were supported by NRCSE at some time during their graduate career.

PhD degrees

Ashley Steel, Quantitative Ecology and Resource Management.

Co-supervisors: Peter Guttorp, Statistics, and Jim Anderson, Fisheries.

Dissertation title: *In-stream factors affecting juvenile salmon out-migration.*

Current position: National Marine Fisheries Laboratory, Seattle, WA

Enrica Bellone, Statistics.

Co-supervisors: Jim Hughes, Biostatistics, and Peter Guttorp, Statistics.

Dissertation title: *Nonhomogeneous hidden Markov models for downscaling synoptic atmospheric patterns to precipitation amounts.*

Current position: Geophysical Statistics Project, NCAR, Boulder, CO

Barnali Das, Statistics.

Supervisor: Peter Guttorp, Statistics.

Dissertation title: *Global covariance modeling: a deformation approach to anisotropy.*

Current position: WESTAT Inc., Washington, DC

Florentina Bunea, Statistics

Supervisor: Jon Wellner, Statistics

Dissertation title: *A model selection approach to partially linear regression.*

Current position: Florida State University, Tallahassee, FL.

MSc degrees

Erin Sullivan, Statistics.

Supervisor: Peter Guttorp, Statistics.

Thesis title: *Estimating the Association between Ambient Particulate Matter and Elderly Mortality in Phoenix and Seattle Using Bayesian Model Averaging.*

Current position: Kaiser Permanente, Portland, OR.

Foreign degrees

Former Center visitors Per Sørensen (Danish Technical University) and Jasha Osterbaan (Ecole des Mines, France) defended their dissertations with external examiners Peter Guttorp and Paul D. Sampson, respectively.



4. Administration

4.1 Executive and advisory committees

4.1.1 Executive committee

The executive committee saw the conclusion of two terms of service: Alison Cullen and David Ford. In a membership election they were each voted in for another three-year term on the executive committee. Due to Dr. Cullen's sabbatical leave during 2000–2001, Loveday Conquest (Fisheries) was chosen as a substitute committee member. Continuing members are Mary Lou Thompson (Biostatistics) and Paul Sampson (Statistics). Much of the work of the executive committee is done via email, although formal decisions and strategic discussions generally require meetings. Meetings are announced with their agendas on the executive committee web page

<http://www.nrcse.washington.edu/people/execom.asp>. The main tasks during this year were the review of internal proposals, and planning for proposals for renewal funding.

4.1.2 Advisory committee

The Advisory committee participated in the internal workshop in January. The discussions focused on directions of change of Center structure, outreach, and research topics.

4.2 Members

Two new members were proposed by the membership and elected by the executive committee during the year. They are Jon Wakefield, Statistics and Biostatistics, and Ray Hilborn, Fisheries.

4.3 Hiring

Using funding from the VIGRE project in Applied Mathematics, Mathematics and Statistics at the University of Washington, we hired Jean-Yves (Pip) Courbois from Oregon State University as a post-doctoral researcher at NRCSE. Dr. Courbois is a specialist in environmental sampling techniques, particularly model-based design.

4.4 Space

There is serious concern that the Statistics department and its various research groups, including NRCSE, will run out of space in the near future. The general space situation at the University is extremely tight, although some of the applied schools and colleges may have some space available. Discussions are presently taking place with Forest Resources, Fisheries, and Applied Physics on providing additional space in the future.



4.5 Relations to other statistical research groups

The Center aims at building a national and international network of environmental statistical research. To that effect we cooperate with a variety of other research groups.

4.5.1 NCAR (National Center for Atmospheric Research)

An NRCSE research assistant, Barnali Das, spent Autumn and Winter quarters 1999–2000 at the Geophysical Statistics Project at NCAR working on the development of statistical methods for data collected on a globe ([section 3.1.8](#)). This visit was jointly funded by NRCSE and NCAR/GSP. The two groups organized a joint workshop on large data sets in July 2000 (see [section 2.3.3](#))

4.5.2 IMPACT

Center members: Peter Guttorp, Paul D. Sampson
Research assistant: Fadoua Balabdaou

NRCSE contributions to the international collaborative “IMPACT” project for the development of tools for “Estimation of Human Impact in the Presence of Natural Fluctuations” are summarized in the project’s first integrated annual report available from <http://www.mai.liu.se/impact/index.html>. Peter Guttorp presented NRCSE research on meteorological adjustment of air quality data at the project’s group meeting in Linköping, Sweden, June 19-20, while Paul Sampson will present results of preliminary analyses of Paris regional air quality data at the project’s group meeting in Fontainebleau, France, Nov 20-21. RA Fadoua Balabdaoui is developing further space-time models of these data in preparation for relating them to meteorological data and for assessing the output of an air quality model providing air quality model predictions for the summer of 1999.

4.5.3 Other research groups

A subcontract with R. Henry at USC will allow a continuation of our work in receptor modeling, focusing on issues of spatio-temporal dependence and comparison to deterministic process models.

5. A view towards the future

During the time following the EPA site visit, considerable effort has gone into rethinking the current form of NRCSE. In the course of developing proposals for continued funding of the Center, we have been considering the educational aspects of it, as well as its role as a national focal point for environmetric work, and the consequent need to involve researchers at other institutions in the work performed here. Several different directions are being considered, and funding will be sought from different sources.



There are several upcoming conferences at which NRCSE are organizing sessions. Sampson, Faustman and Wakefield will be highlighting NRCSE work at the EPA Statistician's meeting in Philadelphia in May. The Joint Statistical Meetings in Atlanta in August will, as has become a tradition, have two NRCSE-organized session. One covers receptor modeling, and the other is concerned with monitoring network design. At the TIES meeting in Portland in August there will also be two NRCSE-organized sessions: one on nonstationary covariance models, and one on meteorological adjustment of air quality data.

In May 2001 there will be a workshop on Spatial Moving Average Models at the University of Washington. This workshop is organized by Dave Higdon, Duke University, and Jay Ver Hoef, Alaska Department of Fish and Game. In June, an NSF-CBMS Regional Conference on Environmental Statistics is planned, also at the University of Washington. The key speaker is Richard Smith from the University of North Carolina, who will be giving ten lectures on the subject. Additional lectures will be given by Peter Guttorp, NRCSE, Doug Nychka, NCAR, Paul D. Sampson, NRCSE, Paul Switzer, Stanford and Jim Zidek, University of British Columbia.



Appendix A. Seminars

Some of the NRCSE seminars during 1999-2000 were videotaped and are available at the Center web page at URL <http://www.nrcse.washington.edu/seminars/nrcse-video.asp>. The following is a list of seminar presentations during the academic year.

Fall Quarter 1999

October 11, 1999: Moshe Pollak, Department of Statistics, The Hebrew University of Jerusalem (Joint with Statistics)

“A Likelihood Approach to Control Charts”

October 21, 1999: Peter Guttorp, Department of Statistics, University of Washington (Joint with the Departments of Statistics and Biostatistics)

“Picture the future—graphical innovation in environmental statistics”

December 3, 1999: Chris Bretherton, Departments of Atmospheric Sciences and Applied Mathematics (Joint with Department of Atmospheric Sciences)

“Statistical Methods for Downscaling GCM Precipitation Predictions over Complex Terrain”

Winter Quarter 2000

January 13, 2000: Thomas Lumley, Department of Biostatistics and NRCSE, University of Washington

“Case-Pseudocontrol Studies – A Free Lunch?”

Spring Quarter 2000

April 27: Stanley Barone Jr., PhD., Research Biologist, Cellular and Molecular Toxicology Branch, Neurotoxicology Division, U.S. Environmental Protection Agency, Research Triangle Park, NC

“Preliminary efforts at incorporating developmental effects of exposure to chlorpyrifos into a biologically-based dose response model”

May 11: Richard J. Jackson, MD, MPH., Director, National Center for Environmental Health Centers for Disease Control

“Public Health and Environmental Protection: Unfortunate Rivals, Unrivaled Partners”



May 18: Douglas Bell, PhD., Head, Genetic Risk Group, Laboratory of Computational Biology and Risk Analysis, National Institute of Environmental Health Sciences

“Polymorphism in Carcinogen Metabolism and DNA Repair: Modulation of Exposure Induced Damage and Disease”

May 25: Rob McConnell, MD, Associate Professor Division of Occupational and Environmental Health, Department of Preventive Medicine, University of Southern California

“Asthma, Lung Function Growth, and Air Pollution: Results from the Southern California Children’s Health Study”

June 1: Dan Costa, Sc.D., Pulmonary Toxicology Branch, Experimental Toxicology Division, National Health and Environmental Effects Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC

“The Toxicology of Ambient Particle Matter: Links to the Epidemiology”

Summer Quarter 2000

August 24: David Vere-Jones, Victoria University of Wellington and Statistics Research Associates, and David Harte, Statistics Research Associates, New Zealand

“Modeling for Earthquake Forecasts: Point Process Models and Associated Software”



Appendix B. Technical reports 1999-2000

TRS number 55 (2000)

Influence of Large Scale Circulation Measures on Precipitation at Local Stations in the South East of the US

Claudia Tebaldi

TRS number 54 (2000)

Estimating the Association between Ambient Particulate Matter and Elderly Mortality in Phoenix and Seattle Using Bayesian Model Averaging

Erin M. Sullivan

TRS number 53 (2000)

The Method of Synthesis in Ecology

E. David Ford and Hiroaki Ishii

TRS number 52 (2000)

Limitations to Empirical Extrapolation Studies: The Case of BMD ratios

Kevin P. Brand, Paul J. Catalano, James K. Hammitt, Lorenz Rhomberg and John S. Evans

TRS number 51 (2000)

Compositional Receptor Modeling

Dean Billheimer

TRS number 50 (2000)

A Comparison on Consistency of Parameter Estimation Using Optimization Methods for a Mixture

Marianne C. Turley and E. David Ford

TRS number 49 (2000)

The Impact of Wavelet Coefficient Correlations on Fractionally Differenced Process Estimation

Peter F. Craigmile, Donald B. Percival and Peter Guttorp

TRS number 48 (2000)

Setting environmental standards: A statistician's perspective

Peter Guttorp

TRS number 47 (2000)

Wavelet-Based Parameter Estimation for Trend Contaminated Fractionally Differenced



Processes

Peter F. Craigmile, Donald B. Percival and Peter Guttorp

TRS number 46 (2000)

ORCA: A Visualization Toolkit for High-Dimensional Data

Peter Sutherland, Anthony Rossini, Thomas Lumley, Nicholas Lewin-Koh, Dianne Cook,
Zach Cox

TRS number 45 (2000)

Compactly Supported Correlation Functions

Tilmann Gneiting

TRS number 44 (2000)

Developing an Efficient Surveillance Scheme for Assessing Compliance with Air Quality
Standards

Ronit Nirel

TRS number 43 (2000)

Multivariate Receptor Modeling for Temporally Correlated Data by Using MCMC

Eun Sug Park, Peter Guttorp and Ronald C. Henry

TRS number 42 (2000)

Quality Assurance of Environmental Models

Alice Shelly, E. David Ford and Bruce Beck

TRS number 41 (2000)

Statistical Issues in the Study of Air Pollution Involving Airborne Particulate Matter

Lawrence H. Cox

TRS number 40 (2000)

Effects of Ambient Fine and Coarse Particles On Mortality In Phoenix, Arizona

Merlise A. Clyde, Peter Guttorp and Erin Sullivan

TRS number 39 (2000)

Bayesian Estimation of Semi-Parametric Non-Stationary Spatial Covariance Structures

Doris Damian, Paul D. Sampson and Peter Guttorp

TRS number 38 (2000)

Mathematical Background for Wavelet Estimators of Cross-Covariance and Cross-
Correlation

Brandon Whitcher, Peter Guttorp and Donald B. Percival



TRS number 37 (2000)
MCMC in $I \times J \times K$ contingency tables
Florentina Bunea and Julian Besag

TRS number 36 (1999)
Ecological Indices and Graphical Modeling of Factors Influencing Benthic Populations in
Streams
Florentina Bunea, Peter Guttorp and Thomas Richardson

TRS number 35 (1999)
Estimating Short-term PM Effects Accounting for Surrogate Exposure Measurements
from Ambient Monitors
Lianne Sheppard and Doris Damian

TRS number 34 (1999)
Determining the Number of Major Pollution Sources in Multivariate Air Quality Recep-
tor Models
Eun Sug Park, Ronald C. Henry and Clifford H. Spiegelman



Appendix C. Conference presentations

Oct. 1999 T. Gneiting: Matheron's Hankel group - an algebraic gem in geostatistics. Fields Institute, Toronto (Canada).

Dec. 1999 T. Gneiting: Correlation models in spatial statistics and positive definite functions. Portland State University, OR.

Dec. 1999 S. M. Bartell, R. P. Ponce, W. C. Griffith, and E. M. Faustman: Temporal Fallacies in Biomarker Based Exposure Inference. Society for Risk Analysis Annual Meeting, Atlanta, GA.

Dec. 1999 S. M. Bartell, J. H. Shirai, C. H. Pierce, and J. C. Kissel: Estimation of Childhood Soil Ingestion Rates Using a Probabilistic Toxicokinetic Lead Model. Society for Risk Analysis Annual Meeting, Atlanta, GA.

Dec. 1999 S. M. Silbernagel, D. A. Grace, and E. M. Faustman: Nuclear Waste Transportation—A Case Study on Identifying Risk Information Needs. Society for Risk Analysis Annual Meeting, Atlanta, GA.

Dec. 1999 W. C. Griffith: Use of semiparametric statistical methods to model environmental transport of contaminants. Society for Risk Analysis Annual Meeting, Atlanta, GA.

Dec. 1999 W. C. Griffith, K. McCarthy, E. Faustman, J. Moore: Evaluation of Hanford Cleanup Certification Packages to Support Records of Decision. Society for Risk Analysis Annual Meeting, Atlanta, GA.

Jan. 2000 S. Liu, J. Koenig, D. Kalman, J. Kaufman, T. Larson, L. Sheppard: PM exposure assessment in high-risk subpopulations. PM 2000: Particulate Matter and Health. Charleston, SC.

Jan. 2000 T. Lumley, D. Levy, L. Sheppard: Design Bias in Case-Crossover Analyses of Acute Health Effects of Air Pollution. PM 2000: Particulate Matter and Health. Charleston, SC.

Jan. 2000 L. Sheppard, D. Levy, H. Checkoway: Teasing Apart the Role of Location and Atmospheric Conditions in Air Pollution Exposures for Health Effect Analyses: Results from the CABS Air Pollution Exposure Substudy. PM 2000: Particulate Matter and Health. Charleston, SC.



Jan. 2000 M. Clyde, P. Guttorp and E. Sullivan: Effects of Ambient Fine and Coarse Particles on Mortality in Phoenix, Arizona. PM 2000: Particulate Matter and Health. Charleston, SC.

Jan. 2000 L. Sheppard, D. Damian, M. S. Kaiser, M. Daniels: Incorporating Spatial Predictions of Ambient Particulate Matter into an Analysis of Asthma Hospital Admissions. PM 2000: Particulate Matter and Health. Charleston, SC.

Jan. 2000 J. Vandenberg, M. Brauer, A. Cullen, E. Fabianova, F. Hrubá, M. Lendacka, E. Mihalikova, P. Miskovic, A. Plzikova: Measuring human exposures to priority air pollutants in Slovakia. PM 2000: Particulate Matter and Health. Charleston, SC.

Jan. 2000 O. Yu, L. Sheppard, T. Lumley, J. Q. Koenig, G. G. Shapiro: Effects of Ambient Carbon Monoxide and Atmospheric Particles on Asthma Symptoms: Results from the CAMP Air Pollution Asthma Study. PM 2000: Particulate Matter and Health. Charleston, SC.

Jan. 2000 T. F. Mar, J. Q. Koenig, T. V Larson, L. Sheppard, R. A. Stier and C.S. Claiborn The association between air pollution and peak expiratory flow in asthmatics in Spokane, Washington. PM 2000: Particulate Matter and Health. Charleston, SC.

Feb. 2000: P. Guttorp: Environmental standards: A statistical approach. UC Santa Barbara, CA.

Mar. 2000 L. Conquest: Incorporating Judgement into Ecological Sampling. University of Uruguay in Montevideo, Uruguay.

Mar. 2000 J. Wakefield: Modeling spatial variation in risk, ENAR meeting, Chicago, IL.

Apr. 2000 J. Wakefield: Modeling spatial variation in risk. Pacific Northwest Statistics Meeting, UBC, Vancouver, Canada.

Apr. 2000 A. Raftery: Inference for Deterministic Simulation Models: The Bayesian Melding Approach. Conference on the Statistical Analysis of Computer Codes, Gregynog, Wales.

Apr. 2000 S. Bates: Bayesian Assessment of Uncertainty in Deterministic Environmental Exposure Models. The Utility of Bayesian Decision Analysis and Environmental Problems. Interface 2000.

Apr. 2000 T. Gneiting: Covariance functions for spatial and spatio-temporal data: recent developments. 6th International Geostatistics Congress, Cape Town (South Africa).



May 2000 P. Guttorp: Setting environmental standards—A statistician's approach. Statistics: Reflections on the past and visions for the future. Conference in honor of C. R. Rao's 80th birthday.

May 2000 P. Craigmile: Wavelet Based Parameter Estimation of Trend Contaminated Long Memory Processes. Bernoulli World Congress, Guanajuato, Mexico.

May 2000 B. Das: Estimating Global Temperature using Anisotropic Global Covariance Functions. Bernoulli World Congress, Guanajuato, Mexico.

May 2000 F Bunea: A Model Selection Approach to Partially Linear Regression. Bernoulli World Congress, Guanajuato, Mexico.

Jun. 2000 L. Conquest: Analysis of Short Repeated Measures Series from Designed Experiments. Oceanic Institute, Waimanalo, HI.

Jun. 2000 T. Gneiting: Criteria of Pólya type for radial positive definite functions. Université d'Angers, France.

Jul. 2000 P. Craigmile: Decorrelation Properties of Wavelet Based Estimators for Fractionally Differenced Processes. Wavelet Applications in Signal Processing minisymposium, 3rd European Congress in Mathematics, Barcelona, Spain.

Jul. 2000 T. Gneiting: Covariance functions for spatial and spatio-temporal data: recent developments. International Conference on Spatial Statistics in the Agro-, Bio- and Geosciences, Freiberg (Germany).

Aug. 2000 S. Bates and A. Raftery: Assessing Deterministic Environmental Exposure Models. Joint Statistical Meetings, Indianapolis, IN.

Aug. 2000 P. Guttorp, P. D. Sampson, D. Damian, S. Mitra and W. Meiring: A Covariance-Based Approach to Assessment of Environmental Air Pollution Models. Joint Statistical Meetings, Indianapolis, IN.

Aug. 2000 E. D. Ford: Assessment of deterministic models in the ecological and environmental sciences. Joint Statistical Meetings, Indianapolis, IN.

Aug. 2000 M. Handcock, J. Sedransk and A. Olsen: Ecological Assessment of Riverine Systems by Combining Information from Multiple Sources. Joint Statistical Meetings, Indianapolis, IN.



Aug. 2000 J.-Y. Courbois: Horvitz-Thompson based estimators for finite population variance components. Part I: The population variance. Joint Statistical Meetings, Indianapolis, IN.

Aug. 2000 T. Lumley: Is it true, is it kind, is it necessary? International Society for Environmental Epidemiology meeting, Buffalo, NY.

Sep. 2000 P. Sampson: Developments in the Modeling of the Nonstationary Spatial Covariance Structure of Environmental Processes. TIES/SPRUCE 2000. Sheffield, UK.

Sep. 2000 J. Wakefield: A critique of ecological studies. Imperial College, London, UK.

Sep. 2000 J. Wakefield: A critique of ecological studies. European meeting on spatial and computational statistics, Ambleside, UK.

Sep. 2000 E. Park: Multivariate Receptor Models and Model Uncertainty. Fourth International Conference on Environmetrics and Chemometrics, Las Vegas, NV.

Sep. 2000 T. Lumley: Visualising high-dimensional data in time and space: ideas and tools from the Orca Project. Fourth International Conference on Environmetrics and Chemometrics, Las Vegas, NV.

Sep. 2000 L. Conquest: Incorporating Judgment in Ecological Sampling. Fourth International Conference on Environmetrics and Chemometrics, Las Vegas, NV.

Appendix D. Workshop agendas

D1. EPA Las Vegas

Tuesday, Dec 14

10:15 Peter Guttorp (Statistics): *Research at NRCSE*

11:00 Thomas Lumley (Biostatistics): *Orca: A toolkit for visualizing structured high-dimensional data*



11:45 Eun Sug Park (NRCSE): *Multivariate receptor modeling for temporally correlated data using MCMC*

2:00 David Ford (Forestry): *Pareto optimal model assessment*

2:45 Adrian Raftery (Statistics): *Inference for deterministic simulation models: The Bayesian melding approach*

3:30 Discussion of statistical issues for site characterization and assessment

Wednesday, Dec 15

9:15 Loveday Conquest (Fisheries): *Use of ranked set sampling in stream research*

10:00 Discussion of statistical issues for landscape ecological assessments

1:00 Alison Cullen (Public Affairs): *Exposure assessment - A tale of two surveys*

1:45 Tom Lewandowski (Environmental Health): *Linked toxicodynamic and toxicokinetic model for developmental neurotoxicity*

2:30 Discussion of toxicokinetic and toxicodynamic modeling



D2. Exposure assessment in environmental and occupational health

Donovaly, Slovakia, October 25-26, 1999

Organizers

Michael Brauer, UBC; Alison Cullen, NRCSE; John Vandenberg, U.S. EPA

Program

Monday, Oct. 25

Session 1: Problem definition and study design

Alison Cullen, Michael Brauer and Frantiska Hrubá

Session 2: Data collection and chemical analysis

Michael Brauer, Eva Mihalíková and Peter Miskovic

Session 3: Data analysis

Alison Cullen, Michael Brauer and Kaja Hrubá

Session 4: Poster display

Tuesday, Oct. 26, 1999

Session 5: Results interpretation and risk characterization

Alison Cullen, Eleonora Fabianova and John Vandenberg

Session 6: Reporting results and risk communication

Eleonora Fabianova and John Vandenberg

Session 7: Discussion and implications



D3. Large Data Sets

NCAR, Boulder, Colorado, July 24-26, 2000

Organizers:

Di Cook, Iowa State, Chris Wikle, University of Missouri, David Madigan, Soliloquy Inc., Doug Nychka, NCAR GSP, Peter Guttorp, NRCSE

Program

Monday, July 24

8:35 - 9:35 - Jerry North, Texas A&M University: "Some Estimation Problems Utilizing Large Climate Data Sets"

9:35 - 10:05 - Lawrence Buja, NCAR "Community Climate System Model (CCSM) Data"

10:30 - 11:30 - Di Cook, Iowa State University "Issues and Approaches for Visualization of Large Multi-Dimensional Data"

1:00 - 2:00 - Padhraic Smyth, U. California-Irvine "Part I: What is Data Mining?"

2:00 - 2:45 - Alexey Kaplan, Lamont Doherty Earth Observatory, Columbia U. "Least-squares optimal analyses of historical climate data sets I: Problem set-up and existing solutions"

3:15 - 4:00 - Alexey Kaplan, Lamont Doherty Earth Observatory, Columbia U. "Least-squares optimal analyses of historical climate data sets II: Difficulties and prospects"

Tuesday, July 25

8:30 - 9:30 - Dan Carr, George Mason University "Several Templates for Looking at Large Georeferenced Data Sets"

9:30 - 10:30 - Marina Meila, Carnegie Mellon University and the University of Washington, "Fast Algorithms for Learning Tree Graphical Models in High Dimensions"

11:00 - Noon - Hsin-Cheng Huang, Institute of Statistical Science, Academia Sinica, "Fast Spatial Prediction of Global Processes from Satellite Data"

1:30 - 2:30 - Mark Gahegan, Penn State University "Using Expertise to Guide Geoscientific Visualization"



National Research Center for
Statistics and the Environment



Wednesday, July 26

8:30 - 9:30 - Dave Higdon, Duke University "Building Dependence Structure for Large Space-Time Datasets"

9:30 - 10:00 - Tim Hoar, NCAR "Getting to know a large dataset: Satellite Observations of surface quantities."

10:30 - 11:30 - Padhraic Smyth, University of California, Irvine "Part II- Data mining: The potential role of data mining in atmospheric and environmental sciences"



D4. Internal workshop

January 20, 2000

SPEAKER SCHEDULE

8:30 am--8:45	Introduction	Peter Guttorp
8:45	Paul Sampson	Covariance modeling
9:10	Thomas Lumley	Extending data visualization to structured data: the Orca project
9:35	Mary Lou Thompson	Standards
10:00--10:45	Coffee break and posters	
10:45 am	Lianne Sheppard	Health effects of PM
11:15--12:15	Small group discussion	
1:45 pm	Alison Cullen	The Slovakia project
2:15	Adrian Raftery	Statistical Inference for Deterministic Simulation Models: The Bayesian Melding Approach
2:45	Ashley Steel	The Truth About Science: A Hands-On Scientific Research Curriculum
3:15--4:00	Coffee break and posters	
4:00--5:00	Small group discussion	
5:00--5:30	Wrap-up	Larry Cox

POSTER PRESENTATIONS:

Enrica Bellone, Jim Hughes and Peter Guttorp
A stochastic model for precipitation amounts at multiple stations

Dean Billheimer
Compositional Receptor Modeling

Elaine Faustman
Linking Toxicokinetic and Toxicodynamic Models for Methylmercury Developmental Toxicity

William Griffith
Temporal Fallacies in Biomarker Based Exposure Inference

Patrick Heagerty
Spatial Transition Models and Forecasting of Gypsy Moth Defoliation

Nick Hedley, Tim Nyerges



Thomas Lumley

Air pollution time series: case-crossover analyses and other difficulties.

Nicolle Mode, Loveday Conquest, and David Marker

Ranked Set Sampling for Ecological Monitoring: Costs, Comparisons and Compromises

Kerrie Nelson

Statistical methods for modeling multiply censored data

Don Percival, Peter Craigmile and Peter Guttorp

Wavelet-based trend detection and estimation

Eun Sug Park

Multivariate Receptor Modeling for Temporally Correlated Data by Using

MCMC

Chris Bretherton

Variations in Pacific Northwest Snowpack and Regional Climate—Past and Future



D5. Teaching Environmental Statistics at the UW

Attendees: Joyce Cooper (Mechanical Engineering), Peter Guttorp (Statistics), June Morita (Bothell Interdisciplinary Arts and Sciences), Don Percival (Applied Physics Laboratory), Marcia Ciol (CQS), Mary Lou Thompson (Biostatistics), Bruce Bare (College of Forest Resources), Loveday Conquest (Fisheries), Craig Zumbrunnen (Geography), Sally Liu (Public Health), Suzanne Withers (Geography), Alison Cullen (Public Affairs), Christy Howard (CQS)

Who is the audience for Environmental Statistics courses?

- Undergraduate Program on the Environment Students
 - Required to take one basic statistics course and one capstone course
- Evans School Master's students
 - Required to take one course in analysis beyond Evans 2 introductory stat courses
 - Roughly 20 students each year are environmental concentrators.
- Natural Sciences grad students - CFR, Fisheries, SMA, Atmospheric Science, QERM, etc.
- Environmental Management Certificate Program graduate students from throughout UW
 - Must take 2 electives in some area of environmental management, analysis, etc.
 - Need abilities in critical study design and critiquing design, not just methods.
 - Develop skills in strategic planning and decision making

I. Offerings at the UW

- A. Geography 426 - intro to use of stat in geography emphasize what's inappropriate to use in analysis weekly computer lab session with instructor who demonstrates how to run computer analysis and interpret the data. Students are given data, they run analysis, and interpret results.
- B. Geography 326 - elementary statistics up to regression, focused on research design.
- C. Statistics Department - Environmental Statistics - taught by Peter Guttorp in past years
 1. Case-based course - ASARCO smelter in Tacoma, Port Townsend Paper Mill, among others.
 2. Students expected to propose research design based on cases, propose research questions, analyze data.
 3. Can be effective as undergraduate or graduate course - Uses same



- cases for both, but adjusts the depth of the issues explored accordingly
4. Not offered recently due to lack of student interest

II. Areas of Need

- A. Spatial Statistics - No undergraduate courses offered currently. Not offered at a low enough level to be accessible.
- B. Correlated Data - There are only 2 biostat courses at the graduate level on correlated data, but they are very specific to medicine.
- C. Time Series
- D. Temporal & spatial Correlation
- E. Multivariate -with applications in Natural Sciences: Ecology, Biology, Fisheries, Forestry
- F. Risk Analysis/Decision-making – proposed as a future course by EM certificate program and also the Evans School of Public Affairs
- G. GIS with integration of spatial statistics and GIS, many GIS courses are offered however

III. Problems in Teaching Statistics

- A. Teach out of date theory from decades ago
- B. Courses can be boring unless examples and cases are aptly chosen
- C. Many courses rely on basic statistical knowledge, for which students tend to be ill-prepared.
- D. In CQS, they teach service courses in statistics
 - 1. Students are heterogeneous - with varied backgrounds and interests
 - 2. These are a requirement for many students, not all want to be there
 - 3. It is hard to make course relevant and interesting for everyone.
 - 4. Instructors seek more connections with researchers/faculty in other department to help find relevant data and research topics for student's term projects.

IV. Suggestions for improvements

- A. Develop web site as a clearinghouse for statistics course information
 - 1. Each department/faculty who teach statistics course submit comprehensive syllabus describing what is taught in the course. The syllabi could then be combined into one master listing and posted on the web.



- B. Make up an information packet with techniques and application topics as a resource for instructors.
- C. More outreach to advisors so they know what statistics courses are available and inform their advisees about them.
- D. Develop a repository of data set on environmental topics and/or relevant published articles.
- E. Especially in the CQS stat course, connect students with someone on campus who has done statistical research in the student's area of interest. Students could receive guidance from this person in conducting their own analyses for their course term project.
- F. Develop a Speakers Bureau to bring in researchers to give talks on how they apply statistics in their work.
 - 1. Make this available and accessible to undergraduate students to give them a sense of why statistics is important and how it is utilized.
- G. Need more money to implement many of these ideas!
 - 1. NRC just released a report that strengthens the link between statistical sciences and mathematical sciences.
 - 2. Important to know about this report and cite it in grant applications, as leverage verifying the importance of statistics at the University.

V. Proposed Program Level Changes

- A. Discussion of offering an Environmental Statistics Certificate Program for graduate students.
- B. CFR is probably offering 3 new courses - one in GIS/Intro to ArcView, and two in spatial analysis.



E. Course description for EPA Region X Risk Assessment Course

Instructors: Elaine Faustman, Scott Bartell and Bill Griffith

Each of the 21 sessions below will devote the first 20-25 minutes to a discussion of the concept followed by 10 minutes of examples of applications, 10-15 minutes for the class members to apply what they have learned to an exercise, and 10-15 minutes of discussion

August 7

Introduction (45 min)

Risk Assessment Framework

How to formulate questions that statistical methods can assist in answering

Describing Populations

1. General Methods (1 hr 30 min)

Sampling

Measures of central tendency and variability

Graphical Techniques

2. Parametric/Nonparametric Methods (1 hr)

General Tools

Detection Levels

Graphical Techniques

Comparison of Populations

3. Multiple Populations (1 hr)

ANOVA

Corrections for multiple comparisons

4. Statistical power to make comparisons (1 hr)

5. Estimating differences between populations and power (1 hr)

6. Upper Confidence limits and interpretations (1 hr)

August 8

Combining Distributions of Populations

7. Monte Carlo Simulation methods (50 min)

8. Applications in Risk Assessment (50 min)

9. Two Dimensional Monte Carlo (50 min)

Nonparametric Methods

10. Classical methods for comparing 2 populations (50 min)

11. Randomization methods for making comparisons (50 min)

12. Graphical methods for describing distributions of measurements (50 min)



Regression

- 13. Classical regression methods (50 min)
- 14. Confidence limits on regression (50 min)

August 9

Regression

- 15. Evaluating Regression models using residuals (50 min)

Toxicology

- 16. Survival Analysis for censored data (50 min)
- 17. Estimating age specific rates using Kaplan-Meier methods (50 min)
- 18. Comparing two populations using Cox models (50 min)
- 19. Meta analysis of multiple studies (50 min)
- 20. Analysis of noncancer studies with multiple endpoints (50 min)
- 21 Benchmark Dose(50 min)

Summary and evaluation