FINAL REPORT: SAMSI PROGRAM ON SPACE-TIME ANALYSIS FOR ENVIRONMENTAL MAPPING, EPIDEMIOLOGY AND CLIMATE CHANGE

Edited by Alan Gelfand, Montse Fuentes and Richard Smith

May 29, 2011

Organizers: Noel Cressie (Ohio State), Michael Stein (University of Chicago), Dongchu Sun (University of Missouri), Jim Zidek (University of British Columbia).

Scientific Advisory Committee: Peter Diggle (Lancaster University), Peter Guttorp (University of Washington), Jesper Møller (Aalborg)

Directorate Liaison: Jim Berger (Richard Smith after 07/01/2010)

Local Scientific Coordinators: Montse Fuentes (NCSU), Alan Gelfand (Duke), Richard Smith (UNC).

National Advisory Committee Liaison: Jun Liu (Harvard).

This 12 month SAMSI program focussed on problems encountered in dealing with random space-time fields, both those that arise in nature and those that are used as statistical representations of other processes. The sub-themes of environmental mapping, spatial epidemiology, and climate change are interrelated both in terms of key issues in underlying science and in the statistical and mathematical methodologies needed to address the science. Researchers from statistics, applied mathematics, environmental sciences, epidemiology and meteorology were involved, and the program promoted many opportunities for interdisciplinary, methodological and theoretical research.

The program began with a Summer School on Spatial Statistics (July 28-August 1, 2009, at SAMSI), which featured invited lectures and hands-on tutorials from Sudipto Banerjee (U. Minnesota), Reinhard Fürrer (U. Zurich), Doug Nychka (National Center for Atmospheric Research) and Stephen Sain (National Center for Atmospheric Research).

The Opening Workshop took place from September 13–16, 2009, and featured the following tutorial lectures:

- Alan Gelfand, Duke University: Hierarchical Modeling for Analyzing Space and Space-time Data
- Jim Zidek, University of British Columbia: Designing Monitoring Networks

- Michael Stein, University of Chicago: Asymptotics for Spatial and Spatial-temporal Data
- Sylvia Richardson, Imperial College London: Introduction to Spatial Epidemiology

These were following by a workshop that included five invited paper sessions on "Climate Change", "Random Fields and Applications", "Spatial Epidemiology", "Spatial and Time-space Point Processes" and "Interface of Deterministic Modeling and Space-time Statistics", as well as two New Researcher sessions.

On Thursday, September 17, 2009, SAMSI was host to a visit from several scientists at the National Climatic Data Center (NCDC). This included presentations from Matt Menne and John Bates of NCDC.

Other workshops organized in connection with the program were:

- SAMSI was joint organizer of the GEOMED 2009 workshop, Nov. 14–16, 2009, at the Medical University of South Carolina
- Climate Change Workshop, at SAMSI — February 17-19, 2010
- One-day Workshop on Objective Bayesian for Spatial and Temporal Models, in San Antonio, TX — March 20-21, 2010
- Statistical Aspects of Environmental Risk, at SAMSI — April 7-9, 2010
- Transition Workshop, at SAMSI — October 11-13, 2010

Three graduate level courses were taught as part of the program:

- Spatial Epidemiology (Fall);
- Theory of Continuous Space and Space-Time Processes (Fall);
- Spatial Statistics in Climate, Ecology and Atmospherics (Spring).

As usual in SAMSI programs, the bulk of the research activity centered around the Working Groups. Nine Working Groups were formed, and remainder of this report is concerned with reporting the research activities of each of these groups.

1 WG1: Paleoclimate

1.1 Goals and Outcomes of the Working group

A major challenge in climate work is the fact that surface temperatures before about 1850 can only be indirectly inferred from temperature-sensitive geological proxy data. The same is true of other climate variables, such as precipitation and the pressure field. An understanding of the climate of the past is necessary in order to understand the natural variability of the system, and to place observed changes in the system over the last century, and changes predicted for the future, in a broader context. Estimates of the past evolution of climate variables such as the surface temperature space-time field, along with estimates of the associated uncertainty, can be used to assess the extent to which currently observed changes are unusual with respect to the natural variability of the system, and to make probabilistic assessments about dynamical hypotheses of how climate varies and changes.

While much work has been done on this sort of problem, there remains room for substantial improvement in the statistical techniques used to reconstruct past climate from proxy data. The working group has three main objectives within which there are many sub-objectives:

1. Develop new and/or improve existing statistical methods for three different reconstruction problems, where each will involve combining incomplete paleoclimatic data and instrumental observations:
   (a) The reconstruction of univariate climate field, such as temperature.
   (b) The reconstruction of a climate index, such as ENSO or the AO, that has major impacts on weather and climate on a global scale.
   (c) The reconstruction of a multivariate climate field, such as temperature and precipitation.
2. Test these methods against existing ones and objectively assess their performance.
3. Produce new paleoclimatic analyses for the last several hundred to several thousand years, beginning with surface temperatures over different spatial domains.

1.2 Summary of working group Activities: 2009 Sep–2010 October

1.2.1 Membership

Core members: Peter F. Craigmile (Ohio State); Murali Haran (Penn State); Bo Li - Purdue University Elizabeth Mannshardt-Shamseldin (Duke); Martin P. Tingley (SAMSI postdoc, now at NCAR); Bala Rajaratnam (Stanford — chair of working group).

All the above were in residence during Fall 2009.

Other members: Priscilla Greenwood (Arizona State); Rajib Paul (Western Michigan)

1.2.2 Summary of administrative working of group activities

- Standard Meeting Time: Thursday 1pm EST
- Weekly meetings during the academic year 2009-2010 and now biweekly meetings
- First semester Activities ("soul searching"): 1) Presentations from both statisticians and environmental/geoscientists and consultations with experts. 2) Identification of problem areas deemed to be important to the paleoclimate community. 3) Explore the feasibility of studying problems that were identified
- Second semester (and Summer) Activities: 1) Preliminary discussions of joint research work

  I. Paleoclimate exploratory data analysis

  II. Discussions of a theoretical nature and presentation of whether in the paleoclimate context one should use separable covariance functions or not

- Contributions at Climate Change workshop (February, 2010)
- Development and Writing of paper that is currently in review
- JSM (2010) representation: Joint session of paleoclimate and extremes working groups

1.2.3 Paleoclimate Working Group Research Output

The group has been extremely productive and there is a tangible, approximately 100 page research output from the working group in the following paper:

Tingley, M.P., Craigmile, P.F., Haran, M., Li, B., Mannshardt-Shamseldin, E. and Rajaratnam B., "Piecing together the past: Statistical insights into paleoclimatic reconstructions". Under review at *Journal of Climate*.

Available at http://statistics.stanford.edu/∼ckirby/techreports/GEN/2010/2010-09.pdf

The paper identifies all the issues (as we see them) in the language of modern day statistics. The paper builds on recent work in the paleoclimate literature using Bayesian hierarchical models — gives a unifying framework. The "grand goal is to move paleoclimate reconstructions away from simple regression based approaches towards hierarchical models." The aims of the paper are:

1. Establish a modeling and notational framework that is transparent to both the earth science and statistics communities.
2. Outline and distinguish between scientific and statistical challenges and indicate where modern statistics can contribute.
3. Offer some suggestions for model construction and explain how to perform the required statistical inference.
4. Identify issues that are important to both the earth science and applied statistics communities, and encourage greater collaboration between the two.

1.2.4 Other Research Activities

Craigmile, Peter and Bala Rajaratnam. DISCUSSION OF: A STATISTICAL ANALYSIS OF MULTIPLE TEMPERATURE PROXIES: ARE RECONSTRUCTIONS OF SURFACE TEMPERATURES OVER THE LAST 1000 YEARS RELIABLE? The Annals of Applied Statistics 2011, Vol. 5, No. 1, 88–90.

Haran, Murali and Nathan M. Urba. DISCUSSION OF: A STATISTICAL ANALYSIS OF MULTIPLE TEMPERATURE PROXIES: ARE RECONSTRUCTIONS OF SURFACE TEMPERATURES OVER THE LAST 1000 YEARS RELIABLE? The Annals of Applied Statistics 2011, Vol. 5, No. 1, 61–64

Nychka, Douglas and Bo Li. DISCUSSION OF: A STATISTICAL ANALYSIS OF MULTIPLE TEMPERATURE PROXIES: ARE RECONSTRUCTIONS OF SURFACE TEMPERATURES OVER THE LAST 1000 YEARS RELIABLE? The Annals of Applied Statistics 2011, Vol. 5, No. 1, 80–82.

Smith, R.L. (2010), Understanding sensitivities in paleoclimate reconstructions. Submitted for publication.

Tingley, Martin P. SPURIOUS PREDICTIONS WITH RANDOM TIME SERIES: THE LASSO IN THE CONTEXT OF PALEOCLIMATIC RECONSTRUCTIONS. DISCUSSION OF: A STATISTICAL ANALYSIS OF MULTIPLE TEMPERATURE PROXIES: ARE RECONSTRUCTIONS OF SURFACE TEMPERATURES OVER THE LAST 1000 YEARS RELIABLE? The Annals of Applied Statistics 2011, Vol. 5, No. 1, 83–87

Tingley, M.P. and Li, B. (2011), Comments on "Reconstructing the NH mean temperature: Can underestimation of trends and variability be avoided?" by Bo Christiansen; under review at *Journal of Climate*.

1.3 List of Working group members

Mark Berliner, Ohio State University
K Sham Bhat, Penn State
Peter Craigmile, Ohio State University
Noel Cressie, Ohio State University
Cindy Greenwood, Arizona State University
Michele Guindani, University of New Mexico
Murali Haran, Penn State
John Harlim, Courant Institute/NCSU
Gardar Johannesson, Lawrence Livermore National Laboratory
Karen Kafadar, Indiana University
Bo Li, Purdue
Ernst Linder, Univeristy of New Hampshire
Elizabeth Mannshardt-Shamseldin, SAMSI/Duke
Doug Nychka, NCAR
Garritt Page, Duke
Rajib Paul, Western Michigan University
Bala Rajaratnam, STanford
Brian Reich, NCSU
Steven Roberts, Australian National University
Julia Salzman, Stanford
Jenise Swall, EPA
Richard Smith, UNC/SAMSI
Carolyn Snyder, Stanford
Martin Tingley, SAMSI/NCAR
Saeid Yasamin, SAMSI
Ian Wong, Stanford
Jun Zhang, SAMSI

2 WG2: Spatial Exposures and Health Effects (Montse Fuentes, Amy Herring and Brian Reich)

2.1 Summary

This working group developed and implemented spatial-temporal statistical models to study the impact under climatic change conditions of air pollution on human health. We introduced Bayesian multivariate spatial dependence structures in the environmental stressors, climatic variables, and health outcomes, while taking into account different sources of uncertainty in models and data. We developed novel spatial quantile regression models for the climatic and pollution variables for better characterization of extremes, tail behavior, and complex dependences between weather and pollution. We developed Bayesian hierarchical shrinkage methods for assessing spatial associations between complex pollutant mixtures and health outcomes. We improved upon existing approaches by simultaneously accounting for different pollutant types, such as ozone and particulate matter (PM) or speciated PM, characterizing the spatial temporal structure of the susceptible periods of fetal development (for pregnancy outcomes) and the exposure lag (for mortality outcome), while taking into account different sources of uncertainty in models and data.

This work had led to several papers, and the submission of an NIH proposal, as well as several conference presentations.

NIH Proposal: Space-time Modeling for Linking Climate Change, Pollutant Exposure, Built Environments, and Health Outcomes. PI: Montserrat Fuentes, co-PIs: Amy Herring, Alan Gelfand, Brian Reich and Marie Lynn Miranda. $1.1M, NIH R-01 submitted

For the remainder of the report of this working group, we describe the outcomes of several individual projects that were conducted as part of the working group's activities.

2.2 Estimating Individual Activity Spaces (Leader: Catherine Calder)

In this collaboration with sociologists and geographers, we are developing methods to estimate individual activity spaces, or the space-time context for human exposures, and examining the consequences of exposure to violence and disadvantage on health and behavioral outcomes.

Papers in preparation:

- Calder, C.A. and Darnieder, W.F. Bayesian Inference for Incomplete Marked Spatial Point Patterns.
- Krivo, L.J., Washington, H.M., Peterson, R.D., Browning, C.R., Calder, C.A., and Kwan, M.-P. Social Isolation of Disadvantage and Advantage: The Reproduction of Inequality in Urban Space.

Presentation:

- C.A. Calder, "Bayesian Estimation of Individual Activity Spaces from Incomplete Activity Pattern Data" Conference on the Dynamics of Space-Time Use, The Ohio State University, Columbus, OH, October 2-3, 2009

- C.A. Calder, "Bayesian Inference for Incomplete Marked Spatial Point Patterns: Estimating Individual Activity Spaces". SAMSI Program on Space-Time Analysis Closing Workshop, October 11-13, 2010

Posters:

- "Bayesian Inference for Incomplete Marked Spatial Point Patterns" by C.A. Calder. Presented at Valencia/ISBA Meeting, Benidorm, Spain, SAMSI Program on Space-Time Analysis Closing Workshop, June 3–8, 2010.

Grants: NSF and NIH

2.3 Effects of Air Pollution Exposure on Cardiovascular Health (Leader: Catherine Calder)

Using spatially-referenced longitudinal data collected as part of the Dallas Heart Study, we are estimating the effects of exposure to particulate matter on cardiovascular health.

Grants: in preparation for the NIH

2.4 Identifying Critical Windows in Studies of Air Pollution and Preterm Births (Josh Warren, Montserrat Fuentes, Amy Herring)

We developed a model for examining the relationship between exposure to PM2.5 and ozone and the probability of preterm birth, with a focus on identifying critical windows of pregnancy during which increased exposure to these pollutants may be particularly dangerous. Our methods have been applied to data from Texas vital records, are currently being applied to a Texas birth defects registry, and will soon be applied a ten-state study that is the largest study of the etiology of birth defects ever conducted, the National Birth Defects Prevention Study. We are currently working on extensions of these methods that relax some distributional assumptions and that look at multivariate outcomes.

Paper:

- Warren, J., Fuentes, M., Herring, A., and Langlois, P. Air Pollution and Preterm Birth: Identifying Critical Windows of Exposure. Under Review.

2.5 Health Impacts of Future Ambient Ozone Levels due to Climate Change (Howard Chang, Jingwen Zhou, Montserrat Fuentes)

To quantify the health impacts of future air pollution due to climate change, we have developed a modeling framework that integrates data from climate model outputs, historical meteorology and ozone observations, and a health surveillance database. We applied this approach to examine the risks of future ozone levels on non-accidental mortality across 19 urban communities in Southeastern United States. We are currently extending the model to address the following statistical challenges: calibrating multivariate climate model outputs, spatial misalignment, and joint analysis of air quality and temperature.

Papers:

- Chang, H.H., Zhou, J. and Fuentes, M. (2010). Impact of Climate Change on Ambient Ozone Level and Mortality in Southeastern United States. International Journal of Environmental Research and Public Health 7: 2866-2880.
- Zhou, J. and Fuentes, M. Calibration of air quality deterministic models using nonparametric spatial density functions. In preparation.

Talks:

- Impact of climate change on ambient ozone level and mortality in Southeastern United States. Multi-pollutant Multi-city Working Group, US Environmental Protection Agency, June 2010. Durham, NC
- Impact of climate change on ambient ozone level and mortality in Southeastern United States. SAMSI Workshop on Statistical Aspects of Environmental Risk, April, 2010.

2.6 Time-to-event Analysis of Air Pollution and Preterm Birth (Howard Chang, Brian Reich, Marie Lynn Miranda)

Most studies of preterm birth use a binary indicator of preterm birth as the response. We propose a flexible spatial survival model to estimate the effect of environmental exposure on the risk of preterm birth. We view gestational age as time-to-event data where each pregnancy enters the risk set at a pre-specified time (e.g. the 32th week). The pregnancy is then followed until either (1) a birth occurs before the 37th week (preterm); or (2) it reaches the 37th week and a full-term birth is expected. Our simulation study shows that this is a more powerful study design than more common time series and case-control designs. Also, by analyzing these data as survival data, we are able to study effects of both acute and long-term exposures jointly. The proposed approach is applied to geocoded births during 2002-2007 in North Carolina and we report an increased risk of preterm birth associated with exposure to fine particulate matter during the first and second trimesters.

Papers:

- Chang, H.H., Reich, B.J., and Miranda, M.L. Spatial time-to-event analysis of fine particulate matter and preterm births. In preparation.
- Chang, H.H., Reich, B.J., and Miranda, M.L. Fine particle air pollution and preterm birth in North Carolina, 2003-2007. In preparation.

Talks:

- Fine particle air pollution and preterm birth in North Carolina, 2001-2005. Third North American Congress of Epidemiology, June 2011, Montreal, Canada
- Spatial time-to-event analysis of fine particulate matter and preterm births. Summer Research Conference, Southern Regional Council on Statistics, June 2010. Virginia Beach VA.
- Time-to-event analysis of preterm birth and fine particulate matter. Joint Statistical Meeting August 2010. Vancouver Canada.

- Time-to-event analysis of preterm birth and fine particulate matter. Environmental Statistics Seminar, Harvard Department of Biostatistics, August 2010. Boston MA.

2.7 Exposure Measurement Error in Time Series Studies of Air Pollution and Health (Howard Chang, Montserrat Fuentes)

In a time series design, the health outcome is only available as daily total number of adverse events in a community. Unbiased risk estimates require the exposure measure to coincide with the true average exposure experienced by all at-risk individuals in the community. We have examined the effects of exposure measurement error due to spatial heterogeneity in the pollutant concentrations, motivated by the health effects of coarse particulate matter. We are currently developing models to incorporate spatial distribution of the at-risk population and personal exposure simulations in the times series design. The proposed approach is used to study the association between personal exposure to fine particulate matter and mortality in the New York City metropolitan area.

Paper:

- Chang, H.H., Peng, R.D., and Dominici, F. Estimating the acute health effects of coarse particulate matter accounting for exposure measurement error. Under Revision.

Talk:

- Estimating the acute health effects of coarse particulate matter accounting for exposure measurement error. ENAR, 2010. New Orleans, LA.

2.8 Additional Activities

- Invited JABES highlight session at JSM 2011.
- 2011 ENAR short course: A Practical Introduction to the Analysis of Geocoded and Areal Health Data.
- Collaboration with EPA on a national multi-site time series analysis of daily mortality and chemical constituents of fine particulate matter. Paper in preparation:

  Chang HH, Baxter LK, Fuentes M, and Neas LM. Daily mortality and chemical constituents of fine particle air pollution.

3 WG3: Interaction of Deterministic and Stochastic Models (Co-leaders: Murali Haran, Pennsylvania State University, and M.J. Bayarri, University of Valencia)

3.1 Overview

Core group members: M. J. Bayarri, University of Valencia, Spain; Jim Berger, Duke; Murali Haran, Penn State; John Harlim, NC State (from SAMSI stochastic dynamics program); Emily

Kang, SAMSI postdoc: Peter Kramer, RPI (from SAMSI stochastic dynamics program); Hans Kuensch, ETH Zurich, Switzerland.

Other key members: Sham Bhat, Penn State: Anabel Forte, University of Valencia, Spain; Robert Wolpert, Duke; Danilo Lopes, Duke; Priscilla Greenwood, Arizona State.

The goal of our group was to work on statistical issues in complex deterministic and stochastic models. We identified the following potential areas:

1. Approximations, emulation and calibration for complex computer models, typically Gaussian process-based models, and inference for stochastic differential equations.
2. Bridging the gap between the "usual" statistics idea of spatial modeling (accounting for dependence) and capturing spatial dynamic processes that are of scientific interest, which we envisioned as a potential avenue for interactions between the spatial program and the SAMSI stochastic dynamics program.
3. Using deterministic models/process to construct flexible space-time models.
4. Deterministic models as surrogate for probability/process model ("computer likelihoods").
5. Embedding deterministic models within a hierarchical framework to learn about parameters of interest.
6. Developing efficient computational approaches for simulating from multiscale models.

This working group was unique among the SAMSI space-time modeling working groups of 2009–2010 in that the research involved a close collaboration between two SAMSI programs in 2009–2010, the stochastic dynamics program, predominantly involving applied mathematicians, and the spatial program, primarily involving statisticians. Two applied mathematicians, Peter Kramer from Rensselaer Polytechnic Institute, and John Harlim from North Carolina State University, have been core members of the group, and have driven the research problem definition in conjunction with the statisticians. The focus of this collaboration was work in multi-scale dynamical systems; to our knowledge this is the first such collaboration between statisticians and applied mathematicians.

Summary of working group activities: The fall 2009 semester involved several research talks and discussions about potential projects by group members and collaborators of group members. Several of these problems included specific scientific applications:

(i) Modeling/inference for a component model for alcohol usage,

(ii) Dengue fever modeling based on a data set from Peru,

(iii) Exploring mathematics underlying a class of spatial dynamics models.

(iv) Studying multiscale dynamical systems or heterogeneous multiscale methods (HMMs).

(v) Emulation/calibration for complex computer models for multivariate spatial processes.

(iv) Bayesian model averaging of climate models for temperature projections.

Our group focused on projects (iv)-(vi); several related projects have either been completed or are ongoing projects involving the working group members. Projects (i)-(iii) represent future avenues for research. Several of our research project proposals will continue through our recently formed collaborations. Others may find their home in future SAMSI programs. In particular, many of the complex computer models problems may fit nicely into the "Uncertainty Quantification" SAMSI program in 2011.

3.2 Specific Activities

3.2.1 Multiscale modeling and simulation, spatio-temporal filtering

Project (iv) on multiscale modeling has been a focal point of research discussions. A core group of working group members listed above is meeting biweekly to discuss multiscale simulation/computation (HMM)-related research. HMMs are used for simulating physical processes that have both "macroscale" and "microscale" dynamics. This has become an active research area in applied mathematics and scientific computing in the past decade and has lots of important scientific applications, for example modeling complex fluid dynamics and climate systems. This appears to be a rich and still unexplored area of research for statisticians. Hence, a significant outcome of our working group's activities is the fact that we have opened up possibilities for a new and potential fruitful line of research.

Research manuscripts are planned and the group intends to submit multi-institutional grant proposals to the National Science Foundation, among other funding sources. The group is working closely with Sorin Mitran, an applied mathematician and participant in the stochastic dynamics program at SAMSI, who is providing the specific examples and associated computer code and data which will act as a testbed for the new methods we develop.

Research manuscripts

J. Harlim, Numerical Strategies for Filtering Partially Observed Stiff Stochastic Differential Equations, J. Comput. Phys., 230(3), 744-762, 2011.

E.L. Kang and J. Harlim, A Fast Filtering Framework for Assimilating Partially Observed Multiscale Systems: The Macro-Micro-Filter, submitted to Mon. Wea. Rev., 2010.

3.2.2 Complex computer models with multivariate spatial output

Project (v) above, on Gaussian process models for computer model emulation and calibration was another area of focus for the group.

Research manuscripts:

Bhat, K. S., Haran, M., Goes, M. (2010) Computer model calibration with multivariate spatial output: a case study in climate parameter learning. In Frontiers of Statistical Decision Making and Bayesian Analysis (Ming-Hui Chen, Dipak K. Dey, Peter Mueller, Dongchu Sun, and Keying Ye, eds.). Springer

Bhat, K. S., Haran, M., Keller, K., Tonkonojenkov, R. (2010) Inferring likelihoods and climate system characteristics from climate models and multiple tracers, submitted.

M.J. Bayarri (2010). A methodological review of computer models. In Frontiers of Statistical Decision Making and Bayesian Analysis (Ming-Hui Chen, Dipak K. Dey, Peter Mueller, Dongchu Sun, and Keying Ye, eds.). Springer, pp 157-168.

Reichert, P., White, G., Bayarri, M.J., Pitman, E.B. (2011). Mechanism-based emulation of dynamic simulation models: concept and application in hydrology. Computational Statistics and Data Analysis (in press).

Submitted grant proposal:

P. Kramer (RPI), Statistical Computing Methodologies in Dynamical Network and Multiscale Applications, co-PI K. Bennett; submitted to National Science Foundation, Division of Mathematical Sciences in December 2010; under review.

3.2.3 Bayesian model averaging for climate projections

Bhat, K. S., Haran, M., Keller, K. (2010) Combining multiple climate models using Bayesian model averaging to project surface temperature, in preparation.

4 WG4: Computation, Visualization, and Dimension Reduction (Frank Zou, NISS; Scott Holan, University of Missouri and Noel Cressie, Ohio State)

Our Working Group (WG) consists of approximately 60 members, though some are not active. It has approximately 10 postdocs and 8 graduate students.

Active postdocs: Sourish Das, Emily Kang, Esther Salazar, Ben Shaby, Xia Wang, and Frank Zou

Active graduate students: Sham Bhat, Matthias Dorit Hammerling, Katzfuss, and Yajun Liu

WG has attracted several outside members, including: Kate Cowles, Hedibert Lopes, and Brian Smith

Meetings started on 9/25/09, with an average attendance of 14 (on site) and 10 (remote).

4.1 Schedule of meetings (September 2009 through March 2010)

9/25/09. Introductory/Organizational meeting

10/2/09. Two presentations, on high performance computing and dimension reduction: Andrew Finley (Michigan State University), Peter Kramer (Rensselaer Polytechnic Institute).

10/9/09. Two presentations, on parallel computing and S4 classes in R: Ginger Davis (University of Virginia) and Blair Christian (EPA)

10/16/09. Two presentations, on regional climate models and spatially varying coefficient models: Ernst Linder (University of New Hampshire) and Scott Holan (University of Missouri, Columbia)

10/23/09. Two presentations, on covariance tapering and petroleum applications: Jo Eidsvik (Norwegian University of Science and Technology) and Orietta Nicolis (University of Bergamo, Italy)

10/30/09. One presentation, on modeling dynamic controls on ice streams: Noel Cressie (The Ohio State University)

11/06/09. Two presentations, on parallel computing and hierarchical spatial process models: Virgilio Gómez-Rubio (Universidad de Castilla-La Mancha) and Sudipto Banerjee (University of Minnesota)

11/13/09. Two presentations on RAMPS — an R package for complex spatio-temporal data: Brian Smith (University of Iowa) and Kate Cowles (University of Iowa).

11/20/09. Two presentations, on independent component analysis for multivariate time series and Monte Carlo strategies for calibration in climate models: David Matteson (Cornell University) and Gabriel Huerta (University of New Mexico)

12/04/09. Two presentations, on ancillary sufficient interweaving scheme and Monte Carlo strategies for calibration in climate models: Yajun Liu (University of Missouri) andMurali Haran (Penn State)

12/11/09. One presentation, on understanding covariance tapering: Ben Shaby (SAMSI)

1/15/10. Two presentations, on spatio-temporal statistics and models that include covariates in covariance structures: Noel Cressie (Ohio State) and Alexandra Schmidt (Universidade Federal do Rio de Janeiro)

1/22/10. One presentation, on spatio-temporal analysis of total nitrate concentrations using dynamic statistical models: Sujit Ghosh (North Carolina State)

1/29/10. Two presentations, on spatial dynamic factor analysis and using temporal variability to improve spatial mapping with application to satellite data: Hedibert Lopes (University of Chicago) and Emily Kang (SAMSI)

2/5/10. Two presentations, on dynamic spatial models and Bayesian methods in syndromic surveillance: Dani Gamerman (Universidade Federal do Rio de Janeiro) and Frank Zou (NISS).

2/12/10. One presentation, on space-time modeling with Markov random fields, plus subgroup updates: Marco Ferreira (University of Missouri, Columbia)

2/26/10. One presentation on Ensemble Kalman Filters, plus subgroup updates: Hans Kuench (Swiss Federal Institute of Technology)

3/12/10. One presentation, on surveillance for detecting emergent space-time clusters, plus subgroup updates: Renato Assuncao (Universidade Federal de Minas Gerais)

3/26/10. One presentation, on spatially varying autoregressive processes, plus subgroup updates: Bruno Sansó (University of California Santa Cruz).

In addition to the above, visits were paids to the CICS group at the National Climatic Data Center (NCDC) by Scott Holan (12/08/09) and by Bruno Sansó and Richard Smith (03/25/10). At these meetings, the potential was discussed for applying the new statistical techniques to new spatio-temporal datasets from NCDC, though it was subsequently decided to concentrate instead on regional climate model output from the NARCCAP group at the National Canter for Atmospheric Research.

In May, 2010, the groups split into four subgroups that subsequently met separately. The work of these subgroups will now be described.

4.2 Covariance Decomposition Using Wavelets (Leader: Orietta Nicolis)

The activities of the working group began by examining relevant papers on methods for reducing the rank of large covariance matrices. Initially, their attention was directed toward methods based on multiresolution techniques and wavelets functions. In particular, each member gave a presentation and wrote a summary of relevant papers.

In this context several issues were discussed including

1. Which wavelet was better than others in approximating spatial covariance functions?
2. How the computational burden can be reduced using a reduced/fixed rank covariances?
3. How to extend these methods using a Bayesian approach.
4. How methods based on spatial multiresolution covariances can be extended to the spatio-temporal domain.

The group decided to propose a covariance model (based on wavelet functions) for analyzing satellite data. They considered the dataset of Aerosol and SO2 from the NASA website:

http://mirador.gsfc.nasa.gov/index.shtml.

The aim was to propose a spatio-temporal model for analyzing the ash plume of the Iceland volcano during the month of April 2010.

The group realized that the data needed a pre-processing before being used and found several problems. The group decided to use simulated data, to start, and has ongoing meetings to discuss their progress.

4.3 Spatio-Temporal Point Processes (Leader: Noel Cressie)

Noel Cressie led the SG on spatio-temporal point processes. Their first activity was to carry out a literature survey. The group read and wrote summaries of about 20 papers, and have posted the summaries on 'SG2 Central,' the subgroup website. After the literature survey, the group looked for an appropriate space-time dataset to analyze. The group eventually decided to use the NARCCAP dataset and to model the pixels of extreme rainfall. In fact, such thresholded pixels are better modeled as a random set whose "germs" are generated by a type of autoregressive (in time) Poisson spatial point process. Random sets are defined by their hitting functions and, as of this writing; the SG is calculating the hitting function of the proposed spatio-temporal set model.

The subgroups intention is to publish a joint paper. They communicate by telecon at a regular time (10am Mon) most weeks, and the minutes of the meetings are posted on SG2 Central.

Submitted paper:

N. Cressie, R. Assunçao, S.H. Holan, M. Levine, O. Nicolis, J. Zhang and J. Zou (2011), Dynamical random-set modeling of concentrated precipitation in North America. Technical Report No. 854, March, 2011, Department of Statistics The Ohio State University, Columbus, OH. Submitted to Statistics and Its Interface.

4.4 Spatial and Spatio-Temporal Classes in R (Leaders: Kate Cowles and Brian Smith)

Brian Smith developed an R class of spatial objects that can be used with both geostatistical models and conditional autoregressive models.

Brian Smith developed the R package "magma" that enables parallelized matrix operations to be carried out on graphical processing units (http://cran.r-project.org/package=magma).

Kate Cowles and summer students beta tested and benchmarked the "magma" package and found speed-ups of up to 40 fold for operations related to spatial data analysis.

Kate Cowles identified relevant funding sources (from both government and industry) for research in spatiotemporal modeling and computation.

4.5 Space-time models for regional climate model data (Leader: Bruno Sansó)

Subgroup 4, which eventually reduced to Xia Wang (NISS), Andrew Finley (MSU), Ingelin Steinsland (ntnu), Dorit Hammerling (UMich), Esther Salazar (Duke), Paul Delamater (MSU), and Bruno Sansó (UCSC), has been meeting regularly to work on space-time models for regional climate output. The following paper has been submitted for publication:

Salazar, E., Sansó, B., Finley, A.O., Hammerling, D., Steinsland, I., Wang, X. and Delamater, P. (2011), Comparing and Blending Regional Climate Model Predictions for the American Southwest.

Bruno Sansó presented a talk based on this paper at the SAMSI spatial transition workshop as well as at Workshop on Environmetrics in Boulder, CO, 10/11-13/2010. Dorit Hammerling presented a poster based on this paper at the 3rd NARCCAP user workshop, April 7-8 2011, Boulder, CO.

4.6 Syndromic Surveillance

Manuscripts:

Zou, J., Karr, A., Heaton, M., Banks, D., Datta, G., Lynch, J. and Vera, F. (2010), Bayesian Methodology for Spatio-Temporal Syndromic Surveillance. Submitted to Statistical Analysis and Data Mining.

Heaton, M., Banks, D., Zou, J., Karr, A., Datta, G., Lynch, J. and Vera, F. (2011), A Spatio-Temporal Absorbing State Model for Disease and Syndromic Surveillance. Submitted to Statistics in Medicine.

Presentations and Posters:

Bayesian Methods in Syndromic Surveillance, QMDNS Meeting, Fairfax, VA, May 2010.

Bayesian Methods in Syndromic Surveillance, DTRA/NSF Algorithm Workshop, Chapel Hill, NC, June 2010.

Modeling the Spatio-temporal Dynamics of Risk with Application to Disease and Syndromic Surveillance, DTRA/NSF Algorithm Workshop, Chapel Hill, NC, June 2010.

Hierarchical Bayesian modeling in syndromic surveillance, ISBA World Meeting, Benidorm, Spain, June 2010.

5 WG5: Spatial Extremes (Richard Smith)

Membership:

Juliette Blanchet (WSL-Institut für Schnee, Switzerland)
Lisha Chen (Yale)
James Christian (EPA)
Dan Cooley (Colorado State)
Peter Craigmile (Ohio State)
Sourish Das (Duke/Samsi)
Anthony Davison (EPFL, Switzerland)
Robert Erhardt (UNC)
Alan Gelfand (Duke)
Souparno Ghosh (Texas A&M)
Radu Herbei (Ohio State)
David Holland (EPA)
Scott Holan (Univ of Missouri)
Gabriel Huerta (Univ of New Mexico)
Janine Illian (Univ of St. Andrews)
Soyoung Jeon (UNC)
Matthias Katzfuss (Ohio State)
Linyuan Li (Univ of New Hampshire)
Kenny Lopiano (Univ of Florida)
Jim Lynch (Univ of South Carolina)
Elizabeth Mannshardt-Shamseldin (Duke/Samsi)
Garritt Page (Duke)
Abel Rodriguez (UC Santa Cruz)
Huiyan Sang (Texas A&M)
Bruno Sansó (UC Santa Cruz)
Marian Scott (Univ of Glasgow, U.K.)
Benjamin Shaby (SAMSI)
Richard Smith (UNC/SAMSI)
Martin Tingley (SAMSI/NCAR)
Anand Vidyashankar (Cornell)
Maria Franco Villoria (Univ Glasgow, U.K.)

Jianqiang Wang (NISS)
Robert Wolpert (Duke)
Linda Young (Univ of Florida)
Jun Zhang (SAMSI)
Zhengjun Zhang (Univ of Wisconsin)
James Zidek (University of British Columbia)

Classical (univariate) extreme value theory revolves around the class of limiting distributions for sample extremes, which may be summarized by the Generalized Extreme Value (GEV) distribution

$$G(x\ ;\ \mu,\psi,\xi) \quad = \quad \exp\left\{-\left(1+\xi\frac{x-\mu}{\psi}\right)_+^{-1/\xi}\right\} \quad (y_+ = \max(y,0)).$$

The extension to higher dimensions, known as *Multivariate Extreme Value Theory*, leads to multivariate distributions with the property that for and dimension $K \geq 1$ and for each $N \geq 1$ there exist constants $\{A_{Nk} > 0,\ B_{Nk},\ 1 \leq k \leq K\}$ such that

$$G^N(x_1,...,x_K) \quad = \quad G(A_{N1}x_1 + B_{N1},...,A_{NK}x_K + B_{NK}).$$

Distributions satisfying this property are known as *Max-Stable Distributions.*

In studying spatial extremes, much attention recently has focussed on a class of processes known as *Max-Stable Processes*: A stochastic process over some index set (typically space or time) such that all the finite-dimensional distributions are max-stable.

In recent years, a number of specific examples of max-stable processes have been examined, including the "Smith model" (originally given in an unpublished 1990 paper), the class of models due to Schlather (2002), an extension of the Schlather model defined by Davison and Gholamrezaee, and the class of "Brown-Resnick" models and their extensions (Kabluchko, Schlather and de Haan, 2009).

Each of these models is defined by a stochastic representation that does not lead to closed-form solutions for the multivariate densities of the process. Therefore, direct methods of statistical estimation, such as maximum likelihood and Bayes, are not applicable. However, for each of the classes of max-stable models specified above, there do exist closed-form expressions for the bivariate densities. This suggests estimation based on maximizing the *composite likelihood*, where the composite likelihood in effect consists of the product of all bivariate densities (Padoan *et al.* 2010).

This more or less defined the starting point for our working group: max-stable processes have been increasingly used over the last few years as a model for spatial extremes, and thanks to the composite likelihood theory, there is now a viable means of estimating them, but that still leaves a lot of open questions, including the development of a "threshold" theory of estimation (comparable with the extension of univariate and later multivariate extreme value theory from the generalized extreme value distributions for sample maxima to univariate and multivariate generalized Pareto distributions for exceedances over thresholds), the development of a Bayesian theory of estimation (with associated applications to, for example, prediction of extreme events), and a wider range of applications. Among the latter were a special focus on paleoclimatic extremes (motivated in part by the work of the parallel group on Paleoclimate, with whom we shared some common members) and the recently developed theory of "single-event attribution" in climate change, which is concerned with the problem of how well we can attribute a single extreme event to human-induced causes of climate change as opposed to natural variability.

After an initial series of meetings of the whole group, it was decided to focus future attention within four subgroups concentrating on these four specific themes.

5.1 Threshold methods

Membership: Dan Cooley, Soyoung Jeon, Elizabeth Mannshardt-Shamseldin, Richard Smith, Robert Wolpert.

The challenge is to develop a version of the composite likelihood approach for max-stable processes based on threshold exceedances rather than the block maxima approach which is the main one considered previously. It turns out that there is a fairly straightforward construction to write down a possible composite likelihood in this setting, but understanding the properties of the resulting estimators is a much harder task. This group benefitted considerably from interactions with the spatial extremes group at EPFL (Lausanne, Switzerland), which included visits to SAMSI by Anthony Davison and Raphael Huser from EPFL; Huser gave a talk at the Transition Workshop on the approach that he has developed in collaboration with Davison. Meanwhile, SAMSI Graduate Fellow Soyoung Jeon worked independently on an alternative approach which uses the the second-order theory of bivariate extremes to develop bias and variance approximations to the composite likelihood estimators in a threshold context; such approximations lead to theoretical expressions for the optimal threshold as well as establishing asymptotic normality of the resulting estimators. A paper is shortly to be submitted (Jeon and Smith 2011).

Independently, Robert Wolpert developed approximations and a simulation methodology for the joint distibution of $k \geq 2$ points from a class of max-stable processes, and for maxima over regions. The method has potential to lead to exact maximum likelihood or Bayesina inference in certain cases (Wolpert 2010).

References:

Jeon, S. and Smith, R.L. (2011), Max-Stable Processes for Threshold Exceedances in Spatial Extremes. Manuscript, shortly to be submitted.

Wolpert, R. (2010), Spatial extremes. Preliminary version of a paper.

5.2 Bayesian methods

Membership: Dan Cooley, Peter Craigmile, Rob Erhardt, Matthias Katzfuss, Ben Shaby, Robert Wolpert.

Since there is no closed-form expression for the joint density of k observations from a max-stable process, when $k > 2$, there is no direct method of Bayesian inference. For frequentist inference, the method of composite likelihood has become fashionable (Padoan, Ribatet and Sisson, 2010). This therefore raises the question of whether the composite likelihood concept can be extended to a Bayesian context. A recent paper by Ribatet, Cooley and Davison (2010) has proposed such a modification. Independently, SAMSI postdoc Ben Shaby found an alternative construction (Shaby

2011). Both methods rely on "correcting" the composite likelihood function in the neighborhood of its maximum to derive a more realistic approximation to the true likelihood function.

An alternative approach to approximating the likelihood by a simpler form is to approximate the process itself. Reich and Shaby (2011) proposed a discrete approximation to one of the well-known classes of max-stable processes, resulting in a construction for which a standard MCMC algorithm is applicable.

A different kind of alternative approach is to abandon the likelihood function altogether and use the "approximate Bayesian computing" framework that has become popular for using Bayesian analysis for problems with intractable likelihoods. The computational demands of this approach are considerable, but UNC graduate student Rob Erhardt has developed an approach using parallel computing that seems very promising. Although vastly more computationally intensive than the composite likelihood approach, recent simulations show that the resulting estimates are superior to those computed via maximum composite likelihood. A paper has been submitted for publication (Erhardt and Smith, 2011).

Papers:

Erhardt, R. and Smith, R.L. (2011), Approximate Bayesian Computing for Spatial Extremes. Submitted for publication.

Reich, B.J. and Shaby, B. (2011), A finite-dimensional construction of a max-stable process for spatial extremes. Submitted.

Ribatet, M., Cooley, D. and Davison, A.C. (2010), Bayesian inference from composite likelihoods, with an application to spatial extremes. Under revision for *Statistica Sinica*.

Shaby, B. (2011), The open-faced sandwich adjustment for estimating function-based MCMC. *In preparation.*

5.3 Paleoclimatic Extremes

Membership: Peter Craigmile, Elizabeth Mannshardt-Shamseldin, Martin Tingley, Jim Zidek.

This group examined the relationship between changing climate and extreme climate events. Possible trends identified in temperature reconstructions lead to several questions about long-term climate behavior, and how to interpret this behavior given the patterns seen in proxy series. For example, it is of interest to use proxy data to address questions such as "Is there evidence that the extreme events of recent decades are more extreme than previous decades?" This group looked at what the statistics of extremes has to offer the field of paleoclimatology through modeling of the original proxy series, and sought to address several of the emerging areas of research that intertwine extreme value analysis and paleoclimatic reconstruction.

Paper in preparation:

Mannshardt-Shamseldin, E., Craigmile, P. and Tingley, M. (2011), Paleoclimatic extremes in proxy data. Shortly to be submitted.

5.4 Operational (Single-Event) Attribution

Membership: Richard Smith, Martin Tingley, Xuan Li (UNC), Matthias Katzfuss, Michael Wehner (Lawrence Berkeley Lab.), Jun Zhang.

Detection and Attribution is a well-established field of climate change research that is concerned with deciding whether observed changes and trends in the climate can be attributed to human causes. In a typical application of the method, a climate model (or several climate models) is run under different combinations or forcing conditions, e.g. control runs (stationary climate; no external forcing), natural forcings (including solar fluctations and volcanoes) and a combination of natural and anthropogenic forcings, where the anthropogenic forcings include greenhouse gases and sulfate aerosols. By regressing an observed climate signal on different combinations of climate models, it is possible to say when the anthropogenic signal is "detected" (usually characterized by a regression coefficient being statistically significantly different from zero), and in that case, the "attribution" refers to the partioning of observed climate change to different forcing factors, as measured by their regression coefficients.

Recently, attention has shifted to the problem of *attribution of single events*, where a single extreme meteorological event may be studied for evidence that it was in some sense "caused by" anthropogenic climate change. Following a landmark paper by Stott, Stone and Allen (*Nature*, 2004), such evidence is often presented in terms of a *fraction of attributable risk* (FAR) of an extreme event that is derived from the anthropogenic components of a climate model (or models).

This working group sought a more rigorous derivation of FAR using extreme value theory. The idea is to estimate the probability of an observed extreme event based on comparative runs of climate models under different combinations of control, natural-forcings or anthropogenic-forcings scenarios; however, the confidence intervals associated with such probability estimates are invariably very wide. Smith and Wehner will shortly submit a paper proposing a profile likelihood solution to this problem; Li (2011) is working for her PhD dissertation on an alternative Bayesian approach.

References:

Li, Xuan (2010), Bias and Variability Assumptions in a Bayesian Analysis of Climate Models. PhD Dissertation Proposal, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill.

Smith, R.L. and Wehner, M. (2011), A Threshold-based Extreme Value Approach to Event Attribution. In preparation.

5.5 Other activity of this working group

SAMSI postdoc Jun Zhang worked on a penalized likelihood approach for estimating return values from multiple precipitation stations, where the penalty function is designed to ensure smoothness in the variation of the GEV coefficients over space. This is an alternative to the popular Bayesian hierarchical approach to problems of this nature.

Montse Fuentes presented a talk on an alternative "nonparametric" approach to joint distributions of spatial extremes, avoiding the assumptions concerned with max-stable processes:

Anthony Davison gave an invited talk at the Climate workshop and he, Juliette Blanchet and Mathieu Ribatet also contributed to working group discussions a number of times by Webex link from Switzerland.

References:

Blanchet, J. and Davison, A. C. (2011) Spatial modelling of extreme snow depth. *Annals of Applied Statistics*, to appear.

Davison, A. C., Padoan, S. and Ribatet, M. (2010) Statistical modelling of spatial extremes. Under revision for *Statistical Science*.

Davison, A. C. and Gholamrezaee, M. M. (2010) Geostatistics of extremes. About to be resubmitted.

Fuentes, M., Henry, J. and Reich, B. (2011), Nonparametric Spatial Models for Extremes: Application to Extreme Temperature Data. *Extremes*, accepted.

Zhang, J., and Zheng, W., (2011), A Penalized Likelihood Approach for m-year precipitation return values estimation. Submitted to Journal of Agricultural, Biological, and Environmental Statistics.

6 WG6: Fundamentals of Spatial Modeling (Leader: Dongchu Sun, University of Missouri)

6.1 participants

Renato Assunçao, Universidade Federal
Susie Bayarri, Universitat de Valencia
Sudipto Banerjee, University of Minnesota
Jim Berger, SAMSI/Duke
Jose M Bernardo, University of Valencia
Howard Chang, SAMSI
Noel Cressie, Ohio State
Marco Ferreira, University of Missouri
Anabel Forte Deltell, Universitat de Valencia
Sherry Gao, Missouri Department of Conservation
Virgilio Gmez-Rubio, Bienvenidos a la Universidad de Castilla-La Mancha
Zhuoqiong He, Uinversity of Missouri
Jay Ver Hoef, NOAA, Alaska
Monica Jackson, American University
Hannes Kazianka, Alpen Adria Universität Klagenfürt
Andrew Lawson, Medical University of South Carolina
Jaeyong Lee, Soeul National University
Michael Levine, Purdue University
Ye Liang, University of Missouri
Yajun Liu, University of Missouri
Desheng Liu, Ohio State

Antonio López Qulez, Universitat de Valencia
Jim Lynch, University of South Carolina
Miguel A.Martinez-Beneito, Conselleria de Sanitat Generalitat Valenciana
Xiaoyi Min, University of Missouri
Shawn Ni, University of Missouri
Garritt Page, Duke
Rui Paulo, Universidade Técnica de Lisboa
Gavino Puggioni, UNC
Bala Rajaratnam, Standford
Brian Reich, NCSU
Cuirong Ren, South Dakota State
Chester Schmaltz, University of Missouri
Paul Speckman, University of Missouri
Dongchu Sun, University of Missouri
Anand Vidyashankar, Cornell
Jianqiang Wang, NISS
Xiaojing Wang, Duke
Chang Xu, University of Missouri
Jun Zhang, SAMSI
Jing Zhang, Miami University
Jian Zou, NISS

6.2 Presentations and Workshop

Weekly meetings took place on Wednesday, 11:00–12:45, during academic year, with irregular meetings thereaafter. There were 12 meetings during Fall 2009; 14 meetinga during Spring 2010. Some of the meetings discussed existing literature; these led to Ongoing research with discussions. Outside experts were invited to give presentations (Sudipto Banerjee, Bala Rajaratnam), as well as other group leaders.

The group organized a one-day Workshop on Objective Bayesian for Spatial and Temporal Models in San Antonio, TX, March 20, 2010.

6.3 Research Themes

6.3.1 Important issues in spatial and temporal models

- MCAR and Smoothed ANOVA with Spatial Effects (Sudipto Banerjee, University of Minnesota)
- Relation between spatial models and graphical models, (Balakanapathy Rajaratnam, Standford University)
- Spatio-temporal smoothing of risks based on spatial moving averages. (Miguel Angel Martnez Beneito)
- MCAR, Jointly modeling of spatial effects with given marginal, correlation between two nested and nonnested variables (Chester Schmaltz, University of Missouri)

- Irregular second order gaussian markov random fields (Paul Speckman, University of Missouri)
- Dynamic Multiscale Spatio-Temporal Models (Marco Ferreira, University of Missouri)
- Non-Gaussian Models, Zero-inflated Bayesian Spatial Models with Repeated Measurements (Jing Zhang, Miami University)
- Bayesian regression based on principal components for high dimensional data, Jaeyong Lee, Soeul National University

An Example: Irregular Second Order Gaussian Markov Random Fields (Paul Speckman)

- Gaussian Markov random fields are standard for analysis of areal data or data on a grid.
- The common models for areal data (e.g., CAR models) are first order.
- On the other hand, there are higher order priors for gridded data.
- A general method is proposed for constructing second order priors for general point-referenced data.
- The priors approximate Gaussian processes arising as solutions to stochastic differential or partial differential equations;
- Can have higher order smoothness properties; They are constructed to have sparse precision matrices;
- The new class of priors generalize higher order Gaussian MRF to arbitrary (gridded or not) data. The method is illustrated with an application to a rainfall data set.

6.3.2 Effects of spatial confounding

Core Members: Garritt Page, Duke; Yajun Liu, University of Missouri; Jaeyong Lee, Seoul National University, Korea; Dongchu Sun, University of Missouri.

Research themes:

- For a spatially-varying error structure, there often exist (possibly unmeasured) covariates that also vary spatially.
- In these scenarios it is hard to distinguish the effect of the covariate from that of the residual spatial variation.
- In a iid normal error setting, it is well know that dependence between a covariate and error produces biased regression coefficient estimates. However, the consequences of this dependence in a spatial setting are not as well understood.
- It is needed to investigate the effect of spatial confounding on covariate estimation.
  - the bias

  - the variance
  - the mean squared error

- Prediction
- Measurement errors
- Extension to the case of multiple (potentially unmeasured) covariates.
- Effect on the Bayes factors

6.3.3 Objective Bayesian and Fiducial Inference for spatial and temporal models

Core Members: Jose M Bernardo, University of Valencia, Spain; Jim Berger, SAMSI/Duke University; Marco Ferreira, University of Missouri; Jan Hannig, UNC; Cuirong Ren, South Dakota State University; Jing Zhang, Miami University; Zhuoqiong He, Uinversity of Missouri; Shawn Ni, Uinversity of Missouri; Dongchu Sun, Uinversity of Missouri.

Other Participants: Victor De Oliveira, University of Texas at San Antonio; Rui Paulo, Universidade Técnica de Lisboa, Portugal; Bruno Sansó, University of California at Santa Cruz; Stefano Cabras, University of Cagliari; Maria Eugenia Castellanos, Rey Carlos University, Spain; Brunero Liseo, University of Roma, Italy; Maria M. Barbieri, Universita Roma, Italy; Siva Sivaganesan, University of Cincinatti.

Research Objectives:

- In the absence of sufficiently quantifiable subjective knowledge concerning the unknown parameters, it is typical to resort to objective noninformative or priors; these priors allow most of the benefits of Bayesian analysis to be achieved, while avoiding the difficulty of obtaining a subjective prior.
- Alternatively, one could use Fiducial Analysis. The Recent work by Jan Hannig (UNC) and his collaborators offers promising methods for spatial analysis.

6.3.4 One-day Workshop on Objective Bayesian for Spatial and Temporal Models at San Antonio, TX, March 20–21, 2010

- The main theme is to explore appropriate prior distributions for parameters of spatial and temporal models.
- Formal objective priors were introduced for AR (1) models by Berger and Yang (1994), and for geo-statistical models by Berger, De Oleveira and Sansó (2001, JASA) and Paulo (2005, Annals of Statistics).
- However, the current results are mainly for situations in which there is no nugget effect and no measurement errors; in practice, both are ubiquitous.

Current and Future Work

- Group members are working on a monography summarize the work during the year

- To be published by Chapman and Hall
- About 10 Chapters, 5 papers on Spatial Models, one on Spatial Confounding, and 4 papers on Objective Bayesian and Fuducial Models

7 WG7: Geostatistics (Sudipto Banerjee, Veronica Berrocal, Brian Reich and Alan Gelfand)

7.1 Working Group

Geostatistics: The Geostatistics working group is led by Sudipto Banerjee (UMN), Brian Reich (NCSU) and Alan E. Gelfand (Duke). This group will explore spatial and spatiotemporal models for geostatistical data analysis with special emphasis on issues relating to preferential sampling.

The participants in this group include (in alphabetical order):

Sudipto Banerjee (University of Minnesota),
Jarrett Barber (University of Wyoming),
Francisco M Beltran (University of California Santa Cruz),
Candace Berrett (Ohio State University),
Veronica Berrocal (SAMSI),
Avishek Chakraborty (Duke University),
Howard Chang (SAMSI),
David Conesa (University De Valencia, Spain),
Sourish Das (SAMSI),
Dipak Dey (University of Connecticut),
Yiping Dou (University of British Columbia),
David Dunson (Duke University),
Jo Eidsvik (Norwegian University of Science and Technology),
Andrew Finley (Michigan State University),
Alan Gelfand (Duke University),
Michele Guindani (University of New Mexico),
Pritam Gupta (University of Wyoming),
Sandra Hurtado Rua (University of Connecticut),
Janine Illian (University of St. Andrews, U.K.),
Karen Kafadar (Indiana University),
Hannes Kazianka (University of Klagenfurt, Austria),
David Kessler (University of North Carolina),
Kenny Lopiano (University of Florida),
Gabriele Martinelli (Norwegian University of Science and Technology),
Amy Nail (Duke University),
Orietta Nicolis (University of Bergamo, Italy),
Garritt Page (Duke University),
Juergen Pilz (University of Klagenfurt, Austria),
Ana Rappold (EPA),
Brian Reich (North Carolina State University),
Huiyan Sang (Texas A& M University),

Alexandra Schmidt (University Federal do Rio de Janeiro, Brasil),
Ron Smith (Centre for Ecology and Economics, Edinburgh, U.K.),
Gunter Spöeck (University of Klagenfurt, Austria),
Ingelin Steinsland (Norwegian University of Science and Technology),
Jay Ver Hoef (NOAA),
Linda Young (University of Florida),
Chengwei Yuan (University of New Hampshire),
Jing Zhang (Miami University),
Zhengyuan Zhu (Iowa State University),
James Zidek (University of British Columbia).

7.2 Research Activities and Findings

After the opening workshop and a series of weekly meetings, the working group identified some important issues for initial research. Given the large size and the broad spectrum of topics, the working group was further subdivided into smaller subgroups, with each subgroup focusing upon one of the following topics.

- *Preferential sampling and point-process modeling:* **(Leader: Alan Gelfand).** This group explored hierarchical spatial and spatiotemporal models that account for stochastic dependencies in settings where the processes generating the data and the sampling locations are dependent. The group planned to explore underlying theoretical connections between preferential sampling and models for missing data as well as the connections between spatial sampling designs and preferential sampling. More generally, the group sought to understand the nature of preferential sampling, modeling issues related to preferential sampling and the effect of preferential sampling upon model performance and inference. Integrating point process models into hierarchical frameworks for large datasets has also been explored by this group. There was substantial overlap in the activities of this subgroup and the group for point-processes.

- *Spatial sampling designs:* **(Leader: Gunter Spöeck).**. This working group carried out methodological developments in spatial sampling design. In particular, this group explored deterministic algorithms based on spectral representations and experimental design ideas to assess the efficiency estimates that demonstrated how good sampling designs actually are using convex design functionals and convex optimization theory. The group also explored spatial sampling designs for non-stationary random fields having non-Gaussian skew distributions generated using spectral representations of random fields and making use of the experimental design theory for GLMMs. These methodologies were explored in applied contexts such as exploring the adequacy of existing pollution monitoring networks for protecting human health.

- *Estimation techniques for geostatistical models:* (**Leader: Jo Eidsvik**) This subgroup focused on fast computation methods for large geostatistical models. There have been a number of suggestions on how to solve the so-called "big-N" problem, i.e. to handle the $O(n^3)$ problem of matrix factorization in large Gaussian or latent Gaussian models. One approach that the group explored was to combine effective modeling aspects and inference methods. For instance, predictive process models provide dimension-reduction while at the same time being amenable to approximate Bayesian inference using Integrated Nested Laplace Approximation (INLA). Since the predictive process models introduce no additional parameters over the regular (full) model formulation, it is relatively straightforward to apply INLA. The speed-up compared with MCMC is large. At every evaluation the computational cost is reduced by using a predictive process model, also providing large speed-ups. The statistical performance is reduced a little for the estimates of covariance parameter, but we see small effects on estimates for regression effects and predictions.

This group also explored parallel computing in the context of geostatistical models. Parallel computing is getting easier on modern computers, and can be implemented on the Graphical processing unit (GPU) as well as on many Central Processing Units (CPU). One could approach this from a Bayesian or a frequentist viewpoint. The frequentist approach is somewhat similar to INLA, since an optimization routine is required. The group focused on composite likelihood models, where the structure is such that parallelization is immediate. The variance of the estimate can be computed from the Godambe sandwich estimate. Similar ideas, using composite likelihood models and the sandwich estimate for variance, hold for prediction at unobserved sites too. This group also explored complex hierarchical models for analyzing large geostatistical datasets arising in forestry and the environmental sciences.

- *Ozone process models:* **(Leader: Amy Nail).** What separated the goals of this working group from many ozone modeling efforts was the focus on developing statistical models of ozone that represent the chemical and physical mechanisms through which ozone is created, destroyed, and transported – that is, to do with a statistical model what is typically done with a differential-equation-based deterministic model.

  The group separated into the 5 sub-projects. The first project was led by Brian Reich and represents a cross-over project with the Non-Gaussian and Non-Stationary work group. The objective was to produce a methodological paper with application to ozone examining a new method for modeling the effects of covariates on the spatial covariance in a space-time model.

  The purpose of the second project was to develop methodologies, with application to ozone, for examining a nonparametric variable selection approaches that account for spatial correlation.

  The third project sought to model the dispersion of NOx from multiple point sources, and to combine such models with information about area-source emissions and transport to predict NOx concentrations in space and time. Statistical challenges included adapting deterministic dispersion modeling techniques (differential-equation-based or otherwise) to a statistical model, tackling the change-of-support problem, performing sequential estimation while managing large spatial covariance matrices, and accounting for covariates in the covariance.

  The group also looked at optimizing code for a dynamic linear model with spatial covariances. Since a process-based space-time model of either ozone or NOx will require the modeling of transport, sequential estimation will ensure a fully-model-based framework. Experimentations were conducted using R code to call C code to implement a dynamic linear model of ozone with spatial covariances, but this approach found computation time to be formidable (especially if this model is to be applied for the whole U.S. or eastern U.S.). Dimension reduction approaches are currently being explored.

  Note: The ozone process modeling workgroup was initiated by Amy Nail, who was employed at the U.S. EPA at the time of heading up the group, but when the group separated into the 5 sub-projects listed below, Brian Reich took charge of the first one.

- *Numerical model validation:* **(Leader: Brian Reich)** The numerical models subgroup focused primarily on developing Bayesian spatial methods for calibrating deterministic forecasts to produce probabilistic forecasts. The work split in two directions: spatial quantile regression and Bayesian spatial model averaging. Numerical weather forecasts are often calibrated by adjusting the mean, and perhaps the variance, of the predictive distribution. For non-Gaussian data such as precipitation, this can lead to underestimation of extremes. To calibrate the entire conditional distribution of precipitation given a forecast, we propose Bayesian spatial quantile regression. Quantile regression allows for a separate the relationship between the forecast and each quantile of the predictive distribution, therefore calibrating the probabilistic forecast at each quantile level. We show this leads to good performance in both the center and tails of the predictive distribution.

  Recognizing the several sources of uncertainty affecting the output of a numerical model, in recent years, there has been a shift from deterministic predictions to probabilistic ones, and runs of a single numerical model have been replaced by runs of ensembles of numerical models. Therefore, the working group also explored statistical methods to calibrate the spatial output generated by an ensemble of numerical models. We extend previous work using Bayesian model averaging (BMA) to produce probabilistic weather forecasts. BMA assumes the predictive distribution is a mixture of density functions centered around each bias-corrected forecast and BMA weights that depend on the past predictive performance of each ensemble member. The proposed method is an extension of BMA in that both the weights and the parameters of each mixture component are assumed to vary in space. This is intuitively appealing as it is expected that the performance of each numerical model are not constant in space, but might vary from location to location.

  The working group produced a productive and hopefully long-lasting collaboration between Ingelin Steinsland, Veronica Berrocal, and Brian Reich. The group also expects two papers to be submitted before the end of the year.

- *Feedback:* **(Leader: Jarrett Barber).** The 'feedback' subgroup was interested in the cutting of edges in graphical models for the purpose of controlling undesirable feedback among model components. Focus was primarily on models of health effects and exposure in a Bayesian framework. Prototype R code was produced to begin simulation experiments to explore feed back control. The subgroup is currently conducting extensive simulation experiments.

The working group – and subsequently the subgroups – continued to meet on a weekly basis and focused on the above issues. The interactions produced several productive collaborations between statisticians, deterministic modelers, scientists and GIS experts who embarked upon several interesting projects. A non-exhaustive list of publications emerging from the Geostatistics group is provided below.

Banerjee, S. and Gelfand, A.E. (2010). Modelling spatial gradients on response surfaces. In *Frontiers of Statistical Decision Making and Bayesian Analysis*, eds. M.H. Chen, D.K. Dey, P. Mueller, D. Sun and K. Ye, New York: Springer

Berrocal, V.J., Gelfand, A.E., and Holland, D.M. (2010). A bivariate space-time downscaler under space and time misalignment. *Annals of Applied Statistics*, (in press).

Luke Bornn, Gavin Shaddick and James V Zidek, The effects non-stationarity and preferential sampling in modeling and measuring environmental fields for health effect analysis. Draft report in preparation.

Song Cai, Xiaoli Yu, Gavin S Shaddick and James V Zidek, Preferential sampling in the monitoring black smoke: a case study Draft report nearly completed.

Chakraborty, A., Gelfand, A.E., Wilson, A.M., Latimer, A.M. and Silander, J.A. (2010). Modeling large-scale species abundance with latent spatial processes. *Annals of Applied Statistics* (in press).

Eidsvik, J., Shaby, B., Reich, B.J., Wheeler, M. and Niemi, J. (2010). Estimation and prediction in spatial models with block composite likelihoods using parallel computing. Submitted.

Eidsvik, J., Finley, A., Banerjee, S., and Rue, H. (2010). Approximate Bayesian inference for large spatial datasets using predictive process models. Under revision, *Computational Statistics and Data Analysis*.

Finley, A.O., Banerjee, S. and MacFarlane, D.W. (2010). A hierarchical model for predicting forest variables over large heterogeneous domains. *Journal of the American Statistical Association* **106**, 31–48.

Guhaniyogi, R., Finley, A.O., Banerjee, S. and Gelfand, A.E. (2010). Adaptive Gaussian predictive process models for large spatial datasets. *Environmetrics* (in review).

Prates, M.O., Dey, D.K., Willig, M.R. and Yan, J. (2010), Intervention Analysis of Hurricane Effects on Snail Abundance in a Tropical Forest Using Long-Term Spatiotemporal Data, Journal of Agricultural, Biological, and Environmental Statistics, 1-15, doi: 10.1007/s13253-010-0039-1.

Prates, M.O., Dey, D.K., Willig, M.R., and Yan, J. (2011), Transformed Gaussian Markov Random Fields and Spatial Modeling. Submitted.

Reich, B., Eidsvik, J., Guindani, M., Nail, A. and Schmidt, A. (2010). A class of covariate-dependent spatiotemporal covariance functions. Accepted by *Annals of Applied Statistics*.

Reich, B. and Berrocal, V. (2010). Local Bayesian model averaging for numerical model calibration. To be submitted to *Environmetrics*.

Spöeck, G., Zhu, Z. and Pilz, J.: Simplifying objective functions and avoiding stochastic search algorithms in spatial sampling design, submitted to: Stochastic Environmental Research and Risk Assessment, (2011).

Spöeck, G., Hussain I.: Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan. In: Statistical Methodology, New York: Elsevier Science Inc, doi: 10.1016/j.stamet.2011.01.004, (2

Steinsland, I., Berrocal, V. and Reich, B. (2010). Bayesian spatial quantile regression for numerical model calibration. To be submitted to *Journal of the Royal Statistical Society: Series C.*

Wang, X., Dey, D.K. and Banerjee, S. (2010). Non-Gaussian hierarchical generalized linear geostatistical model selection. In *Frontiers of Statistical Decision Making and Bayesian Analysis*, eds. M.H. Chen, D.K. Dey, P. Mueller, D. Sun and K. Ye, New York: Springer.

Xiaoli Yu, Gavin Shaddick and James V Zidek, Correcting for preferential sampling in monitoring environmental processes. Draft report of theory completed. Application pending.

Xiaoli Yu, Gavin Shaddick and James V Zidek, The effects of preferential sampling in environmental epidemiology. Draft report partially completed.

8 WG8: Point Patterns Working Group

8.1 Goals

The working group on point patterns coalesced around three primary research goals. The first (which yielded the "Mixture" subgroup) focused on novel ways of introducing covariate information into modeling intensities. Special settings here included mixture models for intensities, both parametric and Bayesian nonparametric as well as generalized Neyman Scott processes with covariate information at parent level to drive locations and spreads for offspring. The second goal (which yielded the "Practical" subgroup) was to bring to a more applied level, spatial point processes with first and second order components (first order is a usual nonhomogeneous Poisson process form $\lambda(s)$, second order is a pairwise interaction form $\gamma(s, s')$, inhibition, attraction). Issues here again included the introduction of covariate information into such settings with interest in a fully model-based implementation with associated technical questions such as integrability for the resulting intensity. The third goal (which yielded the "Model diagnostics" subgroup) considered an under-explored problem for point patterns, the question of model adequacy and model comparison. For instance, how do we compare two nonhomogeneous point processes? How do we compare a nonhomogeneous Poisson process with a pairwise interaction process. What happens if we add dynamics, resulting in space time point processes?

8.2 Active members

As a result of the three-pronged research agenda, three active subgroups emerged:

- "Mixture" group - Micheas, Chakraborty, Assunçao, Matteson, Gelfand
- "Practical" subgroup - Assunçao, Lopes, Waller, Illian
- "Model diagnostics" subgroup - Gomez Rubio, Chakraborty, Gelfand

8.3 More detail on the work of the "Mixture" subgroup

Considering covariates in this setting requires distinguishing spatially indexed covariates $X(s)$ from labels (marks). For instance, $X(s)$ might be elevation which might encourage more or fewer points but this is apart from a mark which might label different species with associated point patterns.

Evidently, the issue is direction of conditioning, i.e., the covariate can be a mark at a location) yielding $X|s$ as $X(s)$ and $\{X(s_i)\}|\theta$.

In this regard, Micheas (with Chakraborty and Gelfand), using $\lambda f(s;\theta)$ form, $\{s_i\}|\lambda,\theta$ for the intensity add the covariate in form of discrete mark. In fact, he extends to a mixture form $\lambda f(s)$ where $f(s) = \sum_l \pi_l f_l(s|\theta_l)$. In fact $f_l(s|\theta_l) = \text{BivN}(\mu_l, \Sigma_l)$

Work of Chakraborty introduced covariates to explain where locations will tend to be. Here, the usual form is $\log\lambda(s) = X(s)^T\beta + \omega(s)$. However, again consider $\lambda f(s)$ where $f(s) = \sum_l \pi_l f_l(s|\theta_l)$ with $f_l(s|\theta_l) = \text{BivN}(\mu_l, \Sigma_l)$. So, μ_l provide "centers" for the intensity and $\pi_l = \frac{e^{X(\mu_l)^T\beta}}{\sum_j e^{X(\mu_j)^T\beta}}$.

Assunçao and Lopes developed a cluster model for point processes, driven by covariates but not a mixture model, rather with interacting neighbors.

Papers:

Assunçao, R. and Lopes, D. (2010), Visualizing Marked Spatial and Origin-destination Point Patterns With Dynamically Linked Windows. *Journal of Graphical and Computational Statistics*, accepted.

Chakraborty, A., Gelfand, A.E., Wilson, A. M., Latimer A. M. and Silander, J.A. 2011 Point pattern modeling for degraded presence-only data over large regions. To appear, Journal of Royal Statistical Society, Series C

Chakraborty, A. and Gelfand, A. E. 2011 Mixture modeling for spatial point patterns with location-specific covariates, manuscript in preparation

Matteson, D.S. and Micheas, A.C., A Finite Mixture Model for Spatio-temporal Poisson Process. In Preparation.

8.4 More detail on the work of the "Practical" subgroup

This group worked on the paper listed below. The context is temporal point process with spatial marks in the form of binary indicators on a lattice. The application is to infection propagation. A very high level of missingness arises; only 6 out of 30 weeks are observed, necessitating a challenging data augmentation approach.

Jean Vaillant, Gavino Puggioni, Lance A Waller, and Jean Daugrois (2011), A spatio-temporal analysis of the spread of sugar cane yellow leaf virus. *Journal of Time Series Analysis*, to appear.

8.5 More detail on the work of the "Model diagnostics" subgroup

Again, we assert that model adequacy and model comparison are underdeveloped for point pattern data. What exists follows the lead of Baddeley, with follow on work from Illian. Here, the widely used Ripley - K function is not what is needed because it is only applicable under complete spatial randomness, not a model of interest.

In comparing intensities we can envision various naive ideas such as binning and counts. We can imagine some version of AIC here. Also possible is discretized integrated squared difference between model intensity and empirical intensity (say kernel intensity estimate). Another option is cross-validation, thinning the point pattern to provide a hold-out sample.

Through normalization, there is a density associated with a point process realization, $f(\mathbf{S};\theta)$.

For homogeneous Poisson process it is $f(\mathbf{S}) = \alpha\lambda^{n(\mathbf{S})}$.

For a nonhomogeneous Poisson process it is $f(\mathbf{S}) = \alpha\Pi_i\lambda(s_i)$.

For pairwise interaction process, $f(\mathbf{S}) = \alpha\Pi_i\lambda(s_i)\Pi_{i,j}\gamma(s_i, s_j)$.

From these, we can define the Papangelou conditional intensity: $\phi(u;\mathbf{S}) = f(\mathbf{S}\cup u)/f(\mathbf{S})$.

In turn, we can create residuals (following ideas of Baddeley).

In particular, there is a realized - $I(B;\theta) = n(\mathbf{S}\cap B) - \int_B \phi(u;\mathbf{S};\theta)du$. Also, we have an estimated — $R(B;\hat{\theta}) = n(\mathbf{S}\cap B) - \int_B \phi(u;\mathbf{S};\hat{\theta})du$. These residuals provide a different path for model assessment.

8.6 Ongoing research problems

Finally, we mention some continuing open research ideas. In the point pattern setting, how can we extend AIC to DIC with spatial random effects; how would we do the computation? From above, can we implement a Bayesian analysis of conditional intensity, hence, of realized residuals? Such work would begin with Cox processes first, then proceed to interaction processes. There seems to be little experience in the literature with thinning and cross-validation ideas.

A novel way to model space time point patterns is through integro difference equations (space-time diffusions) and very recently through integral projection models (population demography). There is promising modeling opportunity here with considerable computational challenge.

There is also potential to revisit recent preferential sampling work (Diggle et al.) with regard to environmental exposure, raising the question of the effect of sampling design as well as where to put covariates in this investigation. There remain many open questions here.

Finally, Gomez Rubio has prepared a grant submission to study point patterns arising from patients with respiratory and circulatory illness in Madrid.

9 WG9: Non-stationary and Non-Gaussian Processes

This working group is led by David Dunson (Duke), Jo Eidsvik (NTNU) and Montse Fuentes (NCSU). The objective is to explore non-stationarity and non-Gaussianity for spatial and spatiotemporal models.

The most active participants in this group include (in alphabetical order): Veronica Berrocal (SAMSI), Kate Calder (Ohio State), David Dunson (Duke), Jo Eidsvik (NTNU), Montse Fuentes (NCSU), Michele Guindani (U of New Mexico), Scott Holan (U of Missouri), Katja Ickstadt (Technical University of Dortmund), Jaeyong Lee (Seoul U), Linyuan Li (U of New Hampshire), David Kessler (UNC), Amy Nail (Duke), Orietta Nicolis (U of Bergamo), Brian Reich (NCSU), Alexandra Schmidt (U Federal do Rio de Janeiro), Ben Shaby (SAMSI), Michael Stein (U of Chicago), Robert Wolpert (Duke).

9.1 Research

After the opening workshop the group had a series of weekly meetings covering a broad spectrum of topics. We learnt a lot during these meetings. They let us define common interests and set the stage for collaborations. In the late Fall the working group decided to split in 3 sub-groups. In the Spring the working group only had a handfull of joint meetings. The 3 sub-groups were defined as follows:

- *Spectral domain methods:*

  Spectral methods are powerful tools to introduce new classes of covariance functions. The subgroup proposes new models for nonstationarity by having a spectrum that is space-dependent. By introducing a spatial filtered periodogram we obtain a nonstationary nonparametric estimate of the covariance.

  To every stationary $Z(s)$ there can be assigned a process $Y(\omega)$ with orthogonal increments, such that we have for each fixed s the spectral representation:

  $$Z(s) = \int_{\mathcal{R}^2} e^{is^T\omega} dY(\omega) \tag{1}$$

  The Y process is called the spectral process associated with a stationary process Z.

  The periodogram I_n estimates the spectral density f of a process Z observed in a grid $n \times n$,

  $$I_n(\omega) = (2\pi n)^{(-2)} \left| \sum_{\mathbf{j} \in \mathbf{J}} Z(\mathbf{j}) e^{-i\omega^T \mathbf{j}} \right|^2$$

  The sub-group looks at approaches for the spectral analysis of non-stationary spatial processes, $Z(x)$, which is based on the concept of spatial spectra, this means spectral functions which are space dependent, $f_{\mathbf{x}}(\omega)$.

  The spectral representation of $Z(\mathbf{x})$ is always interpreted as its representation in the form of superposition of sine and cosine waves of different frequencies ω

  $$Z(\mathbf{x}) = \int_{\mathcal{R}^2} \exp(i\mathbf{x}^T\omega)\phi_x(\omega) dY(\omega)$$

  We propose a nonparametric estimate of the spectral density. We first define $J_x(\omega_0)$,

  $$J_{\mathbf{x}}(\omega_0) = \Delta \sum_{u_1 = x_1 - n_1}^{x_1} \sum_{u_2 = x_2 - n_2}^{x_2} g(\Delta\mathbf{u}) Z(\Delta(\mathbf{x} - \mathbf{u})) \\ \exp\{-i\Delta(\mathbf{x} - \mathbf{u})^T\omega_0\},$$

  where $\mathbf{s} = (s_1, s_2)$, and $\{g(\mathbf{u})\}$ is a filter. We refer to $|J_x(\omega)|^2$ as the spatial periodogram at a location $\mathbf{x}$ for a frequency ω. $|J_x(\omega)|^2$ is an approximately unbiased estimate of $f_x(\omega)$, but as its variance may be shown to be independent of N it will not be a very useful estimate in practice. Then, we estimate $f_x(\omega)$ by "smoothing" the values of $|J_x(\omega)|^2$ over neighboring values of x.

The presented methodology requires gridded data, but it could be also extended to irregularly spaced data, by using a filtered spline representation across space. This is work in progress.

There is need to study of the properties of the proposed nonparametric estimates using fixed domain asymptotics. This is work in progress.

These methods could be generalized to space and time, and they would allow for nonseparability. A reasonable approach would be semiparametric in space-time: with a parametric model for short distances in space, and a spline representation for large distances. This is work in progress.

- *Mixture models:*

  The sub-group builds on much work from Dirichlet processes, stick breaking processes, and other non-parametric models that have become popular because of large flexibility and computational tractability. Dirichlet mixture models are nonGaussian and nonstationary, but without replicates it is difficult to determine if we have lack of stationarity or lack of normality. However, the Dirichlet process (DP)-type of mixture models would allow for both. Mixture models are flexible enough to characterize complex tail behavior, and can be used to model extremes (i.e a DP copula)

  The stick-breaking prior for F is the potentially infinite mixture $F \stackrel{d}{=} \sum_{i=1}^{m} p_i \delta(\theta_i)$, where p_i are the mixture probabilities and $\delta(\theta_i)$, is the Dirac distribution with point mass at θ_i. The mixture probabilities "break the stick" into m pieces, so the sum of the pieces is one, $\sum_{i=1}^{m} p_i = 1$. $p_1 = V_1$, and the subsequent probabilities are $p_i = V_i \prod_{j=1}^{i-1}(1 - V_j)$, where $1 - \sum_{j=1}^{i-1} p_j = \prod_{j=1}^{i-1}(1 - V_j)$ is the probability not accounted by the first $i - 1$ components.

  The sub-group looks at covariance regression, where a covariance matrix is in the support of the space of covariance stochastic processes. Let $\Sigma(s) = \Lambda(s)\Lambda(s) + \Sigma_0$ induced through

  $$y(s_i) = \Lambda(s_i)\eta_i + \epsilon_i, \quad \eta_i \sim N_k(0, I_k), \quad \epsilon_i \sim N_p(0, \Sigma_0)$$

  We can let $\Lambda(s_i) = \Theta\xi(s_i) = \Theta \sim p \times L =$ coefficients relating predictor-dependent factor loadings matrix to dictionary functions comprising the $L \times k$ matrix $\xi(s)$. This model for the nonparametric covariance model can also be modified to build upon a latent segmentation process providing the label indeces at different locations. Then, $y(s_i) = \sum_k I[Z(s_i) \in B_k]\eta_k(s_i)$. The η_k are Gaussian processes, while the labeling can take on many forms. This is related to work on latent stick breaking processes. There are working papers in these directions.

  The nonparametric Bayesian spatial models should be extended to space and time to allow for nonseparable dependence. These models should also be extended to a multivariate setting, to allow for nonstationary cross-dependence.

- *Covariates in the covariance:*

This sub-group looks at non-stationarity models constructed by adding covariates in the covariance function. Typically, the covariance structure is regarded as a nuisance that we do not learn about, except specifying the scale and correlation range in some way. The mean term, on the other hand, is constructed by sophisticated modeling from available covariates. For prediction purposes, the covariates are important only through the predictor, not the prediction variability. We envision better prediction and more useful prediction distributions by letting the covariance also depend on covariates.

One project that has been regarded in the sub-group is mixtures of Gaussian processes with different scale and range, but weighted by covariates. Define the process as

$$y(s_i) = \sum_k w(x(s_i); \alpha_k) \eta_k(s_i), \qquad \eta_k(s_i) \sim N(0, \Sigma_k)$$

, where the weights $w(x(s_i); \alpha_k)$ of the mixture varies as a function of covariates $x(s_i)$. The processes are parameterized by different range and variance in Σ_k. For instance, when there is much wind, the weight could be highest for a process with large correlation range and larger variance. When there is litte wind, the weight could be highest for a process with small correlation range and less variance. The model is similar to the spatially varying coefficient model and other models using a spatially varying weights for a mixture of processes. The sub-group has submitted a paper on this topic (Reich, Eidsvik, Guindani, Nail and Schmidt, 2010, A class of covariate-dependent spatiotemporal covariance functions), applying the resulting model to daily ozone levels in the South-Eastern US. It is currently under revision.

The suggested mixture models for the covariance structures above naturally extends to include covariates in the covariance structure. For instance, the factors could depend on covariates. Further, it is straightforward to let the prior mean of the latent segmentation process $Z(s_i)$ depend on covariates. This would pull the labeling process to a high/low index in parts of the space, and would naturally exhibit in the covariance for locations where the labels become the same or not.

The sub-groups (or parts of sub-groups) continue to meet to finish ongoing work.

9.2 Deliverables

- Sharing knowledge.
- Papers submitted or in progress.
- Several new collaborations.

10 Other Products of the Program

Submitted grant proposal:

Network for Statistics in Atmospheric and Oceanic Sciences (PI: M. Fuentes, NCSU; co-PIs: M. Stein, University of Chicago and P. Guttorp, University of Washington. Submitted to Program

Solicitation NSF 10-584 on Research Networks in the Mathematical Sciences, National Science Foundation, $5 million.

Papers:

Corberan, A. and Lawson, A. (2011), Conditional Predictive Inference for On-line Surveillance of Spatial Disease Incidence. Submitted to *Statistics in Medicine*.

Dennis, Fox, Fuentes, Gilliland, Hanna, Hogrete, Irwin, Rao, Scheffe, Schere, Steyn, Venkatram. (2010). A framework for evaluating regional-scale numerical photochemical modeling systems. Environmental Fluid Mechanics Journal, in press.

Fuentes, M., and Banarjee, S. (2011). Bayesian Modeling for Large Spatial Datasets. WIREs Computational Statistics, accepted

Fuentes, M. and Foley, K. (2011). Ensembles methods, in upcoming Encyclopedia of Environmetrics.

Fuentes, M., Xi, B. and Cleveland, W. (2010) Trellis Display for Modeling Data from Designed Experiments. Journal of Statistical Analysis And Data Mining, in press.

Reich BJ, Fuentes M. Nonparametric Bayesian models for a spatial covariance. Statistical Methodology. Accepted

Reich, B., Fuentes, M. and Dunson, D. (2011). Bayesian spatial quantile regression. JASA Case Studies, accepted.

Reich BJ, Kalendra E, Storlie CB, Bondell HD, Fuentes M., Variable selection for high-dimensional Bayesian density estimation: Application to human exposure simulation. *Applied Statistics*, accepted.

Zhang, J., Clayton, M. K., and Townsend, P. A., (2011), Missing Data and Functional Concurrent Linear Model for Spatial Images, in preparation.

Zhang, J., Clayton, M. K., and Townsend, P. A., (2010), Functional Concurrent Linear Model for Spatial Images, accepted by Journal of Agricultural, Biological, and Environmental Statistics.