

NSF Annual Report

Final Report for 2008-2009

Submitted to the National Science Foundation

Final NSF Annual Progress Report for 2008-2009

As outlined in the terms of grant DMS-0635449, the following is the Final Annual Progress Report for the Statistical and Applied Mathematical Sciences Institute (SAMSI), for the period August 1, 2008 – July 31, 2009. Past activities that concluded during this period and future activities of SAMSI are also discussed.

0. Executive Summary

A. Outline of Activities and Initiatives for 2008-2009 and the Future	3
B. Financial Overview	6
C. Directorate's Summary of Challenges and Responses.....	7
D. Synopsis of Research, Human Resource Development, and Education	10
E. Evaluation by the SAMSI Governing Board.....	24
<i>Annual Report Table of Contents.....</i>	<i>31</i>

A. Outline of Activities and Initiatives

1. 2008-2009 Programs and Activities Schedule

- Algebraic Methods in Systems Biology and Statistics (Fall 2008, Spring 2009)
 - Opening Workshop and Tutorials (9/14/08-9/17/08)
 - Workshop on Discrete Models in Systems Biology (12/3/08-12/5/08)
 - Workshop on Algebraic Statistical Models (1/15/09-1/17/09)
 - Workshop on Molecular Evolution and Phylogenetics (4/2/09-4/3/09)
 - Transition Workshop (6/18/09-6/20/09)
- Sequential Monte Carlo Methods (Fall 2008, Spring 2009, Summer, 2009)
 - Opening Workshop and Tutorials (9/7/08-9/10/08)
 - Mid-program Workshop (2/19/09-2/20/09)
 - Workshop on Adaptive Design, SMC and Computer Modeling (4/15/09-4/17/09)
 - Transition Workshop (11/9/09-11/10/09)
- Environmental Sensor Networks
 - Transition Workshop (10/20/08-10/21/08)
- Summer Program on Psychometrics
 - Tutorials and Opening Workshop (7/7/09-7/10/09)
 - Intensive Research Week (7/11/09-7/17/09)

Education and Outreach

- 2-Day Workshop for Undergraduates (10/31/08-11/1/08)
- Blackwell-Tapia Conference (11/14/08-11/15/08)
- 2-Day Workshop for Undergraduates (2/27/09-2/28/09)
- Interdisciplinary Workshop for Undergraduates (5/18/09-5/22/09)
- Graduate Student Probability Seminar (5/1/09-5/3/09)

- The Industrial Mathematical and Statistical Modeling Workshop for Graduate Students (7/20/09-7/28/09)
- Graduate Courses at SAMSI
 - Sequential Monte Carlo Methods, Fall 2008
 - Algebraic Methods in Systems Biology and Statistics, Fall 2008

2. 2009-2010 Programs and Activities Schedule

- Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change
 - Summer School (7/28/09-8/1/09)
 - Opening Workshop and Tutorials (9/13/09-9/16/09)
 - GEOMED 2009 and Spatial Epidemiology (11/14/09-11/17/09)
 - Climate Change Workshop (2/17/10-2/19/10)
 - Fundamentals of Spatial Modeling Workshop (3/20/10-3/21/10)
 - Workshop on Statistical Aspects of Environmental Risk (4/7/10-4/9/10)
 - Transition Workshop (10/11/10-10/13/10)
- Stochastic Dynamics
 - Opening Workshop and Tutorials (8/30/09-9/2/09)
 - Workshop on Self-Organization and Multi-Scale Mathematical Modeling of Active Biological Systems (10/26/09-10/28/09)
 - Workshop on Theory and Qualitative Behavior of Stochastic Dynamics (2/8/10-2/10/10)
 - Workshop on Molecular Motors, Neuron Models, and Epidemic on Networks (4/15/10-4/17/10)
 - Transition Workshop (9/27/10-9/29/10)
- Summer Program: Semiparametric Bayesian Inference, with Applications in Pharmacokinetics and Pharmacodynamics
 - Tutorials and Opening Workshop (7/12/10-7/16/10)
 - Intensive Research Week (7/19/10-7/23/10)

Education and Outreach

- 2-Day Workshop for Undergraduates (10/30/09-10/31/09)
- 2-Day Workshop for Undergraduates (2/26/10-2/27/10)
- Interdisciplinary Workshop for Undergraduates (5/17/10-5/21/10)
- The Industrial Mathematical and Statistical Modeling Workshop for Graduate Students (7/19/10-7/27/10)
- Graduate Courses at SAMSI
 - Stochastic Dynamics, Fall 2009
 - Theory of Continuous Space and Space-Time Processes, Fall 2009
 - Spatial Epidemiology, Fall 2009
 - Spatial Statistics in Climate, Ecology and Atmospherics, Spring 2010

3. Programs for 2010-2011

- Analysis of Object Data
- Complex Networks

4. Developments and Initiatives

- NSF awarded SAMSI a \$1.065M supplement to enable hiring of additional postdoctoral fellows.
- The addition to the NISS building was completed and SAMSI moved in during November.
- Pierre Gremaud joined the SAMSI Directorate, replacing Ralph Smith
- A part-time Communications Specialist, Jamie Nunnelly, was appointed.
- Cammey Cole Manning was appointed as Interdisciplinary Undergraduate Coordinator
- A search to replace the Director (by summer 2010) was initiated.
- The directorate structure is being redone as indicated below.
- Additional research collaborations with other institutes were initiated to enhance the overall impact of mathematics and statistics:
 - Activities with the National Center for Atmospheric Research continue, including a joint summer school and a joint postdoctoral appointment.
 - The DMS Mathematical Sciences Institutes began a joint postdoctoral program, because of the academic job crisis.

C. Directorate's Summary of Challenges and Responses

It has been a turbulent, but exciting year at SAMSI. The move to the new building was a challenge, but a wonderful challenge; SAMSI now has the space to allow programs to reach their natural size (as detailed below).

After seven years, SAMSI The scientific programs have been of high caliber, and have led to significant new and ongoing research collaborations between, statistics, applied mathematics, and disciplinary sciences. There has been significant human resource development, through the postdoctoral and graduate programs and through involvement of senior researchers in new interdisciplinary areas. Many students across the country have been shown the SAMSI vision through educational outreach programs and courses. We feel that these successes are amply demonstrated throughout the report; some highlights are given in section D of the Executive Summary. This section discusses the challenges that arose this year and the Directorate's response to these challenges.

Building Addition: Our biggest challenge at SAMSI in previous years was a lack of space. This was finally overcome when the new building addition was finished last Fall.

The 11,782 square foot addition effectively doubled SAMSI's office space, allowing more visitors and stronger engagement by program participants from the Research Triangle universities. The addition also contained space to be shared by SAMSI and NISS, including a break room and accompanying rooftop terrace and a 38-seat lecture room. The lecture room allows most workshops to be held on site. It also has state-of-the-art electronic capabilities that enable simultaneous working group meetings involving remote participants. NISS, which undertook the building addition, and especially NISS Director Alan Karr – who was the key figure in this planning and implementation – are owed a great debt of gratitude by SAMSI.

Of course, equipping the new space with furniture and computers was a challenge, as was the process of moving in, but the SAMSI and NISS staff worked tirelessly to make it as smooth a transition as possible.

The additional space makes it possible to dramatically increase the number of visitors and participants in SAMSI programs. For instance, here is a table summarizing the anticipated long-term participants in next year's main programs:

Expected Participants in 2009-10 SAMSI Programs

	Postdocs	Postdoc Associates	Graduate Visitors	Graduate Locals	Visitors	Faculty Fellows	1-2 month Visitors	Totals
Year-Long	14	4	2	11	6	11		48
Fall	1		5		16		6	28
Spring			1		15		8	24

Directorate Reorganization: As SAMSI has grown, the effort in mounting and administering research programs requires a new directorate structure. The issue is being addressed on two fronts.

First, Dr. Pierre Gremaud of NCSU has been appointed as Deputy Director and will be another full-time (at SAMSI) directorate member. Funds are not currently available to pay for a full-time DD, so the plan – for at least the remaining years of the current SAMSI grant – will be to appoint one of the current Associate Directors as the DD; thus half of the salary of the DD will be provided by the home university (as is now done), while the other half will be provided from SAMSI NSF money. Indeed, the plan is for future university AD appointments, which have been for three years, to be an initial 1.5 years in the current AD role, switching to the DD role for the final 1.5 years of the appointment. This plan will begin in 2010.

A second proposed change is to enhance the role of the Local Scientific Coordinator of a program, who currently is primarily involved, during the year of program operation, with coordinating the presence and activities of local scientists; for this, the LSC is provided (by the partner universities) with a course release. Our plan is to also heavily involve the LSC in the year before the program (with the universities providing a course release for that year also) in assisting the directorate liaison with communication with the program leaders. The LSC will, in effect, become an adjunct member of the directorate, charged with helping the interface between the directorate and the program leaders, and between the partner university department and SAMSI

Program Evaluation: Improving the SAMSI evaluation system is a never-ending challenge. A great deal of information is gathered, including an annual survey of past SAMSI postdocs and program participants. The information from this and the real-time evaluation schemes is presented in Section B and Appendices C and E.

Human Resources: We continue to focus on monitoring the *postdoctoral program*. The applicant pool was excellent this year, with 132 candidates, which was up from last year's 103 candidates. In addition, because of the severe academic job shortage, the DMS mathematics and statistics institutes created a joint supplementary postdoctoral program. The features of this program were as follows:

- There was a joint DMS-institutes webpage for applications.
- SAMSI received 187 applications (out of a total of about 700).
- SAMSI received an additional \$1.065M in funding from the NSF for this program, which is enough for 10 people-years of postdoctoral funding, together with necessary research support.
- Efforts are being made to leverage this funding with second-year support from other sources, to make as many 2-year appointments as possible.
- There will be special efforts in the program aimed towards securing long-term jobs for the postdoctoral fellows.

Obtaining second year support for postdoctoral candidates remains a challenge. This year we partly tried to address the issue by looking ahead, hiring two postdocs “in advance” for 2010-11 programs because we could obtain non-SAMSI support for them in 2009-10.

Advances in Communications and Marketing

A press release was distributed to promote the Blackwell/Tapia conference resulting in 22 media placements (newspapers and websites). A press release about the hiring of additional postdoctoral fellows resulted in almost 50 media placements (newspapers and websites). Meeting notices are distributed to statistics and mathematical journals and organizations who may be interested in attending SAMSI workshops.

A newly-designed newsletter was created and had three editions published in 2008-09. SAMSI also adopted a new e-marketing system called Bronto. The system has 2,753 email contacts from people who are in current SAMSI programs or have gone through previous programs. In 2008-09, SAMSI sent out 20 messages using this system and had an open rate of 26%. SAMSI also created an intranet system using WebOffice. The system has all in-house SAMSI residents and staff's contact information, the in-house master calendar, and ways to hold internal discussions, assign tasks and organize files.

Posters were designed and distributed for the two main programs and for the Education and Outreach program. A special poster was designed for the IMSM workshop, which is jointly sponsored by SAMSI, the NSF, CRSC and NCSU. A new banner was purchased to take to trade shows and meetings. New signs were also purchased and are used at SAMSI's workshops.

In order to reach out to the younger SAMSI audience, a SAMSI Education and Outreach group was created on Facebook. The group has 89 members in it. There is also a SAMSI facebook page that has 51 fans. SAMSI has a LinkedIn group with 36 members participating in the group. NISS and SAMSI are sharing a Twitter account name, @NISSSAMSI, and started to do a few messages in 08-09, as of August 09, it had about 27 followers.

D. Synopsis of Developments in Research, Human Resource Development, and Education

In later parts of the report, the extensive developments in research and education that have occurred this past year in the SAMSI research programs are discussed in detail. To give a flavor of these developments, we highlight some of the findings here, focusing on those for which primary activity ended during this past year.

1. Research

a) RISK ANALYSIS, EXTREME EVENTS AND DECISION THEORY

The past half decade has seen great interest and also great scientific progress on risk analysis; both natural and man-caused events have focused on both the science and the modeling of these events and on the analysis of their associated risks.

The working group on *Adversarial Risk* has extended game-theoretic principles and methods to develop foundational theory for adversarial risk analysis. Previously this theory, while conceptually attractive, had been considered irrelevant for practical purposes. Examination of several formulations of adversarial risk problems, with opportunities for opposition, cooperation and negotiation, has led to a unified framework for analysis. The key contribution is a way to build a rational probabilistic model for actions of the adversary that can then feed into a decision analytic model. From this framework, negotiation and arbitration schemes can be formalized that also have been extended to non-convex utility sets and to multiple agents (adversaries).

Computational aspects of risk analysis brought together researchers from the *Adversarial Risk* and the *Service Sector Risk* working groups. Taking cybersecurity as an example, a series of papers addresses the formalization of risk approaches and then proceeds to use a Bayesian approach to consider the specific problems of modeling and forecasting hardware/software system reliability. Standard approaches to risk analysis, based on parameter estimation and then computations from risk models, underestimate uncertainty – a grave weakness for risk analysis and management. So, a new alternative was developed to compute posterior distributions for the parameters, then compute a posterior predictive risk efficiently using reduced order models. This work will constitute the basis for the chapters on Bayesian Risk Analysis and on Bayesian Reliability in a book project, *Bayesian Analysis of Stochastic Processes*, Fabrizio Ruggieri (working group participant) and Mike Wiper, editors.

Theoretical results from the *Multivariate Extremes-Methodology* and *Bayesian Methods for Extremes* working groups develop distributional (and semi-parametric) theory for multivariate extremal data and for mixed data which is extremal only in certain dimensions of the parameter space. Quantile-based elicitation methods were derived for semiparametric functional estimation in this mixed case. These models (mixture models) were applied to extreme values in river flow, both univariate and multivariate cases.

Both the *Environmental Risk Analysis* and the *Multivariate Extremes – Applications* working groups focused on specific applications and the implementation of methodology, both old and concurrently developed by other working groups. The *Environmental Risk* working group concentrated on data on ozone levels from 95 cities. (12 of these were studied by the Environmental Protection Agency in order to consider lowering the US ozone standard for air pollution.) Development and testing of the consequences of three possible new ozone standards required a predictive model for ozone levels (and exceedences). This research assessed the sensitivity of the predictive inferences to the choice of computational approach, to the various model assumptions and to the model uncertainty. The *Multivariate Extremes – Applications* working group concentrated on implementation of the Ledford-Tawn methodology and evaluation of empirical results using this approach. Thus Atlantic hurricanes and sea surface temperatures were modeled successfully as a bivariate time series with a non-trivial correlation structure.

The organizers are working on a volume to be produced for the ASA-SIAM book series, highlighting the research and proposed research directions resulting from the program.

A book *Bayesian Analysis for Stochastic Process Models*, is being produced by SAMSI participants David Rios Insua and Fabrizio Ruggeri, based in large part on SAMSI research. The book will be produced by Wiley and is expected to appear later in 2009.

Program leader David Rios reports the following collaborative projects resulting from his participation in the SAMSI program:

- A major grant from the Spanish ministry of research and innovation, for 2009-2011, on collaborative decision making with applications to counterterrorism. The SAMSI participating researchers are David Rios Insua, Jesus Rios, David Banks, and Fabrizio Ruggeri.
- A Fulbright grant awarded for 2009 for risk analysis modeling in information and communication technologies, to continue collaborative research started while at Samsi: the Fulbright will fund Dipak Dey (Connecticut) and Javier Cano.
- A major grant from the Spanish ministry of industry, for 2009-2013, to establish a center for risk analysis related to information and communication technology solutions for public administration; SAMSI researchers participating are David Rios Insua, Lea Deleris and Jesus Rios (SAMSI postdoc).
- A startup for applying risk analysis for insurance and certification purposes has been initiated, with 50 % of the required investment obtained from the Centre for the Development of Industrial Technology (Spain); SAMSI researchers participating are David Rios Insua and Jesus Rios.
- A follow-up project on fraud detection methods for telecom transactions was funded by Habber TEC (patent pending).
- A follow-up collaborative research project on Bayesian methods for discrete event simulation, involving SAMSI researchers Haipeng Shen, Mircea Grigoriu, David Rios Insua, and Javier Cano has applied for funding.

b) RANDOM MEDIA

In the working group on *Waves and Imaging*, a range of new collaborations were established, including Yvonne Ou with Jean-Pierre Fouque and Josselin Garnier; Gabriel Peyre with Laurent Demanet; and Sava Dediu with Laurent Demanet. One of the first outcomes from these collaborations was a novel method for efficiently solving wave equations in the context of inverse problems in seismology. The backdrop for this effort was the group meeting's extensive discussion on nonlinear sampling strategies in imaging, including compressed sensing, during the Fall of 2007. What became apparent is that the ideas of sparsity and undersampling suggest an entirely different strategy for simulating linear wave phenomena on a large computational scale, using nonlinear synthesis from a few eigenfunctions of the Helmholtz operator, chosen at random. The main mathematical question concerned the number of such eigenfunctions needed for a given accuracy guarantee, and was solved during the random media program. Under mild assumptions, the answer is a remarkable $O(\log(N))$ where N is the desired resolution. More collaborators will join this effort as the potential impact of this discovery in reflection seismology is now clear: the compressive viewpoint yields embarrassingly parallel algorithms that promise to help rethink the main computational bottlenecks of adjoint-state methods on large CPU clusters.

The working group on Heterogeneity in Biological Materials has a number of consequences to report:

- Based on a collaboration started at SAMSI between applied mathematicians, probabilists and statisticians, Scott McKinley (Duke), Lingxing Yao (Utah), Christel Hohenegger (NYU-Courant), Tim Elston (UNC), John Fricks (Penn State), and Gustavo Didier (Tulane) submitted an FRG to NSF-DMS on "Viscoelastic Diffusion". This is a problem of major importance today in materials science, environmental health, and lung disease.
- SAMSI Graduate RAs Ke Xu and Brandon Lindley both have published papers leading to their primary thesis results. Ke graduates this May, and she worked with Isaac Klapper (Montana State) while he visited SAMSI. Brandon graduated this past May 08, and took a position at U. South Carolina to work on biofilms, a topic he was introduced to at SAMSI.
- Mansoor Haider and Greg Forest have followed up on their working group to organize a large mini-symposium at the September regional AMS meeting at NC State, on the research topics of the working group.
- Greg Forest and H. Zhou (Naval Postgraduate School) organized a mini-symposium at the SIAM annual meeting in San Diego this past summer 08 on research from the working group.

c) ENVIRONMENTAL SENSOR NETWORKS

Friend, foe, or something in between?

Sagebrush has expanded its range over decades as a result of overgrazing and fire suppression in western arid landscapes of the U.S. The native bunchgrasses that initially attracted ranchers have dwindled. Though sagebrush has become something of an emblem of the human-induced shifts in Western landscapes, sagebrush has also taken on

a subterranean role in those systems, a role that is not well understood. There is growing evidence that sagebrush (and other deep-rooted plants in arid ecosystems worldwide) is able to facilitate movement of water from moist soil layers to dry soil layers, allowing the plants to, in essence, bank excess water for use during drier periods. For example, when precipitation is plentiful, water flows through sagebrush roots at the surface down through the root system to drier soil at depth; that water is then redistributed upward via the same roots when upper soil layers dry out. Those upper soil layers are where grasses are rooted, and there is some evidence that water redistributed by sagebrush to dry upper layers can make its way into surrounding vegetation. An interdisciplinary group led by Zoe Cardon (Marine Biological Laboratory, MA) has developed tools for hierarchical Bayesian analysis of shifting water content at various soil depths to characterize this process. One surprising outcome is that recharging of dry soils can be due to a combination of monsoon rainfall and redistribution from deep water triggered by atmospheric water content. The monsoon season, with its punctuated periods of high humidity in an otherwise desert-like environment, may be critical not only for its rainfall, but also for its triggering of extensive water redistribution from deep, moist soil, even when there is no rainfall but atmospheric humidity is high.

From data to information:

Wireless sensor network datasets are characterized by various errors in transducers, sensor node hardware, and communication, and they currently require *ad hoc* human analysis to detect, analyze, and “clean” them. As sensor network deployments grow worldwide, the volume of data will grow tremendously, so current techniques will not scale. A group of researchers led by Ernst Linder (University of New Hampshire) is developing a simple yet powerful automated algorithm for anomaly detection and cleaning of sensor network datasets based on median polish algorithms that integrate human intelligence in the learning process. These algorithms can be easily tuned by researchers to optimize separation of faulty data from valid results, dramatically reducing the effort needed to prepare datasets for further analysis.

Give me meaningful data, please:

A fundamental challenge in wireless sensor networks is the minimization of energy usage so that battery lifetime is as long as possible. Transmission suppression schemes are a promising approach to reduce energy use in sensor networks by using predictive models to suppress reporting of predictable data. However, a fundamental problem is that it is difficult to distinguish suppressed data from data lost due to the inherent unreliability of wireless communication. Progressive, or cascaded, suppression involves suppression of more and more data as it is funneled to the network hub, and promises significant reduction in energy consumption. However, it makes failure handling very difficult, because nodes may act on incomplete and incorrect information in turn affecting other nodes, so that decision errors may also cascade. Jun Yang (Duke University) is working with colleagues to develop a cascaded suppression framework that fully exploits temporal and spatial data correlation, and applies coding theory and Bayesian inference to identify and recover missing data.

d) META-ANALYSIS

The most important “take-home message” from the summer Meta-analysis program was that the concept of multiple sources of evidence itself needs to be generalized and applied more generally and creatively through many if not all areas of statistical practice and research. Multiple sources should not just be taken as separate studies or even the possible simple regrouping of subsets of observations within studies but the bringing to bear of seemingly distinct information sources on given question and even the “creation” of multiple sources as in Bayesian Additive Regression Trees (BART) where differing regression trees are purposefully grown to be later advantageously combined. In some fields, terms like data fusion and data integration are being used for this more general sense of utilizing multiple sources of evidence. For instance, a single strategy of no pooling, complete pooling or partial pooling of separate studies needs to give way to adaptive strategies where the degree of pooling is individually chosen for each and every parameter in the joint probability model used to represent all the relevant sources of evidence. In summary, meta-analysis needs to be recast as just an obvious instance of more generally and perceptively dealing with multiple sources of evidence, in both statistical applications and theory.

The following summary of work by Liu, F., Dunson, D.B. and Zou, F. (2008) demonstrates both a generalized concept of multiple sources of evidence (gene coefficient estimates and annotations) and the replacement of a single strategy of no pooling, complete pooling or partial pooling of studies with an adaptive strategy where the degree of pooling is individually chosen for the different coefficients (parameters).

In large scale genetic epidemiology studies that collect massive numbers of single nucleotide polymorphisms (SNPs) or gene expression measurements, it is extremely challenging to identify genes that are predictive of disease phenotypes given the modest sample size of most studies relative to the number of genes. Due to concern about false positive rates, it is crucial to replicate findings about disease genes in multiple studies. Standard approaches take multistage testing approaches in which one tests if genes identified in initial studies are significant in follow-up studies. This strategy is shown to have major disadvantages in terms of power and type I error rates compared with an innovative approach developed in the SAMSI meta-genetics working group based on simultaneous selection through a multi-task relevance vector machine (MT-RVM) procedure. This approach, which is related to methods used in signal processing, borrows information across studies in the degree of shrinkage of gene-specific coefficients towards zero. The method is scalable to large numbers of genes, can accommodate censored data commonly collected in disease recurrence studies, and clearly outperforms common competitors, such as Lasso. In addition, the meta-genetics group is currently pursuing a new procedure that allows information on gene function annotation to be incorporated, while automatically learning how predictive each annotation source is. The annotated relevance vector machine (aRVM) procedure should be very widely useful in machine learning and other applications beyond genetics, as it allows an adaptive targeted search for important predictors enabling an effective reduction in dimensionality and mechanism for borrowing information across disparate studies.

e) SEQUENTIAL MONTE CARLO METHODS

Much of the work of this program will be finalized in the coming months, but research by the working groups has already produced preliminary results of considerable interest.

The *Tracking and Large-scale Dynamical Systems* working group has made a number of significant advances:

- They have developed simulation code and a range of SMC solution methodologies for the hard multi-object tracking problem in clutter, including random appearance and disappearance of objects from the scene, and unknown numbers of objects using sequential variable dimension methods. They have also developed new smoothing techniques for random finite set observation data.
- In the cloud tracking area advances have been made in detection of multiple chemical releases sequentially tracked over time, relying on a newly developed sequential trans-dimensional ABC algorithm. In addition, the group has developed methods for tracking of irregular-shaped dynamical plume data, using novel sequential MCMC procedures. Results have impressed UK government agencies (DSTL) involved in providing alerts for chemical/biological attacks.

The *Theory* group is studying product estimators for achieving provably good integration performance from random samples, whether MCMC or SMC-based.

The *Population MC* group has made some advances in the design of fully adaptive Monte Carlo methods. They have been working on a methodology which allows one to compute on-line efficient cooling schedules, automated design of importance distributions. These algorithms have been used to solve approximate Bayesian computation problems. Current developments include the design of new adaptive MC methods for computing normalizing constants.

The *Particle Learning* working group has made fast progress in parameter learning for some key application models from economics, epidemiology and neurological data. They are demonstrating efficient sequential learning in highly nonlinear, non-Gaussian models that previously could only be estimated with MCMC.

The *Model Assessment and Adaptive Design* group reports exciting developments:

- New methods for high dimensional model selection and modeling: Particle stochastic search for nonparametric variable selection and augmented particle learning for Bayesian distribution regression.
- Sequential learning for dynamical graphical model structures using particle approximations with application to areas such as financial portfolio analysis.

The *Continuous Time* group is making advances in filtering for diffusions using a new least-action approach, filtering for continuous time branching processes and for survival data. In addition they are commencing work on sequential inference for α -stable Levy processes.

f) ALGEBRAIC METHODS IN SYSTEMS BIOLOGY AND STATISTICS

Much of the work of this program will be finalized in the coming months, but research by the working groups has already produced preliminary results of considerable interest.

The *Evolutionary Biology* working group has made several advances:

- We have shown that “There is no caterpillar in a wicked forest.” This settles a conjecture of Degnan and Rosenberg. A (rooted) caterpillar is a type of tree in which there exists an interior node descended from all other nodes. A wicked forest is a set of trees with a particular nasty property. For any pair of tree topologies A and B in a wicked forest, an observation of a high proportion of gene trees with topology A is evidence that the species tree has topology B, and an observation of a high proportion of gene trees with topology B is evidence that the species tree has topology A. The result is that none of the topology in a wicked forest can be a caterpillar topology.
- We introduced the idea of k-interval speciation to quantify the amount of co-evolution between two trees. We prove that two trees satisfying 1-interval cospeciation are, equivalently, separated by one Nearest Neighbour Interchange operation, which has been well-studied in theoretical phylogenetics.
- We present a polynomial-time algorithm for finding the geodesic distance between two trees in tree space. It is based on producing a sequence of paths, where each successive path is formed by “bending” edges of the previous path. These intermediate paths correspond to “sliding” the legs of the path through tree space to successively shorten the path until the geodesic is obtained.

The *Algebraic Statistics and Experimental Design* working group is investigating two major related projects. The first is polynomial representation of probabilities and the second is lifting cumulant theory from finite discrete distributions to continuous distributions using the concept of finite generation. This work would be a generalization of Morris' classification.

The *Network Inference* working group has studied biochemical reaction networks with mass action kinetics defining a system of ODEs with polynomial nonlinear right hand side. We derive conditions for the existence of at least two positive distinct steady state solutions by introducing sufficient conditions for the existence of a transformation that reduces the polynomial system to a linear one.

2. Human Resource Development

SAMSI's impact on human resources is fully discussed in sections I.B and I.C, with impact on diversity highlighted in section I.H. The individual program reports also contain significant insight into human resource development. Here we give a summary of SAMSI's impact on human resource development and highlight specific examples.

SAMSI's postdoctoral fellows and associates again in 2008-09 have embraced the interdisciplinary tenor of SAMSI programs and have engaged with visible enthusiasm in the activities for graduate and undergraduate students. Most of those completing their SAMSI fellowships are explicitly committed to continuation of interdisciplinary collaborative research and/or interdisciplinary research with SAMSI collaborators; the other two taking up positions as assistant professors expect to find opportunities for collaboration when they arrive at their new posts.

As has happened in previous years, many new collaborations were established at SAMSI this year; the highlights above, as well as the program reports, discuss many of these collaborations.

The impact of new technology for remote participation in SAMSI working groups has yet again increased; essentially every working group is actively using remote access to working group meetings to include participants located outside the Triangle area, many located outside the US. In some cases, even the working group leaders are remote. One such leader, Elizabeth Allman, states "We have found both the talks last term and the readings this semester to be very valuable, particularly since we are situated so far away in distance. This really is a "plus" that SAMSI offers."

An unplanned secondary success of incorporation of remote participants is the extension of the lifetime of the working group. Numerous working groups from previous SAMSI programs still operate, utilizing the SAMSI technology, even though none are actually present at SAMSI.

The detailed participant lists for concluded programs provide ample evidence of the national and international draw of SAMSI activities. SAMSI programs attracted 19 long-term visitors (3 months or more), 47 short-term visitors (a week to 3 months), 11 local faculty fellows, 9 postdoctoral fellows and associates, 18 graduate students (7 visiting), and a total participant list of more than 1000 researchers. During 2008-09, 123 researchers participated remotely as individuals in working groups.

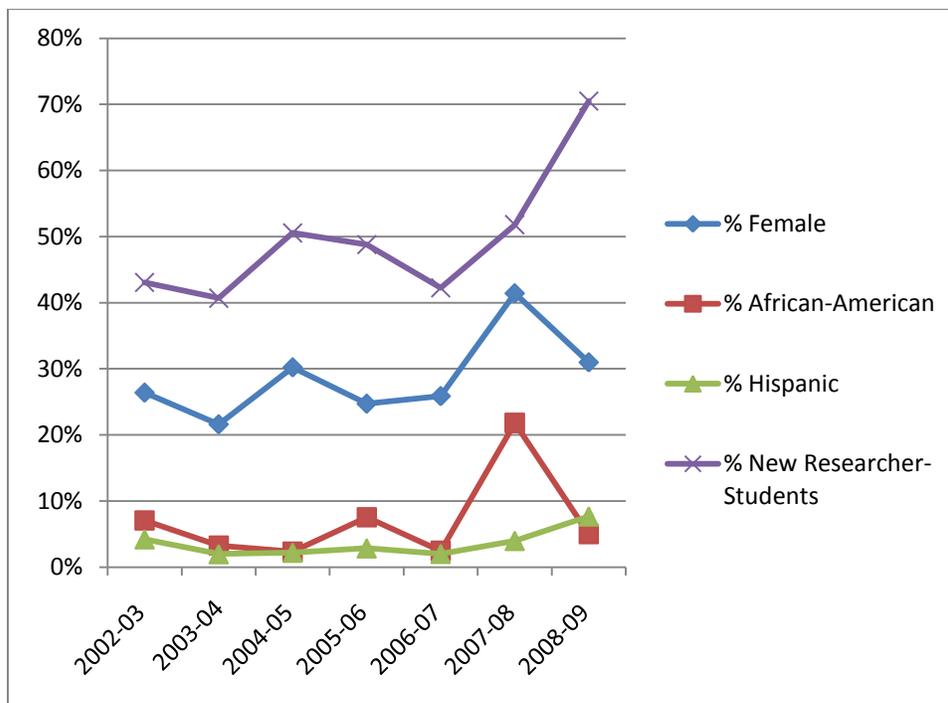
Diversity: SAMSI policy is to give attention to diversity issues throughout all activities, especially in the Postdoc selection process and in the organization and operation of Workshops and Programs. Some highlights of this effort over the past year:

- SAMSI has developed a web page devoted to our diversity activities. The page advertises the various program activities related to minority outreach and has links to other diversity related information outside of SAMSI.
- On Nov. 14-15, 2008, SAMSI hosted the 6th Blackwell-Tapia Conference. This bi-annual event brings together African-American, Native American and

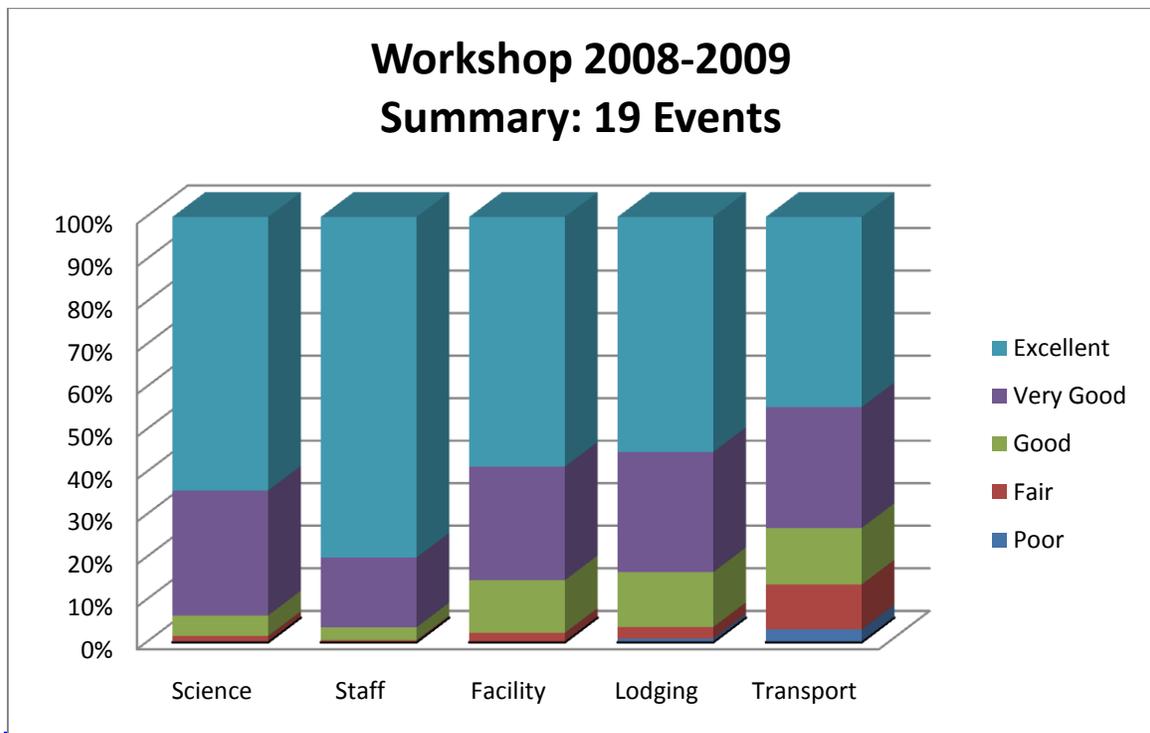
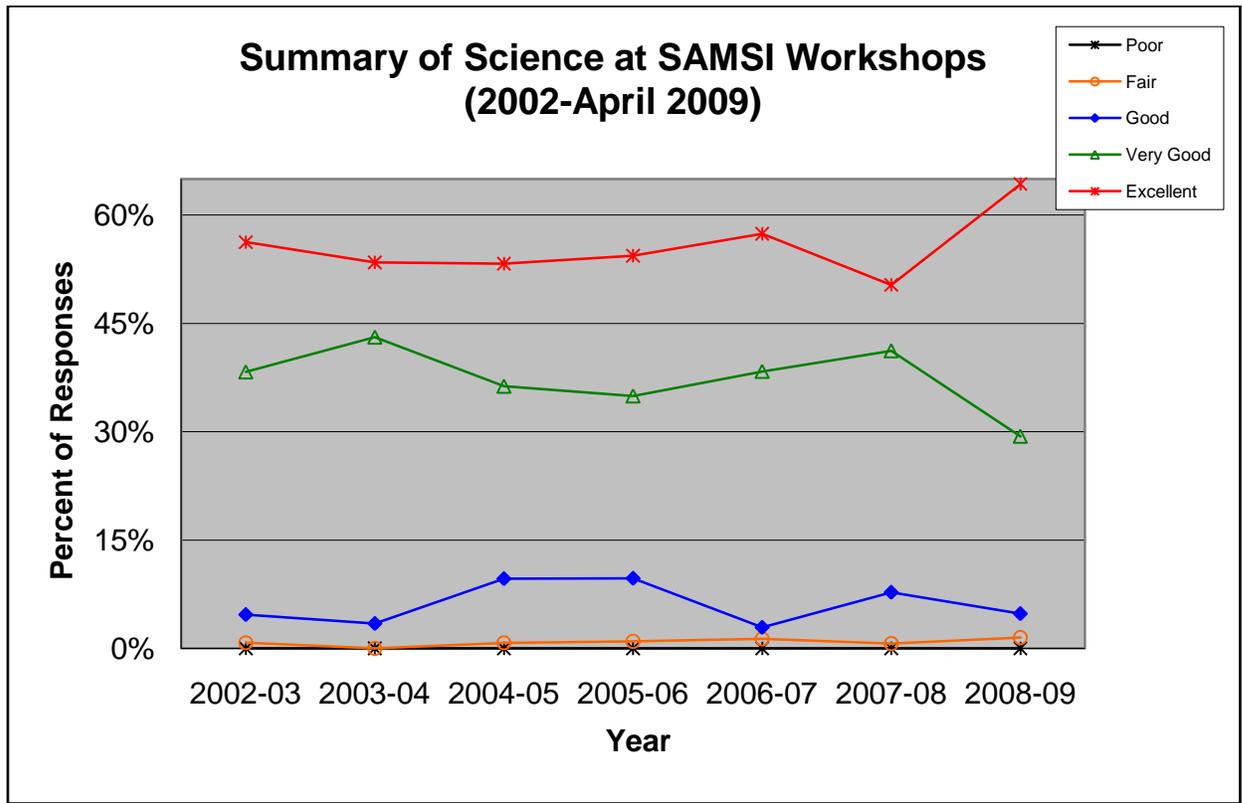
Latino/Latina students, faculty, and researchers from mathematics and statistics. This two day event was attended by over 100 participants, and consisted of research talks, panel discussion of issues relating to minority recruitment, retention, and mentoring, as well as a dinner to honor the 2008 Blackwell-Tapia prize winner Juan Mesa of Lawrence Berkeley Laboratory.

- Michael Minion has been serving as SAMSI's representative to the NSF Institutes' Diversity Coordination Committee which was formed in 2006 by Chris Jones (SAMSI) and Helen Moore (formerly of AIM), and is now chaired by Kathleen O'Hara (MSRI). One effort of SAMSI and this committee was the *Modern Math* program at the 2008 SACNAS National Convention in Salt Lake City. This program was aimed at introducing young scientists to current research topics, providing mentorship and networking opportunities, and recruiting future participants in NSF Institute programs from under-represented groups.

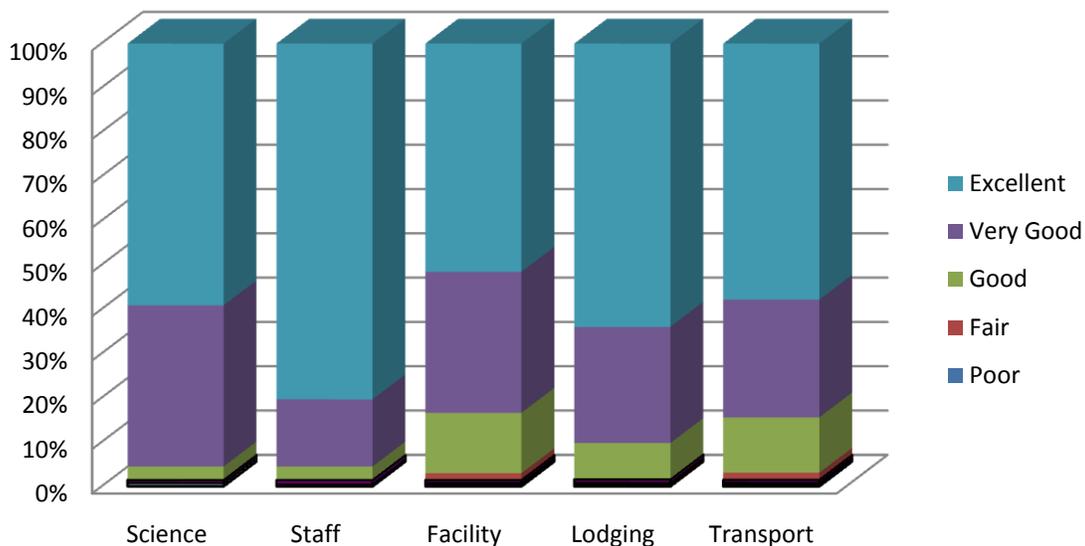
Overall Participation in Workshops by Underrepresented Groups: Here is an overall summary of the participation by underrepresented groups at SAMSI events. Note that large spike in participation by Females and African Americans in 2007-08 was partly due to the Infinite Possibilities Conference, which focused on attracting female African-Americans to a career in mathematics and statistics.



Workshop Evaluations: Detailed evaluations of workshops are given in Appendix F. Here are the summary graphs indicating the satisfaction of participants.



Under graduate Programs 2008-2009 Summary: 4 Events



It is SAMSI's policy always to attract and support the leading scientists, regardless of nationality; but to otherwise focus resources on domestic participants. The table below shows the nationality status of the participants who received some funding from SAMSI.

Year	US Citizen or Permanent Resident	Foreign National Residing in US	Foreign National Not Residing in US	TOTAL
2002-03	209	87	36	332
2003-04	220	90	29	339
2004-05	158	71	21	250
2005-06	217	101	37	355
2006-07	222	146	60	428
2007-08	382	124	45	551
2008-2009	248	112	66	426
TOTAL	1656	731	294	2681
Percentage of all funded participants	61.77%	27.27%	10.97%	

Broadening the DMS research impact: SAMSI’s national impact also depends on Institutional Diversity and the inclusion of participants whose home institutions are not already heavily supported by NSF Funding through DMS. Such inclusion develops the national research base by significantly increasing the number of individuals that can engage in cutting edge research. The SAMSI record in this regard during 2008-09 is excellent, as shown in the following table (for both funded participants and all participants). The ‘Other’ category primarily includes individuals from other disciplines, governmental agencies or laboratories, and industry.

2008-2009 SAMSI Participation			
Funded Participants	Home Institution by DMS Funding Level		
	Top 50 DMS Funded	51-200 DMS Funded	Other
# of Institutions	37	48	74
# of People	148	107	131
% People	38.34%	27.72%	33.94%
All Participants			
	Top 50 DMS Funded	51-200 DMS Funded	Other
# of Institutions	42	61	151
# of People	333	197	231
% People	43.76%	25.89%	30.36%

3. Education

The impact of SAMSI courses and various components of the SAMSI Education and Outreach program are documented in Section I.E. Part 4 and various program reports. We summarize here specific new initiatives and specific highlights of the program.

(i) Two outreach workshops were held to expose undergraduate students from programs around the country to topics and research directions associated with the SAMSI Programs on *Algebraic Methods in Systems Biology and Statistics*, and *Sequential Monte Carlo Methods*. One goal of these workshops was to illustrate the application and synergy between mathematics and statistics which goes far beyond that which students have seen in coursework. The overall objective was to broaden the perspective of students with regard to both future graduate studies and career choices.

(ii) The one-week SAMSI Workshop for Undergraduates encompassed three highly unique components.

- All tutorials and sessions were presented by SAMSI graduate students and postdocs under close supervision of directorate members, members of the Education and Outreach Committee, and local faculty.
- The workshop provided students with an intensive introduction to the synergy between applied mathematics and statistics in the context of physical applications.
- During one of the sessions, the students were introduced to a variety of experiments and each team collected their own physical data.

(iii) The overall goals of the ten-day Industrial Mathematical and Statistical Modeling Workshop for Graduate Students were twofold:

- Expose mathematics and statistics students to current research problems from government laboratories and industry which have deterministic and stochastic components;
- Expose students to a team approach to problem solving.

For the 2008 workshop, research problems were presented by scientists from Glaxo Smith Kline, MIT Lincoln Laboratory, the National Institute of Statistical Sciences, Republic Mortgage Insurance Co and SAS. Each team gave a 30 minute oral presentation summarizing their results on the final day of the workshop and written reports were compiled as the SAMSI Technical Report 2008-11 which can be obtained at <http://www.samsi.info/reports/index.shtml>.

(iv) The Kenan Fellows Program pairs mentors from the SAMSI community with K-12 public school teachers who have been selected to be Kenan Fellows. The program's goals include promoting teacher leadership, developing and disseminating exciting new curriculum in science, technology, and math education, and addressing the problem of teacher retention in public schools.

SAMSI is sponsoring two Kenan Fellows. Danielle DiFrancesca, working with NISS director Alan Karr, associate director Nell Sedransk and assistant director Stanley Young, is developing materials to enable middle school students to think critically about scientific material they encounter on television, on the Internet and elsewhere. SAMSI

Associate Director Michael Minion and his colleague at UNC, Professor Laura Miller, are serving as co-mentors for Ms. Jenny Rucker, a Kenan Fellow from West Cary Middle School sponsored by SAMSI. Ms. Rucker has been working with Minion and Miller to implement curriculum based on her project Pumping and Moving Through Fluids at Different Sizes: Mathematical Models to Describe Fluid Behavior.

E. Evaluation by the SAMSI Governing Board - 2009

(Bruce Carney, George Casella, Donald Estep, James Landwehr, John Simon, Daniel Solomon – Chair)

The Governing Board provides broad oversight for the Institute's administration, finances, and evaluation, and for relationships among the partnering institutions. In recognition of the evolution of the Institute, the Governing Board has elected to modify slightly the set of questions it has historically addressed in its annual report. Our evaluation, as responses to three broad questions follows:

1) What are some outcomes of the synthesis of applied mathematics and statistics?

SAMSI continues to foster interaction between applied mathematics and statistics through the creation of programs focused on topics that involve both disciplines. Working groups established under these programs build teams of researchers consisting of applied mathematicians and statisticians as well as other areas of the mathematical sciences. The results of these efforts lie not only in the production of many papers and reports, but also in the continued interaction among members of the teams after the formal program is completed, and, most importantly, in the culture of multidisciplinary interaction it has established.

Some of the interactions between applied mathematics and statistics (and other disciplines) for which primary activity ended or started in the past year, and that we noted in the annual report, are listed below.

In the program on **Random Media**, the working group on *Heterogeneity in Biological Materials* reports that a collaboration was started at SAMSI among applied mathematicians, probabilists and statisticians. This led Scott McKinley (Duke), Lingxing Yao (Utah), Christel Hohenegger (NYU-Courant), Tim Elston (UNC-CH), John Fricks (Penn State), and Gustavo Didier (Tulane) to submit an FRG to DMS on "Viscoelastic Diffusion". This is a problem of major importance today in materials science, environmental health, and lung disease.

The program on **Environmental Sensor Networks** involved statisticians, environmental scientists, applied mathematicians, engineers, computer scientists, and probabilists. These collaborations have produced three advances:

- *Is Sagebrush a friend or foe?* An interdisciplinary group led by Zoe Cardon (Marine Biological Laboratory, MA) has developed tools for hierarchical Bayesian analysis of shifting water content at various soil depths to characterize the process of water redistribution by sagebrush roots. One surprising outcome is that recharging of dry soils can be due to a combination of monsoon rainfall and redistribution from deep water triggered by atmospheric water content.
- *From data to information:* A group of researchers led by Ernst Linder (University of New Hampshire) is developing a simple yet powerful automated algorithm for

anomaly detection and cleaning of sensor network datasets based on median polish algorithms that integrate human intelligence in the learning process. These algorithms can be easily tuned by researchers to optimize separation of faulty data from valid results, dramatically reducing the effort needed to prepare datasets for further analysis.

- *Give me meaningful data, please:* A fundamental challenge in wireless sensor networks is the minimization of energy usage so that battery lifetime is as long as possible. Transmission suppression schemes are a promising approach to reduce energy use in sensor networks by using predictive models to suppress reporting of predictable data. Jun Yang (Duke University) is working with colleagues to develop a cascaded suppression framework that fully exploits temporal and spatial data correlation, and applies coding theory and Bayesian inference to identify and recover missing data.

Most of the working groups in the **Sequential Monte Carlo Methods** program involved people from a range of disciplines, including applied mathematics and statistics.

- The *Theory* group involved engineers, probabilists and statisticians, and is studying product estimators for achieving provably good integration performance from random samples, whether MCMC or SMC-based.
- The *Population MC* group involved probabilists, engineers, and statisticians, and has made significant advances in the design of fully adaptive Monte Carlo methods. These algorithms have been used to solve approximate Bayesian computation problems. Current developments include the design of new adaptive MC methods for computing normalizing constants.
- The *Model Assessment and Adaptive Design* group involved statisticians and operations researchers and reports these developments:
 - New methods for high dimensional model selection and modeling: Particle stochastic search for nonparametric variable selection and augmented particle learning for Bayesian distribution regression.
 - Sequential learning for dynamical graphical model structures using particle approximations with application to areas such as financial portfolio analysis.
- The *Continuous Time* group involved applied mathematicians, statisticians, and probabilists and is making advances in filtering for diffusions using a new least-action approach, filtering for continuous time branching processes and for survival data. In addition they are commencing work on sequential inference for α -stable Levy processes.

The *Algebraic Statistics and Experimental Design* working group in the program on **Algebraic Methods in Systems Biology and Statistics** involved numerous statisticians and applied mathematicians and is investigating two major related projects. The first is polynomial representation of probabilities and the second is lifting cumulant theory from finite discrete distributions to continuous distributions using the concept of finite generation. This work would be a generalization of Morris' classification. The *Systems Biology* working group in the program involves numerous applied mathematicians and statisticians, and has studied biochemical reaction networks with mass action kinetics defining a system of ODEs with polynomial nonlinear right hand side. They derive conditions for the existence of at

least two positive distinct steady state solutions by introducing sufficient conditions for the existence of a transformation that reduces the polynomial system to a linear one.

2) Is the impact and national recognition of SAMSI on science and human resources commensurate with the scale of SAMSI?

Section D of the Executive Summary describes the developments in research, human resource development, and education. It is clear that SAMSI continues to have a significant impact on disciplinary sciences. We highlight a few areas below. SAMSI programs have also been influencing the research careers of program participants, helping to refocus research directions for some senior researchers and providing formative experiences for post-docs and other junior scientists.

The program on **Risk Analysis, Extreme Events and Decision Theory** has made a number of advances that impact a variety of disciplines.

- The working group on *Adversarial Risk* has extended game-theoretic principles and methods to develop foundational theory for adversarial risk analysis. Previously this theory, while conceptually attractive, had been considered irrelevant for practical purposes. Examination of several formulations of adversarial risk problems, with opportunities for opposition, cooperation and negotiation, has led to a unified framework for analysis. The key contribution is a way to build a rational probabilistic model for actions of the adversary that can then feed into a decision analytic model.
- Computational aspects of risk analysis brought together researchers from the *Adversarial Risk* and the *Service Sector Risk* working groups. Taking cybersecurity as an example, a series of papers addresses the formalization of risk approaches and then proceeds to use a Bayesian approach to consider the specific problems of modeling and forecasting hardware/software system reliability. Standard approaches to risk analysis, based on parameter estimation and then computations from risk models, underestimate uncertainty – a grave weakness for risk analysis and management. So, a new alternative was developed to compute posterior distributions for the parameters, then compute a posterior predictive risk efficiently using reduced order models.
- Both the *Environmental Risk Analysis* and the *Multivariate Extremes – Applications* working groups focused on specific applications and the implementation of methodology, both old and concurrently developed by other working groups. The *Environmental Risk* working group concentrated on data on ozone levels from 95 cities. (12 of these were studied by the Environmental Protection Agency in order to consider lowering the US ozone standard for air pollution.) Development and testing of the consequences of three possible new ozone standards required a predictive model for ozone levels (and exceedences). This research assessed the sensitivity of the predictive inferences to the choice of computational approach, to the various model assumptions and to model uncertainty. The *Multivariate Extremes – Applications* working group concentrated on implementation of the Ledford-Tawn methodology and evaluation of empirical results using this approach. Atlantic hurricanes and sea

surface temperatures were modeled successfully as a bivariate time series with a non-trivial correlation structure.

- The organizers are working on a volume to be produced for the ASA-SIAM book series, highlighting the research and proposed research directions resulting from the program.

For the program **Random Media**, the working group on *Waves and Imaging* developed a novel method for efficiently solving wave equations in the context of inverse problems in seismology. The backdrop for this effort was the group's extensive discussion on nonlinear sampling strategies in imaging, including compressed sensing, during the Fall of 2007. What became apparent is that the ideas of sparsity and undersampling suggest an entirely different strategy for simulating linear wave phenomena on a large computational scale, using nonlinear synthesis from a few eigenfunctions of the Helmholtz operator, chosen at random. The main mathematical question concerned the number of such eigenfunctions needed for a given accuracy guarantee, and was solved during the random media program. Under mild assumptions, the answer is a remarkable $O(\log(N))$ where N is the desired resolution. More collaborators will join this effort as the potential impact of this discovery in reflection seismology is now clear: the compressive viewpoint yields embarrassingly parallel algorithms that promise to help rethink the main computational bottlenecks of adjoint-state methods on large CPU clusters.

The Governing Board delights in reporting that the *Evolutionary Biology* working group in the program on **Algebraic Methods in Systems Biology and Statistics** has shown that “There is no caterpillar in a wicked forest.” This settles a conjecture of Degnan and Rosenberg.

Explanation: A (rooted) caterpillar is a type of tree in which there exists an interior node descended from all other nodes. A wicked forest is a set of trees with a particularly nasty property. For any pair of tree topologies A and B in a wicked forest, an observation of a high proportion of gene trees with topology A is evidence that the species tree has topology B , and an observation of a high proportion of gene trees with topology B is evidence that the species tree has topology A . The result is that none of the topology in a wicked forest can be a caterpillar topology.

The *Tracking and Large-scale Dynamical Systems* working group from the program on **Sequential Monte Carlo Methods** has made a number of significant advances:

- They have developed simulation code and a range of SMC solution methodologies for the hard multi-object tracking problem in clutter, including random appearance and disappearance of objects from the scene, and unknown numbers of objects using sequential variable dimension methods. They have also developed new smoothing techniques for random finite set observation data.
- In the cloud tracking area, advances have been made in detection of multiple chemical releases sequentially tracked over time, relying on a newly developed sequential trans-dimensional ABC algorithm. In addition, the group has developed methods for tracking of irregular-shaped dynamical plume data, using novel sequential MCMC procedures. Results have impressed UK government agencies (e.g. Defence Science and Technology Laboratory) involved in providing alerts for chemical/biological attacks.

The lists of refereed publications associated with SAMSI programs (see Section I.G. of the full report) provide another measure of the impact on the mathematical and disciplinary sciences. There were 25 accepted publications over the year, roughly equally divided between statistics, applied mathematics and other disciplines. There were an additional 71 papers submitted and 98 papers in preparation.

Impact is also measured in the long-term consequences of SAMSI programs. A number of these long-term consequences were mentioned in the Final Report that SAMSI submitted in Fall, 2008 for the first six years of research. These long-term consequences included:

- A collaboration between [Marie Davidian](#) (a statistician from NCSU), [H. T. Banks](#) (an applied mathematician from NCSU), and [Eric Rosenberg](#) (an immunologist clinician from Massachusetts General Hospital), led to major grants and the impetus to form the Center for Quantitative Sciences in Biomedicine (<http://www.ncsu.edu/cqsb>) at North Carolina State University, affiliated with Emory University and Massachusetts General Hospital. Marie Davidian is Director and H.T. Banks is co-director of the center.
- Based on a collaboration and developments arising from a SAMSI program, Eric Ghysels and Rob Engle eventually founded the Society for Financial Econometrics, and Jean-Pierre Fouque (UCSB) founded the Center for Research in Financial Mathematics and Statistics (<http://www.pstat.ucsb.edu/crfms/>) at UCSB; two of the first CRFMS postdocs were heavily involved in the SAMSI program.
- The SAMSI working group on *Granular Materials – Engineering Applications* continued after the program, and M.J. Bayarri (Valencia, Statistics), James Berger (Duke, Statistics), Eliza Calder (Buffalo, Geology), E. Bruce Pitman (Buffalo, Math), Elaine Spiller (SAMSI, Math), and Robert Wolpert (Duke, Statistics) were awarded an NSF Focused Research Group grant to continue the research for three years. Bruce Pitman wrote that the SAMSI program “shifted the direction of my personal research, which in turn gave an expanded range of research projects and opportunities with colleagues in the earth sciences.” In addition to continuing the research, three workshops have been planned under this project, and are hence a direct consequence of the SAMSI program:
 - A workshop in April, 2009 at SAMSI, linking the CompMod research area of adaptive emulation, with the current SAMSI program on Sequential Monte Carlo methodology.
 - A summer school for graduate students and young investigators in applied mathematics, geophysics and statistics, to be held at the Pacific Institute for the Mathematical Sciences (PIMS) after the Joint Statistical Meetings in Vancouver in August, 2010.
 - A workshop in 2011 at an appropriate Geophysical Sciences meeting, to disseminate the research results.

SAMSI’s national impact can also be measured by the activities in other conferences that result from its programs. For instance, in the 2008 Joint Statistical Meetings, there were eight SAMSI motivated sessions.

SAMSI’s strong commitment to the development of human resources in the mathematical

sciences is summarized in Section D.2 of the Executive Summary and detailed in Sections I.B, I.C and I.H of the full report. Videoconference and WebEx technologies have now been adopted by *all* working groups, some with an international reach. Indeed, in some working groups, the majority of participants engage by these means, and some working groups continue to be active after the end of the formal program.

Participation by women and other underrepresented groups is high and appears to be steady or rising, after adjustments for individual year variation due to special events. In particular for 2008-2009, the participants were 31% female, 5% African American and 8% Hispanic. Participation by new researchers and students this year is at an extremely high 71%. The SAMSI website now offers information about its diversity programs at <http://www.samsi.info/about/diversity.shtml>.

The inclusion in SAMSI programs of a substantial number of participants from institutions not heavily supported by NSF-DMS funding is detailed in the full report.

The detailed participant lists for concluded programs provide ample evidence of the national and international draw of SAMSI activities. SAMSI programs attracted 19 long-term visitors (3 months or more), 47 short-term visitors (a week to 3 months), 11 local faculty fellows, 9 postdoctoral fellows and associates, 18 graduate students (7 visiting), and a total participant list of more than 1000 researchers. We also understand that, with the opening of the new wing of the building, programs will be expanding next year to their natural size and that, to date, 8 year-long and 31 semester-long visitors have been approved to visit; this stunning increase indicates the demand for SAMSI programs.

Applications to the postdoctoral program were up to 133 from last year's level of 103. (This does not include the 187 applicants through the joint institutes' stimulus-based postdoctoral process.) The directorate observed that the top statistics and probability candidates have been hearing of the considerable benefits of going through a SAMSI postdoctoral experience, while the top applied mathematics candidates are being attracted by a growing recognition of the importance of integrating applied mathematics and statistics.

3) Is the Directorate meeting the needs of an evolving SAMSI?

After seven years in operation, the directorate model continues to serve SAMSI very well, and transitions in the directorate have gone smoothly. Pierre Gremaud assumed the associate directorship from NCSU, and has integrated quickly with the directorate. NISS Director Alan Karr continued to work closely with SAMSI in connection with the expansion of the NISS building, which opened in November – a signal event in the history of both organizations.

Director James Berger has announced his intention to leave the position in summer, 2010. With NSF approval, the Governing Board has established a nationally prominent search committee to undertake a broad search for his successor and a detailed plan for the search process. This process is well under way, with candidates identified and initial visits to

SAMSI beginning. The partner universities are working together to ensure that there will be a tenured professor position available for the next director if necessary.

Because of the greatly increased size of SAMSI enabled by the new building, it was decided to convert one of the university Associate Directorships into a Deputy Director position, which will be a full-time position with half the salary paid from NSF funds (and half paid by the university, as currently). The Governing Board selected – and the National Science Foundation approved – Pierre Gremaud to become the first Deputy Director, with a term of 1.5 years beginning on January 1, 2010. To his current responsibilities as liaison to NCSU and responsibility for the SAMSI Education and Outreach program, Pierre will be adding the responsibility of oversight of the day-to-day operation of current programs.

The Governing Board itself continues to operate in the expanded structure implemented earlier that now includes two representatives from beyond the four SAMSI partner institutions who are selected by the American Statistical Association (Casella) and the Society for Industrial and Applied Mathematics; this year Don Estep was appointed to begin a three-year term on the Governing Board as the SIAM representative. The Governing Board also includes domain scientist representation from astronomy (Carney) and chemistry (Simon).

The Governing Board Chair and the SAMSI Director continue to have a biweekly telephone conference at which administrative and personnel matters are regularly discussed and issues addressed where they have arisen. There is also excellent cooperation among the partner universities and NISS to ensure that obligations are met and that SAMSI continues to flourish. One recent example is an agreement among the universities to delay funding of equipment purchases for the building expansion, given the delay in the building expansion; all funds were appropriately allocated this year. A second, significant example is the cooperation among the universities and their Human Resources departments in mounting the search for the next SAMSI director.

Table of Contents

0. Executive Summary.....	3
I. Annual Progress Report	32
A. Program Personnel	32
1. List of Programs and Organizers	32
2. Program Core Participants	37
B. Postdoctoral Fellows and Associates	54
1. Overview of the Postdoctoral Fellow Program.....	54
2. 2008-09 Postdoc Activity and Progress Reports	56
3. Postdoc Experience Evaluation.....	73
C. Graduate Student Participation	84
D. Consulted Individuals	88
E. Program Activities.....	89
1. Algebraic Methods in Systems Biology and Statistics	89
2. Sequential Monte Carlo Methods	101
3. Meta Analysis: Synthesis of Multiple Sources of Evidence.....	135
4. Education and Outreach Program	144
F. Industrial and Governmental Participation	152
G. Publications and Technical Reports.....	153
H. Achieving Diversity.....	168
I. External Support and Affiliates	172
J. Advisory Committees.....	175
K. Income and Expenditures	176
L. Report from the Math Institutes Director’s Meeting	179
II. Special Report: Program Plan.....	184
A. Programs for 2009-2010	184
1. Space-Time Analysis for Environmental Mapping, Epidemiology, and Climate Change.....	184
2. Stochastic Dynamics.....	194
3. Psychometrics	201
B. Scientific Themes for Later Years	205
C. Budget for 2009-2010	227
D. Financial Plan for 2009-2010	235
Appendix	
A. Final Program Report: Risk Analysis, Extreme Events and Decision Theory.....	237
B. Final Program Report: Random Media.....	272
C. Final Program Report: Environmental Sensor Networks.....	286
D. Workshop Participant Lists	299
E. Workshop Programs and Abstracts	392
F. Workshop Evaluations.....	502

I. Annual Progress Report

The previous annual progress report was complete in all details only through April, 2008. Hence, we also report activities in Year 6 programs that occurred subsequently and were not itemized in the report. These Year 6 programs were *Risk Analysis, Extreme Events, and Decision Theory; Random Media; and Environmental Sensor Networks*; their final reports are in Appendices A, B, and C, respectively.

A. Program Personnel

1. Program and Activity Organizers

Program Organizers

Program	Name	Affiliation	Field
Random Media	Russel Caflisch	UCLA	Mathematics
	Maarten De Hoop	Purdue U	Applied Math
	Rick Durrett	Cornell U	Mathematics
2007-08 SAMSI Program	Weinan E.	Princeton U	Applied Math
	Josselin Garnier	Universite Paris VII	Mathematics
	George Papanicolaou	Stanford U	Mathematics
	Lenya Ryzhik	U of Chicago	Mathematics
	Ralph Smith	NCSU and SAMSI	Applied Math
	Chrysoula Tsogka	U of Chicago	Applied Math
	Eric Vanden-Eijnden	NYU	Mathematics
	Jack Xin	UC Irvine	Mathematics
	Wojbor Woyczynski	Case Western U	Mathematics
	Hongkai Zhao	UC Irvine	Mathematics & Comp Sci
Risk Analysis, Extreme Events, and Decision Theory	David Banks	Duke U	Statistics
	Vickie Bier	U of Wisconsin	Engineering Physics
	James Broffitt	U of Iowa	Statistics
2007-08 SAMSI Program	Lawrence Brown	U of Pennsylvania	Statistics
	Alicia Carriquiry	Iowa State U	Statistics
	Robert Clemen	Duke U	Decision Sciences
	Dipak Dey	U of Connecticut	Statistics
	Susan Ellenberg	U of Pennsylvania	Biostatistics
	Herbert Hethcote	U of Iowa	Mathematics
	Wolfgang Kliemann	Iowa State U	Mathematics
	Stephen Pollock	U of Michigan	Physics and Oper Res
David Rios Insua	U Rey Juan Carlos	Statistics and Oper Res	

	Nell Sedransk	NISS, SAMSI	Statistics
	Richard Smith	UNC - CH	Statistics
	Robert Winkler	Duke U	Statistics,Mathematics,Econ
	Stan Young	NISS	Statistics
Environmental Sensor Networks	Jim Berger	SAMSI	Statistics
	Zoe Cardon	U of Connecticut	Biology
	Jim Clark	Duke U	Biology
<i>2007-08 SAMSI Program</i>	Jorge Cortes	UCSD	Engineering Mathematics
	Don Estep	Colorado State U	Math and Stat
	Debora Estrin	UCLA	Computer Science
	Paul Flikkema	Northern Arizona U	Elec Engineering
	Alan Gelfand	Duke U	Statistics
	Mark Hansen	UCLA	Statistics
	Bin Yu	UC Berkeley	Statistics
Meta Analysis	Joseph Beyene	U of Toronto	Statistics
	Vanja Dukic	U of Chicago	Biostatistics
	Julian Higgins	UK Medical Res	Statistics
<i>2008 SAMSI Summer Program</i>	Peter Hoff	U of Washington	Biostatistics
	Keith O'Rourke	Duke U	Statistics
	Ken Rice	U of Washington	Biostatistics
	Dalene Stangl	Duke U	Statistics
Sequential Monte Carlo Methods	Jim Berger	SAMSI	Statistics
	Monica Bugallo	Stony Brook	Engineering
	Petar Djuric	Stony Brook	Engineering
<i>2008-09 SAMSI Program</i>	Arnaud Doucet	British Columbia U	Statistics & Comp Sci
	Richard Durrett	Cornell U	Mathematics
	Simon Godsill	Cambridge	Info Engineering
	Michael Jordan	UC Berkeley	Statistics & Comp Sci
	Jun Liu	Harvard	Statistics
	Gareth Roberts	Warwick	Statistics
	Raquel Prado	UC Santa Cruz	Applied Math & Stats
	Neil Shephard	Oxford	Stats & Econometrics
	Simon Tavare	Cambridge	Comp Biology
	Mike West	Duke U	Statistics
Algebraic Methods in Systems Biology and Statistics	Peter Beerli	Florida State U	Comp & Biological Sci
	Andreas Dress	Shanghai	Comp Biology
	Mathias Drton	U of Chicago	Statistics
	Ina Hoeschele	Virginia Tech	Statistics
<i>2008-09 SAMSI Program</i>	Christine Heitsch	Georgia Tech	Mathematics

	Serkan Hosten	SF State U	Mathematics
	Reinhard Laubenbacher	Virginia Tech	Mathematics
	Bud Mishra	Courant Institute	Comp Sci & Math
	Don Richards	Pennsylvania State	Statistics
	Seth Sullivant	NCSU	Mathematics
	Brett Tyler	Virginia Tech	Plant Pathology
	Ruriko Yoshida	U of Kentucky	Statistics
Psychometrics	Charles Lewis	Fordham U	Psychology
	Richard Swartz	U Texas	Biostats
<i>2009 SAMSI Summer Program</i>	Valen Johnson	U Texas	Biostats
	James Berger	SAMSI	Statistics
Education & Outreach	Negash Begashaw	Benedict College	Mathematical Sciences
	Carlos Castillo-Chavez (ex officio)	Arizona State U	Mathematics
<i>2008-09 SAMSI Program</i>	Karen Chiswell	NCSU	Statistics
	Cammeey Cole	Meredith College	Mathematics & CS
	Wei Feng	UNC-Wilmington	Mathematics & Statistics
	Pierre Gremaud (chair)	NCSU	Mathematics
	Marian Hukle	U of Kansas	Biological Sciences
	Negash Medhin	NCSU	Mathematics
	Masilamani Sambandham	Morehouse College	Mathematics
Space-time Analysis for Environmental Mapping, Epidemiology, and Climate Change	Jim Berger	SAMSI	Statistics
	Noel Cressie	Ohio State U	Statistics
	Michael Stein	U of Chicago	Statistics
<i>2009-10 SAMSI Program</i>	Dongchu Sun	U Missouri	Statistics
	Jim Zidek (chair)	U British Columbia	Statistics

Activity Organizers

2007-08 Programs		
Program Year	Activity	Name(s)
<i>Random Media</i>		
2007-08	Random Media Transition Workshop -- <i>May 1-2, 2008</i>	Maarten deHoop (Purdue), Zhilin Li (NCSU), Ralph Smith (NCSU, SAMSI), Hongkai Zhao (NCSU)

Education and Outreach		
2007-08	SAMSI/CRSC Interdisciplinary Workshop for Undergraduates -- <i>May 19-23 2008</i>	Cammeey Cole (NCSU), Ralph Smith (NCSU), Ernie Stitzinger (NCSU), Kim Weems (NCSU)
Risk Analysis, Extreme Events and Decision Theory		
2007-08	Risk Revisited: Progress and Challenges Transition Workshop -- <i>May 21, 2008</i>	Nell Sedransk (NISS & SAMSI), Richard Smith (University of North Carolina)
Summer Program		
2007-08	Summer 2008 Program on Meta-analysis: Synthesis and Appraisal of Multiple Sources of Empirical Evidence -- <i>June 2-13, 2008</i>	Joseph Beyene (U Toronto), Vanja Dukic (U Chicago), Julian Higgins (UK Med Research), Peter Hoff (U Washington), Keith O'Rourke (Duke), Ken Rice (U Washington), Dalene Stangl (Duke)
Sensor Networks		
2007-08	Transition Workshop -- <i>October 20-21, 2008</i>	Jim Berger (SAMSI), Paul Flikkema (N. Arizona State)
2008-09 Programs		
Program Year	Activity	Name(s)
Sequential Monte Carlo Methods		
2008-09	Opening Workshop -- <i>September 7-10, 2008</i>	Arnaud Doucet (U British Columbia), Simon Godsill (U Cambridge), Mike West (Duke U)
2008-09	Mid-Program Workshop -- <i>February 19-20, 2009</i>	Simon Godsill (U Cambridge), Mike West (Duke U)
2008-09	Adaptive Design, Sequential Monte Carlo, and Computer Modeling - <i>April 15-17, 2009</i>	Susie Bayarri (University of Valencia, Duke & NISS), Mike West (Duke); Jim Berger (Duke & SAMSI, Directorate Liaison)
2008-09	Transition Workshop -- <i>November 9-10, 2009</i>	To Be Reported in 2009-10 Annual Report

Algebraic Methods in Systems Biology and Statistics		
2008-09	Opening Workshop -- <i>September 14-17, 2008</i>	Reinhard Laubenbacher (VA Tech), Seth Sullivant (NCSU), Brett Tyler (Virginia Tech), Rudy Yoshida (University of Kentucky)
2008-09	Discrete Models in Systems Biology - <i>December 3-5, 2008</i>	Elena Dimitrova (Clemson), Ilya Shmulevich (Institute for Systems Biology), Brandilyn Stigler (Southern Methodist)
2008-09	Algebraic Statistical Models - <i>January 15-17, 2009</i>	Mathias Drton (University of Chicago), Eva Riccomagno (University of Genova), Seth Sullivant (NCSU)

2008-09	Molecular Evolution and Phylogenetics Workshop – <i>April 2-3, 2009</i>	Peter Huggins (Carnegie Mellon U), Erick Matsen (UC Berkeley), Ruriko Yoshida (U Kentucky)
2008-09	Transition Workshop -- <i>June 18-20, 2009</i>	Reinhard Laubenbacher (VA Tech), Seth Sullivant (NCSU), Rudy Yoshida (University of Kentucky)
Summer Program		
2008-09	Psychometrics -- <i>July 7-17, 2009</i>	Charles Lewis (Fordham U), Richard Swartz (U of Texas M.D. Anderson Cancer Center), and Valen Johnson (U of Texas M.D. Anderson Cancer Center); Directorate Liaison is James Berger (SAMSI).
Education and Outreach		
2008-09	CRSC/SAMSI Workshop for Graduate Students -- <i>July 21-29, 2008</i>	Pierre Gremaud (NCSU), Sharon Lubkin (NCSU), Mette Olufsen (NCSU), Jeff Scroggs (NCUS), Ralph Smith (NCSU)
2008-09	Two-Day Undergraduate Workshop -- <i>October 31-November 1, 2008</i>	Pierre Gremaud (NCSU), Jochen Voss (Warwick U), Jaya Bishwal (UNC)
2008-09	Two-Day Undergraduate Workshop -- <i>February 27-28, 2009</i>	Pierre Gremaud (NCSU), Seth Sellivant (NCSU), Reinhard Laubenbacher (VA Tech)
2008-09	Graduate Student Probability Workshop -- <i>May 1-3, 2009</i>	Changryong Baek, Jessi Cisewski, Xin Liu, Dominik Reinhold, Tiffany Kolba and Rachel Thomas under the supervision of Prof. Amarjit Budhiraja and Prof. Jonathan Mattingly.
2008-09	SAMSI/CRSC Interdisciplinary Workshop for Undergraduates -- <i>May 18-22 2009</i>	Pierre Gremaud (NCSU); Cammey Cole-Manning (Meredith); Kim Weems (NCSU)
2008-09	CRSC/SAMSI Workshop for Graduate Students -- <i>July 20-28, 2009</i>	Pierre Gremaud (NCSU); Ilse Ipsen (NCSU); Ralph Smith (NCSU)
Co-sponsored and Informal Meetings and Workshops		
2008-09	Blackwell-Tapia Conference -- <i>November 15-16, 2008</i>	Michael Minion (UNC & SAMSI), Ricardo Cortez (Tulane), William Massey (Princeton), Carolyn Morgan (Hampton), Cristina Villalobos (Texas - Pan American)

2009-10 Programs		
<i>Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change</i>		
2009-10	Summer School -- <i>July 28 - August 1, 2009</i>	Sudipto Banerjee (U. Minnesota), Reinhard Furrer (U. Zurich), Doug Nychka (National Center for Atmospheric Research), and Stephen Sain (National Center for Atmospheric Research)

2. Program Core Participants

For each of the major programs, the following tables present the key participants for the programs. The participants are categorized and coded as follows:

DL	Distinguished Lecturer	Program affiliated speaker
FF	Faculty Fellow	Teaching release from local university
FA	Faculty Associate	Program affiliated local faculty for which no release time is allocated
GF	Graduate Student Fellow	Student from local university, assigned to a specific program and paid a stipend
GA	Graduate Student Associate	Program-affiliated local student with no stipend
VGF	Visiting Graduate Fellow	Non-local student, paid only expenses
NRV	New Researcher Visitor	Non-local researchers (holding PhD 5 years or less) brought in for short intervals for interaction with program participants
NRC	New Researcher Core Visitor	Non-local researchers (including fellows) who play a major role in program activities
PF	Postdoctoral Fellow	Program-affiliated individual, paid a stipend in association with a local university
PA	Postdoctoral Associate	Program-affiliated individual with appointment shorter than 1 year
SV	Senior Visitor	Researcher (holding PhD 6 or more years) brought in for short intervals for interaction with program participants
RF	Research Fellow	Non-local researchers who play a major role in program activities
WG	Working group Participant	local participants of SAMSI working groups (not fellows, visitors or persons otherwise designated)
WGR	Remote working group participant	remote participants of SAMSI working groups (not otherwise designated)

Grey – is used to indicate funds that are provided by partner university cost sharing.

Note: For visitors who have yet to visit SAMSI or who are still at SAMSI, dollar amount in the tables below are the expense allotment for the visitor.

Sequential Monte Carlo Methods

Program Core Participants

Last Name	First Name	Gender	Affiliation	Department	Status
Argon	Nilay	Female	University of North Carolina at Chapel Hill	Statistics/OR	FF
Armagan	Artin	Male	Duke University	STAT	PA
Bain	Melanie	Female	University of North Carolina	Statistics/OR	GF
Bayarri	Susie	Female	U of Valencia, Duke & NISS	Statistics	RF
Bernardo	Jose	Male	Universitat de València	Statistics	SV
Berrocal	Veronica	Female	Department of Statistical Science - Duke University	STAT	WG
Bishwal	Jaya	Male	UNC Charlotte	Mathematics & Statistics	NRC
Boomer	K.B.	Female	Bucknell University	STAT	WGR
Bornn	Luke	Male	U of British Columbia	Statistics	GF
Briers	Mark	Male	QinetiQ Ltd	Statistics	SV
Bugallo	Monica	Female	Stony Brook University	ENGG	WGR
Carvalho	Carlos	Male	University of Chicago	Graduate School of Business	RF
Chen	Hao	Male	SAS Institute	STAT	WG
Chen	Rong	Male	Duke University – FSB	STAT	WG
Chopin	Nicholas	Male	Bristol U	Statistics	SV
Chorin	Alexandre	Male	UC Berkeley	Mathematics	SV
Clark	Daniel	Male	Institute of Electrical and Electronics Engineers, Inc.	Engineering	RF
Clyde	Merlise	Female	Duke U	Statistics	WG

Coates	Mark	Male	McGill U	Engineering	RF
Colvin	Jacob	Male	University of California, Santa Cruz	STAT	WGR
Corberan	Ana	Female	University of Valencia	Statistics and Operational Res	GF
Cornebise	Julian	Male	SAMSI	Statistics	PA
Dance	Sarah	Female	University of Reading	MATH	WGR
Das	Sourish	Male	SAMSI	Statistics	PF
Djuric	Petar	Male	Stony Brook	ENGG	WGR
Douc	Randal	Male	l'Ecole Polytechnique	Mathematics	WGR
Doucet	Arnaud	Male	U of British Columbia	Statistics	RF
Dunson	David	Male	Duke U	Statistics	WG
Fearnhead	Paul	Male	Lancaster University	Department of Math & Stats	RF
Ferrante	Marco	Male	U of British Columbia	Mathematics	SV
Flury	Thomas	Male	Oxford	Economics	GF
Fokoue	Ernst	Male	Kettering	Statistics	RF
Gning	El hadji Amadou	Male	Lancaster U	Statistics	WGR
Godsill	Simon	Male	U Cambridge	Engineering	SV
Goel	Prem	Male	Ohio State University	STAT	WGR
Gramacy	Robert	Male	U Cambridge	Appl Math & Stat	WGR
Green	Nathan	Male	DSTL	Mathematics	RF
Griffiths	Robert	Male	Oxford U	Statistics	SV
Guerron	Pablo	Male	NCSU	Economics	FA

Hannig	Jan	Male	Academic	PHYS	WGR
Holenstein	Roman	Male	U British Columbia	Statistics	GF
Huber	Mark	Male	Duke University	MATH	FF
Ikoma	Norikazu	Male	Kyushu Institute of Technology	ENGG	WGR
Ito	Kazi	Male	NCSU	Mathematics	FA
Ji	Chunlin	Male	Duke University	STAT	GF
Kang	Min	Female	North Carolina State University	Mathematics	FA
Kantas	Nicolas	Male	U Cambridge	Engineering	GF
Koutsourelakis	Steve	Male	Cornell University	ENGG	WGR
Lawrence	James	Male	Cambridge U	Statistics	WGR
Li	Fan	Female	Duke U	Statistics	WG
Liu	Fei	Female	U Missouri	Statistics	RF
Lopes	Hedibert	Male	U Chicago	Statistics	RF
Loredo	Thomas	Male	University of South Carolina	STAT	WGR
Lynch	James	Male	University of South Carolina	Department of Statistics	RF
Lyubimov	Konstantin	Male	U Georgia	Sociology	SV
Macaro	Christian	Male	Duke & SAMSI	Statistics	PA
Manalopoulou	Ioanna	Female	Duke & SAMSI	Statistics	PF
Mattingly	Jonathan	Male	Duke U	Statistics	WG
McClain	Alex	Male	Duke University	STAT	WG
Mernick	Kevin	Male	New Jersey Institute of Technology	MATH	WGR

Mihaylova	Lyudmila	Female	Lancaster University	ENGG	WGR
Morales	Mario	Male	Hunter College, CUNY	STAT	WGR
Moulines	Eric	Male	Ecole Nationale Supérieure	MATH	WGR
Mukherjee	Chiranjit	Male	Duke University	STAT	WG
Mulder	Joris	Male	Utrecht University	STAT	WGR
Munoz	Maria Pilar	Female	Technical University of Catalonia (UPC)	STAT	WGR
Niemi	Jarad	Male	Duke University	STAT	WG
Obanubi	Olasunkanmi	Male	Imperial College	Mathematics	GF
Papaspiliopoulos	Omiros	Male	U of Warwick	Statistics	RF
Pena	Edsel	Male	U South Carolina	Statistics	RF
Peters	Gareth	Male	UNSW	Statistics	GF
Petralia	Francesca	Female	Duke U	Statistics	GF
Petris	Giovanni	Male	U of Arkansas	STAT	WGR
Prado	Raquel	Female	UC Santa Cruz	Statistics	RF
Robert	Christian	Male	U Paris	Statistics	SV
Rodriguez	Abel	Male	University of California	STAT	WGR
Rogers	Chris	Male	U of Cambridge	MATH	WGR
Roos	Jason	Male	Duke University	SOCL	WG
Roy	Deb	Male	Pennsylvania State	MATH	WGR
Rozgic	Viktor	Male	University of Southern California	ENGG	WGR
Rubenthaler	Sylvain	Male	University de Nice-Sophia Antipolis	Laboratoire J.-A. Dieudonné	RF

Schoolfield	Clyde	Male	University of Florida	MATH	WGR
Schott	Sarah	Female	Duke University	MATH	GF
Septier	Francois	Male	Cambridge University	ENGG	SV
Sethuraman	Jayaram	Male	Florida State U	Statistics	WGR
Shamseldin	Elizabeth	Female	Duke & SAMSI	Statistics	PA
Shen	Bingxin	Female	Stony Brook University	ENGG	WGR
Shi	Minghui	Female	Duke U	STAT	GF
Stroud	Jonathan	Male	George Washington U	STAT	WGR
Sun	Dongchu	Male	U Missouri	Statistics	RF
ter Braak	Cajo	Male	Wageningen University and Research Centre	STAT	WGR
Thomas	Andrew	Male	CREEM	Mathematics and Statistics	RF
Thomas	Len	Male	U St. Andrews	STAT	WGR
Vaswani	Namrata	Female	Iowa State University	ENGG	WGR
Vera	Francisco	Male	U South Carolina	Mathematics	WGR
Vidyashankar	Anand	Male	Cornell University	Statistical Science and Social Statistics	WGR
Vogelstein	Joshua	Male	Johns Hopkins	BIOSCI	WGR
Voss	Jochen	Male	U of Warwick	Mathematics	RF
Wang	Hao	Male	Duke U	Statistics	WG
Wang	Kai	Male	Duke University	STAT	WG
Wang	Quanli	Male	Duke U	Statistics	WG
West	Mike	Male	Duke U	Statistics	FF

White	Gentry	Male	SAMSI / NCSU	Statistics	PA
Wolpert	Robert	Male	Duke U	Statistics	WG
Yardim	Caglar	Male	UCSD	SIO	WGR
Yoshida	Ryo	Male	ISM	Statistics	RF
Zhang	Baqun	Male	NCSU	Statistics	GF

Algebraic Methods in Systems Biology and Statistics
Program Core Participants

Last Name	First Name	Gender	Affiliation	Department	Status
Allman	Elizabeth	Female	U of Alaska	Mathematics	RF
Barker	Brandon	Male	Cornell University	STAT	WGR
Beerli	Peter	Male	Florida State University	LIFE	WGR
Bocci	Cristiano	Male	U of Milan	Mathematics	RF
Casella	George	Male	U of Florida	Statistics	SV
Chen	Wenjie	Female	UNC-Chapel Hill	STAT	GF
Chifman	Julia	Female	U Kentucky	Mathematics	GF
Coleman	Deidra	Female	NCSU	Statistics	GF
Conradi	Carsten	Male	Max Planck Inst.	Mathematics	RF
Cox	Lawrence	Male	National Center for Health Statistics/CDC	MATH	WG
Craciun	Gheorghe	Male	University of Wisconsin	MATH	WGR
Degnan	James	Male	U Michigan	Statistics	RF
Dickenstein	Alicia	Female	U Buenos Aires	Mathematics	RF

Dimitrova	Elena	Female	Clemson University	MATH	WGR
Dinwoodie	Ian	Male	Duke U	Statistics	FF
Drton	Mathias	Male	University of Chicago	STAT	WGR
Falin	Lee	Male	VA Bioinfo	STAT	WGR
Francis	Andrew	Male	University of Western Sydney	MATH	RF
Friedrich	Thomas	Male	U Berlin	Mathematics	GF
Garcia-Puente	Luis	Male	Sam Houston State	Statistics	NRC
Ginestet	Cedric	Male	Imperial College	BIOSTAT	WGR
Gnacadjia	Gilles	Male	Amgen	MATH	WGR
Gopalkrishnan	Manoj	Male	University of Southern California	COMP	WGR
Gunawardena	Jeremy	Male	Harvard	Life	RF
Haney	Richard	Male	Cellular Statistics	STAT	WG
Hara	Hisayuki	Male	University of Tokyo	STAT	WGR
Hinkelmann	Franziska	Female	Virginia Bioinformatics Institute	MATH	WGR
Hosten	Serkan	Male	SF State U	Mathematics	RF
Hower	Valerie	Female	Georgia Institute of Technology	MATH	WGR
Huber	Mark	Male	Duke U	Mathematics	FF
Jarrah	Abdul Salam	Male	Virginia Tech	MATH	WGR
Kahle	Thomas	Male	Max Planck Institute for Mathematics in the Sciences	MATH	SV
Kondor	Imre Risi	Male	Gatsby Unit, University College London	COMP	WGR
Kubatko	Laura	Female	Ohio State University	STAT	WGR

Kuo	Lynn	Female	U Connecticut	Statistics	RF
Laubenbacher	Reinhard	Male	VA Tech	Statistics	RF
Lee	Tong	Male	Hunter College of City University of New York	MATH	WGR
Lewis	Robert	Male	Fordham U	MATH	WGR
Lin	Shaowei	Male	University of California, Berkeley	MATH	WGR
Maini	Philip	Male	U Oxford	Mathematics	SV
Manon	Chris	Male	University of Maryland	MATH	WGR
Maruri-Aguilar	Hugo	Male	London School of Economics	STAT	WGR
Matias	Catherine	Female	CNRS	Statistics	RF
Nagel	Uwe	Male	U of Kentucky	Mathematics	RF
O'Shea	Edwin	Male	U of Kentucky	Mathematics	SV
Owen	Megan	Female	SAMSI	Mathematics	PF
Pantea	Casian	Male	University of Wisconsin – Madison	MATH	WGR
Perduca	Vittorio	Male	Universita' degli Studi di Torino	MATH	WGR
Perez Millan	Mercedes Soledad	Female	Universidad de Buenos Aires	MATH	GF
Petrovic	Sonja	Female	U Illinois	Mathematics	SV
Pistone	Giovanni	Male	Politecnico di Torino	Mathematics	RF
Provan	Scott	Male	UNC	Mathematics	FF
Reading	Nathan	Male	NCSU	Mathematics	FA
Reishus	Dustin	Male	USC	COMP	WGR
Rempala	Greg	Male	Medical College of GA	Mathematics	RF

Rhodes	John	Male	U Alaska	Mathematics	RF
Riccomagno	Eva	Female	U Genoa	Statistics	RF
Schardl	Chris	Male	University of Kentucky	LIFE	WGR
Shen	Jian	Male	Texas State University	MATH	WGR
Shiu	Anne	Female	University of California, Berkeley	MATH	WGR
Siebert	Heike	Female	Freie Universität Berlin	Mathematics	RF
Singer	Michael	Male	NCSU	Mathematics	FA
Slavkovic	Alexandra	Female	Penn State University	STAT	WGR
Solhjo	Soroosh	Male	Johns Hopkins University School of Medicine	LIFE	WGR
Stigler	Brandy	Female	Mathematical Biosciences Institute	Mathematics	RF
Stone	Eric	Male	NCSU	Statistics	FF
Sullivant	Seth	Male	NCSU	Mathematics	FF
Szanto	Agnes	Female	NCSU	Mathematics	FA
Takemura	Akimichi	Male	University of Tokyo	STAT	WGR
Tyler	Brett	Male	Virginia Polytechnic Institute and State University	LIFE	WGR
Tzeng	Jung-Ying	Female	NCSU	Statistics	FF
Uhler	Caroline	Female	UC Berkeley	STAT	WGR
Uwe	Helmke	Male	University of Wurzburg	Mathematics & CS	RF
Veliz-Cuba	Alan	Male	Virginia Tech	MATH	WGR
Vera-Licona	Paola	Female	Rutgers University	MATH	WGR
Wells	Benjamin	Male	NCSU	Statistics	GF

Wynn	Henry	Male	London School of Economics	STAT	WGR
Yamada	Richard	Male	U Michigan	Statistics	RF
Yarahmadian	Shantia	Male	Indiana University, Molecular Biology Institute	MATH	WGR
Yasamin	Ahmad Saeid	Male	SAMSI	Statistics	PF
Yellick	Jason	Male	NCSU	Mathematics	GF
Yoshida	Ruriko	Female	U Kentucky	Statistics	SV
Yoshida	Ryo	Male	Institute of Statistical Mathematics	BIOSTAT	SV
Zou	Yi Ming	Female	U Wisconsin	MATH	WGR
Zuk	Or	Male	Broad Inst. MIT & Harvard	Comp. Physics	RF
Zwiernik	Piotr	Male	University of Warwick	STAT	SV

Summer Program on Meta Analysis

Program Core Participants

Last Name	First Name	Gender	Affiliation	Department	Status
Barrett	Jessica	Female	Medical Research Council UK	Statistics	RF
Basu	Sanjib	Male	Northern Illinois University	Department of Statistics	RF
Bayarri	M.J. (Susie)	Female	University of Valencia	Statistics and Operations Research	RF
Berger	James	Male	SAMSI	Statistics	RF
Bortz	David	Male	University of Colorado	Applied Mathematics	RF
Demidenko	Eugene	Male	Dartmouth Medical School	Statistics	RF
Dukic	Vanja	Female	University of Chicago	Health Studies (Biostatistics)	RF

Dunson	David	Male	National Institute of Environmental Health Sciences	Biostatistics Branch	RF
Gatsonis	Constantine	Male	Brown University	Statistics	SV
Harrell	Leigh	Female	Virginia Tech	Department of Statistics	RF
He	Qianchuan	Male	UNC - Chapel Hill	Department of Biostatistics	GA
Hedges	Larry	Male	Northwestern University	Statistics	SV
Higgins	Julian	Male	Cambridge University	Statistics	RF
Hua	Zhaowei	Female	University of North Carolina, Chapel Hill	Department of Biostatistics	GA
Jackson	Dan	Male	MRC Cambridge	Institute of Public Health	RF
Johnson	Nels	Male	Virginia Tech	Statistics	SV
Kaizar	Eloise	Female	Ohio State University	Department of Statistics	SV
Kim	Yongku	Male	SAMSI	Statistics	PF
Kinney	Satkartar	Female	NISS	Statistics	PF
Kounali	Daphne	Female	University of Bristol	Centre for Multilevel Modelling	RF
Lin	Danyu	Male	UNC	Biostatistics	RF
Liu	Fei	Female	University of Missouri-Columbia	Statistics	RF
Madar	Vered	Female	SAMSI	Statistics	PF
McCandless	Lawrence	Male	Imperial College London	Epidemiology and Public Health	RF
Moreno	Elias	Male	University of Granada	Department of Statistics	RF
Morton	Sally	Female	RTI International	Statistics	SV
Olkin	Ingram	Male	Stanford	Statistics	SV
O'Rourke	Keith	Male	Duke University	Department of Statistical Science	RF

Petricka	Jalean	Female	Duke University	Life	SV
Plante	Jean-Francois	Male	University of Toronto	Department of Statistics	RF
Platt	Robert	Male	McGill University	Statistics	SV
Pungpapong	Vitara	Female	Purdue University	Statistics	SV
Rice	Ken	Male	University of Washington	Statistics	RF
Sedransk	Nell	Female	NISS and SAMSI	Statistics	SV
Shrier	Ian	Male	McGill University	Clinical Epidemiology and Community Studies	RF
Stangl	Dalene	Female	Duke University	Department of Statistical Science	RF
Stevens	John	Male	Utah State University	Mathematics and Statistics	RF
Stuart	Elizabeth	Female	Johns Hopkins Bloomberg School of Public Health	Mental Health, Biostatistics	RF
Sun	Junfeng	Male	University of Nebraska Medical Center	Department of Biostatistics	RF
Thorlund	Kristian	Male	University of Copenhagen / McGill University	Statistics	SV
Tiwari	Ram	Male	Center for Drug Evaluation & Research, FDA	Office of Biostatistics	RF
Trikalinos	Tom	Male	Tufts University	Life	SV
Tzeng	Jung-Ying	Female	NC State University	Statistics	RF
Umbach	David	Male	NIEHS	Biostatistics Branch	RF
Unal	Cemal	Male	Pozen, Inc.	Statistics	WG
Wang	Jen-Ting	Female	NCSU	Statistics	RF
Williams	Matthew	Male	VA Tech	Statistics	GA
Wolpert	Robert	Male	Duke University	Statistical Science	SV
Wouhib	Abera	Male	CDC	Statistics	WG

Xia	Jessie	Female	NISS	Statistics	PF
Young	Stan	Male	NISS	Statistics	SV
Zhang	Lingsong	Male	Harvard University	Statistics	SV
Zhang	Ying	Female	Pozen, Inc.	Statistics	RF
Zhao	Yue	Female	U of North Carolina	Department of Biostatistics	GA
Zhou	Jasmine	Female	NISS	Statistics	PF
Zou	Fei	Female	U of North Carolina	Department of Biostatistics	RF

Summer Program on Psychometrics

Program Core Participants

Last Name	First Name	Gender	Affiliation	Department	Status
Alonzo	Alicia	Female	University of Iowa	Teaching & Learning	SV
Atkinson	Thomas	Male	Memorial Sloan Kettering Cancer	Statistics	WG
Banks	David	Male	Duke University	Statistical Science	RF
Basch	Ethan	Male	Memorial Sloan Kettering Cancer	Other	SV
Bollen	Ken	Male	University of North Carolina	Sociology	SV
Burdick	Donald	Male	MetaMetrics, Inc.	Statistics	RF
Cai	Li	Male	University of California, L.A.	GSE&IS and Psychology	SV
Cao	Jing	Female	Southern Methodist U	Statistics	RF
Cheng	Ying	Female	University of Notre Dame	Psychology	SV
Cho	Sun-Joo	Male	University of California, Berkeley	Statistics	SV
Cleeland	Charlie	Male	University of M. D. Anderson Cancer Center	Life	SV

Cooke	Ben	Male	Duke University	Academic Resource Center	RF
Cui	Ying	Female	University of Alberta	Educational psychology	SV
Das	Sourish	Male	SAMSI, Duke University	Statistics	RF
de la Torre	Jimmy	Male	Rutgers University	Education	SV
Fairclough	Diane	Female	University of Colorado Denver, School of Public Health	Biostatistics and Informatics	SV
Feldman	Betsy	Female	University of California, Berkeley	Graduate School of Education	WG
Finkelman	Matthew	Male	Tufts University School of Dental Medicine	Statistics	SV
Fuentes	Jose	Male	Sandiego State University	Mathenatics and Statistics	SV
Gilligan	Theresa	Female	RTI Health Solutions	Patient Reported Outcomes	RF
Harrell	Leigh	Female	Virginia Tech	Statistics	SV
Hartigan	Brian	Male	University of North Carolina Wilmington	Psychology	RF
Henson	Robert	Male	University of North Carolina, Greensboro	Statistics	SV
Hill	Cheryl	Female	RTI Health Solutions	Patient Reported Outcomes	RF
Huff	Kristen	Female	College Board	R&D	SV
Jang	Eunice	Female	Ontario Institute	Education	SV
Johnson	Matthew	Male	Columbia U	Statistics	RF
Johnson	Valen	Male	U Texas	Statistics	RF
Karelitz	Tzur	Male	Education Development Center	Center for Science Education	SV
Lam	Tsz Cheung	Male	Rutgers University	Educational Psychology	WG
Levy	Roy	Male	Arizona State University	Education	SV
Loye	Nathalie	Female	University of Montreal	Administration et fondements de l'Éducation	SV

Lu	Jun	Male	American U	Statistics	RF
Madden	James	Male	Louisiana State University	Mathematics	WG
McGill	Mike	Male	Virginia Tech	Education	RF
McGowan	Herle	Female	North Carolina State University	Statistics	RF
McLeod	Lori	Female	RTI Health Solutions	Patient Reported Outcomes	RF
Morales	Knashawn	Female	University of Pennsylvania	Biostatistics and Epidemiology	RF
Nelson	Lauren	Female	RTI Health Solutions	Patient Reported Outcomes	RF
Nugent	Rebecca	Female	Carnegie Mellon University	Statistics	SV
Peruggia	Mario	Male	Ohio State University	Statistics	SV
Price	Mark	Male	RTI Health Solutions	Patient Reported Outcomes	RF
Rapkin	Bruce	Male	Albert Einstein College of Medicine of Yeshiva University	Div of Community Collaboration & Implementation	SV
Rijmen	Frank	Male	Educational Testing Service	Psychology	SV
Rivera-Medina	Carmen	Female	University of Puerto Rico	Institute of Psychological Research	RF
Rouder	Jeff	Male	U Missouri	Psychology	SV
Rupp	Andre	Male	University of Maryland	EDMS	SV
Schwartz	Carolyn	Female	Tufts University, School of Medicine	Medicine and Orthopaedic Surgery	SV
Sheng	Yanyan	Female	Southern Illinois University	Other	SV
Sinharay	Sandip	Male	Educational Testing Service	Sociology	SV
Speckman	Paul	Male	U Missouri	Statistics	RF
Sun	Dongchu	Male	Missouri	Statistics	RF
Swartz	Richard	Male	U Texas	Other	RF

Tatsouka	Curtis	Male	Case Western	Statistics	SV
Thissen	David	Male	UNC	Statistics	SV
Tractenberg	Rochelle	Female	Georgetown University Medical Center	Neurology	RF
Uenlue	Ali	Male	University of Augsburg	Institute of Mathematics	SV
Van Zandt	Trish	Female	Ohio State University	Sociology	SV
von Davier	Matthias	Male	Educational Testing Service	Statistics	SV
Wang	Jun	Male	North Carolina State University	Statistics Department	RF
Wang	Xiaojing	Male	Duke University	Statistical Science	RF
Williams	Valerie	Female	RTI Health Solutions	Patient Reported Outcomes	RF
Wilson	Mark	Male	UC Berkeley	Education	SV
Wu	Hao	Male	Ohio State University	Psychology	SV
Yue	Yu	Male	Baruch College, City University of NY	Statistics and CIS	RF
Zhang	Song	Male	U Texas	Comp Sci	RF
Zhang	Jingshun	Male	University of Toronto	Education	WG

B. Postdoctoral Fellows

This section includes the postdoctoral fellow selection and mentoring processes at SAMSI, synopses of the activities of the postdocs from their own perspectives, evaluations of the SAMSI postdoc experience.

1. Overview

The SAMSI Postdoctoral Fellowship experience is designed to bring together Statisticians and Applied Mathematicians in formal integrated research settings (e.g., Working Groups), informal settings (e.g., Lunches, seminars and events for undergraduates), and in opportunities for collaborations with researchers in other scientific disciplines.

Focus on integrating statistical and applied mathematical aspects in SAMSI programs begins with the Postdoc selection process. During the 2007-08 grant year, candidates applied to participate in the 2008-09 SAMSI Programs (Sequential Monte Carlo Methods and Algebraic Methods in Systems Biology and Statistics). The recruiting process involved SAMSI researchers, the SAMSI Directorate, and advertisement in AMSTAT News, ISM News, Mathjobs.com, and the SAMSI website. The 2008-09 Program Leaders and Scientific Advisory Committee were invited to assist in bringing SAMSI opportunities to the attention of promising doctoral candidates working in program-relevant areas of research. Final decision to rest with the SAMSI Directorate following either on-site interview at SAMSI or interview via Webex Teleconference. However, the careful assessment by the Program Leaders was invaluable and, in this case at least, led to happy consensus decisions.

When Postdocs arrive they become part of a Postdoc Community that, in addition to SAMSI Postdocs, includes NISS Postdocs and other young researchers in the NISS-SAMSI complex. This lively Community has monthly Postdoc Lunch Seminars with the Directorate where topics often include the practicalities of an academic or a research career (how to interview successfully for a position, how to plan and write a research proposal, how the publication process works in the mathematical sciences from journal selection through interpretation of written reviews to successful revision). A Biweekly Seminar series for Postdocs and Graduate Students provides a forum for “practice job interview” presentations of research results, serves to refine presentation skills and, serves an interdisciplinary role to inform Postdocs coming from different disciplines and/or working on different SAMSI Programs. Other shared activities within this Community include participation in the SAMSI Education and Outreach Program particularly the Undergraduate workshops, where Postdocs continue to be the most effective presenters for students of this age.

Effective mentoring of Postdocs is an essential part of SAMSI’s mission; therefore each Postdoc acquires two mentors. The Research Mentor, commonly the Working Group Leader of the Postdocs principal Working Group, and the Administrative Mentor, a member of the Directorate to provide knowledge of local issues and SAMSI information. This second mentorship also connects the Directorate in a personal, non-evaluative way

to Postdoc Life at SAMSI. In their comments, SAMSI Postdocs have continued to report that they feel well-supported by this dual-mentor system and by both particular mentors in their personal evaluations.

2008-09 SAMSI Postdocs

Julien Cornebise (Ph.D., Statistics, 2009, Université Pierre et Marie Curie)

SAMSI Program: *Sequential Monte Carlo Methods*

Research Mentor: Arnaud Doucet

Administrative Mentor: Jim Berger

Sourish Das (Ph.D., Statistics, 2008, University of Connecticut)

SAMSI Program: *Sequential Monte Carlo Methods*

Research Mentor: David Dunson

Administrative Mentor: Jim Berger

Christian Macaro (Ph.D., Statistics, 2007, University of Rome Tor Vergata)

SAMSI Program: *Sequential Monte Carlo Methods*

Research Mentor: Hedibert Lopes

Administrative Mentor: Jim Berger

Elizabeth Mannshardt Shamsheidin (Ph.D., Statistics, 2008, University of North Carolina- Chapel Hill)

SAMSI Program: *Sequential Monte Carlo Methods*

Research Mentor: Richard Smith

Administrative Mentor: Nell Sedransk

Ioanna Manolopoulou (Ph.D., Statistics, 2008, University of Cambridge,)

SAMSI Program: *Sequential Monte Carlo Methods*

Research Mentor: Mike West

Administrative Mentor: Nell Sedransk

Megan Owen (Ph.D., Mathematics, 2008, Cornell University)

SAMSI Program: *Algebraic Methods in Systems Biology and Statistics*

Research Mentor: Seth Sullivant

Administrative Mentor: Pierre Gremaud

Saeid Yasamin (Ph.D., Statistics, 2008, University of Indiana)

SAMSI Program: *Algebraic Methods in Systems Biology and Statistics*

Research Mentor: Seth Sullivant

Administrative Mentor: Pierre Gremaud

2. 2008-09 Postdoc Activity and Progress Reports

Julien Cornebise (*Sequential Monte Carlo Methods*)

Activity Report

Course(s): SMC course by Pr. Arnaud Doucet (Fall 2008)

Workshops Attended (and Workshop Support Tasks):

- SMC Mid-Program Workshop (February 2009): remote Webex attendance.
- Algebra Opening Workshop (September 2008): technical briefing of the supporting postdocs.
- SMC Opening Workshop (September 2008): poster presentation and technical support (computer and audio/video system) on Monday.

Postdoc-Grad Student Seminar – Presentation(s):

- Talk *Challenges in the SMC Tracking workgroup*, (October 28th, 2008). This talk has then been re-used by the Tracking workgroup as a recapitulative of its ongoing work.

Undergraduate Workshop(s) – Participation (specifics to be added later):

- Undergraduate Workshop (May 2009): Preparing – with grad student Wenjie Chen, training her for presentation too – the talk *Linear Inverse Problems*.
- Fall Undergraduate Workshop (October 2008): Talk *How to catch a submarine and a plane with the same tool*, October 31st, 2008, and collecting feedback about the expectations of the students in the following discussion.

Other Activities (e.g., teaching)

- Ph.D. dissertation.

As planned in October, after 3 months in SAMSI, I went back to France for 3 months to finish my Ph.D. dissertation, submitted early March 2009. During this period, I kept being involved in SAMSI working groups by regular meetings. I flew back to SAMSI on March 11th, now officially hired as a postdoctoral fellow (no longer a postdoctoral associate).

Since my return to SAMSI a month ago, I have started several research collaborations, exposed in the following of this document. They are of course still in an early stage, but activity is progressing fast and I aim at reaching publication for a good part of them by the end of SAMSI program.

- Informal SMC courses for graduate students

I have been offering advises on SMC methods to several SAMSI graduate students, in informal 1h – 1h30 one-to-one sessions, on topics ranging from Matlab implementation to kernel smoothing of discrete approximations or use of Kullback-Leibler divergence. This resulted from discussions started at working group, seminars, or social events, which turned into mini-courses which were mutually benefiting, as I love teaching and transmitting research experience.

Working Group I, Population Monte-Carlo

Special Tasks for Working Group:

- Organizer of the SAMSI topic contributed session *Population Monte Carlo / SMC Samplers* at the *Joint Statistical Meeting* (August 2009, Washington D.C.)
- Webmaster (year long).
- Chair of working group meetings when Arnaud Doucet (leader of the group) is unavailable.

Presentations to Working Group:

- Adaptive methods in Sequential Importance Sampling, October 17th, 2008.

Research Contributions:

- Working title: *On Auxiliary ABC-SMC*. Joint work with Oliver Ratmann (Imperial College, UK), with possible involvement of Gareth Peters (University of New South Wales, Australia)

This project aims at sharing my experience of SMC and Oliver Ratmann's experience of ABC (Approximated Bayesian Computation) to develop new algorithms in this very active field, by increasing the space with auxiliary variables. This approach eases the mathematical analysis of the existing algorithms and should allow for improved algorithms. The work includes ongoing discussions with Gareth Peters.

ABC - SMC algorithms are a current hot topic in the field of Monte Carlo methods. They aim at handling cases where it is possible to simulate observations but where their likelihood cannot be computed – either because it is intractable, or because the computation is computationally too expensive. These algorithms have been the subject of an impressive number of publications for the sole last two years, including an acute technical controversy on the correction of some bias induced by the methods – however anecdotic this might seem, such ado in the Monte-Carlo community is a testimony of the interest carried to ABC algorithms.

The applications of new algorithms, more efficient and whose critical quantities would be automatically chosen, would be of interest to a wide audience, which spans from Biology community focusing on population genetics where the likelihood can typically not be computed, to signal processing community with defense applications such as source-term estimation and plume tracking based on LIDAR data. Both applicative topics are currently investigated in subgroups of SAMSI's population and tracking working groups, respectively.

- Working title: *Adaptive SMC Samplers*. Joint plan with Arnaud Doucet (Institute of Statistical Mathematics, Tokyo, Japan).

In this work, we plan to extend the results of my Ph.D. from the filtering case to the broader framework introduced by Pierre Del Moral, Arnaud Doucet, and Ajay Jasra in 2006.

Obviously, the complexity added by SMC samplers, compared to SMC filtering, which results from the introduction of the arbitrary sequence of intermediate target distributions – a cooling schedule being only a one of many possibilities – and of the backward kernels, requires extreme care and hence makes adaptive methods all the

more needed. Practitioners peculiarly stressed this need at SAMSI's SMC program opening workshop. We aim at achieving construction of quality criteria based on the importance weights, similar to, e.g., the Coefficient of Variation of the weights, already used to trigger the resampling in the original SMC sampler algorithms. As in the classical SMC case, their asymptotic analysis should make appear function-free risk theoretic well known quantities (namely Kullback-Leibler and chi-square divergences).

The most interesting and challenging part is the design of efficient minimization algorithms for these criteria. Although the parallel with SMC filtering drives the first approach to SMC sampler, the arbitrarily chosen *backward* recursion on the target distributions will require innovative research: it might no longer possible to pick the optimum, and some anticipation on the following iterations will most likely be required.

- Ongoing research joint with SAMSI *Big Data* working group and Duke University, with Artin Armagan (Duke University) and Ioanna Manolopoulou (SAMSI). See presentation in the dedicated forthcoming section.

Working Group II, Tracking

Special Tasks for Working Group

- Chairman of the SAMSI topic contributed session *SMC Tracking* at the *Joint Statistical Meeting* (August 2009, Washington D.C.)
- Webmaster (from September to December 2008, and from April 2009 to end of program).
- Backup webmaster (from December 2008 to April 2009).
- Chair of working group meetings when Simon Godsill (leader of the group) is unavailable.

Presentations to Working Group:

- *Refined metrics on SMC algorithms: first thoughts*, March 23rd, 2009.

Research Area:

- Working title: *On Metrics for Comparing SMC Algorithms*. Joint work with Ernest Fokoue (Ohio State University, USA), François Septier (Cambridge University, UK), and Simon Godsill (Cambridge University, UK).

We aim at building quality criteria to compare distinct SMC filtering algorithms applied to a common problem. It is an extension of my Ph.D. dissertation work which focused on quality criterions for the proposal stage of SMC filtering algorithm. We now want to develop a unified theoretical framework which would similarly take into account the resampling schedule, the possible algorithmic variants, and the choice of the model. The immediate application in sight is multi-target tracking, with Poisson model and Track-Before-Detect models.

The ongoing tracks of thoughts involve investigation of results from several fields, in order to gather them in our SMC context. Model selection, for example, is a long-standing issue in the statistical community, and some recent developments in the late

90s and early 2000s have been aimed at decision-theoretic approaches, such as Gelfand and Ghosh's minimum posterior predictive loss approach, or Spiegelhalter et al.'s Deviance Information Criterion. As another example, the evaluation of the error caused by the discrete nature of the SMC approximation should benefit of existing results from survey theory, where the loss caused by sampling from a wider distribution is a critical issue.

I will visit the two latter collaborators in Cambridge at the end of May 2009 for even more efficient collaboration.

Inter-Working-Groups Research

- Working title: *SMC Samplers for Massive Datasets*. Joint work with Artin Armagan (Duke University) and Ioanna Manolopoulou (SAMSI, *Big Data* working group).

The concentration of SMC researchers, even from different subfields, necessarily gives birth to research interactions transcending the thematic borders of each working-group. This is precisely what how the following research project arose, as a collaboration between the Population Monte Carlo working group mentioned above, the Big Data working group (led by Mike West), and Duke University.

Artin Armagan originally came with questions about static estimation of a mixed-effects parameter based on a massive dataset raising from a longitudinal study of several thousand subjects. Markov Chain Monte Carlo (MCMC) methods cannot achieve correct estimation, as their computational requirements are far too strenuous to deliver an estimate in a decent amount of time.

We therefore investigated the use of SMC samplers, and realized that the issues we were facing were similar to those Ioanna Manolopoulou was dealing with. We therefore gathered our ideas in this ongoing collaboration, and are carefully crafting sequences of intermediate distributions for SMC samplers, which would eventually target the full posterior distribution while keeping a low computation overhead. The tracks currently investigated for this adaptive design of proposal include the instrumental use of variational Bayes methods and incremental subsampling from the data.

Inter-Programs Research

- Working title: *Robustness Assessment of Differential Systems using Monte-Carlo*. Joint work with Carsten Conradi (MPI Magdeburg, Germany, member of SAMSI Algebra Program).

This collaboration was initiated during my stay in the Fall, and pushed further during Carsten Conradi's second stay at SAMSI early April 2009. Carsten Conradi is a research from the Algebra Program, who is focusing on the bi-stability of deterministic systems of differential equations modeling bio-chemical systems. A major question arising in his earlier work is comparison of distinct models for a same bio-chemical system, in terms of robustness of the bi-stability to perturbation of the parameters. He led earlier works on this so-called sensitivity analysis by crude Monte-

Carlo approximations, which induces, however, a large bias whose correction is non-trivial.

Our aim is twofold. We are first formalizing the robustness criteria intuitively approximated so far, in terms of average distance of the points to the boundary of the set of interest in the parameter space. We then devise SMC algorithms to approximate efficiently these quantities.

The peculiarity of this problem lies in the fact that the available oracle, stating whether a given point lies within the set of “stable” parameters, requires as additional input a neighboring point known to belong to the set of interest. Therefore, any stochastic algorithm will only be able to detect an exit from the region of interest –in contrast, most Monte-Carlo approximation (e.g. used in volume assessment problems) require to sample as easily both inside and outside the set of interest.

Beyond this constraint lies an issue peculiar to the set of interest, which might have a null mass for the Lebesgue measure. The design of the criteria as well as the algorithms will therefore have to take into account the specific geometry of the problem, possibly through carefully crafted diffeomorphisms mapping the set of interest to a subset of a space into which it would accept a non-null Lebesgue measure – which would make the inference much easier, specifically as it would allow for easy simulation of a random walk, fundamental tool of our stochastic exploration algorithm.

Other Research Activity

Work on Papers from Ph.D. Research:

- Ph.D. dissertation, *Adaptive Sequential Monte Carlo Methods*, University Pierre and Marie Curie – Paris 6, March 2009.
- *Adaption in SMC filtering by Mixture of Experts*, with Jimmy Olsson (Lund University, Sweden) and Eric Moulines (Télécom ParisTech, France). To be submitted. This article benefited greatly from my stay in SAMSI during Fall, thanks to the numerous conversations with visiting researchers: their needs and interests helped me clarify the purpose of the talk and emphasize certain aspects.
- On the use of the coefficient of variation criterion for sequential Monte Carlo adaptation: a statistical perspective, with Jimmy Olsson (Lund University, Sweden) and Eric Moulines (Télécom ParisTech, France). In preparation.

Continuing Collaborations while at SAMSI:

- Jimmy Olsson, Lund University, Sweden.
- Eric Moulines, Télécom ParisTech, former Ph.D. advisor.

Presentations of Other Research:

- Talk *Recent breakthrough in adaptive sequential Monte Carlo methods*, May 18th, 2009, Parisian seminar of Statistics – inter-university monthly seminar, a major meeting of the French statistical community, where I while have the honor to present the results from my Ph.D.

Sourish Das (*Sequential Monte Carlo Methods*)

Activity Report

Workshops Attended (and Workshop Support Tasks):

1. SMC Opening Workshop, Sept 7-10, 2008
2. Risk Revisited: Progress and Challenges, May 21, 2008
2. Postdoc-Grad Student Seminar – Presentation(s):
 - *Analyzing extreme drinking behavior of patients suffering alcohol dependence disorder using Pareto regression* at the workshop Risk Revisited: Progress and Challenges, May 21, 2008
3. Undergraduate Workshop(s) – Participation (specifics to be added later):
 - *Risk analysis of extreme events: uncertainty in Statistics* at the SAMSI/CRSC undergraduate workshop, May 19, 2008
4. Poster presentation at the SMC Opening workshop September 8, 2008

Other Activities (e.g., teaching)

1. Taught as second instructor with (Jerry Reiter) Stat 101 in the Department of Statistical Science at Duke University, during Fall 2008.
2. Teaching Stat 101 (class of 96 students) in the Department of Statistical Science at Duke University, during Spring 2009.
3. Review papers for Journal of Multivariate Analysis
4. Review papers for Epidemiology
5. Review papers for Computational Statistics and Data analysis
6. Review papers for Journal of Statistical Planning and Inference
7. Organized invited session at ENAR 2009 : Speakers are: (i) Stuart Lipsitz (Brigham and Women’s Hospital), (ii) Xia Wang (University of Connecticut), (iii) Sourish Das (SAMSI, Duke University)
8. Organizing invited session at JSM 2009 (Accepted): Speakers are: (i) Mike Daniels (U of Florida, Gainesville), (ii) Bani Mallick (Texas A & M), (iii) David Dunson (Duke University), (iii) Sourish Das (SAMSI and Duke University)
9. Present an Invited Talk at Bayesian Colloquium of Statistics Department of North Carolina State University: Sep 30, 2008

Working Group I MAAD- Model Assessment

Special Tasks for Working Group: *Webmaster* for SMC- Model Assessment group.

Presentations to Working Group: N/A

Research Area – Plans:

Dunson, Pillai and Park (2007) developed Bayesian method for density regression, by allowing probability distribution to change flexibly with multiple predictors. In such effort, the conditional response distribution is expressed as nonparametric mixture of regression models with mixture distribution changing with predictor. Chung and Dunson (2008) introduced probit stick breaking process (PSBP) as a prior for an uncountable collection of predictor-dependent random probability measure. Our objective is to implement an augmented particle learning algorithm for posterior computation in PSBP mixture models. This involves sequentially updating latent Gaussian variables for each

subject in parallel across a large number of particles. The sampling steps are all straightforward, and the algorithm is currently being coded using a relatively simple mixture of Poissons example. The code will be compared with Gibbs sampling. One of the advantages of PSBP mixture prior is it allows us to specify the predictor dependent prior. This leads us to a develop a very easy variable selection method using kernel distance between the predictors through the model space efficiently. Our sub-working group consists of David Dunson and me.

Manuscript in preparation: *Bayesian Density Regression using Augmented Particle Learning*

Working Group II – Big – Data and distributed computing

Research Area – Plans:

Since our data set, contains massive numbers of predictors. We need to parallelize the algorithm and state of the art knows how about distributed computing from this working group will help us to enhance the efficiency of our algorithm.

Other Research

Work on Papers from Ph.D. Research:

1. *Analyzing extreme drinking behavior of patients suffering alcohol dependence disorder using Pareto regression* with Ofer Harel, Dipak Dey, Jonathan Covault, and Hank Kranzler (Submitted), SAMSI Tech Report # 2008-10
2. *Analysis of 5 Loxin Treatment for Patients with Osteoarthritis in Clinical Trial Using Power Filter* with Dipak Dey (submitted), SAMSI Tech Report # 2008-09
3. *On Bayesian inference of generalized multivariate gamma distribution* with Dipak Dey (submitted), SAMSI Tech Report# 2007-09

Other Research started or continued at SAMSI:

1. *Adaptive Bayesian analysis of binomial proportions* (2009) with Sonali Das (Accpeted in *South African Journal of Statistics*)
2. *Efficacy of Endoscopic Ultrasound (EUS) Guided Celiac Plexus Block (CPB) and Celiac Plexus Neurolysis (CPN) for Managing Abdominal Pain Associated with Chronic Pancreatitis and Pancreatic Cancer: a systematic review and Meta analysis.* (2009) with Kaufman, M. , Singh, G., Das, S., Micames, C., and Gress, F. (Accepted in Journal of Clinical Gastroenterology)
3. *Elicitation of Expert Prior opinion in Context of Presidential Election* (2009) with David Banks (work in progress: Tentative title)

Continuing Collaborations while at SAMSI:

- Sonali Das, CSIR, South Africa
- Marina Kaufman and Gurpreet Singh at SUNY downstate medical center, Brooklyn, NY 11203

Presentations of Other Research:

1. *Hurricane activity in the context of changing environment*, at Interface, 2008
Risk:Reality
2. *On Bayesian inference of generalized multivariate gamma distribution* at JSM 2008, Denver
3. Present an Invited Talk at Bayesian Colloquium of Statistics Department of North Carolina State University: Sep 30, 2008

Research Progress Report & SAMSI Program Final Report

Date: April 27, 2009

Research Contributions – Current Projects (grouped by Working Group)

Research Project Title:

Bayesian Density Regression using Augmented Particle Learning

Collaborator(s) & Mentor(s): David Dunson

Specific Goals & Accomplishments (results):

Dunson, Pillai and Park (2007) developed Bayesian method for density regression, by allowing probability distribution to change flexibly with multiple predictors. In such effort, the conditional response distribution is expressed as nonparametric mixture of regression models with mixture distribution changing with predictor. Chung and Dunson (2008) introduced probit stick breaking process (PSBP) as a prior for an uncountable collection of predictor-dependent random probability measure. We have successfully implemented an augmented particle learning algorithm for posterior computation in PSBP mixture models. This involves sequentially updating latent Gaussian variables for each subject in parallel across a large number of particles. The sampling steps are all straightforward, and the algorithm is currently being coded using a relatively simple mixture of Poisson and mixture of Normal examples. Small simulation study indicates that our method is superior than the default kernel estimator for density estimation. Currently we are running the code for a huge data set with sample size of 100,000. The code will be compared with Gibbs sampling. One of the advantages of PSBP mixture prior is it allows us to specify the predictor dependent prior. This leads us to a develop a very easy variable selection method using kernel distance between the predictors through the model space efficiently.

Research Contributions (publication submissions, articles in preparation, etc.):

The manuscript is in preparation and we expect the manuscript will be ready for submission by June 2009.

Research Area – Plans: Our plan is to continue the research and we can easily expand this method to space-time model for huge data sets. This will be appropriate for the next academic year.

Christian Macaro (*Sequential Monte Carlo Methods*)

SAMSI Activities

Course(s) (fall & spring): Sequential Monte Carlo Methods

Workshops Attended (and Workshop Support Tasks): Opening Workshop, Mid-Program Workshop

Postdoc-Grad Student Seminar – Presentation(s): SMC methods for Long Memory Stochastic Volatility Models

Undergraduate Workshop(s) – Participation (specifics to be added later):

Education and Outreach Program SAMSI Two-Day Undergraduate Workshop, Education and Outreach Program SAMSI/CRSC Undergraduate Workshop

Other Activities (e.g., teaching)

STA103 Duke University, STA293A Duke University.

Working Group

Presentations to Working Group: once every 2-3 weeks.

Research Area – Plans: Develop a sequential Monte Carlo scheme to deal with Long memory in stochastic Volatility models.

Other Research

Work on Papers from Ph.D. Research: Bayesian Non-parametric Signal Extraction for Time Series, Objective Priors for Autoregressive Models.

Other Research started or continued at SAMSI: Bayesian hierarchical non-parametric spectral analysis of magnetic resonance imaging data (with R. Prado).

Continuing Collaborations while at SAMSI: Bi-spectral analysis of long memory Stochastic volatility models (with C. Hurvich)

Research Progress Report & SAMSI Program Final Report

Date: 04/07/2009

Research Contributions – Current Projects (grouped by Working Group)

Research Project Title: Sequential Monte Carlo methods for Long Memory Stochastic Volatility Models

Collaborator(s) & Mentor(s): Hedibert F. Lopes

Specific Goals & Accomplishments (results): Propose to use an alternative representation of an autoregressive model of infinite order as an infinite sum of autoregressive models of order one. This allows to use standard SMC methods without worrying about the degeneracy of the particles. Research Contributions (publication submissions, articles in preparation, etc.): Sequential Monte Carlo methods for Long Memory Stochastic Volatility Models Presentations outside SAMSI (including invitations for future talks): JSM 2009

Future Research Plans (after completion of SAMSI Program)

Research Area – Plans: Bayesian econometrics and financial time series.

Continuing Collaborations (if appropriate): It depends on how quickly I manage to finish the projects I am already involved in.

Elizabeth Mannshardt Shamsheldin (*Sequential Monte Carlo Methods*)

SAMSI Activities

Workshops Attended: Opening SMC Workshop

Postdoc-Grad Student Seminar – Presentation: Mar 31st: “Severe Weather under a Changing Climate: Large Scale Indicators of Extreme Events” – joint work with Dr. Eric Gilleland from the National Center for Atmospheric Research.

One of the more critical issues with a changing climate is the behavior of extreme weather events, as these can cause loss of life, and have huge economic impacts. It is generally thought that such events would increase under a changing climate. However, climate models are currently at too coarse of a resolution to capture the very fine scale extreme events such as tornadoes or hurricanes. One approach is to look at the behavior of large scale indicators of severe weather. Here several factors are considered as large scale indicators of severe weather, including convective available potential energy and wind shear. This presents some interesting statistical issues. Numerous approaches, including the use of the generalized extreme value distribution for annual maxima, the generalized Pareto distribution for threshold excesses, a point process approach, and a Bayesian framework, are examined. Each approach is critiqued and compared for goodness of fit, model robustness, and predictive attributes on both re-analysis data and climate model output data. For the univariate case, it is relatively straightforward to analyze such data though numerous issues must be resolved. These issues include appropriate techniques for threshold selection and prior specification. A bivariate approach can also be considered. In addition, when analyzing weather extremes, one is faced with a spatial field. Predicting extreme weather events is an important, growing area of research and there remain many avenues for further exploration. Acknowledgements to Harrold E. Brooks, Patrick Marsh and Matt Pocerlich.

Graduate Student Workshop: Industrial Mathematical & Statistical Modeling Workshop – co-sponsored by SAMSI and NCSU. Faculty mentor. Developing project on “Severe Weather under a Changing Climate” suitable for 2 weeks of collaboration with graduate students.

Other Activities

Full teaching responsibilities at Duke. Teaching: Introduction to Statistics course Fall 08 and Spring of 09, and special topics graduate course in Extreme Value Theory and Applications for Spring 09.

Working Group

Special Tasks for Working Group: Particle Learning Group - Webmaster

Other Research

Work on Papers from Ph.D. Research: In collaboration with Richard L. Smith “Asymptotic Multivariate Kriging Using Estimated Parameters with Bayesian Prediction Methods for Non-Linear Predictands”.

Other Research started or continued at SAMSI:

Revisions for paper submitted to Annals of Applied Statistics – “Downscaling Extremes: A Comparison of Extreme Value Distributions in Point-Source and Gridded Precipitation Data”

Presentations of Other Research:

- Asymptotic Multivariate Kriging Using Estimated Parameters with Bayesian Prediction Methods for Non-Linear Predictands – Duke Statistical Science Departmental Seminar Series; University of Santa Barbara Colloquium Series – Jan 14th, 2009
- Downscaling of Extremes: A Comparison of Extreme Value Distributions in Point-Source and Gridded Precipitation Data - Colloquium presentation for the University of Virginia

Research Contributions – Current Projects

Research Project Title: Faculty Mentor for the Industrial Mathematical & Statistical Modeling Workshop – co-sponsored by SAMSI and NCSU. Project title: “Severe Weather under a Changing Climate”

Collaborator(s) & Mentor(s): Eric Gilleland (National Center for Atmospheric Research) and Richard L. Smith (UNC-CH)

Specific Goals & Accomplishments (results): to expose graduate students in mathematics, engineering, and statistics to challenging and exciting real-world problems arising in industrial and government laboratory research. Students get experience in the team approach to problem solving.

Research Contributions:

- Finished revisions for Annals of Applied Statistics submission. Will re-submit after other authors offer comments on revisions.
- Paper “Asymptotic Multivariate Kriging Using Estimated Parameters with Bayesian Prediction Methods for Non-Linear Predictands “ in preparation with Richard L. Smith
- Collaboration with Gilleland and Smith on Climate Extremes project for IMSM Graduate Workshop for possible publications after July 2009 workshop.
- “Severe Weather under a Changing Climate: Large Scale Indicators of Extreme Events” – joint working paper with Dr. Eric Gilleland.

Presentations outside SAMSI:

- Downscaling of Extremes: A Comparison of Extreme Value Distributions in Point-Source and Gridded Precipitation Data - Colloquium presentation for the University of Virginia, Nov 21st, 2008
- Asymptotic Multivariate Kriging Using Estimated Parameters with Bayesian Prediction Methods for Non-Linear Predictands – Colloquium presentation for the University of Santa Barbara Colloquium Series, Jan 14th, 2009; Duke Statistical Science Departmental Seminar Series, Jan 12th, 2009
- Severe Weather under a Changing Climate: Large Scale Indicators of Extreme Events – Submitted for presentation at The International Environmetrics Society, July 2009.

Research Area – Plans: Environmental Extremes and Spacial-Temporal Applications

Continuing Collaborations: with Gilleland and Smith. In addition, continued collaborations with projects established during the Space-Time Analysis for Environmental Mapping and Climate Change program.

Presentations outside SAMSI: Submitted abstract for TIES (The International Environmetrics Society) 2009. I am considering possible submission for Bayes (Valencia) 2010. I will discuss options for relevant conference/workshop submissions for appropriate topics developed in next year’s climate change program.

Future Career: I am interested in a tenure-track position in a statistics department at a US research institution.

Ioanna Manolopoulou (*Sequential Monte Carlo Methods*)

SAMSI Activities:

- Attended course on Sequential Monte Carlo methods (Fall ‘08).
- Webmaster of the working group ‘Big data and distributed computing’.
- Webmaster of the postdoc website.
- Webmaster of the Inter-disciplinary Undergraduate Workshop (Spring ‘09).
- Research nugget: 'Needle in a Haystack: Rare Cell Subtypes in Flow Cytometry'.

SAMSI Workshops Attended (and Workshop Support Tasks):

- Inter-disciplinary undergraduate Workshop (May '09): Talk on 'Brief Introduction to the Computing System and MATLAB' and support throughout the workshop.
- Adaptive design, SMC and computer modeling workshop (April '09) planning to give a talk with title 'Adaptive Bayesian Computation for Targeted Learning in Mixture Models'.
- Postdoc-Grad Student Seminar presentation (March '09): 'Targeted re-sampling from very large datasets using mixture modeling'.
- Internal SAMSI SMC workshop (February '09) gave a talk with title 'Targeted re-sampling from very large datasets using mixture modeling'.

- Undergraduate Workshop (October '08): 'Rare event detection in very large datasets'.
- Postdoc-Grad Student Seminar presentation (October '08): 'How to use graphics in presentations'.
- SMC opening workshop (September '08), attended and supported as webmaster of my working group.

Working Group: Big Data and Distributed Computing

The general objective of the group is to investigate the use of Sequential Monte Carlo in very large datasets, specifically in datasets with a large number of observations and/or of high dimensionality, regression on a large number of covariates, clustering in multiple dimensions, rare event detection. SMC methods allow for the use of parallel computing in several levels of the analysis (parallelizing on the particle level, on disjoint areas of the parameter space, on different sets of parameters, or on disjoint sets of observations).

1. One of the projects I have been working on, joint with Professor Mike West, has focused on using Sequential Monte Carlo methods in order to detect rare events in mixture models by means of sequential targeted re-sampling. This study was motivated by an example which arises in flow cytometry, where datasets can be very large, but the parameters of interest related to a region of very low probability in the sample space.

This work will be presented in a theoretical paper with title 'Targeted sequential re-sampling from very large datasets in mixture modelling' (in preparation), and will most likely lead to one or more collaborative biological papers in flow cytometry.

I have given two presentations in the working group. The first gave an overview of different levels of parallelization in Sequential Monte Carlo, based on a number of related papers. Mike West gave an initial presentation on our joint work, and subsequently I gave an updated presentation of the progress of our work.

2. Another project I am working on is with Dr Artin Armagan and Dr Julien Cornebise on SMC methods in very large datasets with complex models. The problem arose recently in a longitudinal mixed effects study lead by Artin Armagan, who initiated a collaboration between Julien Cornebise and me. This is work in progress and combines ideas from the Big Data working group as well as the Population Monte Carlo working group. We are looking into constructing efficient SMC samplers, using both adaptive methods as well as Variational Bayes approaches, in cases where simulating and calculating the exact posterior is computationally very expensive.

Work on Papers from PhD Research

The topic of my PhD thesis was phylogeography, combining phylogenetics and spatial distribution/clustering. Since arriving at SAMSI, I have had the chance to discuss my work with several researchers who work in phylogenetics as part of the Algebraic Methods in Biology workshop.

- Manolopoulou, I., Tavaré, S., 'A Bayesian approach to Nested Clade Analysis' (in preparation), to be submitted (possibly) to Theoretical Population Biology.
- Legarreta, L., Manolopoulou, I., Thebaud, C., and Emerson, B. 'Phylogeography of *Rhinusa vestita* (Coleoptera: Curculionidae) in the Iberian Peninsula: a Bayesian approach'. To be submitted.

Other Research started or continued at SAMSI:

Work with Thomas Kepler, Mike West, Chunlin Ji and Xiaojing Wang on spatial mixture modelling for unobserved point processes with applications in Immunology, to be submitted to Bioinformatics as a paper with title 'Statistical analysis of immunofluorescent histology'. This work is aimed at constructing automated statistical methods for analyzing immunological histological images, and is closely related to the recent paper Chunlin Ji, Daniel Merl, Thomas Kepler and Mike West. "Spatial Mixture Modelling for Partially Observed Point Processes: Application to Cell Intensity Mapping in Immunology." (2008). Using distributed computing techniques employed and presented in the SAMSI 'Big Data and Distributed Computing' working group, the approach is extended by means of dividing images into sub-images and parallelizing computations.

Continuing Collaborations while at SAMSI:

Brent Emerson and Lorenza Legarreta at the Department of Evolutionary Biology, University of East Anglia, UK.

Thomas Kepler, Cliburn Chan, Mikhail Levin at the Centre for Computational Immunology, Duke University, US.

Presentations outside SAMSI:

- Joint Statistical Meeting: Talk with title 'Adaptive Bayesian Computation for Targeted Learning in Mixture Models', as part of the SAMSI Topic Contributed Session.
- Santa Fe Institute: Lectures on 'Histology and Image Analysis' and 'Introduction to Statistics' as part of the Computational Immunology Summer School '09.
- Greek Stochastics Meeting with theme 'Monte Carlo: Probability and Methods'. Talk with title 'Targeted re-sampling from very large datasets in mixture modelling'.

Megan Owen (*Algebraic Methods in Systems Biology and Statistics*)

SAMSI Activities

Course(s): Algebraic Methods in fall 2008

Workshops Attended (and Workshop Support Tasks):

- Tutorials at opening workshop for Sequential Monte Carlo Methods program
- Opening workshop for Algebraic Methods in Systems Biology and Statistics (presented poster, speaker assistance)
- Discrete Models in Systems Biology workshop (speaker assistance)
- Algebraic Statistical Models (speaker assistance)

- Molecular Evolution and Phylogenetics (presented poster, speaker assistance)

Postdoc-Grad Student Seminar – Presentation(s):

- How to present graphics
- Geometry of cophylogeny

Undergraduate Workshop(s) – Participation:

- presentation “Tree Distances and Tree Space”
- developed and ran lab “Interactive Session on Phylogenetic Trees”

Other Activities (e.g., teaching)

- NCSU Algebra and Combinatorics Seminar, Jan. 2009
- NCSU BioMath Seminar, Feb. 2009

Working Group I - Evolutionary Biology

Special Tasks for Working Group: webmaster

Presentations to Working Group:

- Space of Phylogenetic Trees
- discussion on a paper about the edge-product phylogenetic tree space

Research Area – Plans:

I will pursue several research topics related to this working group. I will work on the combinatorial and tree space problems related to the co-phylogeny problem, as posed by the University of Kentucky group of Rudy Yoshida, Chris Schardl, and Jerzy Jaromczyk, and their collaborators. I visited them at the University of Kentucky October 13-18 to discuss this work. I am also interested in finding a biologically relevant distance for the “phylogenetic orange” space, and have had some preliminary discussions with Serkan Hosten and John Rhodes. Finally, there seem to be some connections between the various phylogenetic tree spaces, tropical geometry, and the co-phylogeny problem, which I am also interested in investigating, perhaps with Serkan Hosten, Seth Sullivan, or Rudy Yoshida.

Working Group II - Algebra Network Inference

Special Tasks for Working Group: webmaster

Presentations to Working Group:

Research Area – Plans: I’m interested in the data discretization problem.

Other Research

Work on Papers from Ph.D. Research: Working on converting my thesis into a journal paper.

Research Contributions – Current Projects (grouped by Working Group)

Research Project Title: Geometry of Cophylogeny

Collaborator(s) & Mentor(s): Ruriko Yoshida and Peter Huggins

Specific Goals & Accomplishments (results): The goal is to study the geometry and combinatorics of cophylogeny. We proposed several, biologically-motivated spaces of cophylogenetic trees, as well as a new distance designed for comparing

cophylogenies. We show a connection between this distance and the Nearest-Neighbor Interchange distance.

Research Contributions (publication submissions, articles in preparation, etc.):
Submitted paper “First steps towards the geometry of cophylogeny” to the Bulletin of Mathematical Biology

Presentations outside SAMSI (including invitations for future talks):

- AMS Southeastern Sectional meeting, Invited talk in the Special Session on Applications of Algebraic and Geometric Combinatorics, April 2009
- AWM workshop at the SIAM Annual meeting, July 2009

Research Project Title: Generalization of the space of phylogenetic trees.

Collaborator(s) & Mentor(s): Serkan Hosten

Specific Goals & Accomplishments (results): The space of phylogenetic trees can be viewed at the Bergman fan of the graphic matroid of the complete graph. Using this characterization, we are interested in generalizing the space of phylogenetic trees and the geodesic distance on it.

Research Project Title: Computations in the Space of Phylogenetic Trees

Collaborator(s) & Mentor(s): Scott Provan

Specific Goals & Accomplishments (results): Develop efficient algorithms for calculating such measures as the geodesic distance, centroids, etc. in the space of phylogenetic trees. We found the first polynomial time algorithm for computing the geodesic distance between two trees, as defined by Billera et al. (2001). It had been an open question as to whether a polynomial time algorithm existed. We are currently investigating methods to compute centers of mass in tree space.

Research Contributions (publication submissions, articles in preparation, etc.):

- Preparing paper “Computing the Geodesic Distance in Tree Space in Polynomial Time”

Presentations outside SAMSI (including invitations for future talks):

- 2nd Canadian Discrete and Algorithmic Mathematics Conference (CanaDAM09), May 2009

Research Project Title: Statistics on the space of phylogenetic trees.

Collaborator(s) & Mentor(s): Sayan Murherjee, Katia Koelle, Sean Yuan, and Simon Lunagomez

Specific Goals & Accomplishments (results): We plan to compare different influenza strains using the geodesic distance, investigate the likelihood function on tree space, and develop statistical methods for using tree space and the geodesic distance in biology.

Research Contributions (publication submissions, articles in preparation, etc.):

Presentations outside SAMSI (including invitations for future talks):

Future Research Plans (after completion of SAMSI Program)

Research Area – Plans: I plan to continue studying spaces of phylogenetic trees. As well as pursuing the projects mentioned above, I am also interested in constructing a space of phylogenetic networks.

Continuing Collaborations (if appropriate): I plan to continue all of the collaborations detailed above.

Presentations outside SAMSI:

- 2nd Canadian Discrete and Algorithmic Mathematics Conference (CanaDAM09), May 2009
- AWM workshop at the SIAM Annual meeting, July 2009

Saeid Yasamin (*Algebraic Methods in Systems Biology and Statistics*)

SAMSI Activities

Course(s): Algebraic Statistics

Workshop Attended (and Workshop Support Tasks): Algebraic Methods in Systems Biology and Statistics

Postdoc-Grad Student Seminar – Presentation(s): Algebraic Statistics Group

Working Group I

Special Tasks for Working Group: Algebraic Statistics

Presentations to Working Group: Hypothesis Testing over Symmetric Cones

Research Area – Plans: Maximal Likelihood Estimation for Graphical Models

Working Group II

Special Tasks for Working Group: Network inference

Presentations to Working Group:

Research Area – Plans: Developing some statistical tools for analysis of variance in discrete models.

Other Research

Work on Papers from Ph.D. Research: Hypothesis Testing for Wishart Models, with Steen Andersson

Other Research started or continued at SAMSI: Maximum Likelihood Estimation for Graphical Models, with Seth Sullivant

Continuing Collaborations while at SAMSI: Shape Space Analysis of Symmetric Cones, with Armin Shwartzman from Harvard School of Public Health.

Presentations of Other Research:

Poster presentation at IMA workshop on Multi-Manifold Data Modeling and Applications, October 3, 2008.

Attended the Clifford lecturers on Tropical Geometry at Tulane University November 11-15, 2008.

MSRI workshop on Algebraic Statistics, December 15-16, 2008.

Current Projects (Algebraic Statistics Working Group)

Research Project Title: Maximum Likelihood Inference for Graphical Models

Collaborator(s) & Mentor(s): Seth Sullivant

Specific Goals & Accomplishments (results): Our main goal in this research is to answer these two questions:

1. For a given graphical model, what is the least number of observations needed to obtain the maximum likelihood estimator?
2. For a given data set coming from a (Gaussian) graphical model, how complex the model can be to estimate the concentration parameter?

Presentations outside SAMSI:

1. Department of mathematics, GWU
2. AMS Special Session on Algebraic Methods in Statistics and Probability, March 27-29, 2009

3. Postdoc Experience Evaluation

Julien Cornebise

1. Program Involvement:

Part of the *Tracking* and *Population Monte Carlo* working groups of the Sequential Monte Carlo program. See details in the mid-program report.

2. Interactions with Other Institutions:

Collaborations are ongoing with Duke University, with my former co-author from Lund University, and with my former Ph.D. advisor from Télécom ParisTech, France. See details in the mid-program report.

Past interactions include stays in Computer Science engineering school ESIEA (5 years), Université Paris 6 (4 years), National engineering school Telecom ParisTech (3 years), private pharmaceutical company's statistical company (6 months), Lund University (1 month).

3. High Points at SAMSI:

I am literally amazed by the incredible opportunities to meet and interact with most of the best researcher of the field. People which were so far only (somewhat mythic) names on articles are now blood and flesh persons, researchers, colleagues, which is a tremendous transition from the grad studies to the post-doctorate research. On a scientific point of view, these interactions bring a deep re-evaluation of my earlier work, making it fit in a much broader view of the field, stressing its advantages, seeing how it can be extended to neighboring areas. On a more human point of view, they contribute to mutating a finishing student into a full and mature (though young) member of the scientific community.

An especially remarkable high point is also the very consideration brought to the post-docs as researchers. However rich and benefiting the experience as a grad student may have been, the role of a post-doc is definitely one of a entire researcher, with the associated responsibilities – in terms of research, organization of working groups, advises

to graduate students – and the associated consideration and status. I cannot stress enough the extremely positive impact this change of perspective had on the very nature of my work and my approach to research. However personal this might look, this newly acquired self-confidence, and realizing that, yes, I belong here, I belong to the research community, are something that can hardly be expressed in its full importance. Besides, discussions with several other postdocs have led me to realize that this relief and the former doubts seem to be a common issue amongst young doctors, which help all the most to reduce them and move on to the next stage.

On a more “local” scale, but nevertheless so important, I must thank the whole SAMSI staff, and especially the “Fantastic Four”, “SAMSI’s angels”, Denise Auger, Rita Fortune, Sue McDonald, Terri Nida, for their so precious help, from the first pre-hiring interview to the settling of all the practical and administrative details upon arrival – from lodging to VISA issues, all the more important to a newcomers foreigner –, and still keeping on through the day-to-day life in SAMSI. Never, ever, did I meet such a dedicated, patient, helpful and welcoming team.

4. Suggestions for Improvement:

I hardly see any point that could be improved at SAMSI. Concerning the offered setting and means for doing research, it seems like it cannot be more perfect that it is.

The only point I can find concern the computer resources. SAMSI is gifted with great hardware, the brand new computers are “computational beasts”, and the number of computers is striking for such an institute. However, it could really benefit a centered network administration, which would first and foremost allow remote access – hence making them real work platforms available from anywhere that one could rely on.

Please don’t read me wrong: the current IT responsible, James Thomas, is extremely willing to help, and I here wish to thank him for the reactivity and the help he provides to everyday problems, from the failing printer to Matlab installation, from the need for a new mouse to the addendum of a new software. James Thomas fulfills these IT needs with a dedication that forces the admiration -- I saw him staying until the middle of the night setting up the new computers !

My only suggestion regarding improvements would be to evaluate whether the joint network administration of both NISS and SAMSI networks can be dealt by a single administrator. Though this is probably not the kind of advice expected from a post-doctoral fellow, I would recommend thinking of doubling the resources allocated to this task: SAMSI’s potential is here grand, but still dormant.

My experience so far (both as a user, and as a system administrator for several years in a professional context) is that the network administrators are always a team of at least two people, that hence provide completing competences and knowledge (network administration getting a broader and broader field), as well as a mandatory double reflection and confronting (therefore enriching !) point of views on crucial technical choices – especially in a sensitive networks such as NISS and SAMSI have been outlined to be in the last postdoctoral lunch.

Up to an added increase of security (which has a price that can hardly be measured) and efficiency, the cost would most likely not be as increased as expected, as the spending on a salary would partly be compensated by reduced call to external maintenance and services.

5. Mentoring:

I here would like to express my gratitude for the great comprehension of SAMSI directorate, first and foremost of its director – and my administrative advisor – Jim Berger. SAMSI allowed for exceptional measures to adapt the post-doctoral contract to the late schedule of my dissertation (whose completion was originally planned just before the beginning of SMC program), and permit me to be part of this grand research occasion while still finishing my dissertation.

I will never forget the open-mindedness I have witnessed (and benefited!) when discussing this matter with Jim Berger, comparing efficiently the possible ways to tackle it, and the strong will I perceived to find a solution optimal to both SAMSI and its postdoc. I was afraid that these delays (for which I take entire responsibility) may have put an end to the warm welcome I found in SAMSI. On the very contrary, an offer that was then worked out to finish my 3 months stay here, then getting back to France for a couple of month – time to finish writing the manuscript while keeping up to date to SAMSI's program – then coming back as a full-time postdoctoral fellow. This generous and ideal (from my point of view) agreement was a tremendous incentive to wrapping up the leftover writing as fast as possible, and to give the best of my possibilities to SAMSI, both while back in France to keep up to the progresses of the working groups, but, even more important, now that I am back, done with the dissertation.

For this, again, I would like to renew my thanks to Jim Berger and the directorate of SAMSI.

On the scientific side, the mentoring from Arnaud Doucet has been extremely benefiting. Its long stay during the whole Fall of 2008 was the occasion of fascinating conversations, extremely enriching, as well as occasions to become a reviewer through an article to which editor he recommended me. Beyond that, his mentoring keeps on going now that he is in Japan, by means of email conversations and his support in a grant application to join him at the Institute for Statistical Mathematics in Japan for a two months stay next Fall.

6. SAMSI Benefits for the Future:

The benefits from my SAMSI experience have been outlined at length in the question above, and can be summarized as triggering the key mutation from a finishing grad student to a fulltime researcher, active member of the community.

7. SAMSI in contrast with University Setting:

In my opinion, SAMSI's uniqueness resides in the way it relies on the post-doctorates as the coordinators of its day-to-day scientific life, from working group coherence to

welcoming of “anchor point” for visiting researchers and professors. In a university setting, my guess is that this unique role would be conferred to local assistant professors, rather than postdocs. This responsibility, however, is the key benefit of SAMSI, as it pushes to a greater dynamism and increased interaction with the members of the program.

8. SAMSI in comparison to Other Experiences:

I lack extra-university research experiences to make a comparison, e.g. with private research institutes. However, I can compare with the average French position, when no post-doctoral research position is found after a research, as “Teaching and Research Assistant” (A.T.E.R.), which is a one-year once-renewable contract consisting of a full teaching load. It is commonly reported by young French researchers – though I did not experiment it firsthand, so this is to be taken with caution, while keeping in mind that, happily, this is only a tendency to which a lot of exceptions do exist – that the young A.T.E.R. is mainly filling the gaps in the teaching schedule, and that the only achievable research during this year can be finishing some articles taken from the Ph.D. dissertation.

No need to stress how much this lies at the opposite extreme of SAMSI.

9. Other Research while at SAMSI:

As stated in the “Mentoring” section, during my first 3 months stay, I have both been acting in two working groups and worked on finishing writing my Ph.D. dissertation. Now that I have been back for a month, and as is further detailed in my mid-program report, the only other research planned while at SAMSI consists on finishing two articles out of my dissertation.

Besides, these two articles are deeply related to SAMSI’s SMC program, as my whole Ph.D. is about Adaptive SMC methods. Therefore, this “other research while at SAMSI” still can be seen as part of the working group activities.

Here again, the benefits of being at SAMSI are blatant, in terms of the scientific maturity in the field, gained by the never-ending interactions occurring there.

10. Other Comments:

I can only renew the expression of my joy to be part of SAMSI. From a professional and scientific perspective, it brings incredible assets, and (as far as I can judge) a formidable booster for a research career. On the human perspective, the dynamic and the energy flowing in SAMSI are a precious gift. I already guess that meeting again such a highly stimulating environment will not occur easily !

Christian Macaro

1. Program Involvement:

Particle learning working group within the Sequential Monte Carlo program.

2. Interactions with Other Institutions:

Duke University.

3. High Points at SAMSI:

High quality research opportunity.

Friendly environment.

Nice and quiet office.

Very good computing facilities.

Nice and friendly interactions with directorate and administration.

4. Suggestions for Improvement:

Allow the access to computing facilities from outside SAMSI.

Provide access to on-line journals.

5. Mentoring: N/A

6. SAMSI Benefits for the Future:

Networking.

Possible publications.

7. SAMSI in contrast with University Setting:

SAMSI is a research based institute.

8. SAMSI in comparison to Other Experiences:

I don't have other research experiences which are comparable to SAMSI.

9. Other Research while at SAMSI:

Bayesian hierarchical non-parametric spectral analysis of magnetic resonance imaging data (with R. Prado)

Bi-spectral analysis of long memory Stochastic volatility models (with C. Hurvich)

10. Other Comments: N/A

Elizabeth Mannhardt Shamseldin

1. Program Involvement:

I have been partially involved with the Sequential Monte Carlo program and participated in the Particle Learning Group working group. My involvement with SAMSI programs will increase substantially in the fall when the "Space-Time Analysis for Environmental Mapping, Epidemiology and Climate Change" program begins. This program will focus 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. Through

my contacts at SAMSI, I have started working with Dr. Eric Gilleland from the National Center for Atmospheric Research on a project looking at large-scale indicators for predicting extreme weather events on a fine-scale resolution. This looks at bridging the mathematical gap between the climate models with produce climate predictions on a large, even global scale, and predicting extreme weather events such as severe storms (tornados, hurricanes, etc) on a more localized scale.

2. Interactions with Other Institutions:

I am concurrently a Visiting Assistant Professor at Duke University. This has given me an opportunity to teach both undergraduate and graduate courses, which is an invaluable learning experience. I also continue to have papers in progress with my dissertation adviser, Richard L. Smith, at the University of North Carolina at Chapel Hill. It is especially convenient that the three institutions are within driving distance of one another, as this facilitates communications, etc. I also continue to have contact with various scientists at NCAR, the National Center for Atmospheric Research in Boulder, CO. Most particularly Dr. Eric Gilleland with whom I am collaborating on a research project concerning extreme weather events. This project will assist in a further interaction with North Carolina State University and the Center for Research in Scientific Computation as we develop a project appropriate for graduate student research efforts for the Industrial Mathematical & Statistical Modeling Workshop for Graduate Students.

3. High Points at SAMSI:

The mentoring postdoc lunches are extremely helpful. The topics covered are useful and relevant, and the one-on-one interaction with the experience directorate is invaluable. The social events are also definitely a highlight. Not only do they provide a place for people to relax and have some good food, they encourage interaction and personal relationships across young researchers, experienced professors, and SAMSI staff. The familiarity it provides leads to a family-environment, which encourages relationships and strengthens professional ties.

4. Suggestions for Improvement:

It is difficult to travel back and forth between SAMSI and other supporting institutions, in large part because logging into SAMSI computers remotely is not possible. This requires all work to be done in-house. It would be beneficial to be able to work from Duke, or from one's personal laptop at home. The additional time saved commuting between institutions would also lead to increased production.

5. Mentoring:

Since I was not fully involved in this year's programs at SAMSI, I did not have an official individual mentor through the program. The general mentoring of the directorate is great. All members (Jim, Nell, Michael, Pierre) are approachable and are genuinely concerned for the success of the young researchers. I feel if I have questions or concerns about any topic pertaining to my work at SAMSI – conference/travel issues, applying for permanent positions, etc, or even general research/statistical questions - that I could

approach any member and they would be happy to discuss things with me. That is not always the case in every work setting, and the care and attention that they provide is very much appreciated.

6. SAMSI Benefits for the Future:

The benefits pertaining to my future in academics as a result of my involvement with SAMSI are immeasurable. I have made and will continue to make many valuable contacts pertaining to not only future research possibilities but future employment as well. The opening workshop and courses offered at SAMSI in conjunction with the programs each year are very informative. They provide a great general background as well as up-to-date innovations in the given research areas. I am looking forward to the opening workshop and courses offered next year for the Space-Time Analysis for Environmental Mapping, Epidemiology and Climate Change program. The working groups are a great hands-on learning environment for how to work effectively on collaborative projects. They also provide an informal setting for presenting and discussing the latest research in a more specific area than the courses and opening workshop.

7. SAMSI in contrast with University Setting:

I have had many responsibilities associated with the university setting in my role as a visiting assistant professor at Duke. It is a great opportunity, but there also is a large time commitment associated with teaching courses each semester. The benefits of the SAMSI setting in contrast with the university setting are many. Academics at SAMSI are able to solely focus on research, which greatly enhances productivity and results. Collaborations are also greatly encouraged, and the colloquial environment at SAMSI with the many visitors that come through makes these collaborations not only possible but effective, which is not always the case when collaborators are separated by distance and time-zones. SAMSI also provides a very comfortable setting among peers of one's same level, which creates a unique environment for workplace camaraderie. This makes coming to work every day an enjoyable experience!

8. SAMSI in comparison to Other Experiences:

After completing my undergraduate degree I worked in industry for a year. Also, during graduate school, I worked as a consultant for Constella/SRA, which provides analytical solutions for the health services industry. Both of these experiences offer many areas of comparison for my SAMSI experience. Once again, SAMSI is a very colloquial, collaborative environment. Industry can provide many interesting projects, however the approaches, resources, and timeline are often at the mercy of the client. Many times a project budget runs out long before the problem has been fully addressed. Or the client may be interested in a more common method, which can be explained to supervisors and subsidiaries, rather than a novel technique that may provide a more enriched solution.

9. Other Research while at SAMSI:

I completed the revisions for “Downscaling Extremes: A Comparison of Extreme Value Distributions in Point Source and Gridded Precipitation Data” an article submitted to the Annals of Applied Statistics. I am also communicating with my adviser concerning the papers under development from my dissertation – “Asymptotic Multivariate Kriging Using Estimated Parameters with Bayesian Prediction Methods for Non-Linear Predictands”

10. Other Comments:

I am enjoying my time at SAMSI and am looking forward to enriching my experience through next year’s program on Space-Time Analysis for Environmental Mapping, Epidemiology and Climate Change. The snacks in the break-room are also appreciated!

Ioanna Manolopoulou

1. Program Involvement:

Part of the 'Big data and distributed computing' working group of the Sequential Monte Carlo program.

2. Interactions with Other Institutions:

Collaborations at Duke University, US, University of Cambridge, UK, and University of East Anglia, UK. Currently affiliated with Duke University.

3. High Points at SAMSI:

The working groups and workshops were certainly a great opportunity, especially because SAMSI managed to attract almost all leading researchers in the field. The SAMSI SMC course run in the Fall was an excellent overview which highlighted the challenges and advantages of SMC methods. The webex facilities are great for maintaining lively working group meetings even after researchers leave SAMSI.

In addition, the SAMSI staff have been extremely helpful with everything, allowing us to focus on the research. Especially for people who moved into the area this has been really valuable.

4. Suggestions for Improvement:

The computing resources and computing access can be greatly improved and would be very beneficial. In many cases this is a matter of better-run computing facilities rather than lack of hardware resources.

5. Mentoring:

My scientific mentor and collaborator has been closely following my progress and providing his inexhaustible scholarship, and has been very proactive about the organization and activity of our working group.

6. SAMSI Benefits for the Future:

Aside from having had the chance of getting great insight in the area of Sequential Monte Carlo and attending a variety of stimulating talks, perhaps the most important benefit was establishing future collaborations with some great researchers.

7. SAMSI in contrast with University Setting:

SAMSI has a strong research focus and is an ideal environment for collaborations, because people work on very similar areas, allowing for overlap of methods. The working group meetings are a great inspiration for projects and encourage exchange of ideas.

8. SAMSI in comparison to Other Experiences:

SAMSI is a much more interactive research environment, with several experts in the field attending workshops and visiting for research, putting SAMSI really at the forefront of Sequential Monte Carlo research.

9. Other Research while at SAMSI:

While at SAMSI I have been working on a couple of papers from my PhD thesis, which overlapped with some of the ideas of the 'Algebraic methods in biology' workshop. I have also been working on a project with collaborators at Duke University.

10. Other Comments:

SAMSI was a great experience also because it was a very diverse and lively social environment.

Megan Owen

1. Program Involvement:

I was very involved with the Evolutionary Biology working group. As well as being the webmaster, I actively participated in the working group meetings, including presenting my own work and leading a journal article discussion. Furthermore, I started three different collaborations with members of this working group. I was also one of the webmasters for the Systems Biology working group, and attended the meetings. Finally, I participated in all of the workshops for the Algebraic Methods in Systems Biology and Statistics program, including presenting a poster at the opening workshop and the Molecular Evolution workshop. I will be speaking at the transition workshop.

2. Interactions with Other Institutions:

I attended the weekly Algebra and Combinatorics seminar at NCSU. I have started a collaboration with an inter-disciplinary group at Duke university.

3. High Points at SAMSI:

I greatly enjoyed meeting and interacting with so many other researchers in my area. I had been worried about finding people to collaborate with during my postdoc, but this was not a problem.

4. Suggestions for Improvement:

The graduate student/postdoc seminar could be more productive, if it were run in a more organized manner (i.e. starting on time, imposing strict time limits on presentations and comments afterwards, etc).

5. Mentoring:

It was very helpful to have both a research mentor (Seth Sullivant) and a second, more experienced mentor (Pierre Gremaud) to ask more general questions.

6. SAMSI Benefits for the Future:

The greatest future benefit of having been a postdoc at SAMSI will be the connection I made with other participants in the program. Furthermore, having a year free from teaching responsibilities, and being able to focus solely on research has been extremely helpful.

7. SAMSI in contrast with University Setting:

The main disadvantage that SAMSI had in comparison to a university setting was the lack of computing resources. In particular, very little software was installed on the computers. However, as the SAMSI building is small, it was easy to meet and interact with the people there. This contrasts with a university setting where people with similar interests may be spread throughout the campus.

8. SAMSI in comparison to Other Experiences: N/A

9. Other Research while at SAMSI:

Besides converting my thesis into a journal paper, all of my research was connected with the Algebraic Methods in Systems Biology and Statistics program.

10. Other Comments:

The SAMSI staff has been exceptionally helpful during my time here.

Seaid Yasamin

1. Program Involvement: *Algebraic Methods in Systems Biology and Statistics*
2. Interactions with Other Institutions: Stanford University
3. High Points at SAMSI: Workshops
4. Suggestions for Improvement: Providing better mentoring and guidance for postdocs
5. Mentoring: The first year of my mentoring was very disappointing. My mentor did not give me a clear research project and for the most of the year I had to carry out the research with no reliance on my mentor. On the other hand, this year my mentoring has been extremely helpful and in fewer than four months I have been able to accomplish my first project.
6. SAMSI Benefits for the Future: Excellent research training
7. SAMSI in contrast with University Setting: More emphasis on research productivity. Less academic interaction.
8. SAMSI in comparison to Other Experiences: N/A
9. Other Research while at SAMSI:
 - 1) Maximum likelihood estimation on undirected graphical models
 - 2) Wishart-Type distribution on Bayesian networks
10. Other Comments:

C. Graduate Student Participation

1. Sequential Monte Carlo Methods

Chunlin Ji (Duke University)

(SAMSI RA) is attached to the Tracking (Godsill) working group and participates actively in the Big data group with West on spatial dynamic modeling for biological cell tracking problems. Ji is developing SMC methods in the context of new classes of models. This research has grown out of existing work of Ji & West in static problems, now extended with new dynamic models that will form an additional part of Ji's PhD thesis research, and one initial paper is in draft at the time of this report (see manuscripts section). Ji has led discussions on this work at several Tracking working group and Big data group meetings, gave a talk at the February 2009 mid-program workshop, and will present this work at the 7th Workshop on Bayesian Nonparametrics in Turin, Italy, in June 2009, and at the 2009 Joint Statistical Meetings in Washington DC 2009.

C. Ji & M. West (2009) Bayesian Nonparametric Modeling for Time-varying Spatial Point Processes (Initial draft completed).

C. Ji., S. Godsill, and M. West (2009) Spatial dynamic mixture modeling for multiple extended target tracking (In preparation).

Sarah Schott (Duke University)

Since our working group (Theory - Huber) lacks a postdoc, Sarah has been organizing the meetings and keeping our web page up to date. On the research side, she has been working on the product estimator problem described above, beginning with simulation studies and currently working to extend large deviations inequalities for binomials from sums to products.

Initially I introduced the product estimator as a side algorithm, a participant in the working group asked the question about the tightness of the constant. This raised an interesting point, and as Sarah and I have studied the problem further has proven far more deep a question than at first realized. In addition to this research avenue, I have learned much about SMC methodology over the course of the program, and still hope to utilize some of these methods in improving perfect simulation algorithms (the focus of my research program.)

Gareth Peters (UNSW Australia).

Gareth has participated actively in the tracking and population MC working groups. He has developed new ABC SMC methods and is also working on α -stable models with SMC.

Viktor Rozgik (University of Southern California)

I have found information about SAMSI opening workshop online and since I have been using Sequential Monte Carlo (SMC) methods in my work before I decided to attend it. I found the talks very interesting and I got involved in work of the Tracking Workgroup

which was the good match for the topic of my thesis "Multimodal fusion for tracking and identification in Smart Environments".

Collaboration with the group members, talks I have heard in the workshops and weekly group meetings; and references and papers in progress shared over the group's webpage were very helpful for my thesis work. Besides getting a much better perspective on the state-of-the-art Sequential Monte Carlo Algorithms and understanding the current research directions I had a hands-on- experience in implementation and testing of the SMC algorithms on the synthetic multi-target tracking problem. For this opportunity I feel very grateful to Prof. Simon Godsill and Dr. Francois Septier. I have managed to transplant and adapt part of this work to problems of audio-visual tracking, speaker segmentation and identification in meeting scenarios. Proposed work for the final part of my thesis includes work on multitarget tracking algorithms which is not focused only on the Meeting Monitoring environments and I hope that I am going to be able to continue collaboration with people I have meet during the workshop in the following period.

Papers Submitted:

Multimodal Speaker Segmentation and Identification in Presence of Overlapped Speech Segments, Journal of Multimedia

Papers In Preparation:

Audio-visual tracking and Speaker Diarization for Unknown Number of Meeting Participants,
to be submitted to IEEE Trans. on Multimedia

Ana Corberan (University of Valencia)

Ana has participated in the adaptive design sub-group of the MAaD working group.

Melanie Bain (UNC)

Nilay Argon and Melanie Bain are currently using SMC methods in solving a dynamic control problem that arises in the aftermath of mass-casualty incidents. To be more specific, we consider a mass-casualty event (such as a plane crash or a terrorist bombing) that resulted in several casualties in need of care. Due to the massive number of casualties, the medical resources are overwhelmed and decision makers need to prioritize patients for service. Depending on their injuries, the patients could be in different stages of health. The stage that a patient is in may affect his/her probability of survival and also service requirement. The decision maker cannot observe the true states of patients but can observe certain signals that the patients send (for example, pulse, breathing rate, etc.). Based on these signals, the decision maker decides which patient should be taken into service dynamically with the objective of maximizing the total expected number of survivors. We initially assume that the decision maker knows how the signals and the true states of patients relate. We also assume that the patients conditions degrade according to a discrete time Markov chain with a known transition probability matrix. We first formulated the above problem as a partially observable Markov decision process (POMDP). The POMDP we obtained could have a very large belief state depending on

the number of patients involved and also the number of health stages that we define. Hence, we will need to use an approximate method to solve this problem. We have thus far considered two approaches from the literature. One is by Thrun (2000), where particle filtering is used to reduce the size of the belief space, and the other is by Luo, Fu, and Marcus (2008), which is based on projecting the high-dimensional belief space to a low-dimensional family of parametrized distributions. We are currently implementing Thrun's approach.

Chiranjit Mukherjee (Duke University)

Participates in the Big Data and Distributed Computing group (though is not officially supported by the program). Mukherjee has developed studies of SMC methods for model fitting and comparison in nonlinear dynamical models arising from systems biology (and other applications). These studies involve very long time series but for which most of the underlying states are unobserved, and his work has explored, evaluated and developed novel approaches to SMC using distributed computation. In March 2009, Mukherjee presented and passed his PhD preliminary exam based on this work, and is now defining his thesis topic in this area. He has led several discussions on the topic at the Big Data and Distributed Computing meetings, presented a poster at the February 2009 mid-program workshop, and will present this work in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009.

Francesca Petralia (Duke University)

(SAMSI RA) is attached to the Particle Learning (Lopes) working group but also participates actively in the Big Data and Distributed Computing group. Petralia is (in March 2009) taking an active role in emerging discussions about computer model-SMC studies driven by motivating applications in environmental CO studies - problems that involve very large data sets and will require intense distributed computation - and will begin to work on this project with West in late spring 2009 linked to the Big Data and Distributed Computing working group. Petralia will present a talk on her work with SMC in econometric models in the Particle Learning working group (Lopes, leader) in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009.

Minghui Shi (Duke University)

(50% SAMSI RA) is working on sequential model search methodology for large, discrete model spaces, typified by "large p " regression model uncertainty. With Dunson, Shi is developing novel extensions of shotgun stochastic search that incorporate new ideas from SMC. Shi will present her PhD preliminary exam on this topic in April 2009, and the topic seems likely to then define her thesis area. Shi has led discussions on this work at Big Data and Distributed Computing meetings, presented a poster at the February 2009 mid-program workshop, and will present this work in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009.

Hao Wang (Duke University)

Participates in the Big Data and Distributed Computing group as well as other working groups, (though is not officially supported by the program). Wang is working, in part, on SMC methods for dynamic graphical models with Carvalho and West, has led discussions on the topic at the Big Data and Distributed Computing meetings, presented a

poster at the February 2009 mid-program workshop, and will present this work at the SAMSI workshop on *Adaptive Design, Sequential Monte Carlo and Computer Modeling* in April 2009 as well as in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington, DC 2009.

2. Algebraic Methods in Systems Biology and Statistics

Wenjie Chen (UNC)

Part of the Algebraic Statistics and Experimental Design working group. She studied using discrete dynamical systems and algebraic techniques for fMRI data.

Julia Chifman (University of Kentucky)

Visited SAMSI in November and April, and participated in the evolutionary biology working group. She has worked on problems about phylogenetic invariants for group-based models.

Deidra Coleman (NCSU)

Part of the Algebraic Statistics and Experimental Design working group.

Thomas Friedrich (Free University Berlin)

Visited in SAMSI from October to January. When he visited at SAMSI, he started working on developing phylogenomic tools to characterize ancestral gene pools (species), modeling most-recent common ancestor species (MRCAS) as clusters of associated gene lineages, with related but not identical gene tree topologies using unsupervised kernel-based clustering as well as developing novel statistical methods for Point-Cloud Data Analysis to determine whether sets of gene sequences exhibit co divergence, with Ruriko Yoshida. He will move to University of Kentucky from Mid-April to continue working on kernel methods on gene trees as a graduate student in the department of statistics under direction of Dr. Yoshida.

Benjamin Wells (NCSU)

Part of the Network Inference/ Structure from Dynamics working group.

Jason Yellick (NCSU)

Part of the Evolutionary Biology working group. He also participated in the SAMSI undergraduate workshop. He lead a working group discussion on ancestral recombination graphs.

D. Consulted Individuals

The individuals consulted for the broad selection of topics within programs and workshops were the members of two groups:

- The **Program Organizers**, listed in Section I.A.1
- Members of the **Advisory Committees**, listed in Section I.J

The specific topics that Program Working Groups chose to pursue were, in general, selected by the Working Group participants themselves, according to their combined interests. In almost all cases, however, a Program Leader headed each working group, so that specific research topics remained consistent with overall program goals. In Section II.E, the various Working Groups, and their members, are discussed.

E. Program Activities

1 Algebraic Methods in Systems Biology and Statistics

1.1 Program Overview

In recent years, methods from algebra, algebraic geometry, and discrete mathematics have found new and unexpected applications in systems biology as well as in statistics, leading to the emerging new fields of “algebraic biology” and “algebraic statistics.” Furthermore, there are emerging applications of algebraic statistics to problems in biology. This year-long program provided a focus for the further development and maturation of these two areas of research as well as their interconnections. The unifying theme is provided by the common mathematical tool set as well as the increasingly close interaction between biology and statistics. The program will allow researchers working in algebra, algebraic geometry, discrete mathematics, and mathematical logic to interact with statisticians and biologists and make fundamental advances in the development and application of algebraic methods to systems biology and statistics. The essential involvement of biologists and statisticians in the program provided the applied focus and a sounding board for theoretical research.

1.1.1 Research Foci

Systems Biology: The development of revolutionary new technologies for high-throughput data generation in molecular biology in the last decades has made it possible for the first time to obtain a system-level view of the molecular networks that govern cellular and organismal function. Whole genome sequencing is now commonplace, gene transcription can be observed at the system level and large-scale protein and metabolite measurements are maturing into a quantitative methodology. The field of systems biology has evolved to take advantage of this new type of data for the construction of large-scale mathematical models. System-level approaches to biochemical network analysis and modeling promise to have a major impact on biomedicine, in particular drug discovery.

Statistics: It has long been recognized that the geometry of the parameter spaces of statistical models determines in fundamental ways the behavior of procedures for statistical inference. This connection has in particular been the object of study in the field of information geometry, where differential geometric techniques are applied to obtain an improved understanding of inference procedures in smooth models. Many statistical models, however, have parameter spaces that are not smooth but have singularities. Typical examples include

hidden variables models such as the phylogenetic tree models and the hidden Markov models that are ubiquitous in the analysis of biological data. Algebraic geometry provides the necessary mathematical tools to study non-smooth models and is likely to be an influential ingredient in a general statistical theory for non-smooth models.

Algebraic methods:

Algebraic biology is emerging as a new approach to modeling and analysis of biological systems using tools from algebra, algebraic geometry, discrete mathematics, and mathematical logic. Application areas cover a wide range of molecular biology, from the analysis of DNA and protein sequence data to the study of secondary RNA structures, assembly of viruses, modeling of cellular biochemical networks, and algebraic model checking for metabolic networks, to name a few.

Algebraic statistics is a new field, less than a decade old, whose precise scope is still emerging. The term itself was coined by Giovanni Pistone, Eva Riccomagno and Henry Wynn. That book explains how polynomial algebra arises in problems from experimental design and discrete probability, and it demonstrates how computational algebra techniques can be applied to statistics. The first of these applications have focused on categorical data and include the study of Markov bases and conditional inference, disclosure limitation, and parametric inference, to name a few.

The central idea underlying algebraic statistics is that the parameter spaces of many statistical models are (semi-)algebraic sets. The geometry of such possibly non-smooth sets can be studied using tools from algebraic geometry. Many problems in computational biology can be described within this framework. This is where algebraic statistics joins algebraic biology as a new methodology for solving problems in systems biology.

The unifying theme of the program is the development and use of a particular set of tools from algebra, algebraic geometry, and discrete mathematics to solve problems in statistics and biology.

1.1.2 Organization and Program Leadership

Organizing Committee: Peter Beerli (School of Computational Sciences and Department of Biological Sciences, Florida State University), Andreas Dress (Director, CAS-MPG Partner Institute for Computational Biology, Shanghai), Mathias Drton (Department of Statistics, University of Chicago), Ina Hoeschele (Department of Statistics, Virginia Tech, and Virginia Bioinformatics Institute), Christine Heitsch (School of Mathematics, Georgia

Tech), Serkan Hosten (Department of Mathematics, San Francisco State University), Reinhard Laubenbacher, Committee Chair (Department of Mathematics, Virginia Tech, and Virginia Bioinformatics Institute), Bud Mishra (Departments of Computer Science, Mathematics, and Cell Biology, Courant Institute, NYU), Don Richards (Department of Statistics, Pennsylvania State University), Seth Sullivant (Department of Mathematics, NCSU), Brett Tyler (Department of Plant Pathology and Weed Science, Virginia Tech, and Virginia Bioinformatics Institute), Ruriko Yoshida (Department of Statistics, University of Kentucky).

1.1.3 Major Participants

Long-Term Visitors: Edward Allen (Wake Forest University), Elizabeth Allman (University of Alaska), James Degnan (University of Canterbury), Alicia Dickenstein (University of Buenos Aires), Luis Garcia-Puente (Sam Houston State University), Jeremy Gunawardena (Harvard Medical School), Chris Hillar (MSRI), Serkan Hosten (San Francisco State University), Reinhard Laubenbacher (VA Tech), Catherine Mathias (Universite d'Evry), Uwe Nagel (University of Kentucky), Edwin O'Shea (UNAM, Mexico), Giovanni Pistone (Politecnico di Torino), John Rhodes (University of Alaska), Eva Riccomagno (University of Genoa), Anne Shiu (UC Berkeley)

Postdoctoral Fellows: Megan Owen (Cornell University), Ahmad Saeid Yasamin (Indiana University)

Graduate Students: Wenjie Chen (UNC), Julia Chifman (University of Kentucky), Deidra Coleman (NCSU), Thomas Friedrich (FU- Berlin), Benjamin Wells (NCSU), Jason Yellick (NCSU)

Faculty Releases: Ian Dinwoodie (Duke), Scott Provan (UNC), Eric Stone (NCSU), Seth Sullivant (NCSU), Yung-Jing Tzeng (NCSU)

1.2 Description of Activities

1.2.1 Workshops

The SAMSI program on Algebraic Methods in Systems Biology and Statistics has been bolstered by a number of workshops and special sessions held throughout the year both at SAMSI and nearby locations.

Opening Workshop: The Kickoff Workshop and Tutorial will be September 14–17, 2008. The principal goal of the workshop was to engage a broadly representative segment of the

mathematical, statistical, and life sciences communities to determine research directions to be pursued by working groups during the program. Four working groups were formed that eventually merged down to three working groups.

The workshop covered a very broad range of topics in the interactions of algebraic methods with systems biology and statistics. After introductory tutorials on Sunday, more focused talks directed towards the problem areas to be highlighted throughout the year were given. Highlighted topics included: algebraic statistical models, combinatorics of biological molecules, automata theory and finite dynamical systems, phylogenetics, causal models, and random graph models.

The workshop also contained a number of discussion sessions with the goal of identifying research areas that would be highlighted throughout the yearlong program. After these intended working group target areas were identified, the program members formed breakout sessions to begin discussions of topics that would be discussed in the working groups throughout the year. The particular areas of the working groups are described below.

The tutorial speakers were: Bernd Sturmfels, Reinhard Laubenbacher, and Elizabeth Allman. The keynote speakers were: Mathias Drton, Jeremy Gunawardena, Christine Heitsch, Bud Mishra, Abdul Jarrah, Chris Schardl, Michael Savageau, Gheorghe Craciun, Sumio Watanabe, Brandilyn Stigler, Meera Sitharam, Lior Pachter, Brett Tyler, Niko Beerenwinkel, Eva Riccomagno, and Steve Fienberg.

Discrete Models in Systems Biology Workshop: The discrete models workshop was held December 3-5, 2008 and was organized by Elena Dimitrova (Clemson University), Ilya Shmulevich (Institute for Systems Biology), and Brandilyn Stigler (Southern Methodist University). The workshop focused on the use of discrete models in systems biology. Discrete modeling approaches have been applied to a wide variety of biological contexts, including gene regulatory networks, epidemiology, and ecosystem dynamics. Examples of topics of interest in the workshop were

1. discrete dynamical systems: multi-state models such as Boolean networks, logical models, and finite dynamical systems; random networks; and analytic tools including statistical-mechanical approaches
2. Bayesian networks, including dynamic Bayesian networks and graphical models
3. static networks: interaction networks and graph-theoretic approaches
4. simulation: finite-state machines, agent-based networks, process algebras

A chief goal was to stimulate the organization of working groups focused on addressing key challenges in discrete modeling in computational systems biology, particularly the establishment of unifying themes and principles.

As part of our commitment to foster a synergistic community, we organized three question-and-discussion sessions to encourage interactions among workshop participants and two poster sessions to showcase the work of junior researchers, including postdocs and graduate students.

Algebraic Statistical Models Workshop: Many classical statistical models, in particular Gaussian models from multivariate statistics and models for discrete random variables, exhibit algebraic structure in their parameter spaces. This workshop focused on both algebraic and statistical aspects of such algebraic statistical models. It is intended to complement other mid-program workshops, which focused more on particular application areas.

The workshop was held at SAMSI, January 15–17, 2009 and featured topics by working group participants, as well as outside experts, whose opinions helped to provide new research directions in the program. The organizers were Mathias Drton, Eva Riccomagno, and Seth Sullivant. Focus topics of the workshop included Markov bases, graphical models, algebraic tools for maximum likelihood estimation, identifiability problems, and cumulant methods.

The workshop included talks by Steffen Lauritzen, Thomas Richardson, Donald Richards, Ruriko Yoshida, Elizabeth Allman, Sonja Petrović, Akimichi Takemura, Serkan Hoşten, Hugo Maruri, and Jason Morton and a poster session.

Miniworkshop on Systems Biology: A subgroup of the systems biology working group was formed to focus on software development for parameter estimation for discrete models. As part of the subgroup activities a miniworkshop at SAMSI was conducted February 24–26, 2009. Participants included E. Dimitrova, L. Garcia, F. Hinkelmann, A. Jarrah, R. Laubenbacher, B. Stigler, and P. Vera-Licona. The workshop was focused on the design of the overall architecture of a software package for parameter estimation and simulation for Boolean network models. A significant part of the time was devoted to actual code development.

Molecular Evolution and Phylogenetics Workshop: Recently there has been a marked synergy between modern biology and higher mathematics. A number of important connections have been established between computational biology and the emerging field of “algebraic statistics,” which combines combinatorics, computational algebra, polyhedral geometry and statistical modeling. The primary objective of this workshop, held April 2–3,

2009, was to bring together new and established researchers in mathematics, biology, and statistics in order to discuss the crossover between algebraic statistics, molecular evolution and phylogenetics.

As part of our commitment to foster a synergistic community, we organized several discussion sessions to encourage interactions among workshop participants to actively begin new collaborations, discuss new research directions, and make new connections. For example, we discussed phylogenetic invariants on group-based models, such as Jukes-Cantor model and their applications to tree reconstruction as well as discussion on the phylogenetic mixture models. There were several discussions from coalescent theory, such as the coalescent approach to approximate the distribution of gene trees. Also we discussed inferences about the impact of phenotype on genotype from the ancestral lineage and also some reviews of phylogenetic reconstructions, what are known and unknown.

We invited four researchers to give a one-hour keynote address, and six researchers to give contributed (invited) talks of 45 minutes length (those include 5 minutes for questions at the end) at the workshop. Invited speakers were: Jeff Thorne, Cecile Ané, Junhyong Kim, Tandy Warnow, Seth Sullivant, Eric Stone, Fumei Lam, Laura Kubatko, Jeremy Sumner, Sonja Petrović, and Jesus Fernandez-Sanchez.

AMS Special Session: Mathematics of Biochemical Reaction Networks: The special session "Mathematics of Biochemical Reaction Networks" was held during the Southeastern Section meeting of the AMS at North Carolina State University during the weekend of April 4-5. The organizers were Gheorghe Craciun, Manoj Gopalkrishnan, and Anne Shiu. The idea to organize this workshop in proximity to SAMSI, and close to the evolutionary biology workshop came was formed during the SAMSI opening workshop.

Our intent was to bring together individuals who study biochemical reaction networks, in order to share ideas that range from those building upon the classical Feinberg-Horn-Jackson deficiency theory, to those more recent algebraic techniques that highlight the rich algebraic structure inherent in these networks. For example, much work has focused on predicting dynamics and resolving questions of stability simply from the topological structure of the underlying reaction network. Another topic that was covered was the class of "monotone" systems. In particular, the session featured reports on collaborations that grew out of activities this academic year at SAMSI in North Carolina and the MBI in Ohio.

Special 45-minute talks were given by Martin Feinberg, Jeremy Gunawardena, and Eduardo Sontag. Several talks in particular have a strong algebraic aspect. For example, Greg Rempala talked about an algebraic statistical model for inferring biochemical reaction networks. Ezra Miller discussed results on binomial primary decomposition and its connection

to boundary steady states of a chemical reaction system. Luis Garcia connected Birch's Theorem and chemical reaction network theory to Bziers patches and their generalizations. Alicia Dickenstein shared results that compare detailed balancing to complex balancing in terms of their associated algebraic varieties. In addition, there were speakers from various backgrounds ranging from theoretical computer science to control theory to probability, and talks whose titles included the words "homotopy" and "number theory."

Transition Workshop: The transition workshop, held June 18-20, 2009, was designed to synthesize the year's activities and provide a blueprint to go forward with research in this field. Talks covered a range of topics related to the working group topics. Some talks discussed how to move forward beyond the program year, some talks highlighted successes from the working groups, and some talks were chosen to increase the range of topics present in the program.

Our speakers were Elena Dimitrova, Luis David Garcia-Puente, Gilles Gnacadja, David Haws, Peter Huggins, Paul Kidwell, Reinhard Laubenbacher, Olgica Milenkovic, Betti Numbers, Megan Owen, Anne Shiu, Heike Siebert, Katherine St. John, Seth Sullivant, Marcy Uyenoyama, Henry Wynn, Saeid Yasamin, and Ruriko Yoshida.

Joint Statistics Meetings Session on Algebraic Methods in Systems Biology and Statistics: A session at the 2009 Joint Statistical Meetings was organized by Ian Dinwoodie, in August 2009, to highlight some of the results that emerged from the SAMSI program. Presentations were given on "Algebraic Methods in Statistics" by Ahmad S. Yasamin, "Conditional Independence Models via Filtrations" by Simon Lunagomez, "Trek Separation for Gaussian Graphical Models" by Seth Sullivant, and "Design of Experiments and Inference of Biochemical Networks" by Reinhard Laubenbacher.

1.3 Working Groups

At the end of the opening workshop in September, the afternoon was devoted to the formation of working groups for the year. Based on participant interest and program themes, the topics that emerged were:

1. Systems biology: The relationship between the structure and dynamics of biological networks;
2. Algebraic statistics and experimental design;
3. Evolutionary biology and phylogenetics.

There was significant overlap between topics and membership of the different working groups. For instance, experimental design is a very important topic in systems biology as well.

1.3.1 Relationship between structure and dynamics of biological networks:

The working group leaders are R. Laubenbacher (VT) and Brandilyn Stigler (SMU).

One of the dominant themes at the opening workshop was the relationship between the structure of biological networks and the kind of dynamics this structure supports. The questions about this relationship is appropriate for a variety of biological networks, ranging from molecular pathways to social networks that support the spread of epidemics. For the working group, the focus was entirely on biochemical networks, encompassing two different modeling frameworks: polynomial dynamical systems over finite fields, in particular Boolean networks, and systems of polynomial differential equations. The primary structure of a network is given by a directed graph that indicates the dependence of the network variables on each other. In both modeling frameworks, one of the goals is to infer constraints on the dynamics of the network from properties of this graph. In the other direction, the goal is to infer the structure of the graph from a partial specification of the network dynamics, e.g., through a collection of time course experiments.

Group Activities: Since the background of the working group members varies considerably, a primary focus of the group activities in the fall and part of the spring were on presentations and discussions aimed at establishing a common background in systems biology and the different approaches to modeling and simulation. The group is now at a stage where first results are being presented, beginning with the work of the subgroup on software development, described below.

A major problem with the construction of large-scale algebraic models is that there are no sophisticated tools available, comparable to ODE tools. Most importantly, the tool of fitting ODE model parameters to available data is key in continuous model construction, but completely absent for algebraic models. Also, tools like bifurcation analysis, sensitivity analysis, stability analysis are all unavailable to the algebraic modeler. However, there exist several or all of these tools in the polynomial dynamical systems framework. The goal of this subgroup of the working group is to collect together available software and integrate it in a coherent package scheduled for release in April. Two publications are scheduled for submission in April and May.

Active participants: Carsten Conradi (MPI Magdeburg), Alicia Dickenstein (University of Buenos Aires), Elena Dimitrova (Clemson University), Ian Dinwoodie (Duke University), Lee Falin (Virginia Tech), Thomas Friedrich (TU Berlin), Gilles Gnacadja (Amgen), Richard Haney (Cellular Statistics), Franziska Hinkelmann (Virginia Tech), Serkan Hosten (San Francisco State University), Abdul Salam Jarrah (Virginia Tech), Reinhard Laubenbacher (Virginia Tech), Tong Lee (Virginia Tech), Shaowei Lin (Univ. of California-Berkeley), Megan Owen (SAMSI), Mercedes Soledad Perez Millan (University of Buenos Aires), Anne Shiu (Univ. of California-Berkeley), Heike Siebert (FU Berlin), Brandy Stigler (Southern Methodist University), Seth Sullivant (NCSU), Jung-Ying Tzeng (N.C. State University), Alan Veliz-Cuba (Virginia Tech), Benjamin Wells (N.C. State University), Henry Wynn (London School of Economics), Richard Yamada (University of Michigan), Shantia Yarahmadian (Indiana University), Saeid Yasamin (SAMSI),

1.3.2 Algebraic Statistics and Experimental Design:

The Algebraic Statistics and Experimental Design (ASED) working group in the 2008-2009 SAMSI program Algebraic Methods in Systems Biology and Statistics has approximately 30 members, including many remote participants in England, Italy, Japan, and throughout the U.S. The group was formed during the opening workshop in September, when all members were present and decided upon research themes and meeting times. The working group leaders are Serkan Hosten of San Francisco State University and Ian Dinwoodie of Duke University.

The ASED group works on a range of statistical applications that use computational tools of commutative algebra. These applications include sampling and Monte Carlo methods for discrete data (tables and sequences), experiments (data gathering) and data analysis for reverse engineering of biological networks, disclosure limitation, and foundations of phylogenetic trees in evolutionary and population biology. Each application area emphasizes certain algebraic tools and each has roots in particular research groups with a wide international base. The different mathematical tools and research groups are well-represented in the group members. The working groups benefited from the active and steady participation of individuals with much depth and experience. In particular, Henry Wynn and Giovanni Pistone shepherded the experimental design part of ASED, and worked on collaborations with the Systems Biology working group, where design issues are important for finding wiring diagrams and network connections. Phylogenetic trees were supported by John Rhodes, Elizabeth Allman, and Seth Sullivant, who completed foundational work on identifiable tree models in the course of the year, and shared their work from the Evolutionary Biology work-

ing group. Ongoing work in high-dimensional tables (sampling and disclosure limitation) was well-represented as well, with presentations by researchers and practitioners Akimichi Takemura, Larry Cox, Edwin O’Shea, Adrian Dobra, and others. Also some connections were made with the Sequential Importance Sampling program through research on sequential Monte Carlo methods for statistical inference on Boolean dynamics in biological networks.

The ASED working group met formally on Mondays at noon, in addition to informal collaborations. About half the participants logged-in from remote locations using the Webex networking application. The list of talks and speakers is at www.samsi.info under the ASED working group link, together with supporting materials and documents.

The complete list of ASED working group members is below:

Elizabeth Allman (University of Alaska-Fairbanks), Deidra Coleman (N.C. State University), Lawrence H. Cox (CDC), Elena Dimitrova (Clemson University), Ian Dinwoodie (Duke University), Luis David Garcia-Puente (Sam Houston State University), Hisayuki Hara (Tokyo), Serkan Hosten (San Francisco State University), Thomas Kahle (Leipzig), Imre Risi Kondor (University College London), Reinhard Laubenbacher (Virginia Bioinformatics Institute), Tong Lee (Virginia Tech), Hugo Maruri-Aguilar (London School of Economics), Catherine Matias (CNRS), Uwe Nagel (University of Kentucky), Edwin O’Shea (Avanzados del IPN), Vittorio Perduca, Mercedes Soledad Perez Millan (Universidad de Buenos Aires), Sonja Petrovic (University of Illinois at Chicago), Giovanni Pistone (Torino), Eva Riccomagno (Genoa), Seth Sullivant (NCSU), Akimichi Takemura (Tokyo), Caroline Uhler (Univ. of California-Berkeley), Alan Veliz-Cuba (Virginia Tech), Benjamin Wells (N.C. State University), Henry Wynn (London School of Economics), Richard Yamada (University of Michigan), Ahmad S. Yasamin (SAMSI), Jason Yellick (N.C. State University), Ryo Yoshida (Japan), Yi Ming Zou (University of Wisconsin-Milwaukee), Or Zuk (M.I.T.), Piotr Zwiernik (University of Warwick)

1.3.3 Evolutionary Biology and Phylogenetics:

As part of SAMSI’s 2008-09 program on Algebraic Methods in Systems Biology and Statistics a working group in ‘Evolutionary Biology’ was formed during the opening workshop in September. Broadly speaking, members of this group are interested in finding, understanding, and solving problems arising in evolutionary biology that might require sophisticated mathematical and statistical techniques that have yet to be developed.

The group is lead by Seth Sullivant, Elizabeth Allman, and John Rhodes, and during the opening workshop interested participants indicated that primary areas of common interest included phylogenetics, coalescent theory, population genetics, and comparative genomics.

Working Group Activities: It was immediately clear that group members had widely diverse backgrounds in statistics, mathematics, and biology, and that participants needed a 'common language' and 'common background knowledge' in order to collaborate. During the first semester and spilling over into the beginning of the second semester, the working group met weekly. Each week a particular group member with expertise in one of the areas of common interest, gave a talk at an introductory level to familiarize other group members with the area, discuss his/her research, and suggest possible problems where algebraic techniques might yield results. Typically, after an hour or more of introduction by the speaker, group members discussed the problems and a question-and-answer period began.

The main topics included: the structure of tree space for phylogenetic trees (2 sessions), mixture models in phylogenetics and invariants (3 sessions), the coalescent model (4 sessions), geometry of cophylogeny (1 session), and comparative genomics (1 session).

- 9/23/08 Megan Owen on the space of phylogenetic trees and the geodesic distance:
Space of Phylogenetic Trees
- 9/30/08 John Rhodes on phylogenetic invariants
- 10/07/08 Seth Sullivant: Some algebraic ideas for phylogenetic mixtures
- 10/14/08 Peter Beerli (Part 1 of introduction to coalescent theory series):
Population genetic calculations that do not fit on the back of an envelope
- 10/21/08 Laura Salter Kubatko (Part 2 of introduction to coalescent theory series):
- 10/28/08 Peter Beerli (Part 3 of the introduction to coalescent theory series):
Finding good trees - Simplifying Coalescent trees
- 11/04/08 Serkan Hosten: Extended UPGMA and phylogenetic tree reconstruction
- 11/11/08 Rudy Yoshida: Open Problems in Geometry of Cophylogeny
- 11/18/08 Julia Chifman: Group-based models
- 01/19/09 James Degnan: Gene tree distributions and coalescent histories
- 01/26/09 Or Zuk: Annotating the Human Genome Using Comparative Genomics

For a particularly successful ending to the fall semester, the evolutionary biology meeting consisted of a session in which individuals suggested open problems for the group to work on.

After an organizational meeting in mid-January, the working group decided to focus primarily on reading papers to acquire a deeper understanding of tree space, gene-tree/species-tree problems (coalescent theory), models of speciation, and ancestral recombination graphs. Several talks were also scheduled during the semester while researchers were visiting at SAMSI for collaborations and workshops.

Weekly working group meetings ran differently this term. The idea was to have all group members read the papers for the week, and one person was assigned to lead a discussion. It

was assumed that no one was an expert in the the area, so that the group could learn by reading up on a topic together. This has worked reasonably well, but the best discussions took place when there were more group members physically present at SAMSI.

Group Membership: The official number of group members is quite high, around 35, though the number of active participants is closer to 20. The number of participants on a weekly basis (“the faithful”) was typically about eight to ten. The meetings on the coalescent model were particularly well-attended.

The names of the active participants have been included below:

Elizabeth Allman (University of Alaska-Fairbanks), Elisaveta Arnaudova (University of Kentucky), Peter Beerli (Florida State University), Julia Chifman (University of Kentucky), Luis Garcia-Puente (Sam Houston State University), Serkan Hosten (San Francisco State University), Laura Kubatko (Ohio State University), Jinze Liu (University of Kentucky), Catherine Matias (CNRS, Laboratoire Statistique et Genome), Uwe Nagel (University of Kentucky), Megan Owen (SAMSI), Sonja Petrovic (University of Illinois at Chicago) Scott Provan (University of North Carolina), John Rhodes (University of Alaska-Fairbanks), Chris Schardl (University of Kentucky), Seth Sullivant (N.C. State University), Amelia Taylor (Colorado College), Jason Yellick (N.C. State University), Ruriko Yoshida (University of Kentucky), Or Zuk (M.I.T. Broad Institute) Piotr Zwiernik (University of Warwick)

Several new collaborations were formed among evolutionary biology group members, and the working group format gave the opportunity for extended research interaction to both these new and pre-existing collaborations.

1.3.4 University Courses

Title: Algebraic Methods in Systems Biology and Statistics

Instructors: Seth Sullivant (NCSU) and Reinhard Laubenbacher (VA Tech)

Course Day and Time: Tuesday 4:30-7:00

Course Description: This course will provide an introduction to the algebraic techniques that have emerged as useful tools in biology and statistics. This course is intended to bridge the gap between abstract algebra and the application areas covered in the year-long program.

After providing an introduction to polynomial rings, ideals, and Grobner bases, we will survey a range of applications of these ideas. Possible topics include: Polynomial dynamical systems over finite fields and applications, graphical and hierarchical models, Markov bases for contingency table analysis, phylogenetic models and the space of trees, applications of tropical geometry in MAP estimation.

2 Sequential Monte Carlo Methods

2.1 Program and its Objectives:

This aim of this 12 month SAMSI program was to develop new approaches to scientific/statistical computing using innovative sequential Monte Carlo (SMC) methods. The program addressed fundamental challenges in developing effective sequential and adaptive simulation methods for computations underlying inference and decision analysis. The research blended conceptual innovation in new and emerging methods with evaluation in substantial applied contexts drawn from areas such as control, communications and robotics engineering, financial and macro-economics, among others. Researchers from statistics, computer science, information engineering and applied mathematics were involved, and the program promoted the opportunity for both methodological and theoretical research. The interdisciplinary aspects of the program were substantial.

2.2 Background

Monte Carlo (MC) methods are central to modern numerical modelling and computation in complex systems. Markov chain Monte Carlo (MCMC) methods provide enormous scope for realistic statistical modelling and have attracted much attention from disciplinary scientists as well as research statisticians. Many scientific problems are not, however, naturally posed in a form accessible to evaluation via MCMC, and many are inaccessible to such methods in any practical sense. For example, for real-time, fast data processing problems that inherently involve sequential analysis, MCMC methods are often not obviously appropriate at all due to their inherent "batch" nature. The recent emergence of sequential MC concepts and techniques has led to a swift uptake of basic forms of sequential methods across several areas, including communications engineering and signal processing, robotics, computer vision and financial time series. This adoption by practitioners reflects the need for new methods and the early successes and attractiveness of SMC methods. In such, probability distributions of interest are approximated by large clouds of random samples that evolve as data is processed using a combination of sequential importance sampling and resampling ideas. Variants of particle filtering, sequential importance sampling, sequential and adaptive Metropolis MC and stochastic search, and others have emerged and are becoming popular for solving variants of "filtering" problems; i.e. sequentially revising sequences of probability distributions for complex state-space models. Useful entree material and examples SMC methods can be found at the following SMC preprint site. Many problems and existing simulation methods can be formulated for analysis via SMC: sequential and batch Bayesian

inference, computation of p-values, inference in contingency tables, rare event probabilities, optimization, counting the number of objects with a certain property for combinatorial structures, computation of eigenvalues and eigenmeasures of positive operators, PDE's admitting a Feynman-Kac representation and so on. This research area is poised to explode, as witnessed by this major growth in adoption of the methods.

The SAMSI SMC program focused on:

- Addressing methodological and theoretical problems of SMC methods, including synthesis of concepts underlying variants of SMC that have proven apparently successful across multiple fields, and the development of methodological and theoretical advances.
- Developing the methodological research – with broad opportunities for test-bed examples, methods evaluation and refinement of generic approaches – in the contexts of a number of important applied problems (e.g. data assimilation, inference for large state spaces, finance, tracking, continuous time models).

The program was an opportunity for exchange between communities, helping to shape the future of stochastic computation and sequential methods. It involved statisticians, computer scientists and engineers as core participants as well as others working collaboratively in a range of applied fields.

2.3 Core Group

A core group of researchers have been based at SAMSI, complemented by external participants in the various working groups, which hold weekly meetings via Webex connections to SAMSI.

Local faculty: Mark Huber (Duke), Mike West (Duke), Nilay Argon

Senior researchers (at SAMSI for significant periods of time in Fall 2009): Susie Bayarri (University Valencia), Jaya Bishwal (University North Carolina Charlotte), Carlos Carvalho (University of Chicago), Arnaud Doucet (University British Columbia), Edsel Pena (University South Carolina), Fei Liu (University Missouri), Marco Ferrante (University Pavia), Nathan Green (DSTL), Hedibert Lopes (University of Chicago), Raquel Prado (University Santa Cruz), Sylvain Rubenthaler (University Nice), Yoshida Ryo (Institute Statistical Mathematics), Jochen Voss (University Warwick).

Researchers (Spring and Summer, 2009): Daniel Clark (University Herriott Watt), Mark Coates (McGill University), Paul Fearnhead (University of Lancaster), Andrew Thomas

(University St Andrews), James Lynch (University South Carolina), Ernest Fokoue (Kettering University)

Postdoctoral fellows and associates: Artin Armagan, Julien Cornebise, Sourish Das, Christian Macaro, Ioanna Manolopoulou, Elizabeth Shamseldin, Gentry White

Graduate students: Melanie Bain (Duke), Luke Bornn (University British Columbia), Deidra Coleman (NCSU), Ana Corberan (University of Valencia), Thomas Flury (Oxford University), Roman Holenstein (University British Columbia), Chunlin Ji (Duke), Olasunkanmi Obanubi (Imperial College), Gareth Peters (University New South Wales), Francesca Petralia (Duke), Sarah Schott (Duke), Minghui Shi (Duke), Baqun Zhang (NCSU)

2.4 Program Organization

2.4.1 Opening workshop

The Opening Workshop was held during September 7-10, 2008 at SAMSI, organized by Arnaud Doucet (British Columbia), Simon Godsill (Cambridge) and Mike West (Duke University). This highly successful event engaged significant parts of the statistical, engineering and mathematics community, as well as others in econometrics and sciences, and included themed sessions from all of the main program topics (working groups).

Tutorial talks were given by four world leaders in the various areas of SMC: Pierre del Moral (Bordeaux), Paul Fearnhead (Lancaster), Hedibert Lopes (Chicago) and Jun Liu (Harvard). These were pitched at various levels, allowing useful participation by attendees starting in the area as much as those already expert in one or more topics. Themed conference sessions were arranged to have a good balance between senior invited talks, new researcher talks and panel discussion. Most sessions stimulated very active discussion.

At a break-out session on the final afternoon, Working Group leaders were allocated and a broad declaration of interest was obtained from all workshop participants for their subsequent participation in the program.

2.4.2 Undergraduate workshop

The SAMSI Two-Day Undergraduate Workshop was held from October 31 - November 1, 2008. There were nine technical talks given by Jaya Bishwal, Jochen Voss, Gentry White, Nathan Green, Christian Macaro, Sourish Das, Julien Cornebise, Ioana Maolopoulou and

Francesca Petralia, covering many aspects of SMC from basic methodology to applications in finance and defence. There was also an interactive R session.

2.4.3 Fall SAMSI course on sequential Monte Carlo

This course provided an introduction to sequential Monte Carlo methodology, theory and applications. It was attended by approximately 40 people. Topics covered include: introduction to SMC, advanced SMC methods, SMC methods for parameter estimation in general state-space models, SMC methods as alternative to MCMC methods. The main instructor was Arnaud Doucet and two 'invited' instructors gave some lectures: Christophe Andrieu (Bristol) and Alexander Chorin (Berkeley).

2.4.4 Mid-term workshop

A mid-term workshop was organised on 19-20 Feb 2009 at the SAMSI Institute. This had participants from most of the working groups, including the leaders of the Continuous Time (Fearnhead - Lancaster), Tracking (Godsill - Cambridge), Big Data (West - Duke), Parameter Learning (Lopes - Chicago) and Model Assessment (Carvalho - Chicago) working groups. These leaders gave overviews of progress in the different working groups and other participants gave research updates on SAMSI related work. Three of the 15 talks were delivered successfully by Webex from remote locations. A particular focus of talks and discussion was the Tracking working group, which assembled many of its participants at the workshop.

2.4.5 Adaptive Design Workshop

An Adaptive Design, Computer Modeling and SMC workshop organized by Susie Bayarri and Mike West, was held from April 15-17, 2009. This was a joint workshop of the SMC program and the NISS project on Computer Models for Geophysical Risks (CMGR). The workshop involved SMC researchers working on models and methods for sequential decision and design problems, and researchers working on statistical analysis of computer model data with special focus on adaptive design. The workshop generated considerable discussion between these two communities to define new computational approaches for design in computer modeling as well as stimulating novel algorithmic research in SMC.

The speakers at the workshop were Nilay T. Argon (UNC-Chapel Hill), Derek Bingham (Simon Fraser University), Gardar Johannesson (LLNL), Herbert Lee (Univ. of California-Santa Cruz), Fei Liu (University of Missouri-Columbia), Hedibert Lopes (University of Chicago), Thomas Loredó (Cornell University), Ioanna Manolopoulou (SAMSI), William

Notz (Ohio State University), Steve Sain (NCAR), Hoa Wang (Duke University), and Brian Williams (LANL), and the discussion leaders were Nancy Flournoy (University of Missouri), Dave Higdon (LANL), Angela Patterson (General Electric), and Abel Rodriguez (Univ. of California-Santa Cruz).

2.4.6 Transition workshop

The transition workshop will be held at SAMSI from 9-10 Nov. 2009.

2.5 Working groups

The working groups met weekly throughout the program. The working groups were formed in the following areas:

- Tracking and Large-Scale Dynamical Systems
- Theory
- Population Monte Carlo
- Continuous Time
- Parameter Learning
- Model Assessment
- Big Data

2.5.1 Tracking and Large-scale Dynamical Systems Working Group

The working group leaders were Simon Godsill (whole year) and Nathan Green (Fall 08).

Regular participants included: Mark Briers (QinetiQ, UK), Daniel Clark (Heriot-Watt University UK), Avishi Carmi (Cambridge University UK), Julien Cornebise (SAMSI post-doc), Ernest Fokoue (Kettering University US), Simon Godsill (Cambridge University - program leader), Nathan Green (DSTL UK, long-term Fall 08 visitor) Chunlin Ji (Duke University - SAMSI Grad. student), Sze Kim Pang (Cambridge University UK), Gareth Peters (Univ. New South Wales, SAMSI Fall 08 visitor), Viktor Rozgic (University of Southern California), Francois Septier (Cambridge University, UK), Joshua Vogelstein (Johns Hopkins University), Gentry White (N.C. State University - SAMSI Post-doc), Namrata Vaswani (Iowa State University)

PhD Student: Viktor Rozgic, Advisor Prof. Shrikanth Narayanan, working in the Signal Analysis and Interpretation Laboratory, Viterbi School of Engineering, University of Southern California

Research and Impact to Thesis: I have found information about SAMSI opening workshop online and since I have been using Sequential Monte Carlo (SMC) methods in my work before I decided to attend it. I found the talks very interesting and I got involved in work of the Tracking Workgroup which was the good match for the topic of my thesis "Multimodal fusion for tracking and identification in Smart Environments". Collaboration with the group members, talks I have heard in the workshops and weekly group meetings; and references and papers in progress shared over the group's webpage were very helpful for my thesis work. Besides getting a much better perspective on the state-of-the-art Sequential Monte Carlo Algorithms and understanding the current research directions I had a hands-on-experience in implementation and testing of the SMC algorithms on the synthetic multi-target tracking problem. For this opportunity I feel very grateful to Prof. Simon Godsill and Dr. Francois Septier. I have managed to transplant and adapt part of this work to problems of audio-visual tracking, speaker segmentation and identification in meeting scenarios. Proposed work for the final part of my thesis includes work on multitarget tracking algorithms which is not focused only on the Meeting Monitoring environments and I hope that I am going to be able to continue collaboration with people I have met during the workshop in the following period.

Paper Submitted: Multimodal Speaker Segmentation and Identification in Presence of Overlapped Speech Segments, Journal of Multimedia

Paper In Preparation: Audio-visual tracking and Speaker Diarization for Unknown Number of Meeting Participants, to be submitted to IEEE Trans. on Multimedia

Working group organization: This working group is focusing on methodology for problems in high-dimensional tracking, with applications in computer vision, tracking, meteorology, biological imaging, etc. Standard particle filters do not perform satisfactorily in this scenario and hence we are pushing the methodology further by development of novel approaches. These include elements of Markov chain Monte Carlo-based filters, genetic algorithm approaches and SMC samplers. The goals of the working group are to produce papers on various topics involving multiple participants from the group, leading to future collaborative projects across a number of disciplines. We are currently organised into subgroups and addressing the following areas:

Subgroup 1: Multiple target tracking (lead Francois Septier (Cambridge)):

Participants included Simon Godsill, Francois Septier, Chunlin Ji, Mark Briers, Viktor Rozgic, Ernest Fokoue, Daniel Clark.

We have generated standard datasets, test scenarios and data simulation code for multiple

target tracking with random birth and death of objects and various sensor characteristics based on point process models or pixellated image data. A number of methodologies have been tested on this scenario, including novel MCMC-based particle filters, resample-move filters, SMC samplers (work still in progress), variational Bayes, and the results are being compiled into a survey paper: A Comparative Study of Particle Methods for Multi-Target Tracking. F. Septier, V. Rozgic, M. Briers, D. Clark and S. Godsill, in preparation.

New smoothing methods for random finite set models have also been developed by Dan Clark.

Several papers on these topics will be presented at a JSM special session on tracking:

- Variational Mean Field Approach to Efficient Multitarget Tracking. E. Fokoue. JSM 2009.
- Dynamic Spatial Mixture Modelling and its Application in Bayesian Tracking for Cell Fluorescent Microscopic Imaging. Chunlin Ji, Mike West. JSM 2009
- Sequential Monte Carlo Smoothing with Random Finite Set Observations. D. Clark and M. Briers. JSM 2009

In addition we have done work in detection and tracking of dynamic group objects, using a virtual-leader formulation and adaptations of our previous SDE-based models, which have appeared as:

- Tracking of Coordinated Groups using Marginalised MCMC-based Particle Algorithm. Francois Septier, Sze Kim Pang, Simon Godsill and Avishy Carmi. IEEE Aerospace Conference, March 2009.

Subgroup 2: Biological cell tracking (lead Chunlin Ji):

Participants include: Chunlin Ji, Simon Godsill, Daniel Clark.

This group interfaces also with the multiple target tracking sub-group, being concerned with video imaging data involving fluorescently labelled multiple cells, which move around, grow, divide, etc. They have investigated a number of approaches including point process based methods, including PHD filters, and also pixel-based likelihood functions. Datasets have already been provided by Duke researchers. Two papers are in preparation:

- C. Ji & M. West (2009) Bayesian Nonparametric Modelling for Time-varying Spatial Point Processes (Initial draft completed).

- C. Ji., S. Godsill, and M. West (2009) Spatial dynamic mixture modelling for multiple extended target tracking (In preparation).

A grant application has gone in from Dan Clark to the UK's BBSRC Tools and Resources panel (New Investigator program) on cell tracking work (outcome April 2009)

Subgroup 3: Covert chemical release (lead Nathan Green):

Participants include: Nathan Green, Francois Septier, Avishy Carmi, Simon Godsill, Mark Briers, Gareth Peters.

This topic involves plume tracking and source term estimation, exploring contour tracking, cloud tracking, ABC methods, SMC samplers and other novel techniques. The subgroup has produced models and simulation code for the source term estimation problem, in which the task is to estimate the location of a covert chemical release through sequential monitoring of the pattern of the resulting chemical plume. DSTL have agreed in principle to provide LIDAR data for this problem, and this is still being negotiated at the time of this report. A number of advances have been made in the area.

For the pure cloud tracking problem (without source term estimation), we have studied the problem of sequential inference about complex evolving cloud structures from LIDAR data, presently all simulated from models. A dynamic Gaussian mixture approximation with unknown number of components is used for the cloud intensity. Some very successful results were obtained from very ambiguous thresholded data, which have impressed specialists at QinetiQ UK Ltd. and DSTL UK Ltd. A paper is submitted to the 2009 Fusion conference, and a further paper is in preparation for the SYSID conference:

- Tracking of Multiple Contaminant Clouds. Francois Septier, Avishy Carmi, Simon Godsill. Fusion 2009 (submitted).
- Multiple Object Tracking Using Evolutionary and Hybrid MCMC-Based Particle Algorithms. F. Septier, A. Carmi, S. K. Pang and S. J. Godsill. SYSID 2009.

In addition to this, work on sequential source term estimation has been undertaken using a new trans-dimensional ABC algorithm that is able for the first time to detect multiple unknown covert releases. A survey paper has been submitted already and a paper on the STEM application is also under preparation:

- S.A. Sisson, G. W. Peters, Y. Fan, and M. Briers, Likelihood-free samplers, Journal Submission, Dec 2008.

- G. W. Peters, M. Briers and ... Trans-dimensional ABC for source term estimation, In preparation.

This work has also led to an invite to the ABC workshop in Paris, June 2009. The work has attracted serious attention from the UK's DSTL defence organisation and is very likely to lead to new grant funding in the near future.

A final sub-topic in this area concerns emulation-based methods for approximation of complex source term simulation problems. A paper is in preparation:

- Emulation Based Priors for Source Term Estimation. Gentry White and Nathan Green, in preparation.

This looks into using an emulation based approach to construction priors for use in a sequential Monte Carlo model for source term estimation. This emulation based approach allows for the construction of priors based on prior information from both computer models as well as field data. These priors offer advantages over existing priors in that they avoid degeneracy in the SMC simulation. This work draws on work from the previous SAMSI program on Development, Assessment and Utilization of Complex Computer Models, including work from the Engineering Methodology working group and the paper "Mechanism-Based Emulation of Dynamic Simulation Models: Concept and Application in Hydrology" (Reichert et. al 2009), currently under submission.

Subgroup 4: Neuron tracking (led by Joshua Vogelstein):

This topic involves tracking of multiple neuronal activity measured in living brains and involves learning of sparse connectivity matrices in continuous-time spiking environments. To include continuous time spike modelling, inference for multiple (sparsely connected) neurons, parameter estimation, image models.

The work has not progressed substantially over the last quarter and we plan that this activity will ramp up over the next 6 months.

Regular participants include: Mark Briers (QinetiQ, UK), Daniel Clark (Heriot-Watt University UK), Avishi Carmi (Cambridge University UK), Julien Cornebise (SAMSI post-doc), Ernest Fokoue (Kettering University US), Simon Godsill (Cambridge University - program leader), Nathan Green (DSTL UK, long-term Fall 08 visitor) Chunlin Ji (Duke University - SAMSI Grad. student), Sze Kim Pang (Cambridge University UK), Gareth Peters (Univ. New South Wales, SAMSI Fall 08 visitor), Viktor Rozgic (University of Southern California), Francois Septier (Cambridge University, UK), Joshua Vogelstein (Johns Hopkins University), Gentry White (N.C. State University - SAMSI Post-doc), Namrata Vaswani (Iowa State University)

2.5.2 Theory Working Group

The Theory working group is led by Mark Huber (Duke). The goal is to develop and analyze algorithms arising in SMC and MCMC applications. The plan of attack is to examine several techniques from both fields, and attempt to answer questions such as: 1) when can a method from one field be used in the other, and 2) is it possible to prove something about the running time of these methods as algorithms? This second question typically reduces to questions about rate of convergence, or the variance of estimators.

Participants include Petar Djuric (Stony Brook), Jan Hannig (UNC), Jim Lynch (U. South Carolina), Jonathan Mattingly (Duke), Edsel Pena (U. South Carolina), Gareth Peters (SAMSI), Giovanni Petris (Arkansas), Clyde Schoolfield (Florida), Sarah Schott (Duke), Namrata Vaswani (Iowa State), Anand Vidyashankar (Cornell).

The theory working group has been exploring relationships between Markov chain approximations and SMC methods, with an eye towards provably good methodologies. Unfortunately, most of the existing work relies on having rapidly mixing Markov chains, at which point the use of SMC is not necessary. However, proper use of Monte Carlo samples remains a difficult issue. Therefore, we have started concentrating on a very general methodology called the "Product Estimator" for moving from samples to approximate integration. This method is very versatile and does not rely on having bounded variance of the random variables used in the Monte Carlo algorithm. Current analysis, however, relies on using a medians-of-averages approach. An approach using pure averages should converge more quickly, but this makes the analysis more difficult, requiring study of products of binomial random variables. The eventual goal is a tighter bound on the tails of these distributions, moving from what is now a constant of 16 to a value of 2.

Since the working group lacks a postdoc, graduate student Sarah Schott has been organizing the meetings and keeping our web page up to date. On the research side, she has been working on the product estimator problem described above, beginning with simulation studies and currently working to extend large deviations inequalities for binomials from sums to products.

Impact on Sarah's research: Initially I introduced the product estimator as a side algorithm, a participant in the working group asked the question about the tightness of the constant. This raised an interesting point, and as Sarah and I have studied the problem further has proven far more deep a question than at first realized. In addition to this research avenue, I have learned much about SMC methodology over the course of the program, and still hope to utilize some of these methods in improving perfect simulation algorithms (the focus of my research program.)

2.5.3 Population Monte Carlo Working Group

The population Monte Carlo working group is led by Arnaud Doucet.

Following up the discussions at the kick-off workshop in September 2008, the Population Monte Carlo working group was created to demonstrate the potential of Sequential Monte Carlo (SMC) methods for general stochastic computation problems. Although most of the current work on SMC address on-line inference problems, the objectives of this group is to focus on the development of SMC and its variants to address problems where Markov chain Monte Carlo (MCMC) methods are traditionally used. Standard MCMC are typically inefficient for multimodal target distributions and the objectives of this group is to develop powerful particle alternatives.

We have worked on three specific topics.

Subgroup 1: Adaptive SMC samplers

SMC samplers is a general methodology which can be used as an alternative to MCMC methods. However it requires specifying a cooling schedule and some proposal distributions. We have proposed a new method which allows us to compute on-the-fly a relevant cooling schedule. The resulting algorithms have been used to solve Approximate Bayesian Computation problems and to perform inference in stochastic volatility models. We are currently developing a methodology to design automatically the parameters of the proposal distributions. Two papers have been submitted.

- An adaptive SMC method for approximate Bayesian computation. Pierre Del Moral, Arnaud Doucet and Ajay Jasra, submitted January 2009.
- Inference in Levy-driven stochastic volatility models. Ajay Jasra, Dave Stephens and Arnaud Doucet, submitted February 2009.

Subgroup 2: SMC samplers for Normalizing Constant Calculations

It is possible to use SMC to compute normalizing constants of high-dimensional distributions. In physics this strategy is known as Jarzinsky's equality. Recently an alternative method known as nested sampling has appeared in the literature. This method enjoys several advantages compared to standard techniques. However, it remains inefficient when applied to multimodal distributions. We are currently studying an adaptive SMC version of nested sampling. Our preliminary results indicated that this new method outperforms significantly the original nested sampling algorithm in complex scenarios. We are currently investigating the theoretical properties of the resulting estimate. This will lead to the following paper.

- Particle nested sampling, Arnaud Doucet and Christian P. Robert, in preparation.

Subgroup 3: Particle Markov chain Monte Carlo

Particle MCMC is a new class of methods which allows us to use SMC proposals within MCMC algorithms (Andrieu, Doucet & Holenstein, 2008). There are several open questions to address such as selecting the optimal trade-off between the number of MCMC iterations/number of particles or how to select adaptively the number of particles as a function of the current parameter value so as to ensure that the variance of the marginal likelihood is below a given threshold. We are currently studying theoretically the performance of these algorithms so as to identify the optimal tradeoff; our study relies on new sharp convergence results for SMC estimates of normalizing constants. We have also proposed some extensions of Particle MCMC methods which allow us to solve optimization problems. These extensions rely on new combinatorial identities for SMC schemes. This will be summarized in the following paper.

- Exponential inequalities for unnormalized Feynman-Kac particle models. Christophe Andrieu, Pierre Del Moral and Arnaud Doucet, in preparation.

2.5.4 Particle Learning Working Group

The particle learning working group is led by Hedibert Lopes.

Introduction: I provide current developments of several projects. Some projects fall within one of the four initial working subgroups¹, while other projects use Particle Learning in specific applications or general classes of models. In what follows I detail three of these projects. The final section lists all projects under investigation by members of the Particle Learning Working Group.

Project 1. Particle Learning in Structured AR Models. In this project, Raquel Prado and I merge the algorithms of Liu and West (2001) with Carvalho, Johannes, Lopes and Polson (2008) to sequentially estimate the following AR(p) process x_t plus noise

$$y_t = x_t + \epsilon_t, \quad \epsilon_t \sim N(0, v),$$

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + w_t \quad w_t \sim N(0, w),$$

¹*subgroup 1:* Revisiting Liu and West; *subgroup 2:* Combining LW and PL; *subgroup 3:* Estimation of economic models; and *subgroup 4:* Long memory stochastic volatility Models

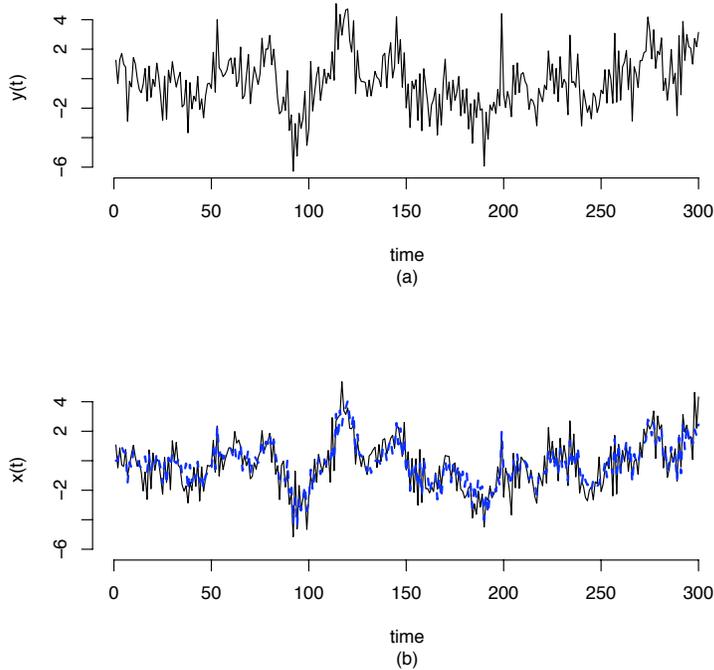


Figure 1: (a) Data (y_t) simulated from an AR(2) plus noise model with two real characteristic reciprocal roots $r_1 = 0.9$ and $r_2 = -0.7$. (b) Solid line: latent process (x_t) simulated from this AR(2) plus noise model. Dotted line: posterior mean of the estimated latent process obtained with the PL algorithm.

where $\phi = (\phi_1, \dots, \phi_p)'$ is the p -dimensional vector of AR coefficients, v is the observational variance and w is the variance at the state level. It is assumed that ϕ , v and w are unknown, and their prior structure will be described below.

We assume a prior structure such that $p(\phi, v, w) = p(\phi)p(v)p(w)$, with standard inverse-Gamma prior distributions for v and w and prior structure on ϕ via the reciprocal characteristic roots (Huerta and West, 1999)

$$\Phi(u) = 1 - \phi_1 u - \dots - \phi_p u^p,$$

where u is a complex number. The AR process is stationary if all the roots of $\Phi(u)$ (real or complex) lie outside the unit circle, or equivalently, if the moduli of all the reciprocal roots of $\Phi(u)$ are below one.

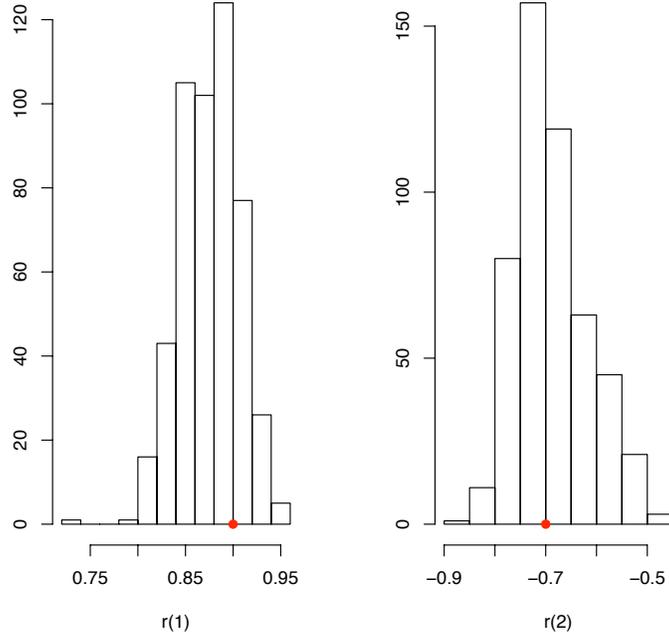


Figure 2: Estimates of $p(r_1|y_{1:T})$ (left plot) and $p(r_2|y_{1:T})$ (right plot) obtained by applying the PL algorithm to the simulated data y_t shown in Figure ???. The dots correspond to the true values of r_1 and r_2 . The results are based on $M = 500$ particles.

Figure 1 displays $T = 300$ observations simulated from an AR(2) plus noise model

$$y_t = x_t + \epsilon_t, \quad \epsilon_t \sim N(0, v)$$

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \omega_t, \quad \omega_t \sim N(0, w),$$

with $\phi_1 = r_1 = 0.9$, $\phi_2 = r_2 = -0.7$, $v = 1$ and $w = 1$. We assume that v and w are known and apply the PL algorithm to achieve on-line filtering and parameter learning. Uniform priors are assumed on r_1 and r_2 , with $r_1 \sim U(0, 1]$ and $r_2 \sim U(-1, 0)$. Figure 2 displays the posterior distribution of r_1 and r_2 .

Project 2. Particle Learning in Epidemic SEIR Models In this paper Vanja Dukic, Nicholas Polson and I present a novel method for classic generalized epidemics models, in the family of the so-called susceptible-exposed-infected-recovered (SEIR) models. The proposed method is based on the particle learning (PL) methodology of Carvalho et al.

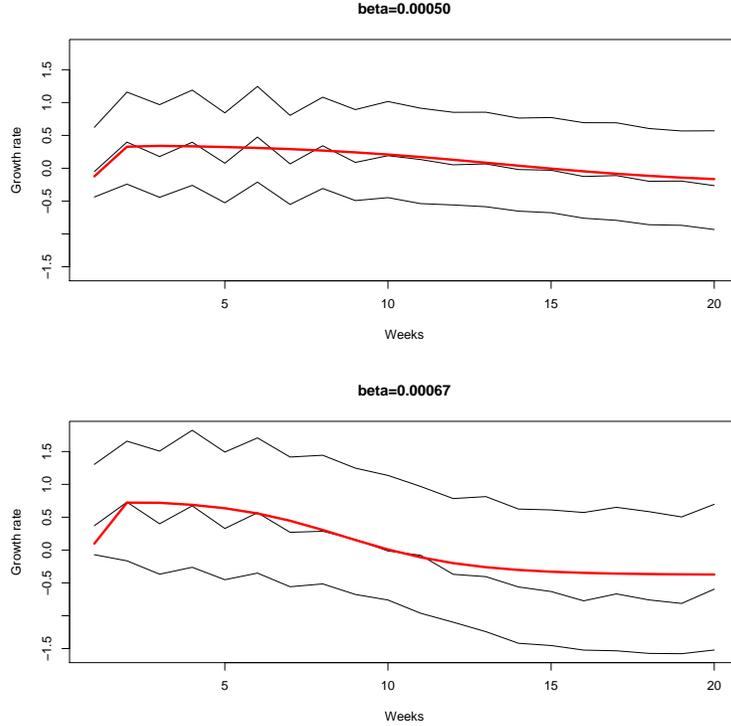


Figure 3: Simulated data. Red line is the true growth rate of the infection.

(2008), which, we argue, is particularly well-suited to on-line learning and surveillance for infectious diseases. In direct comparisons to the widely used MCMC (O’Neil and Roberts 1999, Elderd et al. 2006) and perfect sampling (Fearnhead and Meliglokou 2004), we find the PL method is more efficient, and, in addition, significantly more generalizable to more complex dynamics. We analyze the Google flu trends data for seasons 2003-2008, with the special emphasis on the current season.

The so-called SEIR model (Anderson and May, 1991) is then given as follows: $\dot{S} = -\beta S_t I_t$, $\dot{E} = \beta S_t I_t - \alpha E_t$, $\dot{I} = \alpha E_t - \gamma I_t$ and $\dot{R} = \gamma I_t$, where the dot denotes a time derivative. In this model, the individuals in a finite closed population of size N begin in the uninfected, nonimmune class S and move to the exposed but not yet infectious class E at a rate β . Exposed but not yet infectious individuals move to the infectious class at rate α , while γ is the rate at which infectious individuals I cease to be infectious because of recovery or death. Aligning this with the state-space modeling terminology, the state vector in this model is given by $x_t = (S_t, E_t, I_t, R_t)$ and the parameter vector is $\theta = (\alpha, \beta, \gamma)$.

Figure 3 show the sequential learning of the growth rate of infection, $(I_{t+1} - I_t)/I_t$ in

a simulated example where the population size is $N = 3000$ and the final time horizon is $n = 20$. The parameters are $\alpha = 2$ (latency) and $\gamma = 1$ (recovery), while β (coefficient of transmission) can be either $1.5/N$ or $2/N$. The initial values for S, E, I and R are 3000, 0, 10 and 0, respectively.

Project 3. Particle Learning in DSGE Models Francesca Petralia, Hao Chen, Carlos Carvalho and I apply PL to Dynamic Stochastic Generalized Equilibrium (DSGE) models. DSGE models are now the main tool used by macroeconomists to answer quantitative questions about the aggregate economy. Estimation of those models, however, is a major challenge due to the nonlinearity and non-normality inherent in the likelihood function. Current likelihood based inference either assumes normality (Kalman filter), or uses a particle filter to integrate out the unobserved state variables within a Metropolis Hastings algorithm, ie. *marginally* by looking at $p(\Theta|y^n)$. The goal is to use sequential Monte Carlo algorithms that jointly and sequentially estimate parameter and state learning for DSGE models, ie. *sequentially* by looking at $p(x^t, \Theta|y^t)$.

More specifically, one has to solve a set of equilibrium policy functions, for each parameter value, in order to get a state-space representation

$$U = E_0 \sum_{t=1}^{\infty} \beta^{t-1} \frac{(c_t^\theta (1-l_t)^{1-\theta})^{1-\tau}}{1-\tau}$$

with constraints $y_t = e^{z_t} k_t^\alpha l_t^{1-\alpha}$, $y_t = c_t + i_t$, $k_{t+1} = i_t + (1-\delta)k_t$ and $z_t \sim N(\rho z_{t-1}, \sigma^2)$.

The state variables are (k_t, z_t) , where k_t is capital accumulation and z_t is productivity shock. The observables are

(i_t, y_t) , where i_t is investment and y_t the GDP of the economy. The fixed model parameters are $\Theta = (\alpha, \beta, \rho, \tau, \theta, \delta, \sigma)$. We approximate the solution of the maximization problem with a first order Taylor approximation: $l_t = \alpha_0 + \alpha_1 k_t + \alpha_2 z_t = g(k_t, z_t)$ and $k_{t+1} = \beta_0 + \beta_1 k_t + \beta_2 z_t = h(k_t, z_t)$. Once we estimate $(\alpha_0, \alpha_1, \alpha_2)$ and $(\beta_0, \beta_1, \beta_2)$ we get the transition equations for the states and the measurement equations for the observables. The measurement and transition equations are

$$y_t \sim N(e^{z_t} k_t^\alpha g(k_t, z_t)^{1-\alpha}, \sigma_y^2) \quad \text{and} \quad i_t \sim N(h(k_t, z_t) - k_t(1-\delta), \sigma_i)$$

$$k_{t+1} = \beta_0 + \beta_1 k_t + \beta_2 z_t \quad \text{and} \quad z_t \sim N(\rho z_{t-1}, \sigma^2)$$

where ϵ_y and ϵ_i are normally distributed with mean zero and standard deviations σ_y and σ_i , respectively. The full SMC approach is 1) Draw from an initial distribution N values of Θ ;

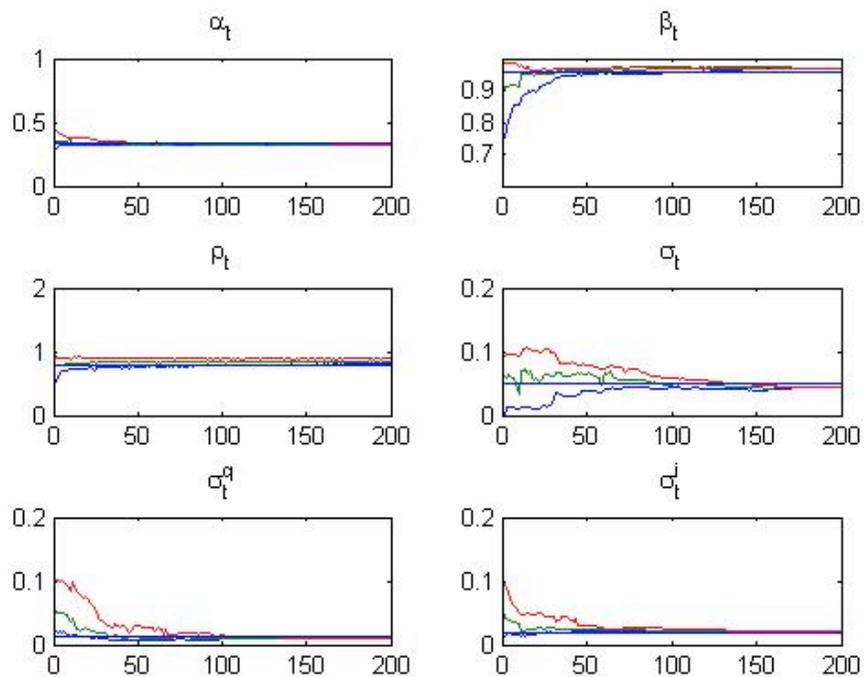


Figure 4: Estimates of $p(r_1|y_{1:T})$ (left plot) and $p(r_2|y_{1:T})$ (right plot) obtained by applying the PL algorithm to the simulated data y_t shown in Figure ???. The dots correspond to the true values of r_1 and r_2 . The results are based on $M = 500$ particles.

2) Solve the model for each set of parameter value; 3) Draw (z_t^i, k_t^i) $i = 1, \dots, N$; 4) Resample step if needed; 5) Draw θ from its distribution and go back to 2.

At first we considered a special case of this model, where there is no labor and δ is equal to 1. This is the simplest model we can think about because we have a unique state space representation of the model. Under these assumptions the measurement equations become $y_t = e^{z_t} k_t^\alpha + \epsilon_y$ and $i_t = \alpha\beta e^{z_t} k_t^\alpha + \epsilon_i$ and the transition equations for the state are $k_{t+1} = \alpha\beta e^{z_t} k_t^\alpha$ and $z_t = \rho z_{t-1} + \epsilon_t$. Figure 4 is based on 15K particles of Liu West + sufficient statistics (for ρ) algorithm when $\delta = 1$.

Conference presentations: The following talks will be presented at the 2009 Seminar on Bayesian Inference in Econometrics and Statistics (SBIES) will take place on May 1-2, 2009 at Washington University in St. Louis, MO:

1. I talk about “Particle Learning for Generalized Dynamic Conditionally Linear Models” and

2. Bruno Lund (my visiting PhD student) talks about “The Role of Options, Stochastic Volatility and Jumps in the Interest Rate Risk Premia Dynamics”, which is fully and sequentially estimated through PL.

JSM 2009. Several members of the PL working group will actively participate in the next edition of the Joint Statistical Meetings in August in Washington, D.C.

1. I talk about “Particle Learning” (Invited Session)
2. I organized and will chair a contributed session entitled “Particle Learning” where
 - (a) Raquel Prado talks about “PL for Autoregressive Models with Structured Priors”,
 - (b) Chiranjit Mukherjee talks about “PL Without Conditional Sufficient Statistics”,
 - (c) Christian Macaro talks about “PL for Long Memory Stochastic Volatility Models”, and
 - (d) Bruno Lund (my visiting PhD student) talks about “The Role of Options, Stochastic Volatility and Jumps in the Interest Rate Risk Premia Dynamics”, which is fully and sequentially estimated through PL.
3. Francesca Petralia talks about “PL for Dynamic Stochastic General Equilibrium Models” in another contributed session.

Other conference talks:

1. I talk about “Particle learning for general mixtures” at the Adaptive Design, Sequential Monte Carlo and Computer Modeling Workshop, SAMSI, April 15-17.
2. I talk about “Particle learning” at the R/Finance 2009: Applied Finance with R meeting , Chicago, April 24-25.
3. I talk about “Particle learning and smoothing” at the X Brazilian School of Time Series and Econometrics, São Carlos, Brazil, 21-24.

Short courses and tutorials

1. I give a short course on “SMC in Stochastic Volatility Models” in the Department of Statistics and Operations Research, Universita Politecnica da Catalunya, June 22 - July 4.

2. I give a tutorial on “Particle Filters” at the X Brazilian School of Time Series and Econometrics, São Carlos, Brazil, July 21-24.
3. I give a short course on “Modern Bayesian Statistics via SMC methods” at the INPE Advanced Course - III Astrostatistics, São Jose dos Campos, Brazil, September 14-18.

Papers under preparation: ²

1. Particle Learning and Smoothing (CaJLoPo)
2. Particle Learning in General Mixtures (CaLoPoT)
3. Particle Filtering and Learning: A Comparison (CaJLoPo)
4. Particle Learning for Autoregressive Models with Structured Priors (PrLo)
5. Particle Learning in Epidemic SEIR Models (DLoPo)
6. Particle Learning Without Conditional Sufficient Statistics (NMuCaLo)
7. Particle Learning for Long Memory Stochastic Volatility Models (MaLo)
8. Particle Learning for DSGE Models (PeChCaLo)
9. Stochastic Volatility Shot-Noise (CaJLoPo)
10. Options, SV and Jumps in the Interest Rate Risk Premia (LuLo)

2.5.5 Model Assessment and Adaptive Design Working Group

The working group leader is Carlos Carvalho.

Summary Goals and Outcomes: Following up the discussions in the kick-off workshop (September 2008) the “Model Selection and Adaptive Design” (**MAAD**) working group was formed with the intent to enhance, explore and demonstrate the potential of particle based methods to address issues related to model uncertainty and sequential design/decision making. The group focuses on applications (listed below) where either the computation of model probabilities or the exploration of model spaces represent an enormous challenge that requires effective computation strategies. The central goal of our efforts is to make use of “state of the art” SMC techniques in trying to tackle these issues.

Since its formation, the group has met weekly at SAMSI for discussions of relevant issues and to report on progress made by many of the participants.

²Ca: Carlos Carvalho; Ch: Hao Chen; D: Vanja Dukic; J: Michael Johannes; Lo: Hedibert Lopes; Lu: Bruno Lund; Ma: Christian Macaro; Mu: Chiranjit Mukherjee; N: Jarad Niemi; Pe: Francesca Petralia; Po: Nicholas Polson; Pr: Raquel Prado; T: Matt Taddy.

Specific Goals and Areas of Focus: At the current stage the group has identified four main areas of focus, as described below:

Subgroup 1. Particle Model Selection: Our goal is to develop a general class of particle methods to accommodate uncertainty in variable selection in high-dimensional settings. There is a rich Bayesian literature on variable selection and stochastic search methods for linear regression models, but very little work has been done for nonparametric models that allow the conditional distribution of a response to change flexibly with predictors. Our initial plan was to develop an efficient particle stochastic search (PSS) approach for high-dimensional variable selection in linear regression, while simultaneously developing a Particle Learning algorithm for posterior computation in probit stick-breaking processes (PSBPs). PSBPs are a recently proposed nonparametric Bayes modeling framework, which allow conditional distributions to change flexibly with predictors. Due to the conjugacy of the PSBP after data augmentation, it should be possible to adapt the Particle Learning algorithm to include a PSS component. This will allow selection of variables having any impact on the conditional distribution of a response, while also accommodating responses having arbitrary scales (continuous, categorical, count, etc). An additional topic that the group will focus on is development of efficient particle methods for calculating Bayes factors for comparing non-nested models. The idea is to initially devote a similar number of particles to each model, but then through resampling as the algorithm progresses, devote increasing numbers of particles to the better models in the list. This will allow accurate posterior computation and estimation of marginal likelihoods for good models, while not wasting computational effort on poor models.

We have made substantial progress in the above areas. Here are specifics of each project:

- **“Particle stochastic search for high-dimensional variable selection”** (Shi and Dunson) - we have continued to make progress in refining our particle stochastic search (PSS) algorithm and have compared the algorithm in a variety of settings to shotgun stochastic search (SSS). We also have results comparing to SSVS for simulated examples and a real data application taken from Hans et al. SSS paper. The paper with the above title is in final preparation stage and will be submitted within a few weeks. We will then move our focus to variable/model selection in nonparametric Bayes regression models, adapting PSS to allow for the inclusion of parameters/latent variables common to the different models in the particles. This will allow variable selection in PSBP mixture models and other interesting cases. We plan to apply this in dynamic mixture model settings as well.

- **“Bayesian distribution regression via augmented particle learning”** (Dunson and Das) - we have continued to make progress in developing and implementing an efficient sequential Monte Carlo algorithm for posterior computation and marginal likelihood estimation in a broad class of mixture models that allow the mixing weights to vary with time, space and predictors. This class of mixture models is referred to as probit stick-breaking mixtures and has the appealing property of facilitating efficient computation through a data augmentation strategy. In particular, for many useful special cases, one can obtain the marginal likelihood in closed form integrating out all of the parameters but conditioning on latent normal variables. Our proposed “augmented particle learning” (APL) algorithm proceeds by sequentially adding subjects in parallel to each of a large number of particles, sampling from the conditional posterior distributions of the latent variables as subjects are added and resampling appropriately. The method avoids the need for sequential importance sampling for updating of particles, instead relying on direct sampling, with marginalization used to improve efficiency. We have primarily code for count regression models which allow the conditional distribution of a count response to change flexibly with a predictor, and have already obtained good results for a mixture of Poisson case with no predictors. A manuscript with the initial results will be submitted within few weeks. The abstract is: To limit assumptions in modeling of conditional response distributions, hierarchical mixtures-of-experts models allow the mixing weights in a regression model to vary flexibly with predictors. Nonparametric Bayes methods can be used to incorporate infinitely-many components, allowing effective model dimension to increase with sample size. However, MCMC algorithms for posterior computation often encounter mixing problems due to multimodality of the posterior. Focusing on a broad class of probit stick-breaking process priors for conditional response distributions indexed by time, space or predictors, we propose an efficient augmented particle filter for posterior computation and approximation of marginal likelihoods. The algorithm sequentially updates random length latent normal vectors within each particle as subjects are added, avoiding truncation of the infinite collection of random measures. Through marginalization after data augmentation, the approach bypasses the need to update parameters, dramatically improving efficiency while avoiding degeneracies. The method can be applied broadly for continuous, count or categorical response variables. The methods are illustrated using simulated examples and an epidemiologic application.

Primary subgroup participants: David Dunson (Duke), Minghui Shi (PhD student, Duke), Sourish Das (SAMSI) and Artin Armagan (SAMSI and Duke).

Subgroup 2. Adaptive Design: For expensive data, as those arising from computer models, astronomy data, destructive experiments, etc, careful designs which contemplate how many data points will be obtained, where and when, is mandatory. These design problems, for these expensive experiments, have to, almost unavoidably, be sequential and adaptive so as to best use the very scarce and expensive information. Sequential decision problems (of which sequential designs are particular cases) involve “look ahead” computations for all possible future observations, which might be computationally challenging for complex models. We intend to explore SMC methods to help with these computations.

The following initial paper is under way: “Adaptive sampling for Bayesian variable selection” (Fei Liu, Fan Li and Dunson). The problem is to sequentially select subjects based on their predictor values, with the response value obtained for the selected subjects and the objective being optimal performance in model selection. We have the methods details worked out and Fei Liu has implemented a couple of simple examples where she demonstrates substantial advantages relative to selecting the subjects in a random order. We have discussed strategies for proving improvements theoretically under the assumption that the number of subjects in the pool to draw from is large, so that we can avoid finite population sampling complications. Fan has found an interesting data example to motivate the approach and the paper should be completed in a month or two depending on Fei’s time. Fei and David have discussed moving on to a “active transfer learning” problem in which there are multiple related regression models and one wants to borrow information in selection of models across the related models.

Primary subgroup participants: Susie Bayarri (Univ. of Valencia and SAMSI), Jim Berger (SAMSI and Duke Univ.), Merlise Clyde (Duke Univ.), Tom Loredon (Cornell Univ.), Ana Corberan (PhD student, Univ. of Valencia), Fei Liu (Univ. of Missouri) and Fan Li (Duke Univ.).

Subgroup 3. Sequential Model Monitoring: In this subgroup we focus on problems of sequential model reassessment and model space exploration as new observations become available. The examples we have been developing so far involve sequential posterior inference about graphical structures underlying the covariance matrix of innovations in dynamic linear models. These models have been applied in large scale sequential portfolio allocation where the graphs provide a regularization tool for the covariance matrix of assets. The development of sequential model selection procedures that address uncertainty about graphs while allowing for on-line updates is an open research area and one of key importance in further applications of DLMS in real forecasting problems. In our first attempt to

solve this problem, we have been using particles systems as discrete approximations for the posterior distribution of models. Hao Wang has been coding some of the ideas discussed and the initial results are promising. We have made significant progress in this area and an first draft of a paper by Hao Wang, Craig Reeson and Carlos Carvalho is ready and should be submitted before the summer. The paper “Sequential Learning in Dynamic Graphical Models” proposes a natural generalization of the dynamic matrix-variate graphical model (Carvalho and West 2007) to time varying graphs. The generalization uses the multi-process modelling idea to introduce sequential graphical model selection procedures that address uncertainty about graphs while allowing for efficient on-line updates. To develop an efficient Bayesian approach for sequentially searching high-dimensional graphical models, we describe a feature-inclusion particle stochastic search algorithm, or FIPSS. The FIPSS algorithm allows parallel exploration of the search space using estimates of edge inclusion probabilities. The model is illustrated using financial time series for predictive portfolio analysis.

Primary subgroup participants: Carlos M. Carvalho (Univ. of Chicago and SAMSI), Hao Wang (PhD student, Duke Univ.) and Craig Reeson (Undergraduate student, Duke Univ.)

Subgroup 4. Dynamic Control: The main objective of this subgroup is to study problems that have a dynamic (sequential) decision making component as well as some uncertainty about the system parameters that would require Bayesian updating. We are particularly interested in problems that arise in health care settings where a decision maker (a doctor/nurse, emergency response officer, hospital management, etc.) will have to give decisions regarding the treatment options of patients or allocation of scarce resources to a group of patients. These decisions are dynamic in nature as the conditions of patients change with time. Such problems have been studied commonly in the Operations Research literature. However, almost all of the earlier studies assume that the decision maker has complete information about the states of patients and the system parameters. There are several situations where such an assumption of perfect information may not be realistic. For example, for rarely observed diseases or disasters that involve nuclear agents, there does not exist sufficient data to estimate parameters that are needed in solving the dynamic control problem. Our objective is to study such dynamic control problems where the decision maker will learn about the disease or the emergency event under consideration as the decisions are made sequentially. As an initial step, we will consider the following problem. Consider a system with several patients in need of care from a single resource (a doctor or an operating room). The patients are affected by the same disease or the same traumatic event but they

could be in different stages of criticality. The stage that a patient is in may affect the cost of keeping that patient waiting, the service requirement of that patient, and also the success probability of the operation. The decision maker cannot observe the true states of patients but can observe certain signals that the patients send (for example, heart rate, blood pressure, etc.). Based on these signals, the decision maker decides which patient should be taken into service next with the objective of maximizing the total expected utility. As we mentioned earlier, the decision maker does not know exactly how the signals and the true states of patients relate and how the patients conditions degrade. When each patient is taken into service, we can observe the true condition of the patient and based on such information collected, we can update the unknown parameters related to the disease/condition. Then, with this updated information, we make the next decision to serve another patient.

Nilay Argon and Melanie Bain are currently using SMC methods in solving a dynamic control problem that arises in the aftermath of mass-casualty incidents. To be more specific, we consider a mass-casualty event (such as a plane crash or a terrorist bombing) that resulted in several casualties in need of care. Due to the massive number of casualties, the medical resources are overwhelmed and decision makers need to prioritize patients for service. Depending on their injuries, the patients could be in different stages of health. The stage that a patient is in may affect his/her probability of survival and also service requirement. The decision maker cannot observe the true states of patients but can observe certain signals that the patients send (for example, pulse, breathing rate, etc.). Based on these signals, the decision maker decides which patient should be taken into service dynamically with the objective of maximizing the total expected number of survivors. We initially assume that the decision maker knows how the signals and the true states of patients relate. We also assume that the patients' conditions degrade according to a discrete time Markov chain with a known transition probability matrix.

We first formulated the above problem as a partially observable Markov decision process (POMDP). The POMDP we obtained could have a very large belief state depending on the number of patients involved and also the number of health stages that we define. Hence, we will need to use an approximate method to solve this problem. We have thus far considered two approaches from the literature. One is by Thrun (2000), where particle filtering is used to reduce the size of the belief space, and the other is by Luo, Fu, and Marcus (2008), which is based on projecting the high-dimensional belief space to a low-dimensional family of parametrized distributions. We are currently implementing Thrun's approach.

Primary subgroup participants: Nilay Tanik Argon (UNC), Abel Rodriguez (Univ. of California, SC), Melanie Bain (PhD student, UNC) and Kai Wang (PhD student, Duke)

2.5.6 Continuous Time Working Group

The working group leaders are Jashya Bishwal, Paul Fearnhed and Jochen Voss. During the initial formative phase, the working group in continuous time models, parameter estimation and finance started a review of the relevant literature. Several articles were presented during the group meetings. Topics covered include the following:

- filtering discrete observations of a continuous time signal
- exact algorithms
- filtering and parameter estimation using a windowed SMC method (in discrete time)
- change point problems for continuous time processes
- filtering for a CIR process given Poisson observations

During this initial phase, several areas were identified where relevant problems could benefit from further study by the group and which the group members intend to investigate further. We give details of two possible application focus areas.

The first focus area concerns filtering/parameter estimation for processes observed via Poisson observations; methodology developed here has applications in credit risk modelling. The underlying mathematical model is as follows: an unobserved signal is modelled by a continuous time, real valued, positive stochastic process, for example a CIR process. The observations consist of one instance of an inhomogeneous Poisson process which has the signal process as its intensity function. The problem is first to recover information about the signal from the observations (filtering) and, in a second step, to utilise the given observation to estimate parameters governing the dynamics of the signal. In the application to credit-risk modelling, the signal describes the intensity of mortgage defaults, the points of the Poisson process are times when individual defaults.

A second focus area is to generalise existing methods to processes driven by fractional Brownian motion instead of Brownian motion. This is of particular interest in finance, where fractional processes can model the long-range dependence that is often observed in (say) volatility series.

Other areas may include changepoint analysis for continuous time processes, and models within genetics.

While focussing on specific application areas, we aim to develop novel, generic methodology. One approach will be to look at developing filtering equations in continuous time, and then considering various approximate ways of implementing these (based on e.g. time discretisation), rather than the standard approach of approximating the continuous time model by a discrete time one, and applying standard SMC methods for the discrete time model.

Research Highlights - Fall 08: The working group has focussed on three areas. Firstly is SMC for hierarchical branching process models, with application for qPCR analysis. Secondly are methods for survival data, where the underlying hazard depends on an unobserved stochastic process. The application for this work is to modelling, analysis and prediction of mortgage default rates. Finally, we have looked at inference for diffusion models via "least-action": defining and calculating a best path for the unobserved diffusion, and constructing Laplace approximations to the transition density of the diffusion. Applications here include models in systems biology.

The working groups is working towards preparation of three papers – one for each of these research areas. All research areas involve collaborations which would not have occurred without the SAMSI program.

Progress - Spring 09: The working group has focussed on four areas. Firstly is SMC for hierarchical branching process models, with application for qPCR analysis. Secondly are methods for survival data, where the underlying hazard depends on an unobserved stochastic process. The application for this work is to modelling, analysis and prediction of mortgage default rates. Thirdly, we have looked at inference for diffusion models via "least-action": defining and calculating a best path for the unobserved diffusion, and constructing Laplace approximations to the transition density of the diffusion. Applications here include models in systems biology. Papers being prepared include:

- Dynamic Latent Factor Model for Mortgage Termination. Paul Fearnhead, James B Kau, Donald C Keenan, Constantine Lyubimov and Anand Vidyashankar (In prep).
- Simulation and Inference for Stochastic Kinetic Models via limiting Gaussian Processes. Paul Fearnhead, Vasilieos Giagos and Chris Sherlock (In prep).

Finally, we are looking at new inference methods for α -stable Lévy processes, and we propose to write a paper on the topic:

1. Monte Carlo inference for α -stable Lévy processes. S. Godsill and P. Fearnhead, in preparation.

The working groups is working towards preparation of three papers – one for each of these research areas. All research areas involve collaborations which would not have occurred without the SAMSI program.

A grant proposal to the UK's EPSRC, which will develop on the least action filtering and α -stable models, is currently being prepared by Rogers, Godsill and West.

2.5.7 Big Data & Distributed Computing Working Group

Summary: Emerging from discussions at the Sept. 2008 opening workshop, the *Big Data & Distributed Computing (BD&DC)* working group was defined by research challenges and themes cutting across numerous applications of stochastic computation and sequential methods: scaling of models and analysis methods for increasing large data sets and in problems with increasing large spaces of underlying parameters and latent variables. Exploiting multi-core, cluster and parallel hardware promotes a need for basic research and innovations in the development of computational algorithms and also of model specification and structuring. The opportunity to make progress in these areas, linked to specific motivating applications, led to the formation of this focused working group. Since inception several BD&DC subgroups have been involved in specific research projects under the general goals, with a series of interconnections with some of the other working groups involving cross-cutting projects.

Goals and Areas of Focus: Exploration, evaluation, development and application of effective computational methods for model fitting and model assessment in problems involving large data sets and high-dimensional latent process parameters (the latter are examples of “big missing data”): sequential Monte Carlo methods, stochastic model search, sequential importance sampling and also annealing/optimization methods. Specific research subgroups are as follows.

BD&DC.1 *Primary subgroup participants:* Manolopoulou, Mihaylova, Mukherjee, Dunson, Das, Shi, West, Yoshida.

Sequential Monte Carlo methods for model learning, estimation and comparison using distributed computing on clusters. This is relevant to several of the specific modelling, methodology and application areas. Activities include study of theory and methods of implementing a variety of SMC methods on clusters, using some of the specific model contexts of interest to this working group for development. Strategies for parallelised and cluster-based computation are being explored for problems involving large data sets and high-dimensional latent processes, i.e., large missing data sets, the latter focused on state-space modelling of long time series. This research interacts with researchers in the Tracking working group.

BD&DC.2 *Primary subgroup participants:* Carvalho, Liu, Mukherjee, Wang, West.

SMC and distributed computing in mechanistic, nonlinear state-space models, with motivating applications in dynamic stochastic models arising in studies of cellular communication in biological networks in systems biology. This area involves model

development and use of customised sequential Monte Carlo methods, and so interacts substantially with some of the other program working groups. Specific characteristics of motivating problems are (a) state-space models with many uncertain parameters that are observed over time, and for which sequential learning is either desirable or necessary; (b) very high-dimensional latent processes. In systems biology problems, models are developed mathematically on very fine, discrete time scales, but actual data is observed at much cruder time scales, so that the fine time scale states become missing data in very high dimensions.

This research area also intersects with studies involving stochastic computation for model fitting in complex computer model emulation, related to the program activities at the intersection of research in computation and computer modelling design and development. This aspect of BD&DC research will be represented in talks at the SAMSI workshop on *Adaptive Design, Sequential Monte Carlo and Computer Modeling* in April 2009 as well as in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009.

BD&DC.3 *Primary subgroup participants:* Manolopoulou, West.

Sequential analysis and decision-guided sample selection and learning about rare events in mixture modelling with very large data sets. Motivating applications come from problems of inference on characteristics of rare sub-populations of biological cells in studies using flow cytometry technology in immunology, vaccine design and other areas. In such studies, a single experiment can easily generate hundreds of millions of observations in, typically 10-20 dimensions, representing marker proteins on the surface of cells. Random sampling to fit models, such as mixture models for classification and discrimination of sub-populations, is standard, though model fitting becomes challenged by sample size and so sequential methods are inherently interesting. Moreover, a specific focus on generating maximal information of rare sub-populations leads to statistical design and biased sampling strategies that are inherently sequential and for which simulation-based methods need to be developed. The interest in and role for distributed computation is evident.

BD&DC.4 *Primary subgroup participants:* Das, Dunson, Li, Liu.

Sequential methods in nonparametric statistical regression and density estimation models, with motivating applications in problems in epidemiology and public health, and in studies of huge data sets in e-commerce and internet traffic research, among others. We are developing new classes of SMC algorithms for posterior computation and

marginal likelihood estimation in a flexible class of mixture models, which allow mixture weights to vary with time, space and predictors. The proposed augmented particle learning (APL) algorithm has had excellent performance in simulation experiments for a variety of data types, avoiding degeneracies common to SMC algorithms through use of marginalization after data augmentation. The algorithm has major advantages over MCMC algorithms in avoiding mixing problems that plague MCMC for mixture models, while also allowing marginal likelihood estimation, which allows testing of competing nonparametric models and parametric vs nonparametric models.

Application areas are numerous. Part of this research is represented in a pending NIH proposal (submitted in early 2009) that proposes the further development of the APL algorithm in applications in statistical genetics and gene-environment interactions studies. This involves models that allow quantitative traits to vary flexibly with high-dimensional single nucleotide polymorphisms and environmental factors.

BD&DC.5 *Primary subgroup participants:* Dunson, Li, Liu.

Intersections of interests with the working group on sequential Monte Carlo in model assessment are developing adaptive strategies for the variable selection problems and efficient Sequential Monte Carlo methods for the evaluation of designs in massive data sets. From the perspective of the SMC and distributed computing, this subgroup specifically focuses on parallelised computing, sequential model updating, and stochastic model space searching in high dimensional variable selection problems. The proposed approach has been shown very efficiently through both simulation studies and several existing health-related data sets. The approach is beneficial to investigators who deals with problems involving model testing and/or model selection.

This aspect of BD&DC research will be represented at the SAMSI workshop on Adaptive Design, Sequential Monte Carlo and Computer Modeling in April 2009, as well as in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009.

BD&DC.6 *Primary subgroup participants:* Dunson, Shi, West, Yoshida.

Stochastic and related deterministic/annealing based methods of search over very large, discrete model spaces such as arising in sparse multivariate factor models with many response variables, and in regression model uncertainty (linear and nonlinear) with many candidate covariates. Advances in computational methods include innovations in annealed entropy methods and Bayesian shotgun stochastic search methods. Some

specific motivating applications are in genomics and public health contexts. One key development involves including model indices within the particles of SMC methods leading to an efficient algorithm for massive dimensional variable selection (Dunson and Shi), another involves a synthesis of entropy annealing based global optimization with stochastic search for very large-scale sparse statistical models (Yoshida and West).

BD&DC.7 *Primary subgroup participants:* Ji, West.

Intersections of interests with the working group on sequential Monte Carlo in compute-intensive tracking problems are generating novel model and computational methods development for tracking problems with prototype applications in monitoring and tracking many cells in systems biology experimental data. Data arising from motivating applications include studies in computational immunology driven by experiments in vaccine design, where the motion of multiple different cell types is monitored by measured fluorescent intensities of cell surface marker proteins. Research here involves novel Bayesian dynamic, non-parametric models for inhomogeneous spatial intensity functions and sequential Monte Carlo methods development for model fitting.

BD&DC.8 *Primary subgroup participants:* Petralia, Mukherjee, West.

A new, emerging area of discussion (March 2009) for a subset of this working group arose during early 2009 from discussion with environmental scientists involve in atmospheric chemistry (CO) monitoring and data synthesis. With a focus on short time-scale inference on improved understanding of the impact of earth surface fires (tropical forest fires, savannah fires, etc) on variations atmospheric CO, a core challenge is integration of *massive* amounts of high-resolution data from new satellites launched in 2009 with predictions from deterministic biophysical simulation models. Sequential Monte Carlo methods are being discussed as part of an initial effort to define a new collaboration with disciplinary computer modelers - an excellent example of a *really* big, and cluster compute-intense BD&DC problem.

Participants: This working group involves local faculty participants, SAMSI postdoctoral fellows, SAMSI visitors, SAMSI and non-SAMSI graduate students, and represents various areas of statistics and computational science. Several junior and female researchers are involved, including some who were quite new to the general area of Sequential Monte Carlo, as well as the specific areas of this working group, prior to the program. See Table 1 for the list of primary and active participants, as well as additional participants who either had some engagement in initial, formative discussions, or are collaborating, or have

occasional ongoing interactions in BD&DC meetings, or who have been short-term SAMSI visitors participating actively.

Name (gender)	Position	Affiliation	Dept/Discipline
<i>(A)</i>			
Carlos Carvalho (m)	Assistant Professor	Chicago	Statistics & Econ.
Sourish Das (m)	Postdoc	SAMSI	Statistics
David Dunson (m)	Professor	Duke	Statistical Science
Chunlin Ji (m)	Graduate RA	Duke & SAMSI	Statistical Science
Fan Li (f)	Assistant Professor	Duke	Statistical Science
Fei Liu (f)	Assistant Professor	Missouri	Statistics
Ioanna Manolopoulou (f)	Postdoc	SAMSI	Statistics
Chiranjit Mukherjee (m)	PhD student	Duke	Statistical Science
Francesca Petralia (f)	Graduate RA	Duke & SAMSI	Statistical Science
Minghui Shi (f)	Graduate RA	Duke & SAMSI	Statistical Science
Ryo Yoshida (m)	Assistant Professor	ISM Tokyo	Statistics
Hao Wang (m)	Graduate student	Duke	Statistical Science
‡Mike West (m)	Professor	Duke	Statistical Science
<i>(B)</i>			
Ernest Fokoue (m)	Assistant Professor	Kettering	Mathematics
Amadou Gning (m)	Postdoc	Lancaster	Communication Systems
Steve Koutsourelakis (m)	Assistant Professor	Cornell	Engineering
Lyudmila Mihaylova (f)	Lecturer	Lancaster	Communication Systems
Mario Morales (m)	Consultant	Emetricz	Engineering/Statistics
Deb Roy (m)	Assistant Professor	Penn State	Statistics
Andrew Thomas	Lecturer	St. Andrews University	Ecology & Statistics
Joshua Vogelstein (m)	PhD student	Johns Hopkins	Neuroscience

Table 1: *(A)* Primary and local participants ([‡] *Working Group leader*); *(B)* Additional researchers (initial participants, collaborators and/or short-term visitors in the BD&DC group)

Student Involvement:

Chunlin Ji (SAMSI RA) is attached to the Tracking (Godsill) working group but also participates actively in the BD&DC group with West on spatial dynamic modelling for biological cell tracking problems. Ji is developing SMC methods in the context of new classes of models. This research has grown out of existing work of Ji & West in static problems, now extended with new dynamic models that will form an additional part of Ji's PhD thesis research, and one initial paper is in draft at the time of this report (see

manuscripts section). Ji has led discussions on this work at several BD&DC meetings, gave a talk at the February 2009 mid-program workshop, and will present this work at the 7th Workshop on Bayesian Nonparametrics in Turin, Italy, in June 2009, and at the 2009 Joint Statistical Meetings in Washington DC 2009.

Chiranjit Mukherjee (Duke graduate student) actively participates in the BD&DC group (though is not officially supported by the program). Mukherjee has developed studies of SMC methods for model fitting and comparison in nonlinear dynamical models arising from systems biology (and other applications). These studies involve very long time series but for which most of the underlying states are unobserved, and his work has explored, evaluated and developed novel approaches to SMC using distributed computation. In March 2009, Mukherjee presented and passed his PhD preliminary exam based on this work, and is now defining his thesis topic in this area. He has led several discussions on the topic at the BD&DC meetings, presented a poster at the February 2009 mid-program workshop, and will present this work in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009.

Francesca Petralia (SAMSI RA) is attached to the Particle Learning (Lopes) working group but also participates actively in the BD&DC group. Petralia is (in March 2009) taking an active role in emerging discussions about computer model-SMC studies driven by motivating applications in environmental CO studies – problems that involve very large data sets and will require intense distributed computation – and will begin to work on this project with West in late spring 2009 linked to the BD&DC working group. Petralia will present a talk on her work with SMC in econometric models in the Particle Learning working group (Lopes, leader) in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009.

Minghui Shi (50% SAMSI RA) is working on sequential model search methodology for large, discrete model spaces, typified by “large p ” regression model uncertainty. With Dunson, Shi is developing novel extensions of shotgun stochastic search that incorporate new ideas from SMC. Shi will present her PhD preliminary exam on this topic in April 2009, and the topic seems likely to then define her thesis area. Shi has led discussions on this work at BD&DC meetings, presented a poster at the February 2009 mid-program workshop, and will present this work in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009.

Hao Wang (Duke graduate student) actively participates in the BD&DC group as well as other working groups, (though is not officially supported by the program). Wang is

working, in part, on SMC methods for dynamic graphical models with Carvalho and West, has led discussions on the topic at the BD&DC meetings, presented a poster at the February 2009 mid-program workshop, and will present this work at the SAMSI workshop on *Adaptive Design, Sequential Monte Carlo and Computer Modeling* in April 2009 as well as in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009.

Other Activities

- BD&DC short-term visitor Andrew Thomas (St. Andrews University) has, as a result of discussions and interactions in the BD&DC group, initiated development of new software for SMC based on the OpenBUGS software, and this is expected to be developed and set up as Open Source software.

Manuscripts Manuscripts linked to research in BD&DC and acknowledging SAMSI/NFS support:

1. D.B. Dunson & S. Das (2009) Bayesian distribution regression via augmented particle learning. In preparation.
2. C. Ji & M. West (2009) Spatial dynamic mixture modelling for unobserved point processes and tracking problems. Initial draft completed.
3. F. Liu, F. Li & D.B. Dunson (2009) Adaptive design for variable selection in normal linear models. In preparation; submission expected in late spring 2009.
4. F. Liu & M. West (2009) A dynamic modelling strategy for Bayesian computer model emulation, *Bayesian Analysis*, **4**(2), - .
5. I. Manolopolou, C. Chan & M. West (2009). Sequential selection sampling for focused inference. In preparation.
6. C. Mukherjee & M. West (2009) Sequential Monte Carlo model fitting and comparison in nonlinear dynamic models. In preparation.
7. R. Yoshida & M. West (2009) Sparse Bayesian inference by annealing entropy. Draft completed and under revision for submission to *Journal of Machine Learning Research*; submission expected in late spring 2009.

Grants

1. A pending NSF grant (January 2009 submission; Li & West Co-PIs) includes proposed computational statistics research that directly come from discussions in this working group related to SMC in nonlinear systems models.
2. A pending NIH grant (February 2009 submission; Dunson PI) on Bayesian methods for assessing gene by environment interactions proposes methodology that relies heavily on SMC approaches that were developed as a result of the working group. Dunson reports that the group has “*had a large impact on the direction of my research and my thinking on how to deal with challenging high-dimensional problems*”.

2.6 Conclusions and expected outcomes

To conclude, the program is progressing well, with much activity across a wide range of topics. We have well-structured Working Groups, holding weekly meetings by Webex. We anticipate that the outcome of the program will be wide dissemination of the results of collaborative research, through published papers. We also plan to produce a journal special issue devoted to the research outputs of the program. The collaborations enabled by the program will no doubt lead to major grant applications in the area of SMC and its scientific application.

3 Summer Program on Meta-Analysis

With the increasing concern in science and medicine for issues such as more complete use of all sources of evidence and reproducibility for single statistical studies, multiple studies and meta-analysis are becoming central to scientific advancement. The Statistical and Applied Mathematical Sciences Institute (SAMSI) addressed this topic through a research program from June 2-13, 2008. It brought together leading statisticians and scientists with interests in meta-analysis, to assess the existing methodology, and develop needed new methodology, and explore pedagogy for bringing the methodology to the broader scientific community.

3.1 Scientific Overview

3.1.1 General Background on Meta-Analysis

Seldom is there only a single empirical research study or source of evidence relevant to a question of scientific interest. However, both experimental and observational studies have traditionally been analyzed in isolation, without regard for previous similar or other closely related studies. A new research area has arisen to address the location, appraisal, reconstruction, quantification, contrast and possible combination of similar sources of evidence. Variously called meta-analysis, systematic reviewing, research synthesis or evidence synthesis, this new field is gaining popularity in diverse fields including medicine, psychology, epidemiology, education, genetics, ecology and criminology.

Statistical methods for combining results across independent studies have long existed, but require renewed consideration, development and wider dissemination by inclusion in the mainstream statistics curriculum. The possibility that the due consideration of all relevant evidence should be accepted as standard practice in statistical analyses deserves investigation. The combination of results from similar studies is often known simply as 'meta-analysis'. Common examples are combining results of randomized controlled trials of the same intervention in evidence-based medicine; of correlation coefficients for a pair of constructs measured similarly across studies in social science; or of odds ratios measuring association between an exposure and an outcome in epidemiology. More complex syntheses of multiple sources of evidence have developed recently, including combined analyses of clinical trials of different interventions, and combined analysis of data from multiple microarray experiments (sometimes called cross study analysis). For straightforward meta-analyses, general least-squares methods may be used, but for complex meta-analyses, the technical statistical approach is not so obvious. Often likelihood and Bayesian approaches provide very different perspectives; and in practice the possible benefits of more complex approaches

may be hard to discern as many meta-analyses are compromised by limited or biased availability of data from studies as well as by varying methodological limitations of the studies themselves.

The presence of multiple sources of evidence has long been a recognized challenge in the development and appraisal of statistical methods - from Laplace and Gauss to Fisher and Lindley. In the 1980s Richard Peto argued that a combined analysis would be more important than the individual analyses, a view taken still further by Greenland and O'Rourke who have suggested that that individual study publications should not attempt to draw conclusions at all, but should instead only describe and report results, so that a later meta-analysis can more appropriately assess the study's evidence fully informed by other study designs and results. Will combined analyses actually replace individual analyses (or at least decrease their impact)? If so, it is time to reexamine the perennial problems of statistical inference in this context.

The concept of multiple sources of evidence itself needs to be generalized and applied more generally and creatively through many areas of statistical research. Multiple sources should not just be taken as separate studies or even the possible simple regrouping of subsets of observations within studies but the bringing to bear of seemingly distinct information sources on given question and even the "creation" of multiple sources as in Bayesian Additive Regression Trees (BART) where differing regression trees are purposefully grown to be later advantageously combined. In some fields, terms like data fusion and data integration are being used for this more general sense of utilizing multiple sources of evidence. A single strategy of no pooling, complete pooling or partial pooling of separate studies perhaps needs to give way to adaptive strategies where the degree of pooling is individually chosen for each and possibly every parameter in the joint probability models used to represent all the relevant sources of evidence.

3.1.2 Specific motivation for the focus of this program

This program comprised two weeks of research, mixing tutorials, research presentations and working group activities on the subject. The goal of this program were three-fold: 1) to bring the area to the attention of statistical researchers, whose expertise is critical to substantiate and clarify the necessary statistical theory and methodology; 2) to nurture the necessary interdisciplinary collaboration and communication between statistical researchers and statisticians who currently work or plan to work with basic and applied science researchers and 3) to provide an entry point into the field to interested students and faculty, and to allow researchers already specialized in the domain to exchange recent results and information.

3.2 Program Structure

3.2.1 Leaders

The program was initiated by Keith O'Rourke. The tutorials in the first week were organized jointly by Vanja Dukic, Ken Rice and Keith O'Rourke. Ingram Olkin opened the program with the lead tutorial followed by Keith O'Rourke, Ken Rice, Vanja Dukic and Julian Higgins. The data analysis sessions were given by Keith O'Rourke and Ken Rice. The working group leaders chosen for the second week of workshops were Dalene Stangl, Ken Rice, Vanja Dukic, Julian Higgins, Keith O'Rourke and David Dunson.

3.2.2 Program Attendance

The program attracted 44 participants in the first week, which featured tutorials and data analysis workshops (see below). A total of 31 participants either continued from the first week or joined the program in the second week; the second week involved working groups and a final summary session (see below). All in all, 55 participants attended some portion of the program.

3.2.3 Tutorials and Opening Workshop

The introductory overview was given by Ingram Olkin, Stanford University. It provided participants with an elementary but thorough introduction to the challenges and opportunities of dealing with multiple studies in the context of biomedical and social science research. Many real application examples were covered to illustrate basic and advanced methods and highlighted the numerous scientific issues and challenges that inevitably arise.

An introductory overview on the likelihood basis for multiple data sources was then given by Keith O'Rourke, Duke University. This provided participants with an elementary introduction to working directly with likelihoods to contrast and combine data based information. This was done both for individual observations - where individual observation likelihoods were contrasted and combined to obtain the usual study estimates - and for studies where study level likelihoods were contrasted and combined. A general meta-analysis approach was then presented in terms of the contrast and combination of likelihoods. Various problems that arise with likelihoods in meta-analysis were then discussed. These problems are largely due to fact that although likelihood concentrates for common parameters (of interest) it expands in dimension for arbitrary (nuisance) parameters and unfortunately there usually are many of these arbitrary (nuisance) parameters.

The advanced tutorials were then started the next day with Keith O'Rourke more thoroughly reviewing the likelihood approach and underlining issues of sparseness with the historical Neyman-Scott examples. Vanja Dukic then covered the integrated likelihood approach as well as some preliminaries for a Bayesian approach. Ken Rice then covered the conditional likelihood approach both from a classical and Bayesian perspective as well as providing material on exchangeability and other more general aspects of meta-analysis. Following this, Vanja Dukic and Ken Rice more fully covered Bayesian approaches to meta-analysis. The tutorial sessions ended with Julian Higgins giving a thorough overview of current (practical) challenges in undertaking meta-analyses in clinical research.

Inter-dispersed with these tutorials, Keith O'Rourke gave a "Data Analysis Session" on likelihood calculations in R and Ken Rice gave one on implementing Bayesian meta-analyses in WinBUGs.

3.3 Working Groups

In the second week, working groups were formed based on the participant research interests. There were six working groups formed comprised of 31 participants, with group sizes ranging from 7 to 17. The Working Groups were

1. Decision theory
2. The role of priors for bias and random effects
3. Bias modeling and information from observational studies
4. ROC and survival analysis
5. Networking, multiple treatments and multivariate
6. Genetics

Here is a summary of their activities.

3.3.1 Decision analysis group

During the week this group explored two questions:

1. In reporting estimates of treatment effect and heterogeneity, is there a loss function for which usual estimates reported are optimal?

2. In non-inferiority trials, how does one choose the delta by which a new treatment is considered "good-enough" relative to the standard treatment and placebo [this question was motivated by an FDA inquiry to the program].

The group studied what is currently done in non-inferiority trials, discussed the difference between random and fixed effects, raised and discussed a concern about only looking at inter-study variability in average treatment effect rather than also being concerned with within study treatment effect variance in choosing between drugs and discussed how "ideally" one would like to address the problems versus how one can take what is currently done and make an improvement that has a chance of being implemented. At the end there seemed to be a consensus that it was necessary to be clear about what the "real" question was and for exactly what "population" so that a full and complete modeling of the decision and its relevant consequences could be undertaken.

As a result of the discussions, some members of the group have written a technical report that has been submitted for publication but is still under review. The reference of the paper is: E. Moreno, F. J. Giron, F.J. Vazquez-Polo and M.A. Negrin (2008). Optimal decisions in cost-benefit analysis. Tech. Report. Dpt. Statistics, University of Granada.

3.3.2 Role of priors for bias and random effects group

This group focused on priors for random effects and bias - two areas in meta-analysis where there is usually a small amount of sample information and hence the choice of priors can be critical.

The discussions around random effects necessarily started with the choice of the parametric distribution of random effects – meant to represent the physical variation in effects from study to study – and then priors for the parameters in these distributions and then non-parametric approaches to random effects. In this group, roughly as in the Decision Analysis group, it was found necessary to be clear about what the "real" question was, what the random effect distribution was meant to represent and exactly what parameters were of inferential interest. A quick review of some current choices for priors for random effects seen in the meta-analysis literature was also undertaken.

The discussion of priors for biases, such as may vary with varying assessed study quality, largely revolved around the possibilities of obtaining empirically motivated priors from the empirical literature. A possibly relevant data set of clinical research studies with various methods of appraising their quality was acquired and a method for investigating quality effects identified in a paper by Greenland and O'Rourke. This likely will become a student project in the near future.

Currently, the group leader, Ken Rice is working on a paper with Keith Abrams entitled Estimating population-averaged contrasts under exchangeability; the role and influence of random-effects distributions.

3.3.3 Bias modeling group

The bias modeling group undertook the challenge of issues and methods for the contrast and combination of biased and confounded sources of information. There ended up being a focus on two main topics - 1) propensity score issues and methods in multiple observational studies and 2) investigations of bias modeling using both RCTs and Non-Rcts together.

The propensity score focus ended up involving two projects, Project A where there was individual-level data available from multiple observational studies and Project B where was only study level data available. Project A , was lead by Elizabeth Stuart of John Hopkins University and B was lead by Robert Platt of McGill University. The motivating question for A was with regard to how propensity scores should be estimated in this setting and the motivating question for B was with regard to whether or not and if so – how propensity score-based subclass estimates from the multiple studies should be combined to get an overall estimate of the effect. Elizabeth Stuart and Robert Platt have since been collaborating on these projects and anticipate involving students in the future.

The investigations of bias modeling using both RCTs and Non-Rcts was lead by Dan Jackson of MRC Cambridge and involved the adaption/extension of methods developed by Steyerberg and the motivating question was with regard to explicating the necessary assumptions and critically assessing their appropriateness. Dan Jackson is continuing to work on the adaption of the Steyerberg method and its extensions and in related work with the Fibrinogen Studies Collaboration [published in *Statistics in Medicine* (2009) – Systematically missing confounders in individual participant data meta-analysis of observational cohort studies]; he has found the discussions at SAMSI useful in his thinking further about applying Steyerberg type methods.

Also of note, Elizabeth Stuart – partly as a result of the SAMSI meeting – is planning to organize a 2010 JSM Invited Session on methods for assessing generalizability.

3.3.4 ROC and survival analysis group

This working group addressed the issues of synthesizing evidence from independent studies about diagnostic test accuracy or survival times – both of which entail individual study and pooled curves or distributions. Both parametric and non-parametric approaches were of interest.

They currently have one paper in preparation, with Jean-Francois Plante of University of Toronto, Vanja Dukic of Chicago University, David Dunson of Duke University and possibly Dalene Stangl of Duke University on Bayesian non-parametric meta analysis of ROC curves. The abstract is as follows: Most standard meta-analytic methods combine information on single parameter, such as treatment effect. For meta-analysis of diagnostic test accuracy, measures of both sensitivity and specificity from different trials are of meta-analytic interest, summarized as a bivariate measure of accuracy, or possibly as a receiver operating characteristic (ROC) curve. Motivated by an analysis of serum progesterone tests for diagnosing non-viable pregnancy, we develop simple fixed-effects and random-effects summary ROC curve estimators, based on a flexible density estimation technique. We compare the performance of the new estimator to the simpler bivariate normal summary ROC estimator.

3.3.5 Network meta-analysis group

Network meta-analysis refers to the situation in which studies brought together for synthesis have compared different subsets from a finite collection of treatments. By exploiting ‘chains’ of evidence, such as making inference on treatment A vs treatment B by contrasting studies of A vs C with studies of B vs C, a network of interrelationships among the studies is created. These meta-analyses are often, and perhaps more appropriately, called multiple treatments meta-analyses (MTM), or mixed treatment comparisons (MTC) meta-analyses.

The working group tackled a variety of problems associated with network meta-analysis. Particular progress was made on methods for illustrating the network graphically. If every study makes a pair-wise comparison – i.e. includes exactly two treatments – then simple graphs with nodes for treatments and lines for comparisons are sufficient to represent the dataset. However, if some studies include three or more treatments, as is typically the case, then such representations do not adequately illustrate the important difference between within-study (direct) comparisons and across-study (indirect) comparisons. In this case, comparisons that come from the data are not independent. The group proposed a diagram in which the distinction is made by using separate lines or shapes for different study designs. Since the workshop, some progress has been made in using graph theory to examine ‘loops’ of evidence in the network.

The importance of separating direct from indirect evidence is largely in order to investigate whether the network of evidence is coherent. Coherence is defined informally as mismatch between direct and indirect sources of evidence, or between two different indirect sources of evidence, on any particular comparison. It is a special kind of heterogeneity between studies that focuses on between-design differences rather than between-study differ-

ences. Two statistical methods for tackling incoherence have been proposed, by T. Lumley (Network meta-analysis for indirect treatment comparisons, *Stat Med* 2002; 21: 2313-2324) and Lu and Ades (Assessing evidence inconsistency in mixed treatment comparisons, *JASA* 2006; 101: 447-459). The former adds a random effect across all studied pair-wise comparisons and tests whether the variance of this random effect is zero. The Lu and Ades approach adds a random effect across each independent evidence cycle, and tests whether the variance of this random effect is zero. There are fewer independent evidence cycles than there are comparisons. However, counting the number of independent evidence cycles is not trivial when there are multi-arm studies. The group discussed other approaches, such as fitting a model than assumes coherences and comparing deviances with a ‘free’ model that makes no assumptions about chains of evidence. Three of the workgroup members (Dan Jackson, Jessica Barrett, Julian Higgins; working with Ian White) have a paper in preparation about some of these ideas.

The working group also discussed technical issues about making inferences in network meta-analyses. Restricted maximum likelihood is often used; Lumley uses the function `lme` in R with a slightly unusual construction for random-effects variances. Inference is less straightforward with multi-arm trials or logistic models. We explored profile likelihood, and inverting the observed information matrix. Plans were made to investigate the use of conditional likelihood, integrated likelihood, and inverting expected information matrix.

3.3.6 Meta genetics group

This working group address issues of multiple sources of evidence for genetics, focusing on Gene Expression Meta Analysis, Meta Analysis for Genetic Association Studies and Accounting for Dependence in High-Dimensional Predictors. Since this summer’s program, the meta-genetics working group has been quite productive. The active core of this group consists of David Dunson at Duke University , Fei Zou at UNC Biostatistics and Fei Liu at the University of Missouri Columbia. They have submitted the following paper to *Biometrics*: Liu, F., Dunson, D.B. and Zou, F. (2008). High-dimensional variable selection in meta analysis for censored data. *Biometrics*, submitted. In addition, they have another paper under way: Liu, F., Dunson, D.B. and Zou, F. (2009). Annotated relevance vector machine with application to polymorphism selection. In preparation.

Their following summary highlights some of the work undertaken to date which represents the most exciting research happening in the program and provides a nice example of both a generalized concept of multiple sources of evidence and the replacement of a single strategy of no pooling, complete pooling or partial pooling of studies with an adaptive strategy where

the degree of pooling is individually chosen for different coefficients.

In large scale genetic epidemiology studies that collect massive numbers of single nucleotide polymorphisms (SNPs) or gene expression measurements, it is extremely challenging to identify genes that are predictive of disease phenotypes given the modest sample size of most studies relative to the number of genes. Due to concern about false positive rates, it is crucial to replicate findings about disease genes in multiple studies. Standard approaches take multistage testing approaches in which one tests if genes identified in initial studies are significant in follow-up studies. This strategy is shown to have major disadvantages in terms of power and type I error rates compared with an innovative approach developed in the SAMSI meta-genetics working group based on simultaneous selection through a multi-task relevance vector machine (MT-RVM) procedure. This approach, which is related to methods used in signal processing, borrows information across studies in the degree of shrinkage of gene-specific coefficients towards zero. The method is scalable to large numbers of genes, can accommodate censored data commonly collected in disease recurrence studies, and clearly outperforms common competitors, such as Lasso. In addition, the meta-genetics group is currently pursuing a new procedure that allows information on gene function annotation to be incorporated, while automatically learning how predictive each annotation source is. The annotated relevance vector machine (aRVM) procedure should be very widely useful in machine learning and other applications beyond genetics, as it allows an adaptive targeted search for important predictors enabling an effective reduction in dimensionality and mechanism for borrowing information across disparate studies.

3.4 Post Program Activities

1. At the Eastern North American Region 2009 meeting of the International Biometric Society most of the working group leaders and some of the participants presented their research. In particular, a session “Advances in Meta-Analysis” was organized, based on the program, with presentations by Eloise Kaizar, Robert Platt, Vanja Dukic, and Dalene Stangl.
2. Professional Courses:
 - Keith O’Rourke gave a two day course on meta-analysis for Statisticians and Students at the University of Alberta in July 2008.
 - Keith O’Rourke gave an Advanced Meta-analysis Short Course at the University of Alberta.

4 Education and Outreach Program

The SAMSI Education and Outreach (E&O) Program encompasses a variety of activities which have achieved national stature for both their scientific and pedagogical content. The annual activities include two-day Undergraduate Outreach Days held both in the Fall in the Spring, a week-long Undergraduate Workshop (UGS) held in May, and the ten-day Industrial Mathematical and Statistical Modeling (IMSM) Workshop for Graduate Students that is held at the end of July. In 2008, SAMSI also hosted the Blackwell-Tapia conference.

4.1 Undergraduate Outreach Days

The two outreach workshops are held annually to expose undergraduates from programs around the country to topics and research directions associated with concurrent SAMSI programs. One goal of these workshops is to illustrate the application and synergy between mathematics and statistics which goes far beyond that which students have seen in coursework. The overall objective is to broaden the perspective of students with regard to both future graduate studies and career choices.

The workshop has evolved through the project. In years 2002-03, 2003-04, and 2004-05, technical presentations directly related to the all on-going SAMSI programs were typically given, together with various tutorials, demos and hands-on activities. While the latter type of activities have been retained, starting in Fall 2005, each workshop has been specifically dedicated to one of the two on-going SAMSI programs for that year. Members of the directorate and SAMSI postdocs typically meet with the students during over dinner one of the workshop to discuss graduate and career opportunities.

4.1.1 Sequential Monte-Carlo Methods

The Fall outreach workshop, held October 31-November 1, 2008, focused on topics from the SAMSI Program on Sequential Monte Carlo methods. The students were provided with an overview of SAMSI by Pierre Gremaud (SAMSI-NCSU) after which program leaders, participants, postdocs and students gave a variety of presentations and tutorials. During the Friday morning session, Jochen Voss (Warwick University) gave a general introduction to Sequential Monte Carlo methods. Gentry White (SAMSI postdoc) then led a tutorial on R. This was followed by two applied presentations where the students were shown how the type of methods under study can be used in practice: Christian Macaro (SAMSI postdoc) discussed stochastic volatility in Finance while Nathan Green (Defense Science and Technology, UK) introduced a problem of tracking the position of a toxic cloud released in an

urban setting. That application was also the object of an interactive R session in the afternoon overseen by Nathan Green, Francesca Petralia (SAMSI graduate fellow from Duke) and Gentry White. Two additional SAMSI postdocs, Julien Cornebise and Sourish Das gave presentations on respectively tracking applications (submarines and planes) and dynamic models. The day was concluded by an open discussion led by Pierre Gremaud on graduate school and career options. During dinner on Friday, members of the Directorate as well as SAMSI visitors and postdocs interacted with students to further discuss career opportunities. Two presentations were given on Saturday morning by Jaya Bishwal (UNC, Charlotte) on stochastic quadratures and financial applications and by Ionna Manolopoulou (SAMSI postdoc) on rare event detection. The workshop was concluded by a MATLAB tutorial given by Chunlin Ji (SAMSI graduate fellow from Duke) and a MATLAB interactive session on financial applications led by Jaya Bishwal together with Melanie Bain, Sarah Schott and Minghui Shi, all SAMSI graduate fellows. Details regarding the workshop can be obtained at <http://www.samsi.info/workshops/2008ug-workshop200810.shtml>.

There were 24 participants which included 12 females, 2 African Americans and 2 Hispanics.

4.1.2 Algebraic Methods in Systems Biology and Statistics

The Spring outreach workshop, held February 27-28, 2009, focused on Algebraic Methods in Systems Biology and Statistics. Following an overview of SAMSI by Pierre Gremaud (SAMSI-NCSU), two general introductory presentations were given, one by Brandy Stigler (Southern Methodist University) on Systems Biology, the other by Seth Sullivant (NCSU) on Algebraic Statistics. Also that morning, Gentry White (SAMSI postdoc) led a tutorial on R. The afternoon started with three connected presentations on algebraic statistical models and design of experiments by Luis Garcia-Puente (Sam Houston State University), Ian Dinwoodie (Duke) and Giovanni Pistone (Politecnico di Torino, Italy). This was followed by an interactive exploring the themes of these three lectures more in depth. The session was led by Ian Dinwoodie, Giovanni Pistone as well as Ben Wells (SAMSI graduate fellow from NCSU) and Saied Yasamin (SAMSI postdoc). The students conducted short studies using R and the package SINGULAR. The afternoon was concluded by an open discussion led by Pierre Gremaud on graduate school and career options. During dinner on Friday, as with the Fall outreach workshop, members of the Directorate as well as SAMSI visitors and postdocs interacted with students to further discuss career opportunities. The theme of the Saturday morning session was Phylogenetics. Jeff Thorne (NCSU) gave the first lecture on Evolutionary Biology. Megan Owen (SAMSI postdoc) then gave a presentation on phylogenetic trees which

was also the theme the interactive session she led following her talk. Jason Yellick (SAMSI graduate fellow from NCSU) helped tutoring the session. Details regarding the workshop can be obtained at <http://www.samsi.info/workshops/2008ug-workshop200902.shtml>.

There were 28 participants which included 8 females, 1 African American, and 2 Hispanics.

4.2 Undergraduate Workshop

The one-week SAMSI Workshop for Undergraduates, held May 18-22, 2009, focused on mathematical and statistical topics pertaining to inverse problems. During the initial sessions, students are introduced to various physical applications and mathematical concepts. Both mathematical and statistical models are typically derived for prototypical systems, and significant attention is focused on estimating material parameters from measured data. The tutorials include substantial exposure to MATLAB and routines for numerical integration and optimization. On the final day of the workshops, each student team presents the results obtained during the week.

The Undergraduate Workshop encompasses three highly unique components.

- All tutorials and sessions are presented by SAMSI graduate students and postdocs under close supervision of a member of the Directorate and local faculty. This allows the undergraduates to interact with peers within educational and research programs they are considering and it provided valuable experience for the presenters, many of whom are considering academic careers.
- The workshop provides students with an intensive introduction to the synergy between applied mathematics and statistics within the context of timely physical applications.
- The students are introduced to a variety of experiments and each team collects their own experimental data. This exposure to data collection illustrate both the physical basis for models and various mechanisms yielding uncertainty or noise. Whereas a number of aspects are listed as highly positive in exit evaluations, the laboratory experience is one of the most highly ranked experiences.

Full documentation regarding the workshop including the presentations, tutorials, software and student presentations can be found at <http://www.ncsu.edu/crsc/events/ugw09/index.php>.

There were 18 participants to the workshop, including 12 females and one hispanic female.

4.3 Industrial Mathematical and Statistical Modeling (IMSM) Workshop

The ten-day Industrial Mathematical and Statistical Modeling Workshop for Graduate Students is currently in its 15th year; the last 8 of these workshop have been supported by SAMSI. The overall goals of the workshop are twofold: (i) expose mathematics and statistics students to current research problems from government laboratories and industry which have deterministic and stochastic components, and (ii) expose students to a team approach to problem solving. During the workshop, the students learn to communicate with scientists outside their discipline, allocate tasks among team members, and disseminate results through both oral presentations and written reports.

Typically, about 40 students participate in the workshop. The attendees are divided into 6 or 7 teams to investigate current research problems presented by scientists and engineers from outside the academic world. Each team gave a 30 minute oral presentation summarizing their results on the final day of the workshop and written reports are compiled.

Both the undergraduate and the graduate workshop share achieve the following goals with respect to intellectual merit:

- Students gain experience in team work. Team work is indispensable in the approach to problem solving, in producing a final written report, and in preparing an oral presentation.
- The students learn to communicate with scientists who are not academic mathematicians.
- The workshops present a unique combination of applied mathematics and statistics that is not part of the usual class work.

The IMSM workshop goes further:

- Students work on genuine industrial research problems. These are not the kind of academic exercises often considered in classrooms. The projects tend to be open-ended and require fresh new insight for both formulation and solution. Sometimes the biggest challenge is to figure out what the real problem is. The students also learn how to derive a useful result under a tight deadline.
- Students acquire crucial insight into the aspects of a non-academic career. Some presenters may know more about their problems and can guide the students away from dead

ends, while other presenters may have brought open-ended problems and are searching along with the students. This combination of approaches exposes the students to the variety of challenges facing scientists in industry.

- The IMSM workshop helps students to decide what kind of career they want. The IMSM workshop provides a unique experience of how mathematics and statistics are applied outside academia. In some cases the help has been in the form of direct hiring by the participating companies.

The broad impact of our Education and Outreach activities is substantial:

- Our workshops help to attract students to and prepare them for a non-academic career, by exposing them to real-world industrial problems.
- The participating students represent a nationally diverse group, with a substantial number of women and minorities.
- The workshops strengthen the interaction between applied mathematics and statistics.
- The workshops, and specifically, IMSM, benefit government and industry research. Often the student teams come up with useful solutions to a project. Several projects initially presented at the IMSM workshop have resulted in long term collaborations between students and faculty on the one side and the companies on the other. Furthermore, several companies have taken advantage of the recruitment opportunity provided through direct contact with some of the most talented students in the mathematical and statistical sciences. Many companies, large and small, have shown continued interest and enthusiasm about the IMSM workshop.

The latest in that series of workshops took place on July 20-28, 2009 at NCSU. The problem presenters were

- Erik Gilleland, National Center for Atmospheric Research,
- John Langstaff, Environmental Protection Agency,
- Jordan Massad, Sandia National Laboratories,
- Frank Meyer, Republic Mortgage Insurance Company,
- John Peach, MIT Lincoln Laboratory,
- Michael Wagner, UNC School of Pharmacy.

Complete information is available at <http://www.ncsu.edu/crsc/events/imsm09/>.

4.4 Other Events

4.4.1 Blackwell-Tapia Conference

This conference, held November 14-15, 2008, was the fifth in a series of biannual conferences honoring David Blackwell and Richard Tapia, two seminal figures who inspired a generation of African-American, Native American and Latino/Latina students to pursue careers in mathematics. Carrying forward their work, this one and a half day conference

- recognized and showcased mathematical excellence by minority researchers,
- recognized and disseminated successful efforts to address under-representation,
- informed students and mathematicians about career opportunities in mathematics, especially outside academia,
- provided networking opportunities for mathematical researchers at all points in the higher education/career trajectory.

The conference included a mix of activities: scientific talks, poster presentations and panel discussions. On Friday afternoon, lectures were given by Jacqueline Hughes-Oliver, a noted statistician at NCSU and Freda Porter who is among the small number of American Indian women who have earned a Ph.D. in mathematics. Porter is President and CEO of Porter Scientific, Inc., in Pembroke, North Carolina. Tim Thorton (University of California, San Francisco) and Angela Gallegos (Tulane University) gave shorter talks. An energetic and successful panel discussion took place on getting undergraduates involved in research. The panelists were Carlos Castillo-Chavez (Arizona State University), Reinhard Laubenbacher (Virginia Tech), Juan Meza (Lawrence Berkeley National Lab), Peter Mucha (University of North Carolina at Chapel Hill) and Michael Shearer (NCSU). The day was concluded by a reception and a poster session. On Saturday morning, lectures were presented by Oscar Gonzalez (University of Texas, Austin) on DNA analysis and Gabriel Huerta (University of New Mexico) on climate models. Rudy Horne (Florida State University), Yolanda Munoz Maldonado (Michigan Technological University), Ulrica Wilson (Morehouse College) and Tanya Moore from Building Diversity in Science presented short talks during the day. Opportunities at the Mathematical Institutes and NSF were also discussed through a presentation; contributors included Jim Berger (SAMSI), Cheri Shakiban (IMA) and Peter March (NSF). A panel discussion on career opportunities in the mathematical sciences took place in the afternoon. The panelists were Carolyn Morgan (Hampton University), Tanya Moore (Building Diversity in Science), Bob Rodriguez (SAS), Nell Sedranks (NISS) and Janet Spoonamore

(ARO). After the panel, Richard Tapia (Rice University) gave a lecture on Optimization and the central place it occupies in contemporary mathematics.

The Blackwell-Tapia Lecture was delivered by Juan Meza who discussed various theoretical and practical issues related to the general field of optimization. Dr. Meza has an exceptionally distinguished record as a mathematical scientist, an accomplished and effective head of a large department doing cutting-edge explorations in the computational sciences, computational mathematics, and future technologies, and a role model and active advocate for others from groups under-represented in the mathematical sciences. In recognition of his numerous achievements, the National Blackwell-Tapia Committee awarded Dr. Meza with the 2008 Blackwell-Tapia Prize. The Conference ended with a reception during which Juan Meza received the award. Further information about the Conference can be found at <http://www.samsi.info/workshops/2008Blackwell-Tapia.shtml>.

There were 62 attendees at the Conference, including 24 females, 26 African Americans and 27 Hispanics.

4.4.2 3rd Annual Graduate Student Conference in Probability

This is the third of a series of probability conferences developed and run by graduate students in probability and statistics. It is sponsored by SAMSI and jointly hosted by the Mathematics Department at Duke University and the Department of Statistics and Operations Research at University of North Carolina, Chapel Hill. The organizing committee members are Changryong Baek, Jessi Cisewski, Xin Liu, Dominik Reinhold, Tiffany Kolba and Rachel Thomas under the supervision of Prof. Amarjit Budhiraja and Prof. Jonathan Mattingly.

The conference objectives are to

- Provide graduate students and postdoctoral fellows with the opportunity to speak on an area of interest within probability; there were 50 such talks at the workshop.
- Foster discussions with a friendly and informal atmosphere.
- Establish connections for potential future collaborations.
- Provide an introduction to recent developments in probability from keynote speakers, David Aldous (UC, Berkeley), Russell Lyons (Indiana University), and Daniel Stroock (MIT).

4.5 Courses

Two courses were offered during the Fall semester in 2008; these are credited 3 credits/units at each of the participating Universities.

Algebraic Methods was linked to the program on Algebraic Statistics and Systems Biology. It was taught by Seth Sullivant (NCSU) and Reinhard Laubenbacher (Virginia Tech.). This course provided an introduction to the algebraic techniques that have emerged as useful tools in biology and statistics. This course was intended to bridge the gap between abstract algebra and application areas in biology. After providing an introduction to polynomial rings, ideals, and Grobner bases, a range of applications were surveyed, among them: polynomial dynamical systems over finite fields and applications, graphical and hierarchical models, Markov bases for contingency table analysis, phylogenetic models and the space of trees, applications of tropical geometry in MAP estimation.

Sequential Monte Carlo Methods was the course linked to the program of the same name. It was taught by Arnaud Doucet (University of British Columbia) and several guest lecturers. The objective was to provide a complete overview of the SMC field. The instructors covered the basics of Monte Carlo methods, importance sampling, sequential importance sampling, auxiliary methods, resampling techniques as well as the most recent adaptive methods. SMC methods were illustrated on a variety of application areas including optimal estimation for non-linear non-Gaussian state-space models, sequential and batch Bayesian inference, computation of p-values, inference in contingency tables, rare event probabilities, optimization, counting the number of objects with a certain property for combinatorial structures, computation of eigenvalues and eigenmeasures of positive operators, PDE's admitting a Feynman-Kac representation and so on. The students were also provided with an introduction to the theory of SMC.

4.6 Diversity

See Section I.H for discussion of the efforts to promote diversity.

F. Industrial and Governmental Participation

Government and industry participation in SAMSI program and activities reflects broad interest in the SAMSI vision. Most SAMSI workshops had extensive participation by individuals from industry and government. Here, we summarize only the more intensive involvements, e.g., participation of such individuals in program working groups.

Risk Analysis, Extreme Events and Decision Theory: This program had working group members from IBM, the Center for Disease Control and Prevention (CDC), and NCAR. Contact was also made to Genesys Lab of Alcatel-Lucent to obtain data for testing methodology.

Environmental Sensor Networks: This program had working group members from government agencies, laboratories, and industry, including EPA, CDC, Marine Biological Laboratory, the IBM Watson Research Center, and the National Institute for Space Research (Brazil).

Sequential Monte Carlo Methods:

The Tracking working group had close interactions with the UK's DSTL governmental defense organization, with Nathan Green being on secondment from there. Mark Briers was on secondment from QinetiQ Ltd. for his visit in Fall 08.

Algebraic Methods in Systems Biology and Statistics:

We had a number of participants who were from government agencies and industry: Lawrence Cox (Amgen), Gilles Gnacadja (Amgen), and Richard Haney (Cellular Statistics).

Education and Outreach Program: In the *Industrial Mathematical and Statistical Modeling Workshop*, the attendees were divided into 6 teams to investigate current research problems presented by scientists from Glaxo Smith Kline, MIT Lincoln Laboratory, the National Institute of Statistical Sciences, Republic Mortgage Insurance Co and SAS.

G. Publications and Technical Reports

1. Random Media

Publications and Technical Reports

- Beale, J.T., D. Chopp, R. J. LeVeque, and Z. Li “*Correction to: ”A Comparison of the Extended Finite Element Method with the Immersed Interface Method...*” [CAMCoS 1 (2006), 207–228]
- Cai, Q., Wang, J., Zhao, H., Luo, R. “*On Removal of Charge Singularity in Poisson-Boltzmann Equation*”, to appear in Journal of Chemical Physics. 2009
- Demanet, L., Gabriel Peyre “*Compressive Wave Computation*”, submitted, 2008
- Fricks, J., Yao, L., Elston, T., Forest, M.G. “*Time-domain Methods for Passive Microrheology and Anomalous Diffusive Transport in Soft Matter*”, SIAM J. Appl. Math., Vol. 69(5), 1277-1308 (2009).
- Hill, D.B., Lindley, B., Forest, M.G., Superfine, R., Mitran, S. “*Experimental and Modeling Protocols for a Micro-parallel Plate Rheometer*”, UNC preprint, to be submitted. 2009
- Hohenegger, C., Forest, M.G., “*Two-point Microrheology, II: Simulation Protocols*”, UNC-NYU preprint, to be submitted. 2009
- Hohenegger, C., Forest, M.G., “*Modeling Aspects of Two-bead Microrheology, Proceedings of XVth International Congress on Rheology*”, Springer, August, 2008, AIP Conference Proceedings, Materials Physics & Applications Series, Vol. 1027 (2008).
- Hohenegger, C., Forest, M.G., “*Two-point Microrheology: Modeling Protocols*”, Phys. Rev. E 78, 031501 (2008).
- Hohenegger, C., Forest, M.G., “*Direct and Inverse Modeling for Stochastic Data in Microbead Rheology, Proceedings in Applied Mathematics and Mechanics (PAMM)*”, Special Issue: Sixth International Congress on Industrial Applied Mathematics (ICIAM07) and GAMM Annual Meeting, Zurich 2007, Published Online: Oct 30 (2008).
- Hou, S., K. Huang, K. Solna, and H. Zhao “*Multi-Tone Imaging*”, submitted, 2008

- Howell, E., Smith, B., Rubinstein, G., Forest, M.G., Lindley, B., Hill, D., Superfine, R., Mitran, S. “*Stress Communication and Filtering of Viscoelastic Layers in Oscillatory Shear*”, J. Non-Newtonian Fluid Mechanics, Vol. 156, 112-120 (2009).
- Huang, K., K. Solna, and H. Zhao “*Generalized Foldy-Lax Formulation*” submitted, 2008
- Ito, K., M. Lai, Li, Z. “*A Well-conditioned Augmented System for Solving Navier-Stokes Equations in Irregular Domains*”, J. Comput. Phys. (2009), doi:10.1016/j.jcp.2008.12.028.
- Jiang, Q., Li, Z., Lubkin, S. “*Theoretical & Numerical Analysis for a Fluid Mixture Model of Tissue Deformation*”, Comm. in Comput. Phys. Vol. 3, 620-634, 2009.
- Leung, S., Zhao, H. “*A New Grid-Based Particle Method for Interface Problems*”, Journal of Computational Physics, Volume 228, Issue 8, 2009.
- Leung, S., Zhao, H. “*A Grid Based Particle Method for Evolution of Open Curves and Surfaces*”, UCLA-CAM 08-72. Submitted. 2009
- McKinley, S.A., Yao, L., Forest, M.G. “*Transient Anomalous Diffusion of Tracer Particles in Soft Matter*”, Duke-UNC preprint, to be submitted. 2009
- Mitran, S., Forest, M.G., Lindley, B., Yao, L., Hill, D. “*Extensions of the Ferry Shear Wave Model for Active Linear and Nonlinear Microrheology*”, J. Non-Newtonian Fluid Mechanics Vol. 154:120-135 (2008).
- Tsynkov, S., “*On SAR Imaging Through the Earth Ionosphere*”, SIAM Journal on Imaging Sciences, 2 (2009) No. 1, pp. 140–182.
- Wan, X., Li, Z., Lubkin, S. “*Mechanics of Mesenchymal Contribution to Cleaving Force in Branching Morphogenesis*”, Biomechanics and Modeling in Mechanobiology, Vol. 7, 417-426, 2008.
- Wang, J., Cai, Q., Li, Z., Zhao, H.K., Luo, R. “*Achieving Energy Conservation in Poisson-Boltzmann Molecular Dynamics: Accuracy and Precision with Finite-Difference Algorithms*”, Chemical Physics Letters, Volume 468, Issues 4-6, 22 January 2009, Pages 112-118.
- Xie, H., Ito, K., Li, Z., Toivanen, J. “*A Finite Element Method for Interface Problems with Locally Modified Triangulations*”, AMS Contemporary Mathematics, Vol. 466, 2008, 179-190.

- Zhong, W. “Energy-preserving and Stable Approximations for Two-dimensional Shallow Water Equations” Submitted to the proceedings of the Abel Symposium 2006, Springer.

Reports in Preparation

- Fouque J.P., Yvonne Ou “*Time Reversal for Elastic Waves*” in preparation, 2008
- Ito K., et al. “*Multi-valued Stochastic Evolution Equations in Hilbert Spaces and Integrable Solution*” in preparation.
- Klapper, I., and M. Grigoriu, “*Micro- and Macro-Scale Material Properties of Heterogeneous Viscoelastic Fluids*” in preparation.
- Zhong, W. “*Parallel Implementation of Material-point Method for Linear Visco-elastic Models*” in preparation.
- Zhong, W., “*High-order Schemes for Generalized Functions in Elliptic Interface Problems*” in preparation.
- Zhong, W., “*High-order Numerical Schemes for 1-D Fluid Mixture Model of Tissue Deformations*” in preparation.

II. RISK ANALYSIS, EXTREME EVENTS AND DECISION THEORY

Publications and Technical Reports

- Cano, J., Rios Insua, D., “*Bayesian Reliability, Repairability and Availability for Hardware Systems through Continuous Markov Chain Models*”, completed.
- Cheng G. and Michael Kosorok “*The Penalized Profile Sampler*” Journal of Multivariate Analysis, 2007 (in review)
- D’Auria, B., Resnick, S.I., “*The Influence of Dependence on Data Network Models of Burstiness*” Cornell University, Tech Report #1449 To appear: Advances in Applied Probability, vol 40, no 1
- Das, S., Dey, D. “*On Bayesian Analysis of Generalized Linear Models: A New Perspective*” Submitted. SAMSI 2007-08
- Das, S., Dey D. “*Analysis of 5 Loxin® Treatment for Patients with Osteoarthritis in Clinical Trial using Power Filter*” Submitted. SAMSI 2008-09

- Das, S., Harel, O., Dey, D., Covault, J., Kranzler, H. “*Analysis of Extreme Drinking in Patients with Alcohol Dependence Using Pareto Regression*” Submitted. SAMSI 2008-10
- Dey, D., Gaioni, E., Ruggeri, F., “*Model Based Prior Elicitation*” (2009)
- Gaioni, E., Dey, D., Grigoriu, M., “*Semiparametric Functional Estimation Using Quantile Based Prior Elicitation*” SAMSI TR2008-06
- Grigoriu, M., Ríos Insua, D., Ríos, J., Shen, H., “*Reduced Order Models for Bayesian Risk Analysis*”
- Kulkarni, V.G., Resnick, S.I., “*Warranty Claims Modelling*” Naval Research Logistics DOI: 10.1002/nav.20287. To appear (2008)
- Li, H., Hosking, J., Jiang, H., “*Environmental Risk Evaluation: a Bayesian Hierarchical Approach for Extreme Temperature over Space and Time*” (2009)
- Nguyen, X., Huang, L., Joseph, A. “*Support Vector Machines, Data Reduction, and Approximate Kernel Matrices*” SAMSI 2008-03
- Nguyen, X., (with Jordan and Wainwright): “*On Surrogate Loss Functions and f -divergences*” Annals of Statistics paper accepted in Feb 08
- Nguyen, X., (with Jordan and Wainwright): “*On Optimal Quantization Rules in Some Sequential Decision Problem*” IEEE Trans on Information Theory paper accepted in January 08
- Nguyen, X., (with Jordan and Wainwright): “*Nonparametric Estimation of the Likelihood Ratio and Divergence Functionals*” IEEE Trans on Information Theory, to be submitted
- Pal, J., Dey, D. “*Bayesian Isotonic Estimation for Exponential Family and Beyond*” Submitted. SAMSI 2008-01
- Pal, J., Banerjee, M. “*Estimation of smooth link function in Monotone response models*” To appear in Journal of Statistical Planning and Inference
- Pal, J. “*Penalized Least Square Regression in Isotonic Regression*” To appear in Statistics and Probability Letters
- Rios Insua, D., Rios, J., Banks, D., “*Adversarial Risk Analysis*” (ARA)
- Rios, J., “*Balanced Increment and Concession Methods for Arbitration and Negotiations*” Paper submitted to Group Decision and Negotiation Journal BIM-BIC (2008)

- Rios, J., “*Supporting Group Decisions over Influence Diagrams*” Paper submitted to Decision Analysis (2008)
- Rios, J., Rios Insua, D., “*Negotiations Over Influence Diagrams*” Completed
- Spiller E.T., and W.L. Kath. “*A Method for Determining Most Probable Errors in Nonlinear Lightwave Systems*” to appear in SIAM Journal on Applied Dynamical Systems (2008)
- Wang X., Dey D., “*A Flexible Skewed Link Function for Binary Response Data*” SAMSI Tech Rep 2008-05
- Wang, X., Dey, D., Banerjee, S., “*Non-Gaussian Hierarchical Generalized Linear Geostatistical Models*” (2009)

Reports in Preparation

- Cano, J., Rios Insua, D., “*Bayesian Reliability Analysis for Hardware/Software Systems*”, almost completed
- Cheng G., “*One-Step M-estimation in Semiparametric Models*” in preparation. (2008)
- Das, S., “*Analyzing Extreme Hurricane Activity using Multinomial-Dirichlet Model*” in preparation
- Gaioni, E., Dey, D., “*Incorporating Expert Opinion into the Joint Modeling of Extreme and Non-extreme Components of River Flow*” sponsored by University of Connecticut, Center for Environmental Statistics and Engineering, in preparation
- Madar, V., “*Bayesian Model Selection for the Generalized FGM Copula in the Bivariate Case when both Marginal Distributions are General Extreme Value*” in preparation
- Madar, V., “*Prior Elicitation in the Bivariate Extreme Value Situation and Some Related Modeling Issues*” in preparation
- Madar, V. “*The Variable-Ratio Simultaneous Confidence Intervals*” in preparation
- Madar, V., Benjamini, Y., and Stark, P.B. “*The Quasi-Conventional Simultaneous Confidence Intervals for Better Sign Determination*” in preparation
- Madar, V. “*The Quasi-Conventional Intervals under Dependence*” in preparation

- Madar, V. “*An Inequality for Multivariate Normal Probabilities of Nonsymmetric Rectangles*” in preparation
- Pal, J. “*Penalized Likelihood Ratio in the Density Estimation Problem*” Invited revision from Scandinavian Journal of Statistics
- Porter, M., “*Discrete Choice Models in Adversarial Risk Analysis*”
- Rios, J., “*Computations in Adversarial Risks*” in preparation
- Rios, J., “*Reduced Order Model for Bayesian Risk Analysis*” in preparation
- Rios, J., “*Bayesian Discrete Event Simulation*” in preparation
- Rios Insua, D., Rubio, J.A. “*Formalisation of Risk Approaches in ICT*” Structured.
- Rios, J., Banks, D. “*Commutativity of Nash Equilibria and Expected Utilities*” Structured and numerical experiments performed
- Ruggeri, F., Wiper, M. Bayesian Analysis of Stochastic Processes
- Spiller, E.T., and G. Biondini “*Importance Sampling for Dispersion Managed Solitons*” in preparation
- Werker, B., Renault, E., “*Causality Effects in Return Volatility Measures with Random Times*” in preparation
- Werker, B., Renault E., “*Appendix to: Causality Effects in Return Volatility Measures with Random Times*” in preparation

III. ENVIRONMENTAL SENSOR NETWORKS

Publications and Technical Reports

- Cardon, Z.G., Flikkema, P., Herron, P.M., Holan, S., Kim, Y., Linder, E., and Stark, J.M. “*A New View of Hydraulic Redistribution of Soil Water During Rainstorms*” To be submitted to Ecology
- Cardon, Z.G., Stark, J. M., Herron, P.M. (2009) “*Hydraulic Redistribution and the Fate of Root-derived Carbon in Soil*” Abstract submitted for Ecological Society of America meetings, August 2009, Albuquerque, NM.

- He, Y. and Flikkema, P. “*System-Level Characterization of Single-Chip Radios for Wireless Sensor Network Applications*” IEEE WAMICON 2009, April 20-21, 2009, Clearwater, FL USA.
- Howard, S. and Flikkema, P. “*Integrated Source-Channel Decoding for Correlated Data- Gathering Sensor Networks*” IEEE Wireless Communications and Networking Conference (WCNC 2008), March-April 2008
- Howard, S. and Flikkema, P. “*Progressive Joint Coding, Estimation and Transmission Censoring in Energy-Centric Wireless Data Gathering Networks*” Fifth IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS 2008), Sept-Oct 2008.
- Kim, Y., “*Modeling Dynamic Controls on Ice Streams: A Bayesian Statistical Approach*” under review on Journal of Glaciology (2008)
- Kim, Y., “*Bayesian Design and Analysis for Superensemble based Climate Forecasting*” in Press, Journal of Climate, V 21, No 9
- Murray, J. “*Median Polish Algorithm for Automated Anomaly Detection in Sensor Networks (MP-Tuner)*” Entry to 2009 Student Computing Competition by the American Statistical Association (Section on Computing and Graphical Statistics).
- Murray, J. “*Median Polish Algorithm for Automated Anomaly Detection in Sensor Networks (MP-Tuner)*” Entry for the 2009 U. of New Hampshire Undergraduate Research Conference. Interactive presentations to be given April 22 and April 24, 2009 (U. of New Hampshire)
- Nguyen, X., Rajagopal, R., Ergen, S., Varaiya, P., “*Distributed Online Simultaneous Fault Detection for Multiple Sensors*” IPSN conference paper accepted for presentation in April 08. Full report to be submitted to IEEE Trans on Signal Processing
- Nguyen, X., Rajagopal, R. “*Theory for Multiple Change-point Sequential Detection*” To be submitted to IEEE Trans on Information Theory.
- Nguyen, X., Huang, L. and Joseph, A. (2008). “*Support Vector Machines, Data Reduction and Approximate Kernel Matrices*” Proceedings of the 19th European Conference on Machine Learning (ECML), September, Antwerp, Belgium.
- Rajagopal, R., Nguyen, X., Ergen, S. and Varaiya, P. “*Theory of Multiple Sequential Changepoint Detection*” To be submitted to IEEE Trans. on Signal Processing.

- Rajagopal, R., Nguyen, X., Ergen, S. and Varaiya, P. (2008). “*Distributed Online Simultaneous Fault Detection for Multiple Sensors*” International Conference on Information Processing in Sensor Networks (IPSN), St. Louis, MO.
- Rajagopal, R., Nguyen, X., Coleri-Ergen, S., and Varaiya, P. (2009). “*Theory of Simultaneous Fault Detection for Multiple Sensors*” Second International Workshop on Sequential Methodologies (IWSM), Troyes, France (invited extended abstract).
- Silberstein, A., Puggioni, G., Gelfand, A., Munagala, K., and Yang, J. “*Suppression and Failures in Sensor Networks: A Bayesian Approach*” Proceedings of the 2007 International Conference on Very Large Data Bases (VLDB '07), Vienna, Austria 2007; 842–853.
- Silberstein, A., Braynard, R., Filpus, G., Puggioni, G., Gelfand, A., Munagala, K., and Yang, J. “*Data-Driven Processing in Sensor Networks*” Proceedings of the 3rd Biennial Conference on Innovative Data Systems Research (CIDR), 2007, Asilomar, California; 10–21.

Reports in Preparation

- Gelfand, A.E. and Puggioni, G. “*Analyzing Space-time Sensor Network Data under Suppression and Failure in Transmission, Statistics and Computing*” (forthcoming).
- Kim, Y., (with L. Mark Berliner) “*A Class of Bayesian State Space Models with Time-Varying Parameters*” in preparation
- Kim, Y. (with L. Mark Berliner) “*Bayesian Diffusion Process Models with Time-Varying Parameters*” in preparation
- Kim, Y., (with L. Mark Berliner) “*Change of Spatiotemporal Scale in Dynamic Models*” in preparation
- Kim, Y., (with L. Mark Berliner) “*Impacts of Approximated Marginal Posterior Distribution of Nuisance Parameter*” in preparation
- Kim, Y., “*Bayesian Inference Based on Superensembles Including Computer Model Experiment Issues*” in preparation
- Kim, Y., “*Statistical Analysis of Atlantic Tropical Storms*” in preparation
- Kim, Y., B. Qaqish and R. Ignaccolo “*An Analysis of the Potential Impact of Various Regulatory Standards for Ozone on the Incidence of Respiratory-related Mortality*” in preparation

- Linder, E., Cardon, Z., Murray, J., Holan, S., Flikkema, P., Ignaccolo R., Kim, Y. “*A Sequential Median Polish for Automated Data Cleaning and Anomaly Detection in Environmental Sensor Networks*” Paper in preparation.
- Nguyen, X., “*Gibbs Posterior for Suppression Design, Dimensionality Reduction, and Model Choices*” Technical Report in preparation
- Nguyen, X., Bell, D., Clark, J., Gelfand, A. and Kim, Y. “*Modeling and Computation of Wireless Sensor Network Data for Environmental Monitoring*” In preparation.
- Nguyen, X., Yang, J., Yang, Y., and Zhu, Z. “*Optimal Sensor Network Design under Budget Constraints*” In preparation.
- Nguyen, X., Holand, S., and Kim, Y. “*A Correlation Process Prior for Anomaly Detection with Functional Data*” In preparation.
- Yamamoto, K. and Flikkema, P. “*Prospector: Multiscale Energy Measurement of Embedded Systems with Wideband Power Supply Signals*” In preparation.

IV. META ANALYSIS

Publications and Technical Reports

- Liu, F., Dunson, D.B. and Zou, F. (2008). “*High-dimensional Variable Selection in Meta Analysis for Censored Data*” Biometrics, submitted
- Moreno, E., Giron, F.J., Vazquez-Polo, F.J., Negrin, M.A. “*Optimal Decisions in Cost-benefit Analysis*” Tech. Report. Dpt. Statistics, University of Granada. In Review (2008)

Reports in Preparation

- Plante, J.F., Dukic, V., Dunson, D., Stangl, D. “*Bayesian Non-parametric Meta Analysis of ROC Curves*”
- Liu, F., Dunson, D.B. and Zou, F. (2009). “*Annotated Relevance Vector Machine with Application to Polymorphism Selection*” In preparation.

V. ALGEBRAIC METHODS IN STATISTICS AND BIOLOGY

Publications and Technical Reports

- Allman, E., Mathias, C., Rhodes, J. “*Identifiability of Latent Class Models with Many Observed Variables*” SAMSI Tech Rep 2008-08 *Annals of Statistics*, to appear (2009)
- Anderson, D.F., Shiu, A. “*Persistence of Deterministic Population Processes and the Global Attractor Conjecture*” Submitted (2009)
- Aoki, S., Takemura, A. “*Some Characterizations of Affinely Full-dimensional Factorial Designs*” Submitted (2009)
- Craciun, G., Pantea, C., Rempala, G.A. “*Dimension Reduction Method for Inferring Biochemical Networks*” Submitted (2009)
- Craciun, G., Pantea, C., Rempala, G.A. “*Algebraic Methods for Inferring Biochemical Networks: a Maximum Likelihood Approach*” Submitted (2009)
- Dickenstein, A., Perez Millan, M. “*How Far is Complex Balancing from Detailed Balancing?*” Submitted (2009)
- Dimitrova, E., Garcia, L., Hinkelmann, F., Jarrah, A., Laubenbacher, R., Stigler, B., Vera-Licona, P. “*Parameter Estimation for Boolean Models of Biological Networks*” To be submitted to *J. Theor. Computer Science* (2009)
- Dimitrova, E., Garcia, L., Hinkelmann, F., Jarrah, A., Laubenbacher, R., Stigler, B., Vera-Licona, P. “*Parameter Estimation for Multi-state Discrete Models of Biological Networks*” to be submitted to *Bioinformatics* (2009)
- Dinwoodie, I., “*Polynomials for Classification Trees and Applications*” SAMSI Tech Rep 2008-07
- Dinwoodie, I., “*Sequential Importance Sampling of Binary Sequences*” SAMSI Tech Rep 2009-04
- Hara, H., Takemura, A., Yoshida, R. “*On Connectivity of Fibers with Positive Marginals in Multiple Logistic Regression*” Submitted (2009)
- Huggins, P., Owen, M., Yoshida, R. “*First Steps Toward the Geometry of Cophylogeny*” Submitted (2009)
- Pistone, G. “*k-exponential Models from the Geometrical Viewpoint*” Submitted to *European Physical Journal B* (2009)
- Pistone, G., Rogantin, M.P. “*Comparing Different Definitions of Regular Fraction*” Submitted to *Journal of Statistical Theory and Practice* (2009)

- Rhodes, J. A. *"A Concise Proof of Kruskal's Theorem on Tensor Decomposition"* SAMSI Tech Rep 2009-01 Submitted (2009)
- Riccomagno, E., Smith, J.Q., Thwaites, P. *"Causal Analysis with Chain Event Graphs"* Submitted (2009)
- Stone, E.A., Griffing, A. *"On the Fiedler Vectors of Graphs that Arise from Trees by Schur Complementation of the Laplacian to Linear Algebra and its Applications"* Submitted (2009)
- Sullivant, S., Talaska, K. *"Trek Separation for Gaussian Graphical Models"* Submitted (2009)

Reports in Preparation

- Allman, E.S., Matias, C., Rhodes, J.A. *"Identifiability of the Affiliation Model and other Models with Hidden Variables"*
- Allman, E.S., Petrovic, S., Rhodes, J., Sullivant, S. *"Identifiability of 2-tree Mixtures for Group-based Models"*
- Allman, E.S., Rhodes, J., Sullivant, S. *"Research Note: 2-tree Mixture Models and Inference"*
- Allman, E.S., Degnan, J., Rhodes, J. *"Clade Probabilities and Identifiability for 5-taxon Species Trees"*
- Allman, E.S., Kubatko, L., Pearl, D., Rhodes, J. *"New Methods for Searching Tree Space"*
- Conradi, C., Flockerzi, D., *"Parametrization of Multistationarity in Mass Action Kinetics"*
- Cox, L. *"Using Linear Programming to Construct Markov Moves in Contingency Tables"*
- Drton, M., Ginestet, C. *"The Role of the Statistical Curvatures in Model Comparison with Application to Directed Acyclic Graphs"*
- Francis, A., *"Counting Bacterial Genome Arrangements"*
- Garcia, L., Sullivant, S. *"Algebraic Causality in Gaussian Graphical Models"*
- Hara, H., Takemura, A. *"Connecting Tables with Zero-one Entries by a Subset of a Markov Basis"*

- Hillar, C., Sullivant, S. “*Finite Grobner Bases in Infinite Polynomial Rings, with Applications*”
- Laubenbacher, R., Szanto, A. “*Incremental Interpolation with Few Function Values*”
- Laubenbacher, R., Sullivant, S., Yoshida, R. “*Algebraic Biology, a Review Article*”
- Malag`o, L., Matteucci, M., Pistone, G. “*Exponential Family Relaxation in Combinatorial Optimization*”
- Maruri, H. “*Fan of Fractional Factorial Designs*”
- O’Shea, E. “*Frequency of Large Gaps in Small Hierarchical Models*”
- Owen, M., Provan, S. “*Computing the Geodesic Distance in Tree Space in Polynomial Time*”
- Pistone, G., Riccomagno, E., Wynn, H. “*Polynomial Algebraic Models*”
- Pistone, P., Wynn, H. “*Finitely Generated Cumulants*”
- Pistone, G., Vicario, G. “*Comparing and Generating Latin Hypercube Designs in Kriging Models*”

VI. SEQUENTIAL MONTE CARLO METHODS

Publications and Technical Reports

- Bishwal, J., Pena, E. A. “*A Note on Inference in a Bivariate Normal Distribution Model*” SAMSI Tech Rep 2009-03
- Del Moral, P., Doucet, A., Jasra, A. “*An Adaptive SMC Method for Approximate Bayesian Computation*” submitted January 2009.
- Jasra, A., Stephens, D., Doucet, A. “*Inference in Levy-driven Stochastic Volatility Models*” submitted February 2009
- Li, S., Lynch, J. “*On a Gibbs Measure Representation for Complex Load-Sharing Parallel Systems*” Submitted to Applied Probability Journals (2009)
- Liu, F., West, M. (2009) “*A Dynamic Modelling Strategy for Bayesian Computer Model Emulation*” Bayesian Analysis, 4(2), - .

- Pena, E. A., Habiger, J. "*Power-Enhanced Multiple Decision Functions Controlling Family-Wise Error and False Discovery Rates*" SAMSI Tech Rep 2009-02
- Pena, E.A., Habiger, J., Wu, W. "*Classes of Multiple Decision Functions Strongly Controlling FWER and FDR*" SAMSI Tech Rep 2009-06
- Rogzic, V. "*Multimodal Speaker Segmentation and Identification in Presence of Overlapped Speech Segments*" Submitted Journal of Multimedia (2009)
- Septier, F., Carmi, A., Godsill, S. "*Tracking of Multiple Contaminant Clouds*" Fusion 2009 (submitted).
- Septier, F., Pang, S.K., Godsill, S., Carmi, A. "*Tracking of Coordinated Groups using Marginalised MCMC-based Particle Algorithm*" IEEE Aerospace Conference, March 2009.
- Sisson, S.A., Peters, G.W., Fan, Y., Briers, M. "*Likelihood-free Samplers*" Journal Submission, Dec 2008.
- Yoshida, R., West, M. (2009) "*Sparse Bayesian Inference by Annealing Entropy*" Draft completed and under revision for submission to Journal of Machine Learning Research; submission expected in late spring 2009. SAMSI Tech Rep 2009-05

Reports in Preparation

- Andrieu, C., Del Moral, P., Doucet, A. "*Exponential Inequalities for Unnormalized Feynman-Kac Particle Models*" In preparation.
- Carvalho, C., Johannes, M., Lopes, H., Polson, N. "*Particle Learning and Smoothing*" In Preparation
- Carvalho, C., Lopes, H., Polson, N., Taddy, M. "*Particle Learning in General Mixtures*"
- Carvalho, C., Johannes, M., Lopes, H., Polson, N. "*Particle Filtering and Learning: A Comparison*"
- Carvalho, C., Johannes, M., Lopes, H., Polson, N. "*Stochastic Volatility Shot-Noise*"
- Clark, D., Briers, M. "*Sequential Monte Carlo Smoothing with Random Finite Set Observations*" JSM 2009

- Doucet, A., Robert, C.P. “*Particle Nested Sampling*” In preparation.
- Dukic, V., Lopes, H., Polson, N. “*Particle Learning in Epidemic SEIR Models*”
- Dunson, D., Das, S. “*Bayesian Distribution Regression via Augmented Particle Learning*”
- Fearnhead, P., Kau, J.B., Keenan, D.C., Lyubimov, C., Vidyashankar, A. “*Dynamic Latent Factor Model for Mortgage Termination*” (In prep)
- Fearnhead, P., Giagos, V., Sherlock, C. “*Simulation and Inference for Stochastic Kinetic Models via limiting Gaussian Processes*” (In prep)
- Fearnhead, P., Vidyashankar, A. “*Bayesian Inference for Quantitation in PCR*”
- Fokoue, E. “*Variational Mean Field Approach to Efficient Multitarget Tracking*” JSM 2009
- Godsill, S., Fearnhead, P. “*Monte Carlo Inference for α -Stable Lévy Processes*” S. Godsill and P. Fearnhead, in preparation.
- Ji, C., Godsill, S., West, M. (2009) “*Spatial Dynamic Mixture Modelling for Multiple Extended Target Tracking*” (In preparation)
- Ji, C., West, M. (2009) “*Bayesian Nonparametric Modelling for Time-varying Spatial Point Processes*” (Initial draft completed)
- Ji, C., West, M. “*Dynamic Spatial Mixture Modelling and its Application in Bayesian Tracking for Cell Fluorescent Microscopic Imaging*” JSM 2009
- Ji, C., West, M. (2009) “*Spatial Dynamic Mixture Modelling for Unobserved Point Processes and Tracking Problems*” Initial draft completed
- Liu, F., Li F., Dunson, D. “*Adaptive Sampling for Bayesian Variable Selection*”
- Liu, F., Li, F., Dunson, D. (2009) “*Adaptive Design for Variable Selection in Normal Linear Models*” In preparation; submission expected in late spring 2009
- Lund, B., Lopes, H. “*Options, SV and Jumps in the Interest Rate Risk Premia*”
- Macaro, C., Lopes, H. “*Particle Learning for Long Memory Stochastic Volatility Models*”
- Manolopolou, I., Chan, C., West, M. (2009) “*Sequential Selection Sampling for Focused Inference*” In preparation.

- Mukherjee, C., West, M. (2009) “*Sequential Monte Carlo Model Fitting and Comparison in Nonlinear Dynamic Models*” In preparation.
- Niemi, J., Mukherjee, C., Carvalho, C., Lopes, H. “*Particle Learning Without Conditional Sufficient Statistics*”
- Peters, G.W., Briers, M. Copsey, K., Lane, R. “*Trans-dimensional ABC for Source Term Estimation*” In preparation.
- Petralia, F., Chen, H., Carvalho, C., Lopes, H. “*Particle Learning for DSGE Models*”
- Prado, R., Lopes, H. “*Particle Learning for Autoregressive Models with Structured Priors*”
- Rogzic, V. “*Audio-visual Tracking and Speaker Diarization for Unknown Number of Meeting Participants*” to be submitted to IEEE Trans. on Multimedia
- Septier, F., Carmi, A., Pang, S.K., Godsill, S.J. “*Multiple Object Tracking Using Evolutionary and Hybrid MCMC-Based Particle Algorithms*” SYSID 2009
- Septier, F., Rogzic V., Briers, M., Clark, D., Godsill, S. “*A Comparative Study of Particle Methods for Multi-Target Tracking*” In preparation.
- Shi, Dunson, D. “*Particle Stochastic Search for High-Dimensional Variable Selection*”
- Vaswani, N., Septier, F., Godsill, S. “*SMC Contour Tracking for Sequential Plume Estimation*” ICASSP 2010 In Preparation
- Wang, H., Reeson, C., Carvalho, C. “*Sequential Learning in Dynamic Graphical Models*”
- White, G., Green, N. “*Emulation Based Priors for Source Term Estimation*” In preparation.

H. Efforts to Achieve Diversity

SAMSI puts considerable emphasis on contributing to the NSF's effort to broaden the participation from underrepresented groups in the mathematical sciences. During the past year, we have organized and co-sponsored many diversity related activities. SAMSI has also developed a web page devoted to our diversity activities. The page advertises the various program activities related to minority outreach and has links to other diversity related information outside of SAMSI.

Blackwell-Tapia Conference

On Nov. 14-15, 2008, SAMSI hosted the 6th Blackwell-Tapia Conference. This bi-annual event in honor of David Blackwell and Richard Tapia brings together African-American, Native American and Latino/Latina students, faculty, and researchers from mathematics and statistics. This two day event was attended by over 100 participants, and consisted of research talks, panel discussion of issues relating to minority recruitment, retention, and mentoring, as well as a dinner to honor the 2008 Blackwell-Tapia prize winner Juan Mesa of Lawrence Berkeley Laboratory.

Participation in the NSF Institutes' Diversity Committee

Michael Minion has been serving as SAMSI's representative to the NSF Institutes' Diversity Coordination Committee which was formed in 2006 by Chris Jones (SAMSI) and Helen Moore (formerly of AIM), and is now chaired by Kathleen O'Hara (MSRI). While Minion was on leave from June-December, 2008, Associate Director Pierre Gremaud assumed these duties. The Institutes Diversity Coordination Committee has been working together to promote diversity in the Mathematical Sciences at national conferences and through other special events.

SAMSI took part in the *Modern Math* program at the 2008 SACNAS National Convention in Salt Lake City. This program was aimed at introducing young scientists to a variety of current research topics, providing mentorship and networking opportunities, and recruiting future participants in NSF Institute programs from underrepresented groups. Pierre Gremaud attended the conference to represent SAMSI, and Gabriel Huerta of the University of New Mexico, who is a participant in the SAMSI program on Space-Time Analysis for Environmental Mapping, Epidemiology, and Climate Change, presented an overview of his research in the program area. SAMSI was again a participant in the Modern Math program in Oct. of 2009, which will be reported in the next Annual Report of SAMSI.

Minority Participation in SAMSI Programs

SAMSI Postdoctoral Positions: Of the five full-time post-doctoral positions associated with the 2008-09 Research Programs, three of the post-docs are female: Ioanna Manolopoulou, Megan Owen, and Elizabeth Mannshardt Shamseldin. For the 2009-10 Research Programs, of 15 post-docs hired, six are women: Emily Kang, Esther Salazar, Xueying Wang, Veronica Berrocal, Yi Sun, and Avanti Athreya, and one is an under-represented minority: Oliver Diaz.

Education and Outreach Programs: SAMSI continues to use its E&O Program to enhance its diversity efforts by active recruitment of under-represented participants. We are actively recruiting from HBCU's for all programs and are continuing to augment the recruitment of Hispanics and Native Americans through the assistance of members of the National Advisory and Education and Outreach Committees. The diversity breakdown in specific E&O workshops is as follows.

- Undergraduate Workshop (May 2007): From the 18 participants, 12 were female, and 3 were Hispanic.
- Industrial Mathematical and Statistical Modeling (IMSM) Workshop (July 2008): From the 37 participants, 15 were female, and 2 were Hispanic.
- 2-Day Undergraduate Workshop (Oct 2008): From the 41 participants, 18 were female, 2 were African American, and 2 were Hispanic.
- 2-Day Undergraduate Workshop (Feb. 2008): From the 38 participants, 10 were female, 1 was African American, and 3 were Hispanic.

Workshop Participation:

There were, of course, numerous workshop participants from underrepresented groups, as indicated in the following table. Also listed are the numbers of new researchers at each of the workshops.

2007-08 Programs						
Program Year	Activity	# Participants	Underrepresented Groups			# New Resrcher-Students
			# Female	# African-American	# Hispanic	
<i>Random Media</i>						
2007-08	Random Media Transition Workshop -- May 1-2, 2008	21	6	0	0	12
<i>Risk Analysis, Extreme Events and Decision Theory</i>						
2007-08	Risk Revisited: Progress and Challenges Transition Workshop -- May 21, 2008	24	9	0	0	12
<i>Education and Outreach Program</i>						
2007-08	SAMSI/CRSC Interdisciplinary Workshop for Undergraduates -- May 19-23 2008	27	16	0	3	27
<i>Summer Program</i>						

2007-08	Meta Analysis -- June 2-13, 2008	66	24	1	2	45
Environmental Sensor Networks						
2007-08	Environmental Sensor Networks Transition Workshop -- October 20-21, 2008	14	4	0	0	8
2008-09 Programs						
			Underrepresented Groups			
Program Year	Activity	# Participants	# Female	# African-American	# Hispanic	# New Resrcher-Students
Sequential Monte Carlo Methods						
2008-09	Sequential Monte Carlo Methods (SMC) Opening Workshop -- September 7-10, 2008	134	28	3	11	87
2008-09	Mid-Program Workshop -- February 19-20, 2009	34	6	1	2	26
2008-09	Adaptive Design, SMC and Computer Modeling - April 15-17,2009	43	8	1	2	27
2008-09	Transition Workshop -- November 9-10, 2009	to be reported in the 2009-10 Annual Report				
Algebraic Methods in Systems Biology and Statistics						
2008-09	Algebraic Methods Opening Workshop -- September 14-17, 2008	119	35	2	5	70
2008-09	Discrete Models in Systems Biology Workshop -- December 3-5, 2008	44	16	0	2	34
2008-09	Algebraic Statistical Models -- January 15-17, 2009	34	9	1	1	22
2008-09	Molecular Evolution and Phylogenetics -- April 2-3, 2009	41	18	0	0	28
2008-09	Transition Workshop -- June 18-20, 2009	33	13	1	2	22
2008-09 Education and Outreach						
2008-09	SAMSI/CRSC Industrial Mathematical & Statistical Workshop for Graduate Students -- July 21-29, 2008	37	15	0	2	37

2008-09	Two-Day Undergraduate Workshop -- <i>October 31-November 1, 2008</i>	41	18	2	2	38
2008-09	Two-Day Undergraduate Workshop -- <i>February 27-28, 2009</i>	38	10	1	3	27
2008-09	Graduate Student Probability Workshop -- <i>May 1-3, 2009</i>	115	31	0	4	109
2008-09	SAMSI/CRSC Interdisciplinary Workshop for Undergraduates -- <i>May 18-22, 2009</i>	36	18	1	1	18
2009-10	CRSC/SAMSI Workshop for Graduate Students -- <i>July 20-28, 2009</i>	40	12	0	1	37
Co-sponsored and Informal Meetings and Workshops						
2008-09	Blackwell-Tapia Conference -- <i>November 15-16, 2008</i>	79	31	33	31	44
Upcoming 2008-09 Meetings and Workshops						
2008-09	Psychometrics Summer 2009 Program -- <i>June 2009</i>	71	23	1	3	36
Upcoming 2009-10 Meetings and Workshops						
2009-10	Space-time Analysis (Spatial) Summer School -- <i>July 28 - August 1, 2009</i>	43	20	0	2	38
2009-10	Stochastic Dynamics Opening Workshop -- <i>August 30 - September 2, 2009</i>	to be reported in the 2009-10 Annual Report				
2009-10	Space-time Analysis (Spatial) Opening Workshop -- <i>September 13-16, 2009</i>	to be reported in the 2009-10 Annual Report				
2009-10	Self-Organization and Multi-Scale Mathematical Modeling of Active Biological Systems -- <i>October 26-28, 2009</i>	to be reported in the 2009-10 Annual Report				
2009-10	Two-Day Undergraduate Workshop -- <i>October 30-31, 2009</i>	to be reported in the 2009-10 Annual Report				
2009-10	Space-time Analysis: GEOMED Spatial Epidemiology Workshop -- <i>November 14-16, 2009</i>	to be reported in the 2009-10 Annual Report				
2009-10	Two-Day Undergraduate Workshop -- <i>February 26-27, 2010</i>	to be reported in the 2009-10 Annual Report				

I. External Support and Affiliates

1. External Support

SAMSI receives extensive support through the home institutions of our long-term visitors. On average, SAMSI pays for approximately 1/3 of a long-term visitors salary in visiting SAMSI; the other 2/3 is provided by the home institution.

Kenan Foundation: provided \$50,000 of supplementary support, mostly directed to the K-12 Kenan Fellows program.

Sequential Monte Carlo Methods: The Adaptive Design Workshop was jointly funded by the NISS project on Computer Models for Geophysical Risks.

Affiliates: Significant support arose from the Affiliates, as discussed in the next section.

2. Affiliate Involvement

2.1. Background

The NISS Affiliates Program and NISS/SAMSI University Affiliates Program are the largest programs of their kind among the DMS-funded mathematical sciences research institutes. NISS director Alan Karr and associate director Nell Sedransk have major responsibility for operation of these programs, but all members of the directorate interact directly with affiliates.

New affiliates in 2008-09 include Bayer HealthCare, PNYLAB, Yahoo! Labs, the Department of Biostatistics, Bioinformatics, and Biomathematics at Georgetown University and the Department of Statistics at Indiana University. A complete listing of affiliates appears below.

As a benefit of membership, affiliates may receive reimbursement for expenses to attend SAMSI workshops as well as NISS events, many of which derive from SAMSI programs.

A central role of the affiliates is as a bridge from SAMSI to the statistics and applied mathematics communities, especially to inform the development of SAMSI programs. To illustrate, the 2007-08 program on Risk Analysis, Extreme Events and Decision Theory, as well as the 2006-07 program on Development, Assessment and Utilization of Complex Computer Models, the National Defense and Homeland Security program in 2005-06, the Latent Variable Models in the Social Sciences (LVSS) program in 2004-05 and the DMML program for 2003-04, all reflect affiliate interest to a significant degree. The upcoming 2009-10 programs both respond to strong affiliate interest and the proposed programs for 2010-11 both include components suggested by affiliates.

2.2 NISS Affiliates and NISS/SAMSI Affiliates

Corporations: Avaya Labs, AT&T Labs Research, Bayer HealthCare, GlaxoSmithKline, Eli Lilly, Merck Research Laboratories, MetaMetrics, Inc., PNYLAB, RTI International, Sanofi-Aventis Pharmaceuticals, SAS Institute, SPSS, Chicago, IL and Yahoo! Labs

Government Agencies and National Laboratories: Bureau of Labor Statistics, Census Bureau, Energy Information National Agricultural Statistics Service, National Center for Education Statistics, National Center for Health Statistics/CDC, National Security Agency, and Office of the Comptroller of the Currency

NISS/SAMSI University Affiliates: University of California Berkeley, Department of Statistics; Carnegie Mellon University, Department of Statistics; Columbia University, Department of Biostatistics; University of Connecticut, Department of Statistics; Duke University, Departments of Mathematics and Statistical Science; University of Florida, Department of Statistics; Florida State University, Department of Statistics; George Mason University; Georgetown University Department of Biostatistics, Bioinformatics, and Biomathematics; University of Georgia, Department of Statistics; University of Illinois Urbana-Champaign, Department of Statistics; Indiana University, Department of Statistics; Iowa State University, Department of Statistics; Johns Hopkins University, Department of Applied Mathematics and Statistics; Medical University of South Carolina, Department of Biostatistics, Bioinformatics & Epidemiology; University of Michigan, Departments of Statistics and Biostatistics; University of Missouri Columbia, Department of Statistics; North Carolina State University, Department of Statistics; North Carolina State University, Department of Mathematics; University of North Carolina at Chapel Hill, Department of Biostatistics; University of North Carolina at Chapel Hill, Department of Mathematics; University of North Carolina at Chapel Hill, Department of Statistics & Operations Research; Oakland University, Department of Mathematics and Statistics; Ohio State University, Department of Statistics; Pennsylvania State University, Department of Statistics; Purdue University, Department of Statistics; Rice University, Department of Statistics; Rutgers University, Department of Statistics; University of South Carolina, Department of Statistics; Southern Methodist University, Statistical Science Department; Stanford University, Department of Statistics; Texas A&M University, Department of Statistics; Virginia Commonwealth University, Department of Biostatistics

2.3 Affiliate Participation

All SAMSI programs and events during 2008-09 had strong affiliate participation, nearing one-half of attendees at some workshops. Expenditures from Affiliates Reimbursement Accounts to attend SAMSI events exceeded \$50,000.

Participation by affiliates in SAMSI programs remains extremely strong. Examples include:

2008-09 Program on Algebraic Methods in Systems Biology and Statistics: Program leaders include faculty from North Carolina State University and Penn State University. Among working group participants is a senior researcher from the National Center for Health Statistics.

2008-09 Program on Sequential Monte Carlo Methods: Program leaders include faculty from the University of California Berkeley and Duke. There was strong participation from the Department of Statistics at the University of Missouri at Columbia, almost one-third of whose faculty will be engaged in 2009-10 programs at SAMSI.

Postdoctoral Fellows: Three of five postdoctoral fellows during 2008-09 received their degrees from affiliated academic departments.

2.4 Plans for the Future

The affiliates programs have instituted a series of *Exploration Workshops* that seek to identify opportunities for the statistical and applied mathematical sciences in emerging areas of science, technology and science. An explicit goal is to examine potential future SAMSI programs. Workshops during the past year addressed “Agent-Based Modeling” and “Statistical Issues in Financial Risk Modeling and Banking Regulation.” Topics planned for 2009-10 include “Financial Risk Modeling” and “Computational Advertising.”

J. Advisory Committees

Committee	Name	Affiliation	Field	Term
Governing Board	Bruce Carney George Casella Don Estep Vijay Nair John Simon Daniel Solomon (Chair)	UNC, Assoc. Dean U of Florida (ASA Rep) Colorado State U NISS Trustees Chair Duke, Asst. Provost NCSU, Dean	Astronomy Statistics Math and Stat Statistics Chemistry Statistics	
National Advisory Committee	Carlos Castillo-Chavez Ricardo Cortez Rick Durrett Jianqing Fan Nancy Kopell Rod Little Jun Liu David Mumford Susan Murphy Daryl Pregibon G.W. Stewart Bin Yu (Chair)	Arizona State U Tulane U Cornell U Princeton U Boston U U of Michigan Harvard U Brown U U Michigan Google, Inc U of Maryland U of CA, Berkeley	Mathematics Mathematics Mathematics Ops Research Mathematics Biostatistics Statistics Applied Mathematics Statistics CS and Statistics Computer Science Statistics	2003-2009 2008-2010 2006-2008 2008-2010 2005-2008 2006-2008 2008-2010 2006-2008 2008-2010 2003-2008 2003-2008 2006-2011
Local Development Committee	David Banks H.T. Banks Lloyd Edwards Gregory Forest Montserrat Fuentes John Harer Sharon Lubkin Sally Morton Richard Smith Butch Tsiatis Mike West	Duke NCSU UNC UNC NCSU Duke NCSU RTI UNC NCSU Duke	Statistics Mathematics Biostatistics Mathematics Statistics Mathematics Mathematics Statistics Statistics Statistics Bioinformatics & Stat	
Chairs Committee	Elizabeth DeLong Patrick Eberlein Alan Gelfand Loek Helmnick Michael Kosorok Vidyadhar Kulkarni Sastry Pantula Mark Stern	Duke UNC Duke NCSU UNC UNC NCSU Duke	Bioinformatics & Stat Mathematics Statistics Mathematics Biostatistics Statistics Statistics Mathematics	
Education and Outreach Committee	Negash Begashaw Carlos Castillo-Chavez (ex officio) Karen Chiswell Anne Fernando Pierre Gremaud (Chair) Leona Harris Gabriel Huerta Marian Hukle Cammey Cole Manning	Benedict College Arizona State U NCSU Norfolk State University NCSU College of New Jersey Univ. of New Mexico U of Kansas Meredith College	Mathematical Sciences Mathematics Statistics Mathematics Mathematics Mathematics Statistics Biological Sciences Mathematics & CS	

II Special Reports: Program Plan

A. Programs for 2009-2010

1. Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change

1.1 Introduction

This year-long SAMSI program will focus 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.

1.2 Research Foci

1.2.1 Environmental Mapping

Spatial or spatial-temporal statistical analysis in environmetrics often entails the prediction of unobserved random fields over a dense grid of sites in a geographical domain, based on observational data from a limited number of sites and possibly simulated data generated by deterministic physical models. In important special cases, spatial prediction requires statisticians to estimate spatial covariance functions and generalized regression tools (also called geostatistical methods).

Many commercially available GIS packages include excellent visualization tools, but a dearth of spatial interpolation tools. In particular, the tools available are often not statistically based, and have been shown to perform very poorly compared to geostatistical tools.

Many standard geostatistical packages have the disadvantage that they do not take into account the variability in estimates due to estimating the covariance function. Most also do not incorporate the modern tools available to represent spatial covariance structures for nonstationary processes.

However, such tools for nonstationary processes have not been extended to multivariate fields except through often unrealistic, simple (Kronecker type) structures. Even more complicated are space-time structures that are non-separable, nonstationary in space and in time, or multivariate with structures that are not temporally symmetric.

Methods for spherical data, especially appropriate for climate research, are currently

being developed, but they need to address complications similar to those that occur for multivariate random fields.

1.2.2 Spatial Epidemiology

Many studies during the past two decades have demonstrated a statistical association between exposure to air pollutants (principally, particulate matter and ozone) with various (mostly acute) human health outcomes, including mortality, hospital admissions, and incidences of specific diseases such as asthma. While a number of different study designs have been used, two dominate. The first, the time series studies, relate variations in daily counts of these adverse health outcomes with variations in ambient air pollution concentrations through multiple regression models that include air pollution concentrations while removing the effects of long-term trends, day of week effects, as well as possible confounders such as meteorology. However, the relative health risks of air pollution are small say compared to smoking. Thus some studies have through Bayesian hierarchical modeling combined the estimated air pollution coefficients for various urban areas to borrow strength.

A different kind of study design is needed for the more challenging problem of estimating the chronic (as against acute) effects of air pollution. This second kind of design involves the use of prospective studies that follow a specific group of individuals for several years or decades, and then relate health outcomes (including mortality, but also specific measures such a heart rate variability) to air pollution after adjusting for personal factors such as age, previous health history, and smoking.

Recently both kinds of studies have been paying more attention than in the past, to spatial effects. Thus, although traditionally, spatial correlations between the cities have been ignored, now multi-city time series studies recognize the increasing evidence pointing to spatially nonhomogeneous associations. As datasets become available that spatially resolve both air pollution and human health outcomes at finer scales, this effect is likely to increase in importance, making it highly desirable to develop spatial and spatio-temporal stochastic processes for the joint distributions of air pollution, human health outcomes and other relevant covariates.

In prospective studies, researchers consider the possible effects of spatially defined covariates such as distance between a residential location and the nearest road. They also recognize the importance of measurement error, in particular the discrepancy between ambient pollution concentrations as measured at monitoring sites and the personal exposure of individuals. In some urban areas, spatial variability in the pollution field is an important component of this error. So some studies have used spatial methods such as kriging and Bayesian prediction to reduce this error by inferring from the ambient measurements, the pollution concentrations at a participant's residence. However, much less work has been done on the logical follow-up question, which is the effect of such variability on the health-effect regression coefficients.

Challenges that face the practitioner of spatial epidemiology, include issues of data availability and quality, confidentiality, exposure assessment, exposure mapping, and study design. Geographic methods of exposure assessment make a number of key assumptions that may limit their applicability in given situations. These include the following:

- equating modeled estimates of exposure (including distance-based measures, or output of EPA exposure numerical models such as SHEDS) with true exposure;
- equating exposure at a point (e.g., place of residence) with total personal exposure, that is, exposure integrated across space and time over the course of daily activities as the individual moves through the spatial exposure field;
- equating group exposure and group exposure-disease relationships with individual exposure and relationships at the individual level, this phenomenon is known as "ecologic fallacy".

Key areas in which further work is needed include:

- Developing methods that account for a subject's movement through spatio-temporal exposure space.
- Developing calibration models whereby spatially sparse direct measurements of exposure can be combined with inexpensive, and therefore spatially dense, surrogates or predictors of exposure, to enable more precise estimation of the true exposure surface.

1.2.3 Climate Change

Much of the case for climate change and the estimation of its deleterious effects has relied on deterministic climate models that embrace physical and chemical modeling. The GCM [General Climate (or Circulation) Model] yields simulated climate data at fairly coarse spatial scales that serves as input to the RGCM (Regional GCM) that runs at finer spatial scales.

These models are at best, approximate representations of the real world, and, hence must be continually assessed. Model errors must be identified and characterized to provide statements about confidence in results. Further the computational overhead of these models mandates trade-offs between the number of realizations of a given model versus number of models used, using both current techniques of experimental design, design of computer experiments, as well as the development of new techniques. The current methods of dealing with this – arguably most important – model validity issue are based on statistical spatial modeling techniques; but these techniques have never been tested for the complexity of climate models.

The results of climate models are extremely multi-dimensional. It is very difficult to present all of this information concisely in a manner that can be understood by decision makers. *Dimension reduction* and *data presentation techniques* are needed for contrasting spatial data, explaining what is being presented, and determining how to describe the confidence of projections from non-random samples.

Also available for assessing climate change are observational data from different measurement platforms (satellites, weather balloons, surface thermometers, *etc.*). Like the simulated data, these can represent very different spatial scales. Many historical time series do not have old data for South America, Africa, or South-east Asia. Even in the satellite era – the most observed period in Earth’s climate history – key observational datasets such as those for lower tropospheric temperatures involve significant uncertainties. Understanding, modeling, and analyzing these spatial and temporal uncertainties, in the context of the massive (but sparse) data and the impact on climate change, requires significant methodological and theoretical advances.

Another key observational data set is the record of changes in ocean heat content. To estimate changes in the heat content of the world’s oceans from sparse data with time-varying biases and coverage, temperature information must be “infilled” over large volumes of the ocean. This is an area where development and fitting of sophisticated space-time models to sparse data is a critical need.

One more crucial need is taking coarse-resolution projections from global and regional climate models down to *estimates for small areas*. [Indeed downscaling and upscaling issues pervade the study of both simulated data and data.] This is not the usual small-area estimation problem. It is actually the opposite: the ‘average’ solution needs to be processed through local climate features – a very poorly understood process.

The potential *effects on humans* from climate change are wide ranging, especially since evidence suggests that *extreme events* are increasing in frequency as a result of global warming. Possible effects include the rise in infectious diseases such as malaria, and deaths caused by heat waves such as occurred in Europe in 2003, or wild fires such as occurred in October 2007 in California. The data that suggests these effects is spatial and, again, the scale of the data and the determination of its causal relationship to climate change require new understandings and methodologies.

1.3 Program Timing and Related Programs

All three of the proposed areas of research in the program are of great current interest to science in general, and statistics and mathematics in particular. For instance, a recent statement by the American Statistical Association highlights the need for improvement in space-time methodology to tackle the difficult problem of climate change (www.amstat.org/news/index.cfm?fuseaction=climatechange).

Two other SAMSI programs have some relationship with this program. The program on *Development, Assessment and Utilization of Complex Computer Models* had one working group which began the consideration of climate change models using space-time methods; indeed, it was partly this work that highlighted the need to conduct a major program in the area. The current program on *Environmental Sensor Networks* considers spatial problems, but only those arising in sensor networks, which have a very different character than those discussed above.

The Mathematical Biosciences Institute had a program in Winter Quarter 2006 on *Spatial Heterogeneity in Biotic and Abiotic Environment and Spatial Evolution*. Both emphases are very different than the types of spatial problems being considered above. For 2010, IPAM is considering a program on *Model and Data Hierarchies for Simulating and Understanding Climate*. We do not yet know the specifics of this program, although the list of organizers is very different than the organizers of the SAMSI program. We will, of course, work with IPAM to ensure that the two programs are synergistic.

1.4 Personnel and Participants

1.4.1 Program Leaders

Program Leaders: Noel Cressie (Ohio State University), Peter Green (University of Bristol), Michael Stein (University of Chicago), Dongchu Sun (University of Missouri), Jim Zidek (University of British Columbia) - Chair

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

Local Scientific Coordinators: Montse Fuentes (N.C. State University), Alan Gelfand (Duke University), Richard Smith (UNC-Chapel Hill)

Directorate Liaison: Jim Berger (SAMSI)

National Advisory Committee Liaison: Jun Liu (Harvard University)

Note: Additional leaders will be appointed from each of the key areas mentioned below, from those who can be long-term visitors.

1.4.2 Postdoctoral Fellows

The postdoctoral fellows and associates for the program are an exciting group of top graduate students in research areas related to the program. Current appointees are Veronica Berrocal, Howard Chang, Sourish Das, Elizabeth Shamseldin, Benjamin Shaby, Martin Tingley, and Jun Zhang. We expect at least one additional appointment through the Math Institutes supplementary postdoctoral program.

1.4.3 Faculty Fellows and Local Researchers

The three partner universities will provide course releases for Jason Fine (UNC), Montse Fuentes (NCSU), Alan Gelfand (Duke), John Harlim (NCSU), Amy Herring (UNC), Brian Reich NCSU), and Richard Smith (UNC) to extensively engage in the program.

Among the other local scientists that will potentially be heavily involved are Jim Berger (Duke), Peter Bloomfield (NCSU), Michael Breen (EPA), Jim Clark (Env., Duke),

Merlise Clyde (Duke), David Dunson (Duke), Chris Frey (Env. Eng., NCSU), Sujit Ghosh (NCSU), Jacqueline Hughes-Oliver (NCSU), Joe Ibrahim (UNC), Ed Iversen (Duke), Alun Lloyd (NCSU), Marie Lynn Miranda (Env., Duke), Haluk Ozkaynak (EPA), Robert Wolpert (Duke), and Helen Zhang (NCSU).

1.4.4 Graduate Students

The three partner universities will provide research assistantships for Avishek Chakraborty (Duke), Sean Cohen (NCSU), Amogh Deshpande (NCSU), Danilo Lopes (Duke), Hongxia Yang (Duke), Danilo Lopes (Duke), and one other TBD to participate in the program.

In addition visiting graduate students that have to date been accepted into the program, for visits of one semester to one year, are Candace Berrett (OSU), Aune Erland (Trondheim), Annabel Fortes (Valencia), Yajun Liu (Missouri), and Gabriele Martinelli (Trondheim).

1.4.5 Long-term Visitors (one semester to one year)

To date, the following researchers have been approved for long-term visits: Sudipto Banerjee (Minnesota), Susie Bayarri (Valencia), Kate Calder (OSU), Lisha Chen (Yale), David Conesa (Valencia), Noel Cressie (OSU), Sarat Dass (MSU), Jo Eidsvik (Trondheim), Marco Ferreira (Missouri), Dani Gamerman (Rio de Janeiro), Virgilio Gomez-Rubio (Castilla-La Mancha), Murali Haran (PSU), Chong He (Missouri), Scott Holan (Missouri), Gabriel Huerta (New Mexico), Monica Jackson (American U.), Gardar Johannesson (LLNL), Jaeyong Lee (Seoul National), Mihails Levins (Purdue), Linyuan Li (New Hampshire), Sakis Micheas (Missouri), Orietta Nicolis (Bergamo), Bala Rajaratnam (Stanford), Ingelin Steinsland (Trondheim), Dongchu Sun (Missouri), and Linda Young (Florida).

There will also be many visitors for periods of weeks to months during the period.

1.5 Workshops and Other Events

1.5.1 Summer School on Spatial Statistics

This summer school will be held July 28 - August 1, 2009 at SAMSI. The instructors will be Sudipto Banerjee (U. Minnesota), Reinhard Furrer (U. Zurich), Doug Nychka (National Center for Atmospheric Research), and Stephen Sain (National Center for Atmospheric Research)

Background: Determining the air quality at an unmonitored location, characterizing the mean summer temperature and precipitation over a region or quantifying the changing incidence of a disease across an urban area are examples where a function of interest depends on irregular and limited observations. Prediction and scientific understanding of environmental and epidemiology data often requires estimating a smooth curve or surface

over space that describes an environmental process or summarizes complex structure. Moreover, drawing inferences from the estimate requires measures of uncertainty for the unknown function. This course will combine ideas from geostatistics, smoothing, and Bayesian inference to tackle these problems.

An important component of the lectures is the use of contributed packages for the R statistical environment (www.r-project.org) for hands-on experience with these methods, analyzing spatial data and practice in problem solving. In addition these open source R packages (e.g. spBayes, fields and spam) provide insight in the computational framework for function fitting and the facility to handle multivariate or large environmental datasets. The overall theme of this course is to illustrate how statistical science requires a blending of the scientific context, statistical modeling and statistical computing to reach a useful solution.

Course Contents: The first part of the course explains a common framework for spatial statistics and splines using ridge regression. This correspondence provides a common computational approach and leads to easy to use methods for Kriging and thin-plate splines. Several case studies will illustrate how these methods work in practice and the class is encouraged to modify the related R code and scripts to explore variations in the analysis. The second part of the course considers multivariate spatial responses and large spatial data sets. Building from the basic methods, these topics extend the R packages either through multivariate covariance functions or sparse matrix methods. The final part of the course will introduce a Bayesian framework for spatial models that not only provides a comprehensive quantification of the uncertainty of the spatial analysis but also provides efficient strategies for dimension reduction in hierarchical models.

In particular, the last part of the course will concentrate on Bayesian methods for spatial epidemiology and other public health applications. Here, data often arise as aggregated summaries over regions (e.g. counts or rates of disease incidence, mortality etc.) and the spatial referencing is done with respect to regions (e.g. counties, census-tracts, zip-codes etc.). While geostatistical models can still be used to model such data, spatial models can now build associations based upon conditional dependencies over the underlying neighborhood structures. These lead to Simultaneously AutoRegressive (SAR) and Conditionally AutoRegressive (CAR) models. Such models will be explained along with existing software resources in R (the spdep and BRUGS packages).

An important part of the course will be blocks of time where students are encouraged to work independently or in teams on the analysis of spatial or space/time datasets. This will not only build skill in statistical computing and the R language but will also be an opportunity for informal presentations and collaboration with other students.

1.5.2 Opening Workshop

This will be held September 13-16, 2009, and will aim to engage as large a part of the statistics, mathematics, and relevant scientific communities as possible, with representative sessions from all of the main program topics.

1.5.3 GEOMED: Spatial Epidemiology 2009 Workshop

This workshop will be held November 14-16, 2009, at the Medical University of South Carolina in Charleston. GEOMED 2009 is the 6th international, interdisciplinary conference on geomedical systems. This meeting is a jointly sponsored event with SAMSI and so the meeting also represents a SAMSI workshop on spatial epidemiology. Today, more and more issues are arising in public health involving geography and medicine. GEOMED brings together statisticians, geographers, epidemiologists, computer scientists, and public health professionals to discuss methods of spatial analysis, as well as present and debate the results of such analyses.

1.5.4 Other Workshops

The working groups will help develop other workshops during the year. The *Transition Workshop*, at the end of the program, will disseminate program results and chart a path for future research in the area.

1.6 Courses

1.6.1 Theory of Continuous Space and Space-Time Processes

This Fall 2009 course will be taught by Montse Fuentes, North Carolina State University and Alan Gelfand, Duke University (with guest lecturers). The course is intended to provide a strong theoretical foundation for space and space-time processes over continuous domains. Topics will include continuous parameter stochastic process theory; spectral methods; spatial asymptotics; nonstationary spatial modeling; dynamic models and spatial time series; nonseparable space-time models; spatial design; space-time data fusion; low rank representations; nonparametric spatial methods; topics in shape analysis.

1.6.2 Spatial Epidemiology

This Fall 2009 course will be coordinated by Montse Fuentes, North Carolina State University, and Richard Smith, University of North Carolina.

Much of modern epidemiology is concerned with relationships between environmental factors and various types of human health outcome. When data are collected at many spatial locations, we may refer to the problem as one of spatial epidemiology. However in most cases, this includes a temporal component as well. Since modeling spatial dependence is often critical to the method of statistical inference, it is necessary to use methods from spatial or spatio-temporal statistics. Very often health data are aggregated (e.g. into zip code or county totals) so models for data at discrete spatial locations, such as Markov random fields, are more appropriate than geostatistical methods. Another kind of problem is exemplified by the NMMAPS study (<http://www.ihapss.jhsph.edu/>): an air pollution-mortality relationship is developed initially for many time series at individual

cities, but inferences are then drawn by combining data across spatial locations. A third kind of problem is when there is uncertainty about the pollution field itself, for example, when data collected at monitors are interpolated to other locations. Sometimes this interpolation is performed by spatial statistics methods, but there is a growing trend to use air pollution models such as CMAQ (the EPA Community Multiscale Air Quality model).

Specific topics for the course are likely to include models for spatially distributed health data; Markov random fields; extensions to spatial-temporal processes; multi-city time series studies; combining data across multiple studies at different spatial locations; measurement error problems that involve spatial interpolation; and use of air quality models.

1.6.3 Spatial Statistics in Climate, Ecology and Atmospheric

This Spring 2010 course will be coordinated by Montse Fuentes, North Carolina State University, with guest instructors.

Much of the case for climate change, weather forecast, and determination of air pollution levels and the impact of all these factors on the ecosystem and human health, has relied on deterministic climate, weather and air pollution models that embrace physical and chemical modeling. These models are approximate representations of the real world, and, hence must be continually assessed. Model errors in atmospheric models must be identified and characterized to provide statements about confidence in results. The results of climate, weather and air pollution models are extremely multi-dimensional. It is very difficult to present all of this information concisely in a manner that can be understood by decision makers. Dimension reduction and data presentation techniques are needed for contrasting spatial data, explaining what is being presented, and determining how to describe the confidence of projections from non-random samples.

Also available for assessing climate change and pollution levels are observational data from different measurement platforms (satellites, weather balloons, surface thermometers, monitoring stations, etc.). Like the simulated data, these can represent very different spatial scales. Understanding, modeling, and analyzing these spatial and temporal uncertainties, in the context of the massive (but sparse) data and the impact on climate change, requires significant methodological and theoretical advances.

In this course we will introduce the statistical methods to characterize uncertainties in climate, weather, ecological and air pollution deterministic models. We will also present statistical frameworks to combine disparate spatial data, from observations and output of deterministic models, and to measure the agreement between an artificially generated climate signal from a climate model and real data as measured by surface observation stations or satellites. We will cover statistical methods for processing ensembles of climate models. We will introduce different spatial temporal modeling approaches to characterize trends in space and time, as well as to estimate dependency structures, and to do space-time prediction for climate, weather, ecological and air pollution data. We will

introduce state-of-the art methods for dimension reduction, spatial extreme events in climate and weather, and impact of climate change on mortality and human health.

1.7 Working Groups

Research working groups will meet throughout the program to pursue particular research topics identified in the kickoff workshop (or subsequently chosen by the working group participants). The working groups consist of SAMSI visitors, postdoctoral fellows, graduate students, and local faculty and scientists. In addition to the three broad research areas indicated above, there is considerable interest in specific methodological issues, such as comparing methods for dealing with nonstationary spatial covariance, and combining data from different sources with model predictions to assess, with reliable error estimates, ocean surface temperature.

1.8 Leveraging

There is great interest in this research area, and we expect that numerous activities will be leveraged with other research organizations. For instance, NCAR is co-sponsoring the Summer School on Spatial Statistics. The environmental mapping research is likely to mesh well with the program on Spatial/Temporal Modeling of Marine Ecological Systems of the Canadian Institute for Complex Data Structures (NICDS). Another possible collaboration with NICDS and SPRUCE is in the area of space-time modeling of wildfires. The EPA and NIEHS have significant interests in spatial epidemiological research, and we will be exploring opportunities for interaction with their researchers.

2. Stochastic Dynamics

2.1 Introduction

This year-long SAMSI program is centered around the broad topic of Stochastic Dynamics with a specific focus on analysis, computational methods, and applications of systems governed by stochastic differential equations. Applications stemming from mathematical biology and the medical sciences will be of particular interest, and issues pertaining to estimation and data assimilation in applications will also be examined.

2.2 Background

The term “Stochastic Dynamics” is one which resonates within many fields in statistics and applied mathematics. The numerical analyst designing algorithms for stochastic differential equations, the math biologist studying transport on the cellular level, the analyst trying to understand the effect of stochastic forcing and data in dynamical systems, the statistician trying to characterize the statistics of dynamic networks, and the mathematical modeler trying to bridge the gap between atomistic and continuum physics are all examples of research in stochastic dynamics. Unfortunately it is too often the case that research being done in one of the above scenarios is not being widely disseminated across the spectrum of statistics and applied math. The aim of the proposed SAMSI program is to bring together experts in different but highly inter-related research specializations under the broader umbrella of stochastic dynamics with the goal of creating collaborations which could potentially lead to exciting advances in particular research areas. Proposed participants come not only from the traditional pool of mathematics and statistics but also from engineering, biology, physics, and health sciences.

Motivated by suggestions from local and national leaders in the field, we have designed a program which will include applied mathematical analysts, probabilists, experts in stochastic and multi-scale computation, and leaders in application areas in which stochastic dynamics play a central role. Each of these groups will have the opportunity to both inform and benefit from the cutting edge research of the other participants. We have identified local experts who are enthusiastic about helping organize and participate in the program, and also individuals from outside the research triangle who have expressed interest in being in residence at SAMSI during 2009-10 and serving as a program organizer. A more detailed description of these research foci follows.

2.3 Research Foci

2.3.1 Stochastic Analysis and Numerical Methods

In recent years it has become increasingly clear that to effectively understand complex stochastic systems, a combination of modern numerical analysis, estimation and sampling techniques, and rigorous analysis of stochastic dynamics is required. Whether one speaks

of path sampling techniques, estimation in complex non-linear dynamics, or simulation of rare-events it is important to bring both sophisticated analytic tools and an understanding of what one can compute efficiently.

A working group in stochastic analysis and numerical methods is partially inspired by a recent workshop, sponsored by [AIM](#) and the [NSF](#), concerning approaches for the numerical integration of stochastic systems which span many temporal-scales. This subject would fit well with other potential working group topics of multi-scale computing, biological applications, and dynamics of networks. Important issues such as the ergodicity of numerical methods for SDEs, the construction of higher order methods for SDEs and SPDEs, the role of holonomic constraints and how to enforce them in numerical methods, or ways to efficiently compute quantities like free energies in chemical kinetic simulations would provide very fertile ground for productive collaboration between mathematicians, statisticians, and computational scientists under the stochastic dynamics banner.

2.3.2 Multi-scale and Multi-physics Computing

The classical continuum equations arising in fluid flow, elasticity, or electromagnetic propagation in materials require constitutive laws to derive a closed-form system. The constitutive laws appropriate for a given set of equations can be derived in two ways: First, from phenomenological considerations such as the linear behavior underlying elastic deformation or Newtonian fluid flow. Second, from averaging of kinetic theory results describing basic molecular dynamics. This has been possible for situations in which the microscopic behavior has been close to thermodynamic equilibrium. In such cases, Gaussian statistics are well verified at the microscopic level and the moments of Gaussian distributions can be computed analytically.

However, many physical processes exhibit significant localized departure from Gaussian statistics. When a solid breaks, the motion of the atoms in the crystalline lattice along the crack propagation path is no longer governed by a Maxwell-Boltzmann distribution. When a material undergoes a phase change, large scale correlations among atoms are formed (or destroyed) which modify the typically Gaussian statistics of the equilibrium phases. Protein folding can be seen as a large-scale modulation imposed by polymer links of the Gaussian statistics of the component atoms. A common characteristic of these situations is that macroscopic features impose the departure from local thermodynamic equilibrium and macroscopic quantities of are practical interest. Crack propagation is initiated by a force acting on a solid and we wish to know how far the solid deforms before it breaks. In solidification, a heat flux evacuates energy from the melt at some rate and we wish to characterize the type of order arising in the material.

In all such cases, a basic problem is how to extract the statistical distributions of physical quantities when the system is away from thermodynamic equilibrium. Knowledge of the distribution would allow local constitutive laws to be formulated. Direct numerical simulation is prohibitively expensive. Continuum level simulation is incomplete due to lack of constitutive laws. Furthermore, while it is clear that higher-order moments

characterizing the microscopic statistical distribution are required, it is not known how many of these moments are needed and what their persistence time might be. The microscopic dynamics are stochastic but subject to multiple macroscopic constraints. One major statistical challenge is how to characterize the microscopic motion in a manner which can be used to derive a constitutive law. A basic computational question is how to advance the system in time efficiently at both the microscopic and macroscopic level. An analysis challenge is how to combine this knowledge and form (e.g. through asymptotic expansions) particular constitutive laws. Thus progress in this area will depend on experts in numerical and stochastic analysis, statistics, and engineering modeling to joining forces and methodologies.

2.3.3 Stochastic Modeling and Computation in the Biological and Medical Sciences

The explosion of interest in mathematical and statistical modeling and computation in the biological and medical sciences, where stochasticity is present at nearly every scale, has been one of the most exciting trends in the biosciences in the last ten years. Math biology has grown from a niche area to a major research group in many US math departments, and graduate programs in mathematics are scrambling to cope with a wave of students seeking to do graduate work in interdisciplinary research areas. Programs in bio-statistics, bio-informatics, and bio-medical engineering are also seeing increased growth. In examples as diverse as bio-chemical networks, diffusion and noise in cellular transport, modeling molecular motors by a stochastic ratchet, the study of epidemics, or the modeling or analysis of neuronal dynamics, stochastic modeling, analysis, and computation permeate the biological sciences. A working group centered in applications of stochastic dynamics in biology and medicine will both allow experts in analysis and computation to be exposed to interesting applications and allow researchers in bio-statistics, math-biology and bio-medical engineering to work together with experts in stochastic analysis and computation.

2.3.4 Dynamics of Biological Networks

Biological network data and processes are distinguished by the inherently heterogeneous dependencies among units. The details of these dependencies, usually represented by binary or more general links, and the dynamical processes describing processes over and involving those links, are critical to a variety of biological phenomena, with stochastic descriptions of flow on static network structures and of dynamically-changing connections. Application areas include networks of neurons, biochemical networks, and social network interactions among animals of the same species (e.g., swarming, epidemic processes) and across species (ecological dynamics).

Among the central issues are:

1. *Modeling*: Dynamic network models are typically descriptive rather than generative. Moreover, many of the existing descriptive tools treat dynamic networks only through the amalgamation of a sequence of static snapshots. More modeling work is needed on both

fronts, both for adequate description but also to attempt to explain the "physics" and "biology" of the network dynamics.

2. *Embeddability*: Many of the processes of interest on networks and of the connectivity in the networks themselves have been explored through discrete-time models. Even for those models using continuous time stochastic processes, the data used to study and further develop the models and their implications often come in the form of repeated snapshots at discrete time points--a form of time sampling as opposed to node sampling--or cumulative network links. Can we represent and estimate the continuous-time parameters in the actual data realizations used to fit models?

3. *Sampling*: Available network data often represents only a subnetwork or subgraph of the full network of interest. This limitation can be considered from both a sample designed-based or a model-based perspective. The consequences of this limitation are understood poorly for static networks, where it strongly impacts the study of stochastic processes between statically-connected nodes, and are essentially not understood at all for dynamic networks.

4. *Prediction*: In dynamic network settings, data generated over time present a series of forecasting problems. Model sensitivity for dynamic processes on and of networks is further complicated by heterogeneities in node roles in the system, raising a number of issues about how to best evaluate alternative predictions from different models.

2.3.5 Estimation and Data Assimilation

For stochastic models to be of practical, real-world importance, it is often necessary to incorporate observations to calibrate and inform the models. Often such "data fitting" is done through general models which are not particularly informed about the underlying dynamics except through the data. In many applications such as molecular dynamics, weather/current modeling, and bio-chemical networks it is particularly important to understand the structure of the dynamics and use it in a holistic way when fitting data. The program will include a working group to look at different ways to better combine sophisticated estimation ideas with understanding of the detailed dynamics.

2.4 Program Timing and Previous Related Programs

The subject of Stochastic Dynamics lies near the center of many application areas of mathematics and statistics and has been the focus of numerous programs and workshops in the recent past. The European Science Foundation is just finishing a 5 year Research Networking Programme *Stochastic Dynamics: fundamentals and application* which has spawned numerous grants and more than 40 workshops and minisymposia. At the NSF workshop in October 2007 on "Discovery in Complex or Massive Data: Common Statistical Themes," there was consensus about an urgent need for models of the dynamics of networks and associated tools for inference. Other relevant recent workshops include *Stochastic Dynamics* in June 2007 at Univ. Paris I, and *The practice and theory of stochastic simulation* in Oct. 2007 at AIM in Palo Alto. Also of significant note are

the two programs *Stochastic Partial Differential Equations* and *Stochastic Processes in Communication Sciences* which will be held in 2009 at the Newton Institute in Cambridge. We hope to be able to run some collaborative activities with the Newton Institute during the Spring semester 2009.

Many of the programs at SAMSI have addressed some aspect of the analysis, computation, or application of stochastic dynamics, for example Large Scale Computer Models for Environmental Systems, Random Media, Multiscale Model Development and Control Design. SAMSI has also had success in the past running programs with a stochastic focus, most notably the programs on Stochastic Computation (2002-03), Inverse Problem Methodology in Complex Stochastic Models (2002-03), and Network modeling for the internet (2003-04). The success of these programs indicates that a full year dedicated to a broader but cohesive set of subjects related to Stochastic Dynamics will be of great interest in both the statistical and applied mathematical communities.

2.5 Organization and Program Participants

Overall Program Leaders: Cindy Greenwood (ASU), Pete Kramer (RPI) , Alejandro Garcia, Peter Mucha (UNC), Jonathan Mattingly (Duke)

Current Scientific Advisory Committee: Hongyun Wang (UC Santa Cruz), Alejandro Garcia (San Jose State), Cindy Greenwood (ASU)

Local Scientific Coordinators: Alan Karr (NISS), Jonathan Mattingly (Duke), Peter Mucha (UNC)

Directorate Liaison: Michael Minion (SAMSI)

National Advisory Committee Liaison: Rick Durrett (Cornell)

Note: Additional working group leader will be appointed during the opening workshop for each the research foci, with special attention paid to ensure diversity.

Confirmed Long-Term Visitors: Cindy Greenwood (ASU), Peter Kramer (Math, RPI), Lea Popovic (Concordia), Gabriel Lord (Heriot Watt), Kevin Lin (Arizona), Robert Pego (Maryland), John Fricks (Penn State), Anna Amridjanova (U Michigan), Carlos Manuel Mora González (Universidad de Concepción)

Postdoctoral Fellows: Graduate training programs in applied mathematics and biostatistics have been working to increase the level of training in stochastic analysis and computation in recent years. Application of these techniques in both math biology and networks are very active research areas and we have recruited post-docs with expertise spanning the research foci of the program including Emily Fox (MIT), Bruce Rogers (Arizona State), Avanti Athreya (Maryland), and Scott McKinley (Duke). We are currently looking to fill one or two more post-doctoral positions from the research increase in funding received from the NSF.

Faculty Fellows and Local Researchers: The three partner universities will provide approximately 6 local faculty to participate in the program. Among the local scientists that will be heavily involved are Jonathan Mattingly (Math, Duke), David Banks (Stat, Duke), Sorin Mitran (Math, UNC), Peter Mucha (Math, UNC), David Adalsteinsson (Math, UNC), Tim Elston (Pharm. UNC), H. T. Banks (Math, North Carolina State), Jason Fine (BioStat UNC) and Amy Herring (BioStat UNC), Kazufumi Ito, (NCSU Math), Alina Chertok, (NCSU Math), Alun Lloyd, NCSU Math, Mike West (Duke).

2.6 Description of Activities

2.6.1 Workshops

Opening Workshop: The opening workshop will be held Aug. 30-Sept. 2, 2009 at SAMSI. This workshop will aim to engage as large a part of the statistics, mathematics, and relevant scientific communities as possible, with representative sessions from all of the main program topics. The first day of the workshop will be devoted to four tutorial sessions on the topics

- Introduction to Stochastic dynamical systems
- Stochastic modeling and applications to Biology and the Medical Sciences
- Numerical Methods for Stochastic Dynamics
- Estimation and data assimilation in Stochastic Systems

The next three days of the workshop will contain five sessions of research talks and panel discussions devoted to the themes

- Qualitative behavior of stochastic dynamical systems and stochastic modeling
- Stochastic Dynamics across many scales
- Challenges in numerical methods for stochastic systems
- Estimation and data assimilations in stochastic dynamics
- Dynamics of biological networks

There will also be sessions devoted to new researchers, a poster session, and a “5 Minute Madness” session wherein speakers will be given five minutes to present relevant research results.

Self-Organization and Multi-Scale Mathematical Modeling of Active Biological Systems: The workshop to be held October 26-28, 2009, will bring together mathematicians, statisticians, biophysicists and engineers to discuss the latest developments in the field of self-organization and multi-scale description of active biological systems, such as suspensions of swimming microorganisms and biofluids, evolving cytoskeletal networks, and many others.

Other Program Workshops: Further workshops will be organized by program participants. Possible workshop themes being discussed are:

A workshop centered around the analysis and computation of multi-scale systems with small scale stochastic forcing. This is a critical topic in areas such as bio-fluid dynamics, meteorology, combustion, and materials science. Engaging mathematical

analysts, statisticians, computational scientists, and application stake-holders interested in this topic could lead to fundamental breakthroughs in this emerging field.

A workshop on the dynamics of networks

A workshop on stochastic modeling in the bio-sciences.

The *Transition Workshop*: The transition workshop will be held in June, 2010 and will disseminate program results and chart a path for future research in the area.

2.6.2 Working Groups

Working groups will meet throughout the program to pursue particular research topics identified in the kickoff workshop (or subsequently chosen by the working group participants). The working groups consist of SAMSI visitors, postdoctoral fellows, graduate students, and local faculty and scientists.

2.6.3 University Courses

During the Fall Semester of 2009, SAMSI will offer the course *Stochastic Dynamics: Theory, Modeling, and Computation*. Cindy Greenwood and Jonathan Mattingly will organize the course and include several SAMSI participants as guest lecturers.

3 Psychometric Modeling and Statistical Inference

3.1 Scientific Overview

Much of current psychometric research involves the development of novel statistical methodology to model educational and psychological processes, and a wide variety of new psychometric models have appeared over the last quarter century. Such models include (but are not limited to) extensions of item response theory (IRT) models, structural equation models (SEMs), cognitive diagnosis models, and generalized linear latent and mixed models (GLLMM). The development of several of these models has been spearheaded by quantitative psychologists, a group of researchers who find their academic homes primarily in psychology and education departments. During the same period, very similar models and methodologies were developed—often independently—by academic statisticians residing in mathematics and statistics departments. The lack of interaction between these two groups has resulted in a substantial duplication of effort and, more importantly, a delay in the development of methodology crucial to both fields. The goal of this program is to bring researchers from both areas together to explore possible avenues for mutual collaboration.

3.2 Program Leadership

The Program Leaders Committee is currently comprised of Charles Lewis (Fordham University), Richard Swartz (University of Texas M.D. Anderson Cancer Center), and Valen Johnson (University of Texas M.D. Anderson Cancer Center); Directorate Liaison is James Berger (SAMSI). Workshop organizers include Jimmy de la Torre, David Banks, and David Thissen.

3.3 Program Participants

We envision that the tutorials, invited contributions, and software demonstrations presented in the first week will attract a diverse group of approximately 50 participants from the psychometric and statistical community. We expect 20-25 participants to remain in residence or to attend the contributed sessions and working-group meetings in the second week. Junior investigators will be actively recruited to participate in the working-group meetings conducted during the second week of the meeting.

3.4 Program Outcome

The goal of this program is stimulate collaborations between researchers in the psychometric and statistical communities. The desired outcome for the program will be a well-defined, concrete list of specific research directions that will facilitate methodological development in related psychometric/statistical models. These will be brought to the attention of the research community via the planned white papers summarizing these directions.

3.5 Program Scope, Timing and Activities

The program will take place within the two-week period between July 7 and July 17, 2008 (Tuesday-Friday (July 7–10) and Monday-Friday (July13–17); no events are planned during the weekend of July 11–12. The following activities are planned:

Week 1: July 7-10

The Psychometric Program will kick-off on Tuesday morning, July 7, 2009. The mornings will be tutorials and the afternoons will be devoted to invited talks by statisticians on topics that relate to the psychometric models presented during the morning tutorials, as well as group discussion of the connections between the approaches. We will also organize demonstrations of software packages frequently used by psychometricians to fit standard psychometric models. Tentative titles and speakers for invited contributions are listed below. Talks will be scheduled for approximately 60-90 minutes.

Tuesday, July 7

- 9:00-12:00** Introduction to Item Response Theory. Yanyan Sheng
- 2:00-3:00** IRT PRO demonstration. David Thissen

Wednesday, July 8

- 9:00-12:00** A Nonlinear Mixed Models Approach to IRT. Mark Wilson, Frank Rijmen.
- 2:00-4:00** Topics in Response Time Analysis. Mario Peruggia and Trish Van Zandt

Thursday, July 9

9:00-12:00 An Introduction to Cognitive Diagnostic Models. Mathias von Davier.

2:00-3:00 CDM Pitfalls and Recommendations. Sandip Sinaray.

Friday, July 10

9:00-12:00 An Introduction to Rater Models. Matthew Johnson.

2:00-4:00 Process de-association and signal detection. Dongchu Sun and Jun Lu.

Week 2: July 13-17

The second week of the conference will consist of a mixture of contributed talks and working-group discussions. A tentative list of working groups and their activities follows:

The **Peer Review Working Group** will meet during the second week of the program. Talks currently scheduled during this week are listed below. Additional talks will be added to the schedule during the course of the Psychometric Program. A white paper will be prepared by the working group and summarized at a joint meeting at the end of the week.

Monday, July 13

10:00-11:00 An Overview of *Journal of the American Statistical Association* Article Reviews. David Banks.

1:30-2:30 An Overview of NIH R01 Peer-Review Scores. Valen E. Johnson.

Tuesday, July 14

10:00-11:00 A Bayesian Approach to Ranking and Rater Evaluation: An Application to Grant Reviews. Jing Cao.

1:30-2:30 A Bayesian Hierarchical Model for Multi-rater Data with Fine Scales. Song Zhang.

Friday, July 17

9:00-12:00 Discuss draft of white paper.

12:00 Adjourn.

The **Patient Reported Outcome Working Group (PRO WG)** will meet during the second week of the program. Tentative titles for talks currently scheduled during this week are listed below. Additional talks and the working group agenda will be finalized on Monday, July 13. A white paper will be prepared by the working group and summarized at a joint meeting at the end of the week.

Monday, July 13

10:00-11:00 Practical Issues in the use of Patient Reported Outcomes. Charlie Cleeland.

11:00-12:00 Issues in longitudinal analysis of Patient Reported Outcomes. Bryce Reeves.

2:00-4:00 Group discussion of weeks agenda. Charlie Cleeland and Bryce Reeves.

Tuesday-Friday Group collaboration on consensus agenda.

Friday, July 17

9:00-12:00 Discussion of the draft of white paper.

1:00-4:00 Finalize white paper.

4:00 Adjourn.

The **Applications and Challenges of Cognitive Diagnostic Models Working Group (CDM WG)** will meet during the second week of the program. Tentative titles for talks currently scheduled during this week are listed below. Additional talks and the working group agenda will be finalized on Monday, July 13. A white paper will be prepared by the working group and summarized at a joint meeting at the end of the week.

Monday, July 13

10:00-12:00 Discussion of cognitive diagnostic models and identification of agenda by working group members. Jimmy de la Torre

Tuesday-Friday Group collaboration on consensus agenda.

Friday, July 17

9:00-12:00 Discussion of the draft of white paper.

1:00-4:00 Finalize white paper.

4:00 Adjourn.

B. Scientific Themes for Later Years

1. Analysis of Object Data

1.1 Introduction

This will be a year-long SAMSI program for 2010-2011 on the analysis of complex data types that are an extension of Functional Data Analysis where one considers methods to analyze data samples of complex objects.

Program Leaders: Hans-Georg Müller (Davis), Jane-Ling Wang (Davis)

Local Scientific Coordinator: Steve Marron (UNC)

Program Co-Leaders: Ian Dryden (South Carolina), Jim Ramsay (McGill)

Directorate Liaison: Nell Sedransk (SAMSI)

National Advisory Committee Liaison:

Additional leaders will be appointed for the various key areas from the long-term visitors.

Modern science is generating a need to understand, and statistically analyze, populations of increasingly complex types. The term “Analysis of Object Data” (AOD) is aimed at encompassing a broad array of such methods. The proposed SAMSI program seeks to bring together a diverse group of researchers (from statistics, other parts of mathematics, and related sciences) to explore the common structure that underlies such methodologies, and to use this knowledge in turn to motivate and synthesize new approaches.

1.2 Program Overview

AOD is an extension of the very active research area of Functional Data Analysis (Ramsay and Silverman 2002, 2005). It essentially generalizes the fundamental FDA concept of *curves as data points*, to more general *objects as data points*. Examples include images, shapes of objects in 3d, points on a manifold, tree structured objects, and various types of movies.

As noted in Wang and Marron (2007), specific AOD contexts can be grouped in a number of interesting ways. A grouping of perhaps mathematical interest is considered first. This is in terms of the type of space in which the data objects lie:

- Euclidean, i.e., (constant length) vectors of real numbers.
- Mildly non-Euclidean, i.e. points on a manifold and shapes.
- Strongly non-Euclidean, i.e. tree or graph structured objects.

Euclidean data objects are quite ubiquitous in a variety of AOD contexts. One focus will be on *Functional Data Analysis* (FDA), viewing curves as data. These curves are commonly either simply digitized, or else decomposed by a basis expansion, which gives a vector that represents each data curve. Many examples of this type of data appear in the Ramsay and Silverman (2002, 2005) books. Evolutionary biology and longitudinal applications will be important drivers of the FDA and shape analysis considered in this program. Especially the social and biological

sciences provide many examples for longitudinal studies which can be modeled as functional data.

A second focus is *Time Dynamics Data*, with an emphasis on differential equations and dynamic systems as drivers of fully or incompletely observed samples of stochastic processes. This will also include point and marked point processes as data objects. Applications can be found in control, engineering, biological modeling of growth or cell kinetics and in e-commerce, where the analysis of auction dynamics is of great interest. In the social sciences repeated events such as child births of a woman and lighting times of cigarettes by a smoker have been studied. Other examples include asthma attacks in medical studies, the dynamics of HIV infections, and the dynamics of gene expression and relations with gene networks.

Another focus will be *Shape Analysis and Manifold Data*, where for example 2 or 3 dimensional locations of a set of common landmarks are collected into vectors that represent shapes. While these vectors are just standard multivariate data, they frequently violate standard multivariate assumptions, such as the sample size being (usually much) larger than the dimension. Research in the direction of High Dimension Low Sample Size (HDLSS) issues will be a major emphasis of the proposed SAMSI program. In addition, the landmarks may be invariant to certain transformations such as location, rotation and scale, and Kendall's shape analysis of such objects leads to non-Euclidean distances being the most natural. Further recent examples include analysis of shapes of unlabeled points, especially on curves, surfaces and images. The closely related manifold data also are based on non-Euclidean distances.

Data which naturally lie in a manifold have been in the statistical literature for some time in the form of directional data (data points which are circular or spherical angles) and play an increasingly important role for the analysis of shapes.

Modern *Image Analysis* applications where the data consist of a sample of images will be another program focus. Such data can be often understood as being located on manifolds. These include medial representations for shape objects (involving a mix of real numbers and angles as parameters), diffusion tensor imaging (a branch of magnetic resonance imaging, which represents directionality of fluid flow using tensors), and diffeo-morphisms (a powerful mathematical approach to studying warpings of space that address non-affine registration challenges.). While manifold data present major statistical challenges (because most statistical methods are very Euclidean in nature), they are termed “mildly non-Euclidean”, because manifolds admit tangent plane approximation, so that (at least when the data are sufficiently concentrated near the point of tangency) approximate Euclidean methods have been employed to good effect. A wide open research area, that will be a major focus on the SAMSI program, is the development of “intrinsic” methodologies, where the statistical analysis is carried out really inside the manifold, which thus avoid distortion problems for manifold data that are not concentrated in a small area.

A fifth focus concerns *Tree and Graph Structured Data*. These objects are “strongly non-Euclidean”, because the data space admits no tangent plane approximation. Thus, there is no apparent approach to adapting even approximate Euclidean methodologies, and statistical analysis must be invented from the ground up. The first workable methodology of this type

appears in Aydin et al (2008). But this field is in its infancy, with large potential as a context for the development of new ideas. Thus it will be another focus of the SAMSI program.

Another way to group AOD contexts is in terms of mathematical areas involved, which highlights the potential synergies that we aim to develop through this SAMSI program. These include:

- Statistics – this is a common theme to all parts of the proposal. Statistics itself as a discipline will be benefitted through the invention of new ways of understanding statistical methods. A clear example of this will be HDLSS asymptotics, which are anticipated to both inform, and be driven by, the methodological component of the program.
- Optimization – in most contexts above (especially manifold and tree structured data) statistical ideas result in optimization problems that can be very challenging to solve. This is anticipated to lead to the development of new ideas for addressing optimization problems. Furthermore, the SAMSI collaboration is intended to lead to a deeper interaction between statisticians and optimizers at all stages of the method development.
- Geometry – there are major geometric challenges, especially in the area of manifold data. The SAMSI program will seek to move beyond the current mode of “statisticians using geometric ideas”, to serious collaboration between statisticians and geometers, again at all stages of method development, seeking connections with the emerging fields of computational topology and metric geometry.
- Probability – there were very early strong connections between statistics and probability that have languished somewhat recently. This program will provide an opportunity to replenish this link between areas. In particular, important open questions are the development of appropriate, e.g. “normal” probability distributions for data lying on manifolds, or tree structured data.
- Differential Equations – As noted in Ramsay and Silverman (2002) there already has been strong application of differential equation ideas in FDA. Another important interface is that a very promising approach to the generation of “normal” distributions on exotic space, is the heat diffusion equation approach. Finally, dynamical systems have become a very active research area in the modeling of biological and other temporal and spatio-temporal phenomena and there exists a natural link with functional data analysis methodology that has not been explored yet. Developing this link will lead to better understanding of such systems and new directions for AOD.
- Topology – an emerging new statistical field is topological data analysis, which seeks to understand structure in very high dimensions, via reducing high dimensional density estimates to focus on informative topological aspects.

Finally AOD contexts can be grouped in terms of application areas:

- Image Analysis has provided a number of driving problems for AOD. Modern images are frequently in 3-d, and the current research focus is on populations of images (as opposed to early challenges, such as denoising a single image). A central problem is registration, e.g. across images handling the problem that organs of interest will be in different locations. There are a variety of approaches to this,

all of which involve AOD at some level. One approach is registration via diffeomorphisms (which themselves naturally lie in a manifold), and these can also be used to analyze population variation. Another is medial representations, which yield a different type of manifold data. Finally, Diffusion Tensor Imaging is naturally analyzed as yet another type of manifold data. A completely different type of AOD image data is trees as data, as discussed in Aydin et al (2008), which are strongly non-Euclidean as noted above. One more challenging data AOD data type comes from Functional Magnetic Resonance Imaging, where each data object is a movie (over time) of 3-d images.

- Bioinformatics data, including microarrays (for gene expression), SNP arrays, proteomics and metabolomics, provides another rich source of driving problems for AOD. While such data sets are typically Euclidean, severe challenges exist because of their HDLSS nature. Major challenges to be investigated during the SAMSI program include data fusion, where the goal is to extract joint information from several of these modalities at once.
- Evolutionary biology has recently actively engaged in FDA methodologies. Examples include the evolution of character traits that correspond to random functions or biodemographic trajectories of mortality, reproduction and other behaviors that are shaped by evolution. The SAMSI program aims to engage with this community, and extend the range of data types, while at the same time developing new methodologies, which can be used in other contexts.
- The emerging area of e-commerce and more generally econometrics has fairly recently made contact with AOD. The strongest connection has been in terms of full transcripts of online auction (e-Bay) bids being viewed as FDA data objects or trajectories of box office receipts of movies after opening day, for example with the goal to predict the overall receipts to be expected for a movie. The proposed SAMSI program aims to carry this research forward, through increased contact with FDA researchers, and through exploring the application of advanced data structures, such as tree or graph structured objects, in this context.
- Psychiatry, psychology and social sciences also have strong connections with AOD. In particular, both autism and schizophrenia have been associated to sizes and shapes of a variety of brain structures. Longitudinal studies often with irregular sampling designs are common in the social sciences. In the presence of nonlinear structures, FDA methodology provides promising alternatives to classical parametric models with random effects. There are also often multivariate time courses and the modeling of complex interactions between their components is then of interest. AOD provides an ideal framework and way of thinking about populations of objects of this type.

1.3 Research Foci and Key Participants

Many of the prospective participants are bridging several of the research areas that will be included under the AOD theme and this will help to generate increased synergies and interactions between these areas, and encourage the potential for interactions between

researchers using different approaches. Nevertheless, we provide a rough grouping below of the key long-term visitors expected in each of the five areas:

1. Functional Data Analysis
2. Analysis of Time Dynamics
3. Shape Analysis and Manifold Data
4. Analysis of Image Data
5. Tree and Graph Structured Data

(FDA)

Last Name	First Name	Affiliation
Ding	Jimin	Wash U, St. Louis
Hall	Peter	U Melbourne
Senturk	Damla	Penn State
Stadtmueller	Uli	U Ulm
Yao	Fang	U Toronto
Kneip	Alois	U Bonn
Boente	Graciela	University of Buenos Aires
Munoz	Yolanda	Michigan Tech
Cao	Jiguo	Simon Fraser

2 (Dynamics)

Dowd	Michael	Dalhousie
Hooker	Giles	Cornell
King	Aaron	U Michigan
Wu	Rongling	Penn State
Wu	Huilin	Rochester
Girolami	Mark	U Glasgow
Stuart	Andrew	U Warwick
Brunel	Nicolas	University d'Evry
Campbell	David	Simon Fraser

3 (Shapes and Manifolds)

Huckeman	Stephan	Goettingen, Germany
Kent	J.T.	U of Leeds
Le	Huilin	Nottingham
Patrangenaru	Vic	Florida State
Wood	Andy	Nottingham
Hotz	Thomas	Inst Mathematical Stochastics

4 (Images)

Aston	John	Warwick
Chiou	Jeng-Min	Academia Sinica, Taipei
Joshi	Sarang	Utah

Morris	Jeff	U Texas
Olhede	Sofia	UCL
Panaretos	Victor	Lausanne, Switzerland

5 (Trees)

Ahn	Jeongyoun	U Georgia
Park	Byeong	Seoul National University
Wang	Haonan	Colorado State
Whitaker	Ross	U Utah
Kim	Yongdai	Yonsei University
Srivastava	Anuj	Florida State

1.4 Participants and Personnel

Key leaders and participants have already been discussed above. Here we mention other categories of participants.

Postdoctoral Fellows: There are numerous graduate students being trained in analysis of Object data so that we expect there to be great interest in the program among graduating Ph.Ds.

Faculty Fellows and Local Researchers: The three partner universities will provide approximately 6 local faculty to participate in the program. Among the local scientists that will potentially be heavily involved are Christina Burch (UNC), Jim Damon (UNC), Herbert Edelsbrunner (Duke), Jason Fine (UNC), Joel Kingsolver (UNC), Katia Koelle (Duke), Hamid Krim (NCSU), Mauro Maggioni (Duke), Steve Marron (UNC), Sayan Mukerjee (Duke), Steve Pizer (UNC), Scott Schmidler (Duke), Nell Sedransk (SAMSI, NISS), Haipeng Shen (UNC), Young Truong (UNC), Mike West (Duke), Hong-Tu Zhu (UNC).

Graduate Students: The three partner universities will provide research assistantships for approximately 6 students to participate in the program. Of course, a number of visiting graduate students are also expected.

1.5 Description of Activities

Workshops: The *Opening Workshop* will be held September 12-15, 2010 at SAMSI. This workshop will aim to engage as large a part of the statistics, mathematics, and relevant scientific communities as possible, with representative sessions from all of the main program topics. The *Transition Workshop* at the end of the program will disseminate program results and chart a path for future research. There will also be workshops relating to the five focus areas.

Working Groups: Working groups will meet throughout the program to pursue particular research topics identified in the kickoff workshop (or subsequently chosen by the working group participants). The working groups consist of SAMSI visitors, postdoctoral fellows, graduate students, and local faculty and scientists, and will be structured along the five focus areas.

References:

- Aydin, B, Pataki, G., Wang, H. N., Bullitt, E. and Marron, J. S. (2008) Tree-line analysis of populations of tree structured objects; submitted.
- Ramsay, J. O. & Silverman, B. W. (2005) *Functional Data Analysis*, 2nd Edition, Springer, N.Y.
- Ramsay, J. O. & Silverman, B. W. (2002) *Applied Functional Data Analysis*, Springer, N.Y.
- Sen, S., Foskey, M., Marron, J. S. and Styner, M. (2007) Shape Analysis using Novel Classification Methods Developed for Data on Manifolds: An Application to M-reps, submitted to ICCV.
- Wang, H. and Marron, J. S. (2007) Object data analysis: sets of trees, *Annals of Statistics*, 35, 1849-1873.

2 Complex Networks (2010-2011)

2.1 Introduction

This year-long program will focus on the emerging area of *network science*. This highly interdisciplinary field is characterized by novel interactions in the mathematical sciences which are occurring at the interface of applied mathematics, statistics, computer science, and statistical physics, as well as those areas with network-oriented thrusts in biology, computer networks, engineering, and the social sciences.

A network is a set of items (vertices) with connections (edges) between them. The mathematical study of networks goes back to (at least) Euler (1735) [2] with the solution to the famous problem of the seven bridges of Königsberg. This result is often regarded as the beginning of graph theory under which wide umbrella the present program belongs. Based on *empirical* studies of specific applications such as the Internet, social, biological and technological networks, significant progress has been made in recent years regarding our understanding of such systems [1, 4, 5]. Various types of quantitative measurements have been proposed and studied; distinctive statistical signatures characterizing specific types of networks are starting to emerge.

The above observations have led to *analytical* efforts aimed at explaining network structures and predicting their capabilities. These theoretical studies tend to focus more on large-scale statistical properties of graphs and include work on Markov graphs, small-world models and models of network growth. Finally, *applied* studies focus on the behavior of processes on networks such as the spread of infection over networks (social or computer), the effect of node failures on communication networks and properties and behavior of various dynamical systems on networks. Gaining a better understanding of the networked systems we encounter in nature or build for technological purposes is the ultimate goal in this field of research.

In spite of many successes, the study of complex networks is still in its infancy in many ways. It is proposed to use four interconnected research foci as a mean to identify and explore the common key mathematical and statistical issues which underlie the empirical, analytical and applied approaches described above.

Current Overall Program Leaders: Eric Kolaczyk (Boston U.), Alessandro Vespignani (Indiana U.)

Current Scientific Advisory Committee: Pierre Degond (Institut de Mathématiques de Toulouse), Stephen Fienberg (Carnegie Mellon U.), Martina Morris (U. of Washington)

Local Scientific Coordinators: Alun Lloyd (NCSU), Peter Mucha (UNC)

Directorate Liaison: Pierre Gremaud (NCSU)

National Advisory Committee Liaison: Bin Yu (UC Berkeley)

2.2 Research foci

2.2.1 Network modeling and inference

Potential leaders and key participants: Mark Handcock (U. of Washington), Eric Kolaczyk (BU)

Mathematics participants: Don Estep (Math/Stat, Colorado State), Reinhard Laubacher (Virginia Tech)

Statistics participants: Edoardo Airoldi (Harvard), Peter Bühlmann (ETH), Sourav Chatterjee, Hugh Chipman (Acadia U.), Nial Friel (U. College Dublin), Haylan Huang (UC Berkeley), Susan Holmes (Stanford), Elizaveta Levina (Michigan), Crystal Linkletter (Brown U.), Sach Mukerjee (U. Warwick), Stanley Wasserman (Indiana U.), Patrick Wolfe (Harvard), Wing Wong (Stat. and Health Research, Stanford), Bin Yu (UC Berkeley)

Participants from other disciplines: Albert-László Barabási (Physics, Notre-Dame), Aaron Clauset (Computer Sc., Santa Fe Institute), Mark Newman (Physics, U. of Michigan), Marco Saerens (Machine Learning, U. Catholique de Louvain).

The analysis of network data has become a major endeavor across the sciences, and network modeling plays a key role. Frequently, there is an inferential component to the process of network modeling i.e., inference of network model parameters, of network summary measures, or of the network topology itself. For most standard types of data (e.g., independent and identically distributed, time series, spatial, etc.), there is a well-developed mathematical infrastructure guiding modeling and inference in practice. In the context of network data, however, such an infrastructure is largely lacking.

To date, the majority of the energy on network modeling has been devoted to the specification of network models (e.g., through classes of random graphs or through generative mechanisms). There has been some work in recent years noticeably advancing our understanding about fitting parameters for certain classes of network models (i.e., exponential-family random graph and latent space models primarily, but also a bit involving generative models). There also is a substantial older literature on inference of network summary measures (e.g., triad censuses, centrality measures, etc.), under various sampling designs, and a small but active recent literature picking up some of the older threads in the modern context. Nevertheless, both areas are arguably still in their early and formative stages, falling short of what we would like to demand of them in practice. Moreover, while these two areas have developed along largely distinct paths in the literature to date, the incorporation of the sampling-based perspective of the latter with the model-based perspective of the former is clearly needed in many practical contexts. Finally, there is a large and growing body of literature on the inference of network topology. However, while this area is rich in methodology, it is poor in the supporting concepts and mathematics necessary to carefully quantify issues relating to validation of inferred networks.

Current limitations in this area can perhaps be traced in no small part to the inherent tension between the simplicity of network models needed for tractability (e.g., of simulation, interpretation, and mathematical study) and the complexity needed to accurately describe reality. Realistically, the tasks of model specification and model inference need to be more closely tied together, with each being informed by the other.

2.2.2 Flows on networks

Potential leaders and key participants: Reka Albert (Physics/Bio, Penn State), Pierre Degond (Math., Institut de Mathématiques de Toulouse)

Mathematics participants: Béla Bollobás (U. Memphis), Rick Durrett (Cornell U.), Oliver Riordan (Oxford), Pieter Swart (LANL), Jean Paul Watson (Sandia), Chris Wiggins (Columbia U.)

Statistics participants: Peter Bickel (U.C. Berkeley), Jan Hanning (UNC), George Michailidis (Stat and ECE, U. Michigan),

Participants from other disciplines: Dirk Helbing (Sociology, ETH), Ravi Kumar (CS, Yahoo), Marathe Madhav (CS, Virginia Tech), Michael Mahoney (Computer Sc., Stanford), Robert Nowak (ECE, Wisconsin), Guy Theraulaz (CICT), Josh Socolar (Physics, Duke), Zoltan Toroczkai (Physics, Notre Dame)

In their simplest form, network flows are defined on directed graphs. Each edge receives a flow in an amount that cannot exceed the capacity of the edge. Many transport applications correspond to network flows: hydraulics and pipeline flows, rivers, sewer and water systems, traffics and roads, supply chains and cardiovascular systems, to name but a few. Several by now classical problems for network flows such as maximum flow have been solved for static flow [3]. These results only partially carry over to dynamic flows (time extended networks) and much remains to be done. Some applications such as communication systems typically split data into packages. There are obvious technical limitations regarding the fineness of such decompositions that have to be taken into account when seeking (quasi-) optimal solutions. Several of the relevant open questions fall under the umbrella of combinatorial optimization.

Transport problems on networks may behave in unexpected ways due to interactions between their different components. For instance, networks containing closed loops/circuits may exhibit phenomena such as localized and/or sustained oscillations; even simple networks, such as Boolean networks, may exhibit phase transition from ordered to chaotic dynamics. Similarly, the statistical modeling and analysis of various types of network flow measurements includes a number of highly ill-posed but sparse inverse problems. The very topology of the networks therefore plays a fundamental role in the behavior of the problems defined on them. Examples abound such as for instance the properties of 3D networks built by social insects and how the network's topology and geometry influence the traffic organization of insects inside the structure. Neither theory nor numerical methods can/should be devised

for such applications by simple “superposition” of existing results or methods for problems in standard domains. The construction of efficient mathematical, numerical and statistical tools for such applications is an important challenge.

2.2.3 Network models for disease transmission

Potential leaders and key participants: Alun Lloyd (Math, NCSU), Lauren Meyers (Bio/Math, U.T. Austin) (to be confirmed)

Mathematics participants: David Bortz (Colorado), Matt Keeling (Biological Sciences and Math., U. of Warwick), Peter Mucha (UNC)

Statistics participants: Tom Britton (Stockholm U.), Andrew Lawson (USC)

Participants from other disciplines: Marc Girolami (CS, U. of Glasgow), Brian Grenfell (Ecology, Princeton), Vincent Jansen (Bio, U. of London), Svante Janson (Uppsala U.), Frederic Liljeros (Sociology, Stockholm U.), Martina Morris (Sociology, U. of Washington), Michael Stumpf (Bioinformatics, Imperial College), Alessandro Vespignani (Physics, Indiana U.), Sharon Weir (UNC, Epidemiology).

Network models provide a natural way to model many infectious diseases. Many diseases, such as sexually transmitted infections (STIs), have long been studied in terms of networks, but in recent years the approach has been adopted in a wider range of disease settings, including acute rapidly-spreading infections. Disease transmission networks are highly dependent on the infection of interest: the sexual partnership network across which an STI spreads has a quite different structure to the social network on which a respiratory infection (such as influenza) would spread. Even in the same population, different diseases see different networks.

Broadly speaking, network-based disease studies have either involved detailed modeling of a specific infection in a particular setting (tactical models) or attempted to elucidate general principles of how particular network structures impact disease transmission (strategic models). Tactical models require a great deal of information about the structure of the relevant network (in addition to the biological details of the disease transmission process). They draw heavily on statistical efforts to quantify real-world networks. Strategic studies, on the other hand, typically focus on one aspect of network structure (such as distance or clustering) and examine its impact on the spread of infection. Simulation-based studies of this kind are reliant on algorithms that generate prototypical networks having a specified property (e.g. the Watts-Strogatz small world network, the Barabasi-Albert scale-free network or Newmans clustered network). Disease network models stand to benefit from advances in statistical methodologies for sampling and quantifying networks as well as in network-generation algorithms.

A number of analytic approaches have been used to study the spread of infection on networks. The use of percolation and branching process theory has been particularly fruitful,

providing results on epidemic thresholds, outbreak probabilities and outbreak size distributions. Moment closure approaches have also been widely used to capture the impact of various aspects of network structure, such as clustering and local spatial structure. Much of this work, however, has assumed that the transmission network is static. In some settings this static picture is unlikely to provide a good description: in a monogamous population, the persistence of a sexually transmitted infection relies upon the break-up and formation of partnerships. Dynamic network structure can be important even for acute, rapidly spreading, infections. The development of analytic techniques that can describe the spread of infection on dynamic networks is a major and important challenge.

2.2.4 Dynamics of networks

Potential leaders and key participants: Raissa D’Souza (Engineering, UC Davis), Stephen Fienberg (Stat, CMU), Peter Mucha (Math, UNC)

Mathematics participants: H. T. Banks (NCSU), Jonathan Mattingly (Duke), Mason Porter (Oxford), Juan Restrepo (Sandia)

Statistics participants: David Banks (Duke), Lisha Chen (Yale), Alan Karr (NISS), Eric Kolaczyk (B.U.), Mark Handcock (U. of Washington), Tom A.B. Snijders (Politics, Statistics, Oxford), Stanley Wasserman (Indiana U.), Mike West (Duke)

Participants from other disciplines: Lada Adamic (Information, Michigan), Tanya Berger-Wolf (CS, Illinois), Dave Blei (CS, Princeton), Skyler Cranmer (Political Science, UNC), James Fowler (Political Science, UCSD), Simon Levin (Biology, Princeton), Michael Macy (Sociology, Cornell), James Moody (Sociology, Duke), Brian Skyrms (Logic & Philosophy, UCI), Chris Volinsky (AT&T).

The changing structure of networks over time is inherent in the study of a broad array of phenomena. Examples for which a *static* transmission network is inadequate abound: from disease transmission to communications networks with changing landscape of connections to political networks where associations and voting similarities vary from one legislative session to the next. While the nature of the underlying processes differs, the flow of generalized information, for all three examples, depends in a nontrivial way on the changes in the node roles, in the structure of communities and in other coarse structural units. The importance of dynamics in networks has been long recognized. The increasing accessibility of network data has led to renewed interest in this area; examples of data include longitudinal data waves and financial correlations with strengths of connection defined over moving windows in time.

Turning now to specific applications, recent progress in bioengineering technologies has made possible the measurement at increasingly high-resolution of dynamic data on complex cellular networks at multiple scales. As a result single-cell molecular studies have become a critical emerging area due to their potential for providing the opportunity for controlled

experimentation and bionetwork design. Further, this type of study can emulate key aspects of mammalian gene networks central to all human cancers.

So far, most of the theoretical modeling work done on the dynamics of networks has been focused on the statistical equilibria of those models (e.g., growing networks by preferential attachment) or on one-time disruption events (e.g., the effect of knocking out hubs). At the same time, computational tools for analyzing and visualizing time-varying networks remain relatively few in number, especially as compared to the wealth of advances in methods for modeling and analyzing static networks. There is thus both need and opportunity for more thorough mathematical and statistical analysis and modeling of dynamic networks.

2.3 Program timing and previous related programs

The IMA had a thematic year on Probability and Statistics in Complex Systems: Genomics, Networks, and Financial Engineering in 2003-04 (<http://www.ima.umn.edu/complex/>). Most of the emphasis was on the structure of the Web, the genome and financial applications which are themes that the present proposed program is not designed to emphasize at this time.

MBI had a year on Mathematical Neuroscience in 2002-03 with main emphasis on signal representation and neuronal dynamics. It also will have a year in 2009-2010 on Molecular interactions within the cell: network, scale and complexity <http://mbi.osu.edu/sciprograms/scientific2009.html>. The MBI program is focused on the specific application of network biology to the modeling of intracellular phenomena. The SAMSI program will take a much broader approach to network science in general. The biological applications being currently considered, namely epidemiology, are obviously very different from the MBI topic.

The Sante Fe Institute also has a recurring theme in Physics of Complex Systems with a sub-theme of Networks: social, biological and technological <http://www.santafe.edu/research/topics-physics-complex-systems.php#3>. Possible interactions and common activities with the Sante Fe Institute will be explored.

IPAM had a one-week workshop on Flows and Networks in Complex Media in Spring 2009 <http://www.ipam.ucla.edu/programs/ktws3/> as part of its year on Quantum and Kinetic Transport: Analysis, Computations, and New Applications. This week long event did have some common aspects with the applied mathematics side of the proposed theme on “Flows on networks”. We expect some of the participants to that event to get involved in the current program and its more systemic approach, even though the intersection of the two list of participants is currently empty.

One of the research foci of the 2009-2010 SAMSI program on “Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change” is spatial epidemiology. The issues to be studied there (essentially effects of air pollution) are fundamentally different from those under consideration here (disease transmission).

This proposed SAMSI program represents thus a unique and timely opportunity for mathematicians, statisticians and scientists to make significant advances in the fast moving research area of network science and its applications. The issues at hand and progress to

be made are indeed in perfect agreement with SAMSI's vision and mission to achieve a new synthesis of the statistical sciences and the applied mathematical sciences with disciplinary science to confront the very hardest and most important data- and model-driven scientific challenges.

2.4 Participants and Personnel

Key leaders and participants have already been discussed above. Here we mention other categories of needed participants.

Potential long term visitors: Edoardo Airoldi (Harvard), Joe Blitzstein (Harvard), David Bortz (Colorado), Lisha Chen (Yale), Aron Clauset (Santa Fe), Raissa D'Souza (Stanford), Pierre Degond (Toulouse), Vanja Dukic (U. Chicago), Don Estep (Colorado State), Mark Handcock (U. Washington), Dirk Helbing (ETHZ), Eric Kolaczyk (BU), Andrew Lawson (U. South Carolina), Liza Levina (Michigan), Crystal Linkletter (Brown), Michael Mahoney (Stanford), Lauren Meyers (Texas), Martina Morris (U. Washington), Mason Porter (Oxford), Marco Saerens (U. Catholique Louvain), Tom Snijders (Oxford), Alex Vespignani (Indiana), Haiyan Wang (UC Berkeley), Chris Wiggins (Columbia)

Postdoctoral fellows: The four themes outlined above cut across active fields of research in both mathematics and statistics. Further, due to its highly interdisciplinary character, we expect the program to generate great interest among graduating students. Three of the incoming SAMSI postdocs starting in 2009 (Oliver Ratmann, Bruce Rogers and Yi Sun) have already expressed strong interest in participating in the program.

Local researchers: The three partner universities will provide approximately 6 local faculty to participate in the program. Among the local scientists that will potentially be heavily involved are John Aldrich (Duke), David Banks (Duke), H.T. Banks (NCSU), Amarjit Budhiraja (UNC), Thomas Carsey (UNC), Skyler Cranmer (UNC), Pierre Gremaud (NCSU), Lisa Hightow-Weidman (UNC, Infectious Diseases), Alun Lloyd (NCSU), Jan Hannig (UNC), Alan Karr (NISS), Jonathan Mattingly (Duke), James Moody (Duke), Peter Mucha (UNC), Audrey Pettifor (UNC, Epidemiology), Charlie Smith (NCSU), Josh Socolar (Duke), Mike West (Duke).

Graduate students: The three partner universities will provide research assistantships for approximately 6 students to participate in the program. We have already identified several graduate students with potential interests in this area. A number of visiting graduate students are also expected.

2.5 Leveraging

There is great interest in this research area, and we expect that numerous activities will be leveraged with other research organizations. For instance, the Network Dynamics and Simulation Science Laboratory from the Virginia Bioinformatics Institute at Virginia Tech has

already expressed interest in common activities. Other opportunities for leveraged activities will be explored such as for instance possible common activities with the Sante Fe Institute and the Canadian national research network MITACS.

Preliminary contacts have also been established with Thomas Carsey (Political Science, UNC) in order to coordinate some of the program's activities with the annual academic conference of the Society for Political Methodology which may take place in the Triangle in 2011.

2.6 Description of activities

Working groups: Working groups will meet throughout the program to pursue particular research topics identified in the kickoff workshop (or subsequently chosen by the working group participants). We expect to have one or two working groups per thematic area. The working groups consist of SAMSI visitors, postdoctoral fellows, graduate students, and local faculty and scientists.

Course: One graduate course will be taught by Eric Kolaczyk. The course will initially cover the basics of Network Theory following for instance [4]. It is expected that during the second half of the semester these initial concepts will be illustrated by examples drawn from the rich array of applications covered by the field. Specifically, guest lecturers will introduce the students to the various themes of the program. The possibility of having a separate course on social networks is also under consideration.

Workshops: The *Opening Workshop* will be held on 8/29/10-9/1/10. This workshop will aim to engage as large a part of the statistics, mathematics, and relevant scientific communities as possible, with representative sessions from all of the main program topics. The Transition Workshop at the end of the program will disseminate program results and chart a path for future research in the area. Other workshops being considered include a workshop on *Network modeling and inference* as this is the main theoretical theme underpinning the entire program. Each of three remaining themes could lead to one workshop by itself. The possibility of combining themes for common workshops will be studied as the Program evolves.

References

- [1] A.-L. BARABÁSI AND E. BONABEAU, *Scale-free networks*, Scientific American, 288 (2003), pp. 50–59.
- [2] L.P. EULER, *Solutio problematis ad geometriam situs pertinentis*, Commentarii academiae scientiarum Petropolitanae, 8 (1741), pp. 128–140, see <http://math.dartmouth.edu/~euler/docs/originals/E053.pdf>.

- [3] L.R. FORD AND D.R. FULKERSON *Maximal flow through a network*, Canadian J. of Math., 8 (1956), pp. 399–404.
- [4] E.D. KOLACZYK, *Statistical Analysis of Network Data*, Springer, 2009.
- [5] M.E.J. NEWMAN, *The structure and function of complex networks*, SIAM Review, 45 (2003), pp. 167–256.

3. Summer Program on Semiparametric Bayesian Inference: Applications in Pharmacokinetics and Pharmacodynamics July 12-23, 2010

3.1 Introduction

3.1.1 Background

Pharmacokinetics (PK) is the study of the time course of drug concentration resulting from a particular dosing regimen. PK is often studied in conjunction with pharmacodynamics (PD). PD explores what the drug does to the body, i.e., the relationship of drug concentrations and a resulting pharmacological effect. Pharmacogenetics (PGx) studies the genetic variation that determines differing response to drugs. Understanding the PK, PD and PGx of a drug is important for evaluating efficacy and determining how best to use such agents clinically.

Hierarchical models have allowed great progress in statistical inference in many application areas. Hierarchical models for PK and PD data that allow borrowing of strength across a patient population are known as population PK/PD models. These models have allowed investigators to learn about important sources of variation in drug absorption, disposition, metabolism, and excretion, allowing the researchers to begin to tailor drug therapy to individuals. Newer Bayesian non-parametric population models and semi-parametric models offer the promise of individualizing therapy and discovering subgroups among patients even further, by freeing modelers from restrictive assumptions about underlying distributions of key parameters across the population. The purpose of this program is to bring together a mix of experts in PK and PD modeling, non-parametric Bayesian inference, and computation.

Modeling for PK/PD data is also a traditional research problem in applied mathematics, and participants from applied mathematics will be sought.

3.1.2 Program Outcome

The aims of the program and workshop are

- to identify the critical new developments of inference methods for PK and PD data;
- to determine open challenges;
- to establish inference for PK and PD as an important motivating application area of non-parametric Bayes. We believe that this goal is particularly important for new and promising researchers.

3.2 Personnel

3.2.1 Organizers

Gary Rosner and Peter Mueller, Department of Biostatistics, M.D. Anderson Cancer Center, Houston, TX. The leadership will be augmented by a program committee of roughly five area experts to organize the workshop program.

3.2.3 Expected Participants

Below is the list of interested individuals, grouped as PK/PD/PGx, Statistics, and Local. About 10 spaces will be reserved for students and young investigators. We will invite confirmed participants to nominate qualified students.

PK/PD/PGx:

- Aarons, Leon unsure about 2 weeks
University of Manchester,
School of Pharmacy and Pharmaceutical Sciences
Leon.Aarons@manchester.ac.uk
- Bayard, David unsure about 2 weeks
Jet Propulsion Lab, NASA
david.bayard@jpl.nasa.gov
- Conti, David V unsure about 2 weeks
Univ. Southern California
Dept of Preventive Medicine
david.conti@keck.usc.edu
- D'Argenio, David only 1 week
Univ. Southern California
Dept of Biomedical Engineering
dargenio@bmsrs.usc.edu
- Gillespie, William R agreed to 2 weeks
Metrum Institute
billg@metruminstitute.org
- Holford, Nick unsure about 2 weeks
Univ. Auckland, School of Medical Sciences
n.holford@auckland.ac.nz
- Jelliffe, Roger W only 1 week
Univ. Southern California
Laboratory of Applied Pharmacokinetics
jelliffe@usc.edu
- Karlsson, Mats unsure about 2 weeks
Uppsala Univ., Dept. of Pharmacometrics
mats.karlsson@farmbio.uu.se
- Krzyzanski, Wojciech agreed to 2 weeks
State Univ. of NY, Dept. of Pharmaceutical Sciences
wk@buffalo.edu

Lavielle, Marc unsure about 2 weeks
INRIA (Institut National de Recherche en Informatique
et Automatique) Saclay &
Universite Paris-Sud, Dept of Mathematics
Marc.Lavielle@math.u-psud.fr

Li, Lang agreed to 2 weeks
Indiana Univ School of Medicine,
Dept. of Medicine, Division of Biostatistics
lali@iupui.edu

Mentre', France No commitment
INSERM - Univ. Paris Diderot
france.mentre@inserm.fr

Neely, Michael unsure about 2 weeks
USC
mneely@usc.edu

Ottesen, Johnny unsure
Roskilde University, Denmark
Dept of Mathematics and Physics

Rosner, Gary agreed to 2 weeks
Univ Texas M.D. Anderson Cancer Center,
Dept of Biostatistics
glrosner@mdanderson.org

Schumitzky, Alan agreed to 2 weeks
Univ. of Southern California,
Dept of Mathematics
schumitzky@aol.com, schum@usc.edu

Thomas, Duncan C agreed to 2 weeks
Univ. of Southern California,
Dept of Preventive Medicine
dthomas@usc.edu

Vinks, Alexander A agreed to 2 weeks
Cincinnati Children's Hospital,
Div. of Clinical Pharmacology
sander.vinks@cchmc.org

STATISTICS:

Basu, Sanjib agreed for 2 weeks
U. Northern Illinois, Statistics
basu@niu.edu

Dahl, David agreed for 2 weeks
Texas A&M Univ. - Dept. of Statistics
dahl@stat.tamu.edu

Escobar, Michael tentative
U. Toronto, Public Health Sciences & Statistics
m.escobar@utoronto.ca

no definite commitment yet, but please invite
Hanson, Tim agreed for some time
Univ. Minnesota - Division of Biostatistics No
hanson@biostat.umn.edu.

Leanna House agreed for 2nd week only
Virginia Tech, VA
lhouse@vt.edu

Lee, Ju Hee 2 weeks
OSU
juheelee@stat.osu.edu

Ickstadt, Katja 2 weeks
U. Dortmund
ickstadt@statistik.tu-dortmund.de

de Iorio, Maria first 10 days
Imperial College
m.deiorio@imperial.ac.uk
can pay her air ticket, but might need local support

Jain, Sonia week 1
Univ. California San Diego - Moores Cancer Center
sojain@ucsd.edu

Johnson, Wes agreed to 1st week
Univ. California Irvine - Dept. of Statistics
wjohnson@ics.uci.edu

MacEachern, Stephen agreed to 2 weeks
Ohio State Univ. - Dept. of Statistics
snm@stat.osu.edu

Mueller, Peter agreed to 2 weeks
Univ Texas M.D. Anderson Cancer Center
pm@wotan.mdacc.tmc.edu

Bhramar, Mukherjee tentative
U. Michigan Biostatistics
bhramar@umich.edu
needs to work out child care

Petrone, Sonia tentative
Bocconi U, Milano
sonia.petrone@uni-bocconi.it

Sivaganesan, Siva agreed to 2 weeks
Univ. Cincinnati - Dept. of Statistics
sivagas@UCMAIL.UC.EDU Yes

Kottas, Thanasi, agreed to 2 weeks
UCSC
thanos@ams.ucsc.edu

Tadesse, Mahlet tentative for 2 weeks
Georgetown U.
mgt26@georgetown.edu
subject to her getting summer funding

might need some travel funding for the program
Thall, Peter agreed to 2 weeks
MDACC
rex@mdanderson.org
Guindani, Michele agreed to 2 weeks
U NM, Department of Mathematics
michele@stat.unm.edu
Vannucci, Marina tentative
Rice U.
marina@stat.rice.edu
Xu, Xinyi agreed to 2 weeks
OSU
xinyi@stat.osu.edu

LOCAL:

Davidian, Marie only 1 week
North Carolina State Univ - Dep't of Statistics
davidian@stat.ncsu.edu
Leary, Bob agreed to 2 weeks
Pharsight
bleary@pharsight.com
Ghoshal, Subhashis agreed for 2 weeks
NCSU, Statistics
ghoshal@stat.ncsu.edu
Ibrahim, Joe agreed for week 1
Joe is away after 7/17
UNC-CH, School of Public Health
ibrahim@bios.unc.edu
Fox, Emily agreed to 2 weeks
Duke U.
fox@stat.duke.edu
Dunson, David agreed for 2 weeks
Duke DSS
dunson@stat.duke.edu
suggests two students/postdocs:
Anirban Bhattacharya (sp Bayes); and Hongxia Yang

We also expect many other local faculty interested in nonparametric Bayesian methods to participate, as well as individuals from the local pharmaceutical industry, EPA, and NIEHS, many of whom use physiologically based PK/PD models (PB-PK) to extrapolate from animal to human data in toxicology studies and other areas.

Applied Mathematics: Although several of the listed PKPD people are applied math, we would very much appreciate suggestions for additional mathematicians who might be interested in participation in the program.

3.3 Program Structure

The program will begin with a week of tutorials and workshop activities. The focus will be on building to a second week of research working groups, that will tackle particular research problems in the area. Specific potential themes for these working groups and workshop sessions are

- semi-parametric population PK/PD models
- dose individualization
- probability models for PK/PGx networks
- joint inference for PK and PD data.

The final day of the program will be a reprise by the working groups of progress, and plans for completion of the research.

Appendix A: Final Report of the Program on Risk Analysis, Extreme Events and Decision Theory

1 Introduction

Over the past several years, there has been a wealth of scientific progress on risk analysis. As the set of underlying problems has become increasingly diverse, drawing from areas ranging from national defense and homeland security to genetically modified organisms to animal disease epidemics and public health to critical infrastructure, much research has become narrowly focused on a single area. It has also become clear, however, that the need is urgent and compelling for research on risk analysis, extreme events (such as major hurricanes) and decision theory in a broader context. Availability of past information, expert opinion, complex system models, and financial or other cost implications as well as the space of possible decisions may be used to characterize the risks in different settings. Integration of expertise developed by researchers in different scientific communities on each of these facets is the objective of this SAMSI program. Risk analysis and extreme events also carry a significant public policy component, which is driven in part by the increasing stakes and the multiplicity of stakeholders. In particular, policy concerns direct attention not only to the dramatic risks for huge numbers of people associated, for example, with events of the magnitude of Hurricane Katrina or bioterrorism, but also to “small-scale” risks such as drug interactions driven by rare combinations of genetic factors.

From the Opening Workshop in the publication by Wiley of *Bayesian Analysis for Stochastic Process Models* (anticipated in August 2009) and also a volume (as yet untitled) for the ASA-SIAM book series, the SAMSI program on Risk has encompassed four principal workshop, three dedicated sessions at JSM 2008 and an invited session at ISI 2009 as well as many individual presentations at professional meetings here and abroad. In all, 11 SAMSI postdoctoral fellows and postdoctoral associates, 12 graduate students (5 from outside the local universities), 22 new researchers and 44 other visitors to SAMSI have been engaged in 7 Working Groups. Total participation in Working Groups (local and remote) has been 89 and the total participation in the program through one or more activities is 167. As of April 2009, research articles submitted for publication total 41 (see bibliographic listing).

Highlights following the program year include an invited session at the International Statistical Institute 2009 meeting in Durban, South Africa and several grants awarded for continuing collaboration among SAMSI program researchers with still others pending as of April 2009. Major grants include a Fulbright award for risk analysis modeling in information and communication technologies (Dipak Dey, University of Connecticut and Javier Cano) and two awarded grants from the Spanish government plus pending applications to support continuing international collaborations involving Risk program researchers in Spain (Universidad Rey Juan Carlos) and in the the US (variously Cornell University, University of Connecticut, Duke University and IBM).

1.1 Research Foci

The aim of this full-year program was to address fundamental issues in risk analysis and the linked problems associated with extreme events and decision theory. By engaging researchers from the statistical sciences, applied mathematical sciences including actuarial science, and the decision sciences, including operations research, the goal was to set research agendas that can produce genuine impact on the practice of risk analysis and assessment as well as on theory and methodology for extreme events and decision theory. Interdisciplinary working groups were formed around both kinds of events and critical research tasks in theory and methodology, following the already identified interests and the existing momentum. Critical research tasks for this program included theoretical development of extreme value theory, implementations of methodologies that integrate expert opinion with data and with models, risk assessment and prediction with applications to high-impact events.

2 Program Organization

The program leaders were Dipak Dey (Univ. of Connecticut), David Ríos Insua (Universidad Rey Juan Carlos), Richard L. Smith (Univ. of North Carolina, Chapel Hill) and Nell Sedransk (SAMSI Associate Director).

The following Scientific Committee provided advice as needed on specific components: David Banks (Duke), Vickie Bier (Univ. of Wisconsin), James Broffitt (Univ. of Iowa), Alicia Carriquiry (Iowa State), Robert Clemen (Duke), Susan Ellenberg (Univ. of Pennsylvania), Herbert Hethcote (Univ. of Iowa), Wolfgang Kliemann (Iowa State), Robert Winkler (Duke), Stan Young (NISS).

3 Workshops

The workshops organized in connection with this program were:

1. Opening Workshop, September 16–19, 2007. Held at the Radisson RTP.
2. RISK: Perception, Policy and Practice, October 3–4, 2007. Held at the Radisson RTP.
3. EXTREMES: Events, Models and Mathematical Theory, January 22-24, 2008 Held at the Radisson RTP.
4. RISK Revisited: Progress and Challenges, May 21, 2008. Held at the Marriott Durham in association with the 2008 Interface.

In addition to workshops organized as part of the SAMSI program, three sessions at the Joint Statistical Meetings in August 2008 were organized around the research accomplished during this SAMSI program: “Risk Analysis for Industry and the Environment,” “Bayesian Modeling of Extreme Events” and “SAMSI Program on Data Analysis, Extreme Events and Decision Theory.” The full programs for these workshops have been documented elsewhere.

4 Research Goals and Activities

4.1 Adversarial Risk Analysis Group

4.1.1 Summary

Game theory has long been considered of little relevance for practical risk management decision-making. This viewpoint has recently become less dogmatic because:

- High-profile terrorist attacks have demanded significant national investment in protective responses, and there is public concern that not all of these investments are prudent and/or effective.
- Key business sectors (especially finance, e-commerce, and software) have become much more mathematically sophisticated, and are now using this expertise to shape corporate strategy for auction bidding, timing of product release, lobbying efforts, and other decisions.
- Regulatory legislation must balance competing interests (for growth, environmental impact, safety) in a way that is credible and transparent.
- The on-going arms race in cybersecurity means that the financial penalties for myopic protection are large and random.

These challenges cross many fields (Statistics, Economics, Operations Research, Engineering, etc.) and are characterized by the fact that there are two or more intelligent opponents who make decisions for which the outcome is uncertain. Collectively, we call this problem area Adversarial Risk Analysis (ARA) and represent a combination of statistical risk analysis and classical game theory.

Traditional statistical risk analysis grew in the context of nuclear reactor safety, insurance, and other applications in which loss was governed by chance rather than the malicious (or self-interested) actions of intelligent actors. But in ARA, one needs to have some model for the decision-making of all the participants. This model might be classically game-theoretic, with (non-cooperative) Nash equilibria as core concept or it might be more psychological, reflecting either a Bayesian formulation or empirical studies of game behavior.

4.1.2 Research Foci

The group addressed both fundamental and applied issues within this new field of adversarial risk analysis. At a fundamental level, the primary objective was to provide a unified approach and new solution concepts, ways to model the beliefs of the adversaries, algorithms to compute the new solutions, together with integration with negotiation analysis methodologies. At an applied level, research focused on the fields of auctions, antiterrorism modeling and cybersecurity.

4.1.3 Main Participants

David Banks, Duke University
Betsy Enstrom, Duke University
Jesus Ríos, Concordia University
David Ríos Insua, Universidad Rey Juan Carlos
Lea Deleris, IBM
Mike Porter, NCSU
Matt Heaton, Duke University
Justin Shows, NCSU
Huiyan Sang, Duke University
Nabendu Pal, Louisiana State University
Javier Cano, Universidad Rey Juan Carlos
Jose Antonio Rubio, Universidad Rey Juan Carlos

4.1.4 Activities

Meetings: The group met regularly on Thursday from 11.30 till 13:00 to discuss research progress, and propose new topics.

4.1.5 Research output

The research output of this group is summarized here under a number of headings.

1. Foundations of adversarial risk analysis

Topics in foundations of adversarial risk analysis covered both theoretical and computational approaches. Work generated the following papers.

- Title: Adversarial risk analysis. (ARA) Authors: D. Ríos Insua, J. Ríos, D. Banks. In this paper, we describe several formulations of adversarial risk problems, providing a unified framework for analysis. We also discuss the research challenges that arise when dealing with these models, illustrate the ideas with examples from sealed auctions, and point out relevance to national defense. The key contribution is a way to build a rational probabilistic model of the actions of the adversary, which is then used to feed a decision analytic model.
- Balanced increment and concession methods for arbitration and negotiation support, (BIM-BIC) J. Ríos, D. Ríos Insua. In this paper, we study arbitration schemes and develop negotiation support methods from the perspective of cooperative bargaining theory. We discuss Raiffa's solution of balanced increments and, based on that idea, propose another solution based on balanced concessions. We also consider negotiation support processes based on the application of these solution concepts. The most notable feature of the proposed schemes is that they allow the consideration of non-convex utility sets for problems with more than two agents, a topic not sufficiently considered in the bargaining theory literature. A risk sharing negotiation problem illustrates the discussion.

- Commutativity of Nash equilibria and expected utilities, J. Ros, David Banks.
In our discussions, we observed that expected utility and Nash equilibria operators do not commute; this creates conceptual difficulties to simulation based approaches in this area. We are identifying appropriate ways to integrate both operations; and both structured and numerical experiments provide examples.

2. Other Modeling

- Discrete choice models in adversarial risk analysis. Mike Porter.
In this paper an alternative model for rationally choosing a probabilistic model of the actions of the adversary has been proposed based on discrete choice models. Since the assumptions introduced prevent an analytic solution, results are obtained via simulation based approaches.

3. Computations in adversarial risk analysis

- Negotiations over influence diagrams, J. Ríos, D: Ríos Insua.
We discuss issues concerning negotiations over influence diagrams. We base our discussion on a modification of the balanced increment method. As in standard decision analysis texts, we deal first with negotiation tables, then with negotiation trees and, finally, with (negotiation) influence diagrams. We show by example that a naive application of the balanced increment method may lead to an inferior solution. Our strategy proposes therefore computing first the nondominated alternatives and then negotiating over such set.
- Computations for adversarial risk analysis.
As basic modelling and communication tools we are using influence diagrams. Here we extend these to a new class of adversarial IDs; the solution concepts appear in the ARA and BIM-BIC papers, and are implemented here using MCMC and other simulation methods.

4. Auctions

- Adversarial risk analysis. D. Ríos Insua, J. Ríos, D. Banks.
The key application in the ARA paper is auctions; and the results there derive from this research.
- Bayesian methods for auction participation support.
We believe we have been very successful in proposing a novel Bayesian approach to first price sealed bid auctions leading to, on one hand implementations for a realistic case and on the other, to extensions to other types of auctions.

5. Terrorism

- Adversarial risk analysis for terrorism prevention.
Having already applied our ARA approach to the so called Defend-Attack, Attack-Defend and Defend-Attack-Defend models, we then extended it to more general problems, modeled as adversarial IDs. TO be successful this required application of computations also developed as part of this adversarial risk analysis project. We would also like to sketch solutions with continuous time asynchronous conflicting interactions, possibly with stochastic adversarial differential equations.

6. Cybersecurity

- An adversarial risk analysis framework for cybersecurity.
This line of research was proposed by Lea Deleris with a qualitative description of the issues involved. The key issue here is that n (members of an interconnected network) versus m (cyberattackers), with possible cooperation among both sides. We extended original ARA model, 1 vs 1, to 1 vs m and then used the ideas in the BIC-BIM paper to consider cooperation in n vs m .
- Formalisation of risk approaches in ICT. D. Ríos Insua, J.A. Rubio.
This actually started from a class discussion at the SAMSI course. In it, we concluded that most approaches to ICT Risk analysis are not well founded and we are trying to formalize one of the most successful approaches. This requires the development of some novel reliability modeling approaches as described in the next two papers.
- Bayesian reliability analysis for hardware/software systems, J. Cano, D. Ríos Insua
We provide a class of models to evaluate and forecast the reliability of complex hardware/software systems, described through Reliability Block Diagrams. Blocks referring to hardware components are modelled through 'pending' continuous time Markov chain models, whereas blocks referring to software components are modelled through a mixture of software reliability growth models. Inference and forecasting tasks with such models are described, and illustrated with an example.
- Bayesian reliability, repairability and availability for hardware systems through continuous Markov chain models, J. Cano, D. Ríos Insua
Hardware systems are present in many fields of human activity. Markov models are sometimes used in hardware reliability, availability and maintainability (RAM) modeling. They are specially useful in situations in which the system we want to analyze may be modeled with several states through which the system evolves, some of them corresponding to ON states, the rest to OFF states. We provide here RAM analyses of such systems within a Bayesian framework. But the computations are too involved and we are devising new computational strategies as in
- Reduced order models for Bayesian risk analysis, M. Grigoriu, D. Ríos Insua, J. Ríos, H. Shen
Standard approaches to risk analysis based on estimating parameters and performing the corresponding risk analysis computations will typically underestimate uncertainty. An alternative Bayesian approach computes posterior distributions

for the parameters and then performs a posterior predictive risk analysis computation. This may be extremely involved computationally requiring some type of approximation. Reduced order models have been recently proposed to approximate given distributions and then perform predictive computations. In this paper we explore the relevance of reduced order models for Bayesian predictive computations, especially in a Bayesian risk analysis context. We consider a simple application in queueing models and a complex application in continuous time Markov chain based reliability models. General conclusions are drawn suggesting the effectiveness of this methodology.

NOTE: This work was done in collaboration with the Service Risk group as part of a broader effort to expand Bayesian discrete event simulation.

4.1.6 Horizontal topics

1. Basic concepts in stochastic processes
2. Basic concepts in Bayesian Analysis
3. Discrete time Markov chains and extensions
4. Continuous time Markov chains and extensions
5. Poisson processes and extensions
6. Continuous time processes
7. Queueing analysis
8. Reliability and maintenance
9. Discrete event simulation
10. Risk Analysis

4.1.7 Other Activities

Research has been invited from multiple working group members for presentation at meetings of national and international professional societies.

- Interface meeting 2008
- Joint Statistical Meetings 2008
- Group Decision and Negotiation 2008
- INFORMS 2008
- Probabilistic Graphical Models 2008
- International Statistical Institute 2009

4.2 Bayes Risk Group

4.2.1 Organization and Membership

Kobi Abayomi, Duke University
David Banks, Duke University

Susie Bayarri, University of Valencia/SAMSI
Jim Berger, SAMSI
Sourish Das, Univ. of Connecticut
Dipak Dey , Univ. of Connecticut
Ian Dinwoodie, Duke University
Betsy Enstrom, Duke University
Elijah Gaioni, Univ. of Connecticut
Mircea Grigoriu, Cornell University
Feng Guo, Virginia Tech
James Hammitt, Harvard University
Huitian Lu, South Dakota State University
Christian Macaro
Vered Madar, SAMSI
Cuirong Ren, South Dakota State University
Abel Rodriguez, Duke University
Fabrizio Ruggeri, CNR-IMATI
Richard Smith, UNC-Chapel Hill
Gentry White, N.C. State University
Dabao Zhang, Purdue University
Iris(Xiaoyan) Lin, University of Missouri-Columbia

4.2.2 Description of Activities

Workshops: The Opening Workshop was held on September 16, 2007 - September 19, 2007. Its principal goal was to engage a broadly representative segment of the statistical, applied mathematical and decision analysis/operations research communities in formulation and pursuit of specific research activities to be undertaken by the Program Working Groups, discussed above. Mid-program workshops focused on specific topics, the first of these took place in October: Risk: Perception, Policy and Practice. A workshop on Extreme Events: Theory, Prediction and Cost was held in late January. Other workshops were organized by the working groups; and a Transition Workshop, at the end of the program, disseminated program results to chart a path for future research in the area.

Courses: Team-taught courses were taught at the NISS/SAMSI building during the fall semesters. The fall semester course began with an introduction to decision theory as a foundation for risk assessment and management; it was continued with a systematic approach to risk analysis, and then concluded with an introduction to expert opinion elicitation and modeling.

Working Groups: The working groups met regularly throughout the program to pursue particular research topics identified during the Opening Workshop and during the January workshop. Each working group consisted of SAMSI visitors, postdoctoral fellows, graduate students and local faculty and scientists. In addition the working group meetings were continued remotely from University of Connecticut.

Presentations: The following presentations were made at the working group meetings;

Sep 27 First planning meeting

Oct 11 Huiyan Sang presented her work on “Hierarchical Modeling for Extreme Values Observed over Space and Time”

Oct 15 Mircea Grigoriu: Large Scale Stochastic Equations - A special lecture
 Oct 18 Discussion on river flow data (Elijah Gaioni); discussion and some modeling issues on Hurricane data (Sourish Das); some thoughts on Bayesian modeling of Multivariate extremes (Dipak Dey)
 Oct 25 Short introduction to dynamic linear models (Gentry White); Large Scale Stochastic Equations - Bayesian Framework (Mircea Grigoriu).
 Nov 8 Kobi Abayomi: Fitting multivariate extreme value dist to multi-'hazard' environmental data.
 Nov 15 Sourish Das: Some modeling issues on Hurricane data.
 Nov 29 Elijah Gaioni: Modeling River Flow: Flash Floods and Mixture Distributions.
 Dec 6 Vered Madar: Some Thoughts on Bayesian Modeling of Bivariate Extremes.
 Feb 4 Elijah Gaioni: Semiparametric functional estimation using quantile based prior elicitation
 Feb 11 Jose Bernardo, University of Valencia
 Mar 3 Fabrizio Ruggeri: Model-based prior elicitation: a possible approach?
 Mar 24 Susie Bayarri, University of Valencia/SAMSI
 Mar 31 Sourish Das

4.2.3 Research Outcomes

Expert opinion. Data inadequacies were perhaps the most clearly identified theme for modeling extreme data. For some rare events whose risk must be assessed, there are no data; more often there are data of mixed degrees of relevance and reliance on experts' opinions is needed to avoid rigid specifications of parameters and/or functional forms within risk models that cannot be documented. Various Bayesian methodological techniques were implemented using prior elicitation and models incorporating expert opinion to produce accurate estimates of parameters of interest. Examples include modeling hurricane intensity and floods.

4.3 Extreme Values

Extreme Value Theory has its origins in papers by Fréchet (1927), Fisher and Tippett (1928) and Gnedenko (1943), who established the existence of special families of extreme value distributions, defined as limiting distributions of maxima and minima in independent, identically distributed sequences of random variables. The theory immediately found applications in practical risk assessment, for example through the work of Gumbel on hydrological extremes or Weibull on strength of materials. During the last thirty years, the scope of both the theory and applications have greatly expanded. The earlier statistical methods that were based on directly fitting the extreme value distributions to data have for many applications been replaced by methods based on threshold exceedances, which have in turn focused attention on new families of distributions (in particular, the generalized Pareto distribution). There is an ever expanding theory of extremes in stochastic processes, which has found particular application in the field of finance. Statistical methods for extremes have become increasingly elaborate, for example using second-order approximations for threshold selection or bias reduction, using robustness concepts, and (especially) a rapidly increasing interest in the use of

Bayesian methods. Applications have ranged over many areas, including finance and insurance, meteorology, hydrology and oceanography.

A particularly significant development of the last thirty years has been the development of a whole field of multivariate extreme value theory. The original papers were concerned with extending the classification of extreme value distributions to cover joint distributions of maxima in dependent processes – for example, a landmark paper of de Haan and Resnick (1977) established domain of attraction conditions for multivariate extremes and the connection with multivariate regular variation. The earliest papers on statistical inference for multivariate extremes started at around the same time, but this research greatly expanded in the late 80s and early 90s. During the last 15 years, two new formulations of multivariate extremes have been proposed. The first originated in papers of Ledford and Tawn (1996, 1997), and was concerned with dependence measures for bivariate extremes that are more sensitive to different kinds of asymptotic behavior than the traditional bivariate extreme value distributions. For instance, bivariate normal variables with correlation in $(0,1)$ are asymptotically independent under the traditional formulation, but the Ledford-Tawn approach captures the hidden dependence that still exists at very high threshold levels. However, this approach has so far been limited to the case of bivariate extremes. A second approach due to Heffernan and Tawn (2004) was based on classes of conditioned limit theorems as one component (but typically not all components) become extreme. However at the moment, this approach is still too new and too poorly understood for its full implications to be appreciated.

The SAMSI program on Risk Analysis, Extreme Events and Decision Theory allowed many of these issues to be analyzed in depth. Talks at the Opening Workshop ranged across the spectrum from theory to applications, from such topics as the role of multivariate regular variation in determining theoretical properties of GARCH and stochastic volatility process in finance, through to a very applied discussion of the role of extreme events in the current mortgage crisis. At the end of this workshop, it was agreed to make multivariate extreme value theory the primary focus of two working groups, one oriented towards new methodological developments and the underlying mathematical theory, the other focused on applications. These topics were further cemented at the January workshop entitled *Extremes: Events, Models and Mathematical Theory*. At this workshop, talks were given by a number of the world's top experts in extreme value theory and its applications.

4.3.1 Theoretical Developments

1. **Classical Univariate EVT.** Two talks at the January workshop highlighted recent developments in the estimation of the tail-index parameter, that determines the rate of growth of extremes. **Chen Zhou** discussed second-order tail conditions and their implications for asymptotic properties of estimators of the tail index, including non-regular cases where classical maximum likelihood theory breaks down. In contrast, **Debbie Dupuis** focussed on robustness, highlighting a “weighted prediction error” criterion for reconstructing the upper tail of a distribution.

From a different perspective, **John Nolan** gave a talk about estimation of stable distributions, which in some contexts are an alternative to fitting an extreme value distribution to long-tailed data.

2. **Extremes in Stochastic Processes.** Three talks at the January workshop discussed particular topics in extreme value theory for (univariate) stochastic processes. **Vicky Fasen** discussed the extreme value theory of “threshold autoregressive” processes, which are a widely used class of nonlinear time series models that have recently found application in the field of financial time series. **Ross Leadbetter** gave a talk about the “capsize risk” problem for ships, used to illustrate the general principle that a naïve approach to extreme value theory may be inadequate for characterizing upcrossings and other significant properties of random processes. **Gennady Samorodnitsky** started his talk with the observation that regular variation of the upper tail of a distribution is preserved under linear filters, and discussed the *inverse problem* of determining when regular variation of the output of a linear filter implies regular variation of the input.

The theory of extremes in *random fields* is still much less well developed than that of one-parameter stochastic processes, but in the special case of continuous Gaussian processes, a rich theory now exists. **Yimin Xiao** gave an excellent overview of this topic in one of the working group meetings.

3. **Multivariate Regular Variation.** **Richard Davis’s** talk at the Opening Workshop covered several aspects of extreme value theory as applied to commonly used models for financial time series, such as the popular GARCH(1,1) model, and as an alternative, a stochastic volatility (SV) model. His starting point was the question “Do fitted models actually capture the desired characteristics of the real data?” He then presented a number of real financial time series, focussing on clustering properties of the extreme values and on the behavior of the sample autocorrelation functions (ACFs) of the log returns and their squares and absolute values. He then surveyed recent developments in multivariate extreme value theory, focussing on the property that regular variation is preserved when forming linear combinations of the data, and discussing a result of Basrak, Davis and Mikosch on conditions for the converse statement to be true. He then discussed applications of this result to GARCH and SV processes, including the theoretical properties of sample ACFs and clustering properties of extremes — for example, a GARCH process typically has extremal index in $(0,1)$, which implies clustering of extreme values, whereas a SV process has extremal index 1, which implies no local clustering. He then returned to some of the real-data time series, discussing how their empirical properties match up with theoretical properties of GARCH and SV processes. Although he did not commit himself to a firm statement about which model fits better, the implication was that in many cases these considerations favor the SV model.

The theme of multivariate regular variation was continued in **Thomas Mikosch’s** talk at the January workshop. In a wide-ranging talk he also discussed the preservation of regular variation under formation of linear combinations, and generalizations of the result to processes in $\mathbb{D}[0,1]$, including results for the Ornstein-Uhlenbeck and Lévy processes. He went on to discuss max-stable and stable random fields, models for spatial and spatio-temporal processes, and large deviations theory for stochastic processes. The final part of the talk covered ruin processes and their multivariate generalization.

4. **Classical Multivariate Extreme Value Theory.** In the Opening Workshop, **Holger Rootzén** discussed the bivariate generalized Pareto distribution, a recent development

in threshold-exceedance methods for bivariate extremes. He gave an example based on insurance claims for windstorm damage to buildings and forest, comparing an analysis in which the two types of claims are considered separately with one in which they are treated as a bivariate pair. He concluded “bivariate analysis may give the most correct evaluation of the real uncertainties”.

Another development presented by Holger was the use of stable laws as a mixture distribution to generate new classes of multivariate extreme value distributions. This was based on joint work with **John Nolan** and Anne-Laure Fougères. As an application, he discussed a problem about pitting corrosion, where the object of interest is maximum pit depth, where the possible presence of common environmental factors means that depth of different pits are not necessarily independent. To solve this problem, he proposed a flexible class of “logistic” models with Gumbel marginal distributions, where the distribution of maxima of all kinds of sets are also Gumbel.

The question of *bivariate measures of extremal dependence* was discussed by **Ishay Weissman** in the January workshop. In this talk he discussed two measures of dependence that have been proposed in previous literature, denoted τ_1 and τ_2 , and presented a number of new identities and bounds.

The field of *multivariate stable distributions* was also discussed by Nolan in one of the working group meetings. Although this leads to different distributions from the traditional multivariate EVT distributions, for many practical applications they may be a suitable alternative.

5. **Max-Stable Processes.** Max-stable processes are the generalization of multivariate extreme value theory to infinite dimensions. In a working group presentation following the January workshop, the originator of the whole concept, **Laurens de Haan**, surveyed the current state of the theory as it appears in his 2006 book with Ana Ferreira. Theoretical developments were presented by **Zhengjun Zhang** and **Stilian Stoev** in the opening workshop. Applications to spatial statistics were presented by **Tailen Hsing** in a discussion, and **Dan Cooley** gave a talk on prediction theory for max-stable processes, in effect the analog of kriging in traditional spatial processes.
6. **Alternative Models for Multivariate Extremes.** As noted in the introduction, two alternative formulations of multivariate extremes have been proposed during the past decade, one initially developed in two papers by Ledford and Tawn (1996, 1997), the other stemming from Heffernan and Tawn (2004). These papers and some recent extensions formed the topic of **Sid Resnick’s** talk at the January workshop, and several working group discussions.

Anthony Ledford gave one working group presentation remotely from Oxford, based on recent work by him and **Alexandra Ramos**. This work contains a reformulation of the original Ledford-Tawn work, with more clearly defined statistical properties, and the potential for extensions to multivariate cases, most of the existing theory being for bivariate models. Other working group talks by **Xiao Qin** and **Richard Smith**, and an opening workshop presentation by **Jonathan Hill**, discussed other aspects of the theory of these models and their relation with classical bivariate extreme value theory.

The more recent model of Heffernan and Tawn leads to a class of *conditioned limit theorems*, in which one component becomes extreme but the objective is to establish conditional limit theorems for the other component(s). This theory is still too recent to have been subjected to many practical tests, but two talks by **Luis Pericchi** presented joint work with Beatriz Mendes that discussed an application to flooding in Puerto Rico.

In his talk at the January workshop and a series of subsequent presentations to the working group, **Sid Resnick** discussed the mathematical relationship among classical bivariate extreme value theory, the Ledford-Tawn approach, and conditioned limit theorems. We have already noted that classical bivariate (or multivariate) extreme value may be characterized in terms of multivariate regular variation. In a series of papers over the last 6 years, Sid and co-authors have shown that the key mathematical condition for the Ledford-Tawn approach is *hidden regular variation*, which is equivalent to regular variation on a cone. In recent papers with Janet Heffernan and Bikramjit Das, Sid has extended this work to cover also the case of conditioned limit theorems. Some key statistical questions remain, however. For example, a key step in all of these limit theorems is standardization of the marginal distribution to unit Fréchet. The traditional approach is through a semiparametric estimator of the index of regular variation (Hill's estimator is the best known of many proposals for this), but Heffernan and Resnick (2005) preferred a nonparametric "rank transform" approach. It remains an open question which of the two is better.

These issues were the centerpiece of an invited session on multivariate extremes at the May 2008 Interface, when **Janet Heffernan** was one of the invited speakers.

4.3.2 Applications of Extreme Value Theory

Some applications have been interwoven into the above theoretical discussion, for example, **Holger Rootzén's** work on pitting corrosion, and **Luis Pericchi's** application of the Heffernan-Tawn method to floods. However, a number of applications received extensive examination in their own right during the course of the program.

1. **Finance.** In recent years, many of the liveliest applications of extreme value theory have been in the area of finance, and the SAMSI program reflected that. As already noted, **Richard Davis's** talk at the opening workshop was motivated by the problem of distinguishing between GARCH and stochastic volatility models for financial time series. Other workshop presentations touching on financial extremes included **Yacov Haimmes's** talk on the Partitioned Multiobjective Risk Method (PMRM) to portfolio selection; **Zhengjun Zhang's** talk on testing and modeling extreme dependence in the financial markets; and **Bas Werker's** talk at the January workshop, on integer-value time series models for financial data. In the January workshop, **Dominik Lambrigger** discussed new measures of *Value at Risk*, focussing on subadditivity and superadditivity properties.
2. **Insurance.** In the Opening Workshop, **Dougal Goodman** presented a broad-ranging review of how risk analysts approach extreme events, from the contrasting points of view of government, industry and regulators. On a much more technical level, **Shyamal**

Kumar talked about phase-type distributions in actuarial science, and their application to ruin theory and related problems.

3. **Energy Pricing.** In the January workshop, **Pilar Muñoz** discussed volatility modeling and risk assessment in electricity markets. Her main approach was a stochastic volatility model for prices, using a state space approach, combined with extreme value theory to model the probability of extreme jumps, conditional on the volatility process. Fitting this model was tried using a particle filter algorithm, and also a modification of the sampling-importance-resampling method that she called SIRJ. The possibility of using multivariate extreme value theory to improve the analysis was posed as an open question.
4. **Meteorology and Hydrology.** Some of the oldest applications of extreme value theory have concerned assessing probabilities of extreme floods or extreme meteorological events, so it was not surprising that these themes emerged several times during the SAMSI program. In a provocative talk at the January workshop, **Jery Stedinger** touched on several key points of the application of extreme value theory to hydrological extremes, including the relationship among maximum likelihood, L-moment and Bayesian approaches to the estimation of extreme value parameters; the relationship between the threshold exceedance approach and older methods based on annual maxima; and a new approach to the regionalization problem (combining data from multiple stations in a region to improve the estimation of extreme value parameters) using a new Bayesian GLS approach. Applications to meteorology included **Laurens de Haan's** presentation at the January workshop about spatial modeling of precipitation extremes in the Netherlands; **Huiyan Sang's** presentation to one of the working group meetings about a spatial hierarchical model for precipitation extremes; and **Elizabeth Shamseldin's** poster presentation on the change of scale problem for precipitation extremes. In the January workshop, **Francis Zwiers** presented a broad overview of how extremes are viewed by climate scientists, focussing on the very wide range of spatial and temporal scales that must be considered; the use of "simple" indices of extremes and some of the pitfalls that can occur with them; the difficulty of reconciling observations and climate models; and finally, the growing problem of "operational attribution", which refers to the extent to which extreme events can be attributed to external forcing factors, in particular greenhouse gases versus natural causes such as solar fluctuations
5. **Volcanoes.** In a presentation at the January workshop, **Elaine Spiller** discussed the work of a large group of SAMSI researchers on pyrostatic flows. The work combined an elaborate differential equation model for flows, the GASP technique for statistically interpolating parameters of the flow model, and extreme value theory to extend the model to encompass the possibility of extremely large eruptions.
6. **Hurricanes.** Although this does not involve extreme value theory as usually defined, several discussions during the program included statistical modeling of hurricane or tropical storm count data. At the opening workshop, **Tom Knutson** reviewed the difficulties of inferring a trend from long-term time series of tropical storms and hurricanes, and also presented some of the conflicting evidence on whether climate models predict an increase in the frequency of hurricanes as greenhouse gases continue to rise. This

gave rise to several statistical projects. Sourish Das’s work is discussed in more detail in the Bayes Risk section of this report. **Yongku Kim** has been working on determining the optimal relationship between hurricane and tropical storm counts and the spatial distribution of sea-surface temperatures (SSTs). Since hurricane counts are discrete, there is really a need for discrete-data time series models, and **Vangelis Evangelou** suggested an approach to this based on models for Poisson time series in recent papers of Davis, Dunsmuir and Streett. He and **Richard Smith** are working on bivariate time series models for the joint evolution of storm counts and SSTs.

4.3.3 Working Group on Multivariate Extremes — Methodology

Participants:

Susie Bayarri, University of Valencia and SAMSI

Jaya Bishwal, UNC-Charlotte

Michela Cameletti, SAMSI

Guang Cheng, SAMSI and Duke University

Dan Cooley, Colorado State University

Sourish Das, University of Connecticut

Dipak Dey, University of Connecticut

Ian Dinwoodie, Duke University

Evangelos Evangelou, UNC-Chapel Hill

Elijah Gaioni, University of Connecticut

Eric Gilleland, NCAR

Dougal Goodman, The Foundation for Science and Technology (UK)

Laurens de Haan, Erasmus University Rotterdam (Netherlands) and University of Lisbon (Portugal)

Jonathan Hosking, IBM

Rosalba Ignaccolo, SAMSI

Huijing Jiang, Georgia Institute of Technology

Myron Katzoff, Centers for Disease Control

Yongku Kim, SAMSI

Lada Kyj, Rice University

Anthony Ledford, Man Investments (UK)

Huitian Lu, South Dakota State University

Wenbin Lu, N.C. State University

Vered Madar, SAMSI

Pilar Munoz, Technical University of Catalonia

XuanLong Nguyen, SAMSI

John Nolan, American University

Jayanta Pal, DUKE Univ. and SAMSI

Luis Pericchi, University of Puerto Rico, Rio Piedras

Xiao Qin, University of North Carolina, Chapel Hill

Cuirong Ren, South Dakota State University

Abel Rodriguez, Duke University

Paul Schuette, Meredith College

Nicoleta Serban, Georgia Institute of Technology
Kazuhiko Shinki, UW-Madison
Richard Smith, UNC-Chapel Hill
Neil Shephard, Oxford (UK)
Huixia Wang, N.C. State University
Ishay Weissman, Technion (Israel)
Gentry White, N.C. State University
Robert Wolpert, DUKE University
Yimin Xiao, Michigan State University
Fei Xu, Renmin University of China
Saeid Yasamin, Indiana University
Dabao Zhang, Purdue University

Schedule of meetings:

Sept 27: Initial group discussion
Oct 11: Richard Smith gave a tutorial on multivariate extreme value theory
Oct 18: Dan Cooley on Spatial Extremes
Oct 25: Jaya Bishwal (remotely from Charlotte) on Financial Extremes
Nov 8: Group discussion, primarily to agree on an outline program for future meetings
Nov 15: Vered Madar on multiple comparisons and possible links with extreme value theory
Nov 29: Nicoleta Serban on high-dimensional wavelets and extremes
Dec 6: Xiao Qin on Dependence Modelling in Multivariate Extremes
Dec 13: Richard Smith on possibilities for extending the Ledford-Tawn models to higher dimensions
Jan 14 2008: Vangelis Evangelou presented an overview of Davis, Dunsmuir and Streett (2003) *Biometrika* paper
Jan 22–24: Workshop on EXTREMES: Events, Models and Mathematical Theory, January 22-24, 2008
Jan 28: Laurens de Haan on Extremal Processes
Feb 4: Group discussion
Feb 11: Richard Smith on statistical models for hurricane counts
Feb 25: Sidney Resnick on Regular Variation, Extreme Value Theory, Hidden Regular Variation and Conditioned Limit Laws (part I of a multi-part talk)
March 3: Anthony Ledford (remotely from Oxford) on “A new class of models for bivariate joint tails” (joint work with Alexandra Ramos)
March 17: John Nolan (remotely from Washington) on Multivariate Stable Laws.
March 24: Resnick presentation part II
March 31: Yimin Xiao on Extreme Value Theory of Gaussian Random Fields
April 14: Resnick presentation part III

4.3.4 Working Group on Multivariate Extremes — Applications

Participants:

Kobi Abayomi, Duke University
Michela Cameletti, SAMSI

Guang Cheng, SAMSI/Duke University
Dan Cooley, Colorado State
Evangelos Evangelou, UNC-Chapel Hill
Eric Gilleland, NCAR
Rosalba Ignaccolo, SAMSI
Yongku Kim, SAMSI /Duke
Wenbin Lu, N.C. State University
Vered Madar, SAMSI
Pilar Munoz, Technical University of Catalonia, Spain
XuanLong Nguyen, SAMSI/ Duke
John Nolan, American University
Nabendu Pal, University of Louisiana
Xiao Qin, UNC-Chapel Hill
Huiyan Sang, Duke University
Paul Schuette, Meredith College
Richard Smith, UNC-Chapel Hill
Nikita Tuzov, Purdue Univ
Huixia Wang, N.C. State University
Robert Wolpert, Duke University
Fei Xu
Zhengjun Zhang, University of Wisconsin

Schedule of meetings:

Sep 27: Organizational meeting. Aims of the working group were discussed and a list of references compiled.

Oct 11: Paul Schuette on “Power laws and extreme values”.

Oct 18: Pal Nabendu on estimation and testing with (univariate) EVD.

Oct. 25: Discussion of the papers by S. Poon, M. Rockinger and J. Tawn (2004), Extreme value dependence in financial markets: Diagnostics, models and financial implications. *Review of Financial Studies* **17**, 581–610; and J.L. Geluk, L. de Haan, and C. G. de Vries (2007), Weak and Strong Financial Frailty, Tinbergen Institute Discussion Paper TI 2007-023/2.

Nov 8: Kobi Abayomi on EVD-multiple environment hazard: World Bank hostpots report; Zhengjun Zhang on Testing and modeling extreme dependencies in financial markets

Jan 28: Kobi Abayomi discussed *Multivariate Models and Dependence Concepts*, by Harry Joe (1997).

Feb 25: Pilar Muñoz on Daily Spanish electricity prices and other variables associated with them: Univariate and bivariate approaches.

Mar 10: Evangelos Evangelou on Description and models for five stock prices.

Mar 24: Nikita Tuzov on Applying EVT analysis to US energy prices.

Apr 7: Jen Ting on US energy prices.

4.3.5 New Research Stimulated by the Program

At the time of writing (April 2009), research begun within the program and research stimulated by the program have been presented and continue to develop.

1. **Extreme Value Distributions.** Xiaoyan Lin (graduate student, visiting from University of Missouri) continues her work on reference priors for extreme value distributions; see Section 6.7 for a more detailed description of her work.

2. **Regular Variation and Multivariate Extremes.**

Sidney Resnick (Cornell University, visiting SAMSI) is working on the connection between regular variation and different formulations of multivariate extreme value theory. Regular variation on cones can be specialized in at least 3 different directions giving (a) classical extreme value theory; (b) hidden regular variation (the Ledford-Tawn approach); and (c) limit approximations for the distribution of a random vector given one component is extreme (the Heffernan-Tawn approach). In each of the first two cases, there exists a reduction to a one dimensional criterion which allows detection of the phenomenon. For case (c), we have taken the initial steps to find a criterion that a conditioned limit law exists to a one dimensional condition that can be statistically confirmed. Done first in an important special case progress has been made on the generalization.

Xiao Qin and **Richard Smith** are working to develop alternative forms of bivariate and multivariate distributions consistent with the Ledford-Tawn-Ramos approach to characterizing extremal dependence. Xiao presented a Topic Contributed Paper on this subject at the JSM in August, 2008.

3. **Max-Stable Processes.** Zhengjun Zhang (University of Wisconsin) has revised and resubmitted his paper “On Approximating Max-stable Processes and Constructing Extremal Copula Functions.”

In addition, Zhengjun has presented this work at the JSM in August 2008, and at the International Conference on Financial Econometrics, June 21-23, Chengdu, China.

XuanLong Nguyen’s (postdoc, SAMSI and Duke) work on estimation methods in max-stable processes (e.g., M4 processes) using empirical process theory and concentration of measure techniques has been drafted into a paper.

4. **Spatial and Space-Time Processes.** Huiyan Sang (PhD student, Duke University) presented “Extreme Value Modeling for Space-time Data with Meteorological Applications” at the International Indian Statistical Association Conference (May 22-25, 2008, Storrs, CT).

Huiyan also worked with Yongku Kim on extreme value modeling for explaining sea surface temperatures observed in space and their impact on hurricane data.

Zhengjun Zhang is preparing a paper “Nonlinear and Extremal Spatial Dependencies of Precipitations in Continental USA.”

Cuirong Ren (Department of Plant Science, South Dakota State University, visiting SAMSI) has written two papers on objective priors in spatial statistics:

- (a) “Objective Bayesian Analysis for a Spatial Model with Nugget Effects” (Cuirong Ren, Dongchu Sun and Zhuoqiong He).

Summary: We often need to consider geostatistical data with nugget effects. In this paper, we have systematically studied the Jeffreys priors and various reference priors, derived by both “exact” and asymptotic marginalization. Interestingly, not all Jeffreys and reference priors yield proper posterior distributions. We have found the conditions under which the corresponding posteriors are proper. Finally, we conduct a simulation study to compare the objective priors by frequentist coverage probabilities of the one-sided credible intervals.

- (b) “Objective Bayesian Analysis for a Spatial Model with Correlated Repeated Measurements” (Cuirong Ren, Dongchu Sun, Jing Zhang and Zhuoqiong He).

Summary: Geostatistics is an important part of Spatial analysis, and has been widely used in case studies. Using the Bayesian hierarchical modeling not only facilitates to count all the variabilities of the parameters, but also helps decompose the problem into several levels, and hence makes the model more flexible and improves the estimation of parameters as well as the prediction of new locations. In this paper the reference priors and Jeffreys priors for a Spatial model with repeated measurement are developed and comparisons are made based on frequentist coverage probabilities of the one-sided credible intervals.

Cuirong also presented a paper at the IISA in May 2008 at the University of Connecticut.

5. Extreme Values in Finance and Insurance.

Dougal Goodman (Director of the Foundation for Science and Technology, London, UK) writes: “I found it invaluable to attend the opening workshop of the programme last September. The workshop stimulated me to think of new ways in which extreme statistics can be applied to policy questions within government departments. I used as an example in my talk the sudden failure of the Northern Rock bank in the UK. The succession of further failures in the banking system particularly Bear Sterns since the workshop raise many interesting questions about how extreme statistics methods could be used to assist managers and regulators in assessing risks in the financial services sector. Multivariate methods surely have an application to these problems.”

Xiao Qin (PhD student from Beihang University, China, visiting UNC-Chapel Hill) has written a paper using extreme value theory for the identification of currency crises. The paper is submitted to *Journal of International Money and Finance*, and was also the subject of a poster presentation at the opening workshop.

Xiao is also using the Ledford-Tawn-Ramos approach to bivariate extremes to model the coincidence of two specific types of financial crises, i.e., banking system crises and currency crises (the “twin crises” in economic literature). She has submitted an abstract to the 2009 Annual Meeting of the American Economic Association.

Zhengjun Zhang has taught a seminar course on “Statistics for Financial Markets and Insurance” at the University of Wisconsin drawing on materials that are closely related to presentations in the SAMSI Risk program.

6. **Energy Markets.** **Pilar Muñoz** (Technical University of Catalonia, Spain, visiting SAMSI) is working on applying univariate and bivariate extreme value theory to daily

Spanish electricity prices and other variables associated with them. She has also started a collaboration on Energy Markets with Nikita Tuzov, PhD student of the Department of Statistics, Purdue University.

7. Meteorology and Hydrology Applications.

Mendez B. and Pericchi L.R. (2008) “Assessing Conditional Extremal Risk of Flooding in Puerto Rico”. *Stoch. Environ. Res. Risk Assess.* (in press).

Luis Pericchi also gave the talk (co-authored with Beatriz Mendes, Abel Rodriguez and Scott Sisson) “Experiences with Modeling in Multivariate Extremes”, Joint Statistical Meetings, Denver, August, 2008.

8. **Hurricanes.** **Yongku Kim**’s work on statistical modeling for Atlantic tropical storms based on climate factors such as northern (spatial) Atlantic sea surface temperature, global surface temperature and Atlantic multidecadal oscillation has yielded promising preliminary results.

9. **Inference on Networks.** **Ian Dinwoodie** (Duke University, visiting SAMSI during 2006/07) has submitted two papers: “Statistical Estimation of Available Bandwidth” by Ian H. Dinwoodie (*Journal of Statistical Computation and Simulation*, September 2007) and “Markov chains, quotient ideals, and connectivity with positive margins” by Yuguo Chen, Ian H. Dinwoodie and Ruriko Yoshida (to appear in a volume dedicated to G. Pistone, Cambridge University Press). He gave a talk “Network Inference from Indirect Measurements” at the Department of Statistics, UIUC.

4.4 Environmental Risk Analysis (ERA) Working Group

4.4.1 Organization and Membership

This group formed during the Opening Workshop, inspired particularly by the talk given by Dr. Anne Smith during that workshop. The Environmental Protection Agency (EPA) is charged under the Clear Air Act with promulgating air pollution standards that are “requisite to protect the human health”. Commonly regulated pollutants include particulate matter, ozone, sulfur dioxide, nitrogen dioxide and carbon monoxide. Part of the process of setting air pollution standards is an assessment of the scientific literature to assess the adverse health effects of air pollution. Many statisticians and epidemiologists are involved in this work. Another part of the EPA review, in which statisticians have been less involved, is the “risk assessment”, in which quantitative estimates of health effects are translated into specific scenarios of health outcomes under various proposed forms of the air pollution standard. This process has become particularly important this year because of the EPA’s review of the ozone standard, which resulted in a new standard being announced in March 2008 (75 parts per billion for the maximum daily 8-hour average ozone, down from 84 ppb under the previous standard). The statistical assumption underlying this risk assessment, however, are poorly understood, especially regarding the uncertainty of the resulting estimates. The overall aim of this group was to investigate and quantify several aspects of this risk assessment.

Regular participants were:

David Bell, Duke
Michela Cameletti, SAMSI
Rosalba Ignaccolo, SAMSI
Yongku Kim, SAMSI/Duke
Amy Nail, N.C. State University
Bahjat Qaqish, UNC-Chapel Hill
Richard Smith, UNC-Chapel Hill

4.4.2 Activities

Oct 15: Amy Nail. Quantifying local creation and regional transport using a hierarchical space-time model of ozone as a function of observed NO_x, a latent space-time VOC process, emissions, and meteorology.
Oct 22: Michela Cameletti. Computer intensive procedure for mapping and modeling a spatio-temporal process and its uncertainty
Oct 29: Bahjat Qaqish. Review of NMMAPS.
Nov 5: Yongku Kim. Change of Spatiotemporal Scale in Dynamic Models
Nov 12: Rosalba Ignaccolo. Review of the paper Everson, PJ and Morris, CN (2000). Inference for multivariate normal hierarchical models. *J.R.Statist. Soc. B* **62**, 399–412.
Nov 19: Eric Gilleland (NCAR) on a review of Wikle, CK and Cressie, N (1999), A dimension-reduced approach to space-time Kalman filtering. *Biometrika* **86**, 815–829.
Nov 26: David Bell, review of Chen et al. (2007), Outdoor air pollution: ozone health effects. *Am.J.Med.Sci.* **333** (4), 244–248.
Dec 3: Richard Smith. Reanalysis of NMMAPS database on ozone and mortality.
Dec 10: Group discussion on NMMAPS data
Jan 28: Yongku Kim on a rollback application to NMMAPS ozone data
Feb 6: Rollback and Programing Issues
Feb 11 and Feb 13: Ozone Risk Modeling including Rollback
Feb 25: Ozone Risk Modeling including seasonal issues
March 3: Ozone Risk Modeling : continued
March 17: Relative risk analysis and other modeling issues

4.4.3 Research Outcomes

The EPA works to develop and enforce regulations that implement environmental laws enacted by Congress. The EPA is responsible for researching and setting national standards for a variety of environmental programs. A recent example is the decision to lower the ozone standard to 75 parts per billion (ppb).

In addition to reviewing the available literature on the health effects of ozone, the EPA did its own analysis of the potential impact of new regulations by analysis of 12 large metropolitan areas. The main aim of the ERA Working Group is to carry out an extensive analysis using data from 95 cities, including the 12 used in the EPA's analysis. The aim is to look at several issues related to the potential impact of lower standards.

To assess the impact of lower standards, a model is needed for how ozone levels would change if they were to meet a new standard. For this purpose the EPA uses "roll-back

functions” that predict those changes. Mainly, three roll-back functions are used: proportional (with or without a threshold level); quadratic roll-back; Weibull roll-back.

Issues to be addressed include: 1) The extent of variability of the estimates for each city; 2) The sensitivity of the analysis to various risk models including the adjustment for PM10. 3) The ozone measure used in the regression model (daily average, daily maximum, maximum 8-hour average). 4) The inclusion or exclusion of days with high temperatures. 5) The different roll-back functions.

The plan is to estimate the effect of different standards including the recently approved standard of 75 ppb in addition to other possible standards such as 70, 65 and 60 ppb.

The analysis outlined above aims to assess not only the various forms of statistical variability involved in assessing the impact of various regulation, but also the sensitivity of the analyses to various model assumptions, and uncertainty about the model itself.

4.5 Service Sector Risk

4.5.1 Organization and Membership

Tim Bedford, Strathclyde Business School

Lea Deleris, IBM

Jonathan Hosking, IBM

David Ríos Insua, Universidad Rey Juan Carlos, Spain

Huijing Jiang, Georgia Institute of Technology

Jesus Ríos, SAMSI

Fabrizio Ruggeri, CNR-IMATI, Italy

Huiyan Sang, Duke University

Nicoleta Serban, Georgia Institute of Technology

Farhad Shafti, Strathclyde Business School

Haipeng Shen, UNC-Chapel Hill

Lesley Walls, University of Strathclyde

Saeid Yasamin, Indiana University

4.5.2 Research Outcomes

The group is currently working on two papers:

1. Reduced order models for Bayesian risk analysis. A technical report has been written, and one additional numerical example need to be finished.

It’s conceivable that the group will follow up this paper with another one that focuses on Bayesian discrete even simulation with application to workforce management in labor-intensive service systems such as call centers or emergency rooms.

2. Statistical service classification for risk management.
 - (a) A pilot dataset has been compiled. Initial statistical analysis has been performed, and shows promising results. The findings from this paper would be of interest to companies such as IBM. There is a difficulty getting real industrial data but efforts are being made in this direction.

- (b) An abstract has been submitted for presentation as the Frontier of Services Conference to be held in Washington DC this October. Acceptance notification will be sent out early May.
- (c) The group had also accepted an invitation to present the project at this year's INFORMS conference (again) in DC this October.
- (d) One hope is that the presentations can lead the group to some data-holders that are interested in sharing their data.

The intention is to write up the paper and submit for publication in *Technometrics* or some similar journal.

As part of the data quest efforts, the group also outreached to various business providers to get them interested in their projects and eventually willing to contribute data. Some examples include IBM Europe and Genesys Labs of Alcatel-Lucent.

5 Other Activities

5.1 Courses

Two graduate courses were held at SAMSI associated with the Risk Program.

5.1.1 Fall Course

Course Title: *Decision Theory and Risk Analysis*. Instructors: Dipak Dey, University of Connecticut; Larry Brown, University of Pennsylvania; David Rios, Universidad Rey Juan Carlos.

Short Course Description: Fundamental concepts for decision theory and use of expert opinion as applied to risk analysis. Exponential families: sufficiency, minimaxity, admissibility. Decision rules and risk: loss functions, convexity, risk analysis. Estimation, analysis and model selection: Minimax, shrinkage, Bayes, hierarchical Bayes, empirical Bayes, data and opinion as prior information.

5.1.2 Spring Course

Course Name: *Extremes and Case Studies in Risk Analysis*. Instructor: Pilar Muñoz.

This was taught as an Independent Study course.

5.2 JSM

The Risk program has organized three Topic Contributed Paper sessions at the 2008 Joint Statistical Meetings, sponsored by the ASA's Section on Risk Analysis.

Session 1: Bayesian Modeling of Extreme Events.

Organizer(s): Dipak Dey, University of Connecticut

Chair(s): Bani K. Mallick, Texas A&M University

Wednesday August 6 2008, 2:00-3:50 pm.

1. A Bayesian Framework For Adversarial Risk Analysis — Jesus Ríos, SAMSI; David Ríos, Universidad Rey Juan Carlos; David Banks, Duke University
2. Semiparametric Functional Estimation Using Quantile Based Prior Elicitation — Elijah Gaioni, University of Connecticut; Dipak Dey, University of Connecticut; Mircea Grigoriu, Cornell University
3. Bayesian Hierarchical Modeling For Extreme Values Observed Over Space And Time — Huiyan Sang, Duke University; Alan Gelfand, Duke University
4. Thresholding for Multivariate Extreme Values — Kobi A. Abayomi, Duke University
5. Bayesian Model Selection Of The Farlie-Gumbel-Morgenstern Copula For Describing Two Generalized Extreme Value Variables — Vered Madar, SAMSI

Session 2: Risk Analysis For Industry And The Environment

Organizer: Richard L. Smith, The University of North Carolina at Chapel Hill

Chair: Elizabeth C. Shamseldin, University of North Carolina

Sunday August 3 2008, 2:00-3:50 pm.

1. Quantifying Local Creation And Regional Transport Using A Hierarchical Space-Time Model Of Ozone As A Function Of Observed NO_x, A Latent Voc Process, Emissions, And Meteorology — Amy J. Nail, North Carolina State University; John F. Monahan, North Carolina State University; Jacqueline Hughes-Oliver, North Carolina State University
2. An Analysis Of The Potential Impact Of Various Ozone Regulatory Standards — Rosalba Ignaccolo, Universita' degli Studi di Torino/SAMSI; Yongku Kim, Statistical and Applied Mathematical Sciences Institute; Bahjat Qaqish, University of North Carolina at Chapel Hill; Michela Cameletti, Universita' degli Studi di Bergamo/SAMSI; Richard L. Smith, The University of North Carolina at Chapel Hill
3. Multivariate Generalized Linear ARMA Processes: An Application To Hurricane Activity — Evangelos Evangelou, University of North Carolina; Richard L. Smith, The University of North Carolina at Chapel Hill; Amy Braverman, Jet Propulsion Laboratory
4. Probabilistic Risk Analysis For ICT Industry — Jose A. Rubio, Universidad Rey Juan Carlos; David Rios Insua, Universidad Rey Juan Carlos
5. Seismic Risk Analysis — Mircea Grigoriu, Cornell University

Session 3: The Samsi Program On Risk Analysis, Extreme Events, And Decision Theory

Organizer: Richard L. Smith, The University of North Carolina at Chapel Hill

Chair: Nell Sedransk, National Institute of Statistical Sciences

Tuesday August 5 2008, 10:30 am – 12:20 pm

1. Extreme Co-Movements And Extreme Impacts In High Frequency Data In Finance — Zhengjun Zhang, University of Wisconsin
2. Modelling multivariate extreme dependence — Xiao Qin, Beihang University; University of North Carolina; Richard L. Smith, The University of North Carolina at Chapel Hill; Ruoan Ren, Beihang University
3. Multivariate Analyses Of Extremes — Luis R. Pericchi, University of Puerto Rico, Rio Piedras; Beatriz Mendes, Universidade Federal de Rio de Janeiro; Scott Sisson, University New South Wales; Abel Rodriguez, University of California, Santa Clara
4. Downscaling Extremes: A Comparison Of Extreme Value Distributions In Point-Source And Gridded Precipitation Data — Elizabeth C. Shamseldin, University of North Carolina; Richard L. Smith, The University of North Carolina at Chapel Hill; Stephan Sain, National Center for Atmospheric Research; Dan Cooley, Colorado State University; Linda O. Mearns, National Center for Atmospheric Research
5. Hurricanes And Global Warming — Richard L. Smith, The University of North Carolina at Chapel Hill; Evangelos Evangelou, University of North Carolina; Gabriel A. Vecchi, Geophysical Fluid Dynamics Laboratory; Thomas R. Knutson, Geophysical Fluid Dynamics Laboratory

6 Education and Outreach

In this section we include individual reports from the postdoctoral fellows and graduate students supported by the SAMSI Risk program, and of the undergraduate workshop that was held in November 2007.

6.1 Guang Cheng (Postdoctoral Fellow, SAMSI and Duke)

Guang Cheng completed his postdoc in Summer 2008. He is now an assistant professor position in the Department of Statistics, Purdue University.

6.1.1 Completed Papers

1. Guang Cheng (2007) Semiparametric Additive Isotonic Regression (Under Revision)
2. Guang Cheng and Helen Zhang (2008), Efficient Estimation and Consistent Variable Selection for Partial Spline Models (under revision, to be submitted to *Annals of Statistics*).
3. Guang Cheng (2007), One-Step M-estimator for Semiparametric Models (In progress)
4. Guang Cheng, Yufeng Liu and Helen Zhang, (2008) Linear or Nonlinear Automatic Selection for Partial Linear Models (In Progress)

6.1.2 Other Activities

1. Invited talk about “Semiparametric Additive Isotonic Regression” in Nonparametric Conference 2007, Columbia, SC. Also to be presented at the JSM in August 2008.
2. I have begun research collaboration with Prof. Nicoleta Serban at Georgia Tech when she visited SAMSI in the fall semester of 2007. Our collaboration focuses on Hierarchical Functional Data Modelling.
3. I have also worked on the theoretical problem proposed by Prof. Richard Smith about multivariate extensions of the Ledford-Tawn approach. I hope to have some results by the summer of 2008.

6.2 Jesus Ríos (Postdoctoral Fellow)

6.2.1 Research interests

Risk analysis, Decision analysis, Negotiation analysis, Game Theory

6.2.2 PhD Program

University/Department: Rey Juan Carlos University (Spain), Department of Statistics and Operation Research

Dissertation Advisor: David Ríos

Year of Ph.D.: May 2006

6.2.3 SAMSI Research

SAMSI Research Mentor: David Ríos

6.2.4 Course(s) (fall and spring)

Decision theory and risk analysis

6.2.5 Workshops Attended (and Workshop Support Tasks)

1. Opening workshop;
2. Risk: Perception, policy and practice
3. EXTREMES: Events, Models and Mathematical Theory (poster presentation)
4. RISK Revisited: Progress and Challenges (talk presentation)

6.2.6 Special Tasks

Webmaster (September 2007 December 2007)

6.2.7 Talks and presentations

10/17/2007: Analyzing Adversarial Threats

Two-Day Undergraduate Workshop: November 9-10, 2007

1. Discovering Influence Diagrams with GeNIe: Decision analysis for risk management;
2. Discovering Game theoretic concepts with Gambit for adversarial risk analysis

6.2.8 Working Group I: Adversarial risk

Special Tasks for Working Group: webmaster

Presentations to Working Group: 10/11/2007: Modelling the others: Game theory Rationality vs. Bayesian approach

10/18/2007: Some adversarial risk models

10/25/2007: A possible alternative approach to adversarial risk analysis

11/15/2007: Asymmetric information in adversarial risk analysis

01/31/2008: Our framework for ARA: The assessment problem. Example: Bidding in a Auction

02/07/2008: Random games and the commutativity issue

02/21/2008: The Auction problem

6.2.9 Research Area and Plans

Application of game theory, risk analysis and portfolio theory to adversarial decision settings, like in terrorism, business competition... Emphasis on issues related with how to model adversarial dynamic decisions, external uncertainties and modeling adversaries behavior as well as on computational issues.

6.2.10 Research Progress Report and SAMSI Program Final Report

Research Project Title: Foundations of adversarial risk analysis, with David Banks, David Ríos

Review of ideas from game theory, decision analysis and probability risk analysis that have been applied in adversarial decision making. We propose an improved approach and illustrate it with examples in antiterrorism and corporate auction biddings

Research Contributions (publication submissions, articles in preparation, etc.):

1. Paper submitted to Group Decision and Negotiation journal: Balanced increments and concessions methods for arbitration and negotiations
2. Paper completed Adversarial risk analysis

Presentations outside SAMSI (including invitations for future talks):

1. Presentation scheduled at GDN 2008 in Coimbra in June 08
2. Presentation scheduled at JSM 2008 in Denver, Colorado in August 08

Research Project Title: Computations for adversarial risk analysis, with David Ríos

Specific Goals and Accomplishments (results): It project focuses in computational issues for finding nondominated solution in a collaborative framework (eg, two countries collaborating for managing risks by sharing resources to mitigate terrorist attacks or natural disasters), Nash equilibria in adversarial settings and prescriptive recommendations based on a Bayesian/Game theoretic analysis of adversarial actions (following our framework proposed in our first project)

Research Contributions (publication submissions, articles in preparation, etc.):

1. Paper submitted to Decision Analysis: Supporting group decisions over influence diagrams
2. Paper in preparation: Computations in adversarial risks (skeleton of the paper prepared, all required reading done)

6.2.11 Future Research Plans (after completion of SAMSI Program)

I have a new appointment from April, 1st 2008 at Aalborg University (Denmark).

6.3 Vered Madar (Postdoctoral Fellow)

Dr. Madar graduated from Statistics and OR, Tel-Aviv University, Tel-Aviv, Israel, (PhD, 2007) working under Professor Yoav Benjamini. At SAMSI she has been working in the program on Risk Analysis, Extreme Events and Decision Theory, under the mentorship of Dipak Dey and Nell Sedransk.

6.3.1 SAMSI Activities

- Attended Risk Analysis course (fall)
- Attended all SAMSI's 2007/08 workshops (fall and spring)
- Postdoc-Grad Student Seminar: Bayesian Modeling of Bivariate Extremes with Applications (Nov. 7).
- Poster at SAMSI Extremes Workshop (January 23)

6.3.2 Undergraduate Workshop

Specifics to be added later

6.3.3 Bayes Risk Working Group

- Special Tasks for Working Group: Webmaster
- Presentation to Working Group: "Some Thoughts on Bayesian Modeling of Bivariate Extremes (Dec, 6)

- Research Area (1): Bayesian Model Selection for the Generalized FGM copula in the bivariate case when both marginal distributions are general extreme value.
- Research Area (2): Prior elicitation in the bivariate extreme value situation and some related modeling issues.

6.3.4 Multivariate Extremes (Methodology) Working Group

- Presentations to Working Group: “Introduction to Multiple Comparisons (Nov. 15)
- Planned Research: NonBayesian Copula Selection when both marginal distributions are general extreme value.

6.3.5 Other Research

Papers from Ph.D. Research (work in progress):

- The Variable-Ratio Simultaneous Confidence Intervals (self)
- The Quasi-Conventional Simultaneous Confidence Intervals for Better sign Determination (with Yoav Benjamini and Philip B Stark)
- The Quasi-Conventional Intervals Under Dependence (self)
- An inequality for multivariate normal probabilities of nonsymmetric rectangles.

Presentations of Other Research: UNC stat seminar, January 14: The Quasi-Conventional Simultaneous Confidence Intervals.

6.4 Sourish Das (Graduate Student)

Mr. Das is PhD Student, University of Connecticut, Department of Statistics, working under Dr. Dipak Dey. His expected completion date of PhD is Summer 2008. His mentor at SAMSI has been Dipak Dey.

6.4.1 Activities attended

- Opening Workshop (Sep 16–19, 2007)
- Workshop on Risk: Perception, Policy and Practice: October 3–4, 2007.

6.4.2 Presentations

- Postdoc-Grad Student Seminar: Hitchhikers Guide to Presentations
- Postdoc-Grad Student Seminar: Analysis of Hurricane Activity in West Pacific and Indian Ocean; 11/8/2007
- Undergraduate Workshop: Presented Analysis of Hurricane Activity

- Undergraduate Workshop: Helped Prof. Dey and Prof. R. Smith organizing the session on hands on experience. I gave them a data set on Hurricane Activity at Atlantic Ocean since 1851 to 2006. Students analyzes that data set using R.
- Graduate Fellow Presentation Poster (title and abstract to be added later)

6.4.3 Report on Research

The main area of research is Bayesian Extreme Value Theory.

I am developing Bayesian Method of analyzing extreme category in Multinomial-Dirichlet model, especially, in the context of the Hurricane data of Indian Ocean (southern hemisphere region) and Pacific Ocean (West pacific region). Here the storms are categorized into 5 category; where estimating the probability of rare category (that is category 5 hurricane) is challenging. This work will be a part of the 3rd chapter of my Ph.D. dissertation.

6.5 Elijah Gaioni (Graduate Student)

Elijah Gaioni is completing two papers that have come out of discussions that arose during the Bayes Risk working group meetings. Both papers address the problem of inadequate numerical data by incorporating quantile-based expert information into the statistical modeling framework. The first paper is a joint work with Mircea Grigoriu, Elijah and myself, entitled Semiparametric functional estimation using quantile based prior elicitation. The first draft of this paper has been completed, and it will be submitted to a peer-reviewed journal shortly. The second paper models river behavior where the emphasis is on the joint modeling of the extreme and non-extreme components of the process. This paper is nearly finished and will also be submitted to a peer-reviewed journal when it is completed later this semester.

(a) Semiparametric functional estimation using quantile based prior elicitation. (Dipak Dey, Mircea Grigoriu, Elijah Gaioni)

(b) Incorporating expert opinion into the joint modeling of extreme and non-extreme components of river flow. (Elijah Gaioni, Dipak Dey)

The extreme river flow work will continue to be sponsored by the Center for Environmental Statistics and Engineering through the current semester and possibly future semesters.

6.5.1 Report on Research

This report summarizes my activities and research related to the Bayes Risk group at SAMSI.

There are three main research projects Ive been involved in. The first has resulted in the paper entitled Semiparametric functional estimation using quantile based prior elicitation, which is a joint work with Mircea Grigoriu and Dipak Dey. The second, which is nearing completion and will also be written up as a paper, deals with extreme values in river flow phenomena. The third is an extension of this second paper to the multivariate case and is a work in progress. All papers will be submitted to peer-reviewed statistics journals for publication. Further, since these topics are highly interrelated each will contribute one chapter towards my Ph.D. thesis.

The first paper addresses the problem of incorporating vague prior information, as specified through a small number of quantiles, into marginal distribution estimation. An optimal

prior distribution consistent with this information is sought in a semiparametric framework. The functional of interest may then be used for predictive purposes. In order to overcome computational difficulties an innovative means of nonparametrically representing the prior distribution is employed. The statistical software package R is being used to implement this methodology.

The second avenue of research mentioned pertains to the study of extreme river flow events. These events were modeled as mixtures of gamma and extreme value distributions in a Bayesian framework. Both the extreme and non-extreme components of such processes were jointly modeled. The decision to tackle this particular problem arose out of a working group discussion held shortly after the Risk Analysis, Extreme Events and Decision Theory program. In particular, we explore flash flooding in Texas using response and covariate information obtained from the United States Geological Survey (USGS) website. The covariates are introduced through a generalized linear model and serve to enhance the predictive capacity of the model.

Preliminary results for both of the first two papers mentioned above have already been presented at numerous working group meetings, and during SAMSI's graduate student seminar, and at the University of Connecticut student seminar. Talks at the New England Statistics Symposium, INAR, and JSM are also planned.

Much of the mathematics for the third paper, which deals with the multivariate extension of the case mentioned above, has already been completed. The correlation structure between the different multivariate responses is introduced through the mixing parameters, and it naturally accommodates responses that are measured on different scales. The implementation details have yet to be completed, though they will build on the R code used for the univariate version.

In addition to the presentations mentioned above, I participated in an undergraduate workshop. During this workshop on November 9th I gave a presentation covering some of the basic statistical elements that could be incorporated into an analysis of the extreme component of river flow. Subsequently, an interactive session was conducted during which undergraduate students applied what they had learned using the `extRemes` package in R. At the end of the one-day workshop, the graduate and undergraduate students spoke over dinner about possible careers in the mathematical sciences.

As I continue my studies at the University of Connecticut, support for the second and third papers mentioned above will be provided by the Center for Environmental Statistics and Engineering. Weekly meetings through WebEx provide the basis for continued joint collaboration.

6.6 Evangelos Evangelou (Graduate Student)

I participated in the "Risk Analysis, Extreme Events and Decision Theory Program at SAMSI as a graduate fellow. Being a graduate student, I am still in the process of learning and familiarizing myself with new research ideas and topics, and my involvement in the program has greatly contributed towards expanding my research horizons. The courses offered, the seminars and the working groups at SAMSI have had a significant impact to my research.

My course work at SAMSI included two courses, one in each semester. The first course introduced us into new issues such as prior elicitation and adversarial risk. The latter constituted the topic of my class project. Under the guidance of Dr. Ríos Insua, I developed an

idea for modeling actions that result to random payoff. A classic example is a terrorist attack where the government is placing resources to defend its region while the terrorist chooses an action for attacking. In my project, I suggested modeling the loss as a beta distributed random variable times a constant and then look at the expected loss. For the second course, I focused on modeling financial time series. I worked together with Dr. Munoz on modeling five stocks from the European market. For these series, we found that the models that fit best are GARCH or E-GARCH with t distributed errors.

My contribution to the working groups consisted in participating in discussions and holding two presentations. In the “Multivariate Extremes Applications working group I presented the earlier mentioned financial time series project. I also participated in the “Multivariate extremes Methodology working group where I presented a paper for analyzing time series data following the Poisson distribution. This paper was proposed by Dr. Smith as a method to be used to analyze hurricane occurrences in the Atlantic and investigate the correlation with sea surface temperature; his idea was to analyze the two variables as a bivariate time series to remove the autocorrelation and then test for correlation between them.

During the SAMSI undergraduate workshops, I had the opportunity to provide students with an introduction to the methodology for extreme value analysis. At the same time, I guided students in the use of computer software in the practice session. During breaks I had the opportunity to talk to them and answer their questions regarding graduate studies.

Among other activities, I also attended the SAMSI seminars, where I became familiar with typical extreme value analysis topics such as modeling the dependence on extreme values for different variables and estimation of the parameters of M4 processes. Overall, these seminars have nurtured and greatly expanded my interest and knowledge in extreme value theory, both at a theoretical and practical level.

6.7 Xiaoyan Lin (graduate student)

Xiaoyan Lin is a graduate student from the University of Missouri, Columbia, who is visiting SAMSI from February to May. Following is a report of her current research.

The idea is to get reference prior under partial invariance structure and to prove the reached prior at least has a proper posterior.

Reference priors under partial invariance structure

Theorem

Suppose $(\boldsymbol{\theta}, \boldsymbol{\xi})$ is the parameter, where

- a component of $\boldsymbol{\theta}$ is the parameter of interest;
- for each fixed $\boldsymbol{\xi}$, $p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\xi})$ has the same group invariance structure with the reference prior being the right-Haar prior $\pi^{RH}(\boldsymbol{\theta})$;
- natural compact sets are of the form $\Theta_c \times \Xi_c$,

the reference prior is then

$$\pi(\boldsymbol{\theta}, \boldsymbol{\xi}) = \pi^{RH}(\boldsymbol{\theta})\pi^R(\boldsymbol{\xi} \mid \boldsymbol{\theta}_0),$$

where $\pi^R(\boldsymbol{\xi} \mid \boldsymbol{\theta}_0)$ is the conditional reference prior given some fixed $\boldsymbol{\theta}_0$; this will not depend on the chosen value of $\boldsymbol{\theta}_0$.

As a special case, consider a family of densities,

$$p(x \mid \mu, \sigma, \boldsymbol{\xi}) = \frac{1}{\sigma} g\left(\frac{x - \mu}{\sigma}, \boldsymbol{\xi}\right), x \geq \mu, \quad (1)$$

where $\sigma > 0$ and $\boldsymbol{\xi} \in \Xi \subset \mathbb{R}^k$. Here μ is a location parameter and σ is a scale parameter. g is a known density depending on $\boldsymbol{\xi}$ only. Suppose we are interested in $\boldsymbol{\theta} = (\mu, \sigma)$. The right Haar prior for $\boldsymbol{\theta}$ is

$$\pi^{RH}(\mu, \sigma) \propto \frac{1}{\sigma}.$$

It is easy to see that the reference prior $\pi^R(\boldsymbol{\xi}) = \pi^R(\boldsymbol{\xi} \mid \boldsymbol{\theta}_0)$ for $\boldsymbol{\xi}$ can be derived from the model $\{g(y, \boldsymbol{\xi}), \boldsymbol{\xi} \in \Xi\}$. Obviously, the generalized Pareto distribution and the generalized extreme value distribution belong to the family.

Current Results

1. The three parameter Pareto distribution

$$f(x \mid \mu, \sigma, \xi) = \frac{1}{\sigma} \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1 - \frac{1}{\xi}}, \quad (2)$$

where the support is $x \in (\mu, \infty)$, if $\xi \geq 0$, and $x \in (\mu, \mu - \frac{\sigma}{\xi})$, if $\xi < 0$.

- when $\xi > -1/2$, the derived reference prior is

$$\pi(\mu, \sigma, \xi) \propto \sigma^{-1} [(1 + \xi)(1 + 2\xi)]^{-1/2}.$$

Note that it is different from the Jeffreys prior

$$\pi(\sigma, \xi) \propto \sigma^{-1} (1 + \xi)^{-1} (1 + 2\xi)^{-1/2}$$

in Castellanos & Cabras (2007). To ensure the valid inference using these two priors, the posterior propriety is required. In Castellanos & Cabras (2007), they have proved that using the Jeffreys prior will lead to a proper posterior. However, there seems a mistake (?) in their proof.

- when $\xi < -1/2$, there's no fisher information. Following the general formal definition, we derived that the reference prior for the standardized generalized Pareto distribution is $-1/\xi$. However, the numerical reference prior seems quite different when ξ goes to $-\infty$. Therefore, I need to check the prior derivation carefully later to see if there's any mistake.

2. The three parameter generalized extreme value distribution has CDF

$$F(y) = \exp \left[- \left\{ 1 - \frac{\xi(y - \mu)}{\sigma} \right\}^{1/\xi} \right] \quad (3)$$

where the support is $x \in (-\infty, \mu + \sigma/\xi)$, if $\xi \geq 0$, and $x \in (\mu + \sigma/\xi, \infty)$, if $\xi < 0$.

– When $\xi < 1/2$, the Jeffreys prior for the standardized GEV is

$$\pi(\xi) \propto \sqrt{\frac{1}{\xi^2} \left\{ \frac{\pi^2}{6} + \left(1 - \gamma - \frac{1}{\xi}\right)^2 + \frac{2q}{\xi} + \frac{p}{\xi^2} \right\}},$$

where $p = (1 - \xi)^2 \Gamma(1 - 2\xi)$, $q = \Gamma(2 - \xi) \{ \psi(1 - \xi) - (1 - \xi)/\xi \}$, $\gamma = 0.5772157$ is Euler's constant, $\Gamma(r)$ is the gamma function and $\psi(r) = d \log \Gamma(r) / dr$.

– When $\xi > 1/2$, there's no fisher information. At current stage, I only have some numerical reference prior. In future, the theoretical reference prior will be explored.

3. The three parameter Weibull (μ, η, β) distribution with the density

$$p(x | \mu, \eta, \beta) = \frac{\beta(x - \mu)^{\beta-1}}{\eta^\beta} \exp\left(-\frac{(x - \mu)^\beta}{\eta^\beta}\right), \quad x > \mu. \quad (4)$$

Under the partial invariance rule, the reference prior

$$\pi(\mu, \eta, \beta) \propto \frac{1}{\eta^\beta}.$$

For the two parameter Weibull when μ is known, the reference prior is again

$$\pi(\eta, \beta) \propto \frac{1}{\eta^\beta}.$$

I have proved that when not all of x_i 's are equal, the posterior distribution of (η, β) are proper for the two parameter Weibull distribution. In future, I will explore the posterior propriety for the three parameter Weibull using the prior $1/(\eta^\beta)$.

6.8 Undergraduate Workshop

A two-day undergraduate workshop, organized around the themes of the Risk program, was held at SAMSI, November 9–10 2007. Presentations were delivered by:

1. Richard Smith — Statistics of extremes: Assessing the probabilities of very rare events
2. Elaine Spiller — Models of volcano avalanches: Constructing a risk map for pyroclastic flows
3. Interactive student session on extreme value modeling. Led by Evangelos Evangelou and Guang Cheng.
4. Dipak Dey — Bayesian modeling geared towards extreme events
5. Huiyan Sang — Hierarchical Bayesian modeling of extreme precipitation
6. Sourish Das — Analysis of hurricane data
7. Elijah Gaioni — Modeling river flow data and floods

8. Interactive student session, led by Jayanta Pal and Vered Madar
9. Ralph Smith — Discussion of Graduate School and Career Options
10. David Banks — Game theory and risk analysis: A smallpox application
11. Jesus Ríos, Betsy Enstrom, and Matt Heaton — Discovering game theoretic concepts useful for risk analysis
12. Jesus Ríos, Betsy Enstrom, and Matt Heaton — Discovering influence diagrams with Genie: Decision analysis and risk analysis
13. Mike Porter — Intelligent site selection models for asymmetric threat prediction and decision making

Appendix B: Final Report of the Program on Random Media

1 Introduction

Random media is a classical field which is presently receiving widespread attention as new theory, approximation techniques, and computational capabilities are applied to emerging applications. Due to the breadth of the field, its inherent deterministic, stochastic and applied components have typically been investigated in isolation. However, it is increasingly recognized that these components are inexorably coupled and that synergistic investigations are necessary to provide significant fundamental and technological advances in the field. The SAMSI Program on Random Media provided a forum to investigate statistical and deterministic components of random media for applications.

The goal of the program was to bring together researchers investigating a variety of phenomena pertaining to random media. Specific research directions were drawn from the following topics: random media including scattering theory in highly discontinuous and random media, time reversal, interface problems, imaging problems, scattering theory, porous Media, imaging in random media and related applications. The program addressed a number of fundamental issues in model development, analysis, and numerical approximation. The inherent synergy between deterministic, statistical, and physical analysis necessitates a concerted collaboration between applied mathematicians, statisticians, engineers, geologists, and material scientists which is too often absent but is necessary to provide fundamental advances to the field.

2 Program Organization

2.1 Program Leaders

The program leaders were Russel Caffisch (UCLA), Maarten De Hoop (Purdue University – co-Chair), Rick Durrett (Cornell University – NAC Liaison), Weinan E (Princeton University), Josselin Garnier (Universite Paris VII), William Kath (Northwestern University), George Papanicolaou (Stanford University), Lenya Ryzhik (University of Chicago), Ralph Smith (SAMSI, Directorate Liaison), Chrysoula Tsogka (University of Chicago), Eric Vandendijnden (NYU), Jack Xin (UC Irvine), Wojbor Woyczynski (Case Western Reserve University), and Hongkai Zhao (UC Irvine – co-Chair).

2.2 Local organizers:

The following individuals were the main local organizers for the program: Kazufumi Ito, Zhilin Li, and Ralph Smith, all from North Carolina State University.

2.3 Major Participants

Long and Short Term Visitors: The following individuals spent between a month and semester at SAMSI participating in the program: Yu Chen (Courant Institute, New York University), Laurent Demanet (Stanford), Maarten De Hoop (Purdue University), Josselin Garnier (Paris VII), Isaac Klapper (Montana State University), Xiaofan Li (Illinois Institute of Technology), John Strain (UC Berkeley), Hongkai Zhao (UC Irvine), Guowei He (Iowa State University and Chinese Academy of Science, short term), Ping Lin (University of Dundee, UK, short term).

Postdoctoral Fellows: Elaine Spiller (Mathematics, SUNY-Buffalo), Weigang Zhong (Mathematics, Maryland).

Graduate Students: Qunlei Jiang (North Carolina State University), Brandon Lindley (University of North Carolina), Hui Xie (North Carolina State University), Ke Xu (University of North Carolina), Jason Wilson (Duke University), Sarah Olson (North Carolina State University), Elizabeth Bouzarth (University of North Carolina), Qin Zhang (North Carolina State University).

Other Participants: Jinru Chen, Yushun Wang (Visiting scholar, North Carolina State University), Zhonghua Qiao (postdoc at North Carolina State University), Martin Hiller (North Carolina State University).

Faculty Releases: Tom Beale (Mathematics, NCSU), Greg Forest (Mathematics, UNC), Kazi Ito (Mathematics, NCSU), Chuanshu Ji (Statistics, UNC), Zhilin Li (Mathematics, NCSU), Mauro Maggioni (Mathematics, Duke).

2.4 Working Groups

The working groups met weekly either throughout the year or in the Fall semester, 2007, to pursue their particular research topics. These were identified in the kickoff and midprogram workshops and/or subsequently chosen by the working group participants. A few working groups had their activity concentrated in a shorter period of time. As usual at SAMSI, the working groups consist of SAMSI visitors, postdoctoral fellows, graduate students, and local faculty and scientists. A number of working group members did not reside at SAMSI nor in the area, and took active part on the meetings via teleconferencing and Webex access. The working groups had active web pages in which material, notes, agendas and members were regularly posted.

Heterogeneity in Biological Materials: Led by Greg Forest(UNC)

The active participants were Greg Forest (UNC), Weigang Zhong (SAMSI), Isaac Klapper (Montana State), Brandon Lindley (UNC), Ke Xu (UNC), Elizabeth Bouzarth (UNC). Scott McKinley (Duke), Mircea Grigoriu (Cornell), Lingxing Yao (UNC), Mansoor Haider (NCSU), Chuanshu Ji (UNC), Lisa Fauci (Tulane remote), Robert Dillon (Washington State), and Christel Hohenegger (NYU).

Stochastic PDE: Led by Kazufumi Ito (NCSU)

The participants were Jim Berger (Duke and SAMSI), Mircea Grigoriu (Cornell), Martin Hiller (SAMSI/NCSU), Kazufumi Ito (NCSU), Min Kang (NCSU), Shengtai Li (Los Alamos), Elaine Spiller (SAMSI), John Strain (UC-Berkeley), Yimin Xiao (Michigan State), Jack Xin (UC-Irvine), Qin Zhang (NCSU).

Interface Problems: Led by Thomas Beale (Duke) and Zhilin Li (NCSU)

The participants were Thomas Beale (Duke), Jinru Chen (NCSU & China), Kazufumi Ito (NCSU), Qunlei Jiang (NCSU), Isaac Klapper (Montana State), Xiaofan Li (IIT), Zhilin Li (NCSU), Zhonghua Qiao (NCSU), John Strain (UC-Berkeley), Jason Wilson (Duke), Hui Xie (NCSU), Wenjun Ying (Duke), Qin Zhang (NCSU), Hongkai Zhao (UC-Irvine), Weigang Zhong (SAMSI & NCSU).

Waves and Imaging: Led by Laurent Demanet (Stanford) and Maarten de Hoop (Purdue)

The participants were Yu Chen (NYU), Laurent Demanet (Stanford), Maarten de Hoop (Purdue), Kazufumi Ito (NCSU), Mauro Maggioni (Duke), Vahagn Manukian (NCSU), Yvonne Ou (UCF), Hongkai Zhao (UC-Irvine).

3 Research Foci

The SAMSI Program on Random Media provided a forum to investigate statistical and deterministic components of random media for applications including, but not limited to, time reversal, interface problems, imaging in random media, and scattering theory for discontinuous media.

Time Reversal: The component on time reversal built upon recent analysis and experimental observations that time reversal of waves propagating in disordered media permit refocusing. This somewhat unexpected property has profound ramifications in domains such as wireless communications, medical imaging, nondestructive evaluation, and underwater acoustics. Whereas the behavior of one-dimensional acoustic waves is mathematically and statistically understood, questions regarding multidimensional media remain widely open with the exception of the paraxial wave equation.

Interface Problems: Interface problems arise in a diverse range of applications including multiphase flows and phase transitions in fluid mechanics, thin film and crystal growth simulations in material science, and mathematical biology problems modeled by partial differential equations involving moving fronts. In computational fluid dynamics, electromagnetic scattering and ground water flows, efficient numerical approximation are essential for quantifying the effective property of the medium due to fluctuating inhomogeneous and random medium. The level set method has proven to be an extremely versatile tool for tracking deformations in shape geometries, moving interfaces, and free boundaries in a number of related applications, and one facet of the program will focus on extensions of this approach to include the effects of random media and stochastic processes. Other aspects of the interface component will focus on modeling and analysis of random interface growth processes including crystal growth and

solidification, Monte-Carlo Wiener-Chaos expansion and homogenization methods for stochastic partial differential equations, and level set methods and Lagrangian formulations (particle approaches) for random media simulations.

Imaging: Imaging problems in random media arise in a number of applications including biomedical imaging and seismic analysis. In the latter category, a detailed knowledge of earth medium heterogeneities is necessary for oil and gas recovery, earthquake and volcanic predictions, and environmental analysis. One fundamental issue involves the multiscale relation between large scale structures, which are considered as deterministic, and small scale heterogeneities which are considered to be random fluctuations from the deterministic structures. A related issue concerns the analysis of coupled processes.

Scattering Theory: Whereas mathematical scattering theory for one-dimensional regimes is fairly mature, little of the analysis extends to multidimensional media with the exception of the baraxial wave equation. Hence this facet will focus primarily on the development of theory, numerical methods and validation techniques pertaining to scattering theory for multidimensional media.

4 Specific Activities and Publications

4.1 Heterogeneity in Biological Materials Working Group

The working group on Heterogeneity in Biological Materials developed a variety of focused projects that continue to the present. The projects were driven by applications to lung biology and biofilms, where mucus and related viscoelastic materials play vital functions. The most exciting outcome is the broad project on adapting the ideas of the immersed boundary method as a means to impose microstructure throughout a complex fluid. This investigation is not yet complete, but significant progress has been made.

Additional collaborative projects that arose from the working group and are being actively pursued include one on stochastic methods for diffusive transport, including inverse characterization from experimental data and mean passage time for anomalous diffusion, and a second on new numerical methods for heterogeneous biological media that merge the immersed boundary method and fluid solvers. These collaborations involve participants that are local and remote (the latter being Lisa Fauci, Robert Dillon, Isaac Klapper, and Mircea Grigoriu).

The working group on Heterogeneity in Biological Materials has a number of consequences to report:

- Based on a collaboration started at SAMSI between applied mathematicians, probabilists and statisticians, Scott McKinley (Duke), Lingxing Yao (Utah), Christel Hohenegger (NYU-Courant), Tim Elston (UNC), John Fricks (Penn State), and Gustavo Didier (Tulane) submitted an FRG to NSF-DMS on “Viscoelastic Diffusion”. The proposal is still pending.
- SAMSI Graduate RAs Ke Xu and Brandon Lindley (advised by Forest) both have published papers leading to their primary thesis results. Ke graduates August, 2009 and

she worked with Isaac Klapper (Montana State) while he visited SAMSI. Brandon graduated in May 08 and took a position at U. South Carolina to work on biofilms, a topic he was introduced to at SAMSI.

- Greg Forest and H. Zhou (Naval Postgraduate School) organized a mini-symposium at the SIAM annual meeting in San Diego this past summer 08 on research from the working group.
- Greg Forest and Qi Wang organized a minisymposium at the SIAM Computational Sciences and Engineering meeting in Miami, FL on complex fluids, attended by Lisa Fauci, Robert Dillon and grad students and postdocs from our working group.
- Mansoor Haider and Greg Forest followed up on their working group to organize a large mini-symposium at the regional AMS meeting at NC State on April 4-5, 2009, again attended by members of the SAMSI working group (Hohenegger, McKinley).
- Greg Forest, Brandon Lindley and Qi Wang organized a minisymposium at the regional SIAM meeting in Columbia, SC on April 4-5 on complex fluids, attended by members of the SAMSI working group.
- Weigang Zhong was introduced to the Immersed Boundary method in our working group. He contacted Boyce Griffith at NYU by recommendation of Lisa Fauci, and learned how to use the parallel IB code. His new job at Corning, Inc. is on problems that are proprietary, but related to the IB method.

Publications:

1. D.B. Hill, B. Lindley, M.G. Forest, R. Superfine, S. Mitran, *Experimental and modeling protocols for a micro-parallel plate rheometer*, UNC preprint, to be submitted.
2. C. Hohenegger, M.G. Forest, *Two-point microrheology, II: simulation protocols*, UNC-NYU preprint, to be submitted.
3. Scott A. McKinley, Lingxing Yao and M. Gregory Forest, *Transient Anomalous Diffusion of Tracer Particles in Soft Matter*, Duke-UNC preprint, to be submitted.
4. J. Fricks, L. Yao, T. Elston, M.G. Forest, *Time-domain methods for passive microrheology and anomalous diffusive transport in soft matter*, SIAM J. Appl. Math., Vol. 69(5), 1277-1308 (2009).
5. E. Howell, B. Smith, G. Rubinstein, M.G. Forest, B. Lindley, D. Hill, R. Superfine, S. Mitran, *Stress communication and filtering of viscoelastic layers in oscillatory shear*, J. Non-Newtonian Fluid Mechanics, Vol. 156, 112-120 (2009).
6. C. Hohenegger, M.G. Forest, *Modeling aspects of two-bead microrheology*, Proceedings of XVth International Congress on Rheology, Springer, August, 2008, AIP Conference Proceedings, Materials Physics & Applications Series, Vol. 1027 (2008).

7. C. Hohenegger, M.G. Forest, *Two-point microrheology: modeling protocols*, Phys. Rev. E 78, 031501 (2008).
8. S. Mitran, M.G. Forest, B. Lindley, L. Yao, D. Hill, *Extensions of the Ferry shear wave model for active linear and nonlinear microrheology*, J. Non-Newtonian Fluid Mechanics Vol. 154:120-135 (2008).
9. C. Hohenegger, M.G. Forest, *Direct and Inverse Modeling for Stochastic Data in Microbead Rheology*, Proceedings in Applied Mathematics and Mechanics (PAMM), Special Issue: Sixth International Congress on Industrial Applied Mathematics (ICIAM07) and GAMM Annual Meeting, Zrich 2007, Published Online: Oct 30 (2008).

4.2 Stochastic PDE Working Group

This working group held several meetings over the Fall of 2007, and Spring 2008.

The Stochastic PDE working group had weekly meetings and discussed joint collaborations and works on Random Field Theory and its applications in Communications and Image Classification. Specifically, the following topics were discussed and presented

- the existence of solutions to the stochastic heat and wave equations with non-Lipschitz but monotone nonlinearity and the temporal and special statistic properties of solutions based using the random field theory,
- the fiber communication system modelled by the randomly perturbed dispersion-managed nonlinear Schroedinger and using the corresponding soliton solutions,
- the interacting particle system and applications to network communications and data flow analysis.

The students associated to the working group worked on three specific projects. Over the Spring 2008 semester, presentations were given for each of these projects.

The first project entailed classification of a random surface based on the random pattern. The motivation comes from steel fabrication. When a sheet of steel is fabricated, it is often far from perfect. In flawed regions, the molecular arrangement may differ from the ideal steel regions. The molecular patterns of the steel appear to have very little structure and a homogenous appearance similar to noise. After viewing the random pattern present in the flawed steel, and the pattern present in the good steel, it becomes apparent that these regions have a different random pattern. The flawed steel appears to have a more heterogeneous mixing pattern than the good portions have. Leveraging this insight, we focused on discriminating between the two regions based on local covariance statistics then classifying based on classification trees. Of particular interest in the study are the vector autoregressive statistics.

The second project involved simulations of some interacting particle systems. To begin with, we simulated the Totally Asymmetric k-Exclusion Process. The dual of the results of the simulations were then used to test the theoretical upper and lower bounds in hopes of finding more precise bounds for the process.

The third project dealt with quantum probability theory, quantum filtering theory, and the stabilizing feedback control for quantum spin systems. Based on the quantum filtering theory,

our focus was to construct a stabilizing continuous feedback control for quantum filtering equation in quantum spin systems.

Publication: K. Ito et al., *Multi-valued Stochastic Evolution Equations in Hilbert Spaces and Integrable Solution*, in preparation.

4.3 Interface Problems Working Group

This group held regular meetings over the Fall of 2007, and Spring 2008. A web page

<http://www.samsi.info/200708/ranmedia/wg/het-random/if-index.html>

describes the topics covered and some presentations for the working group:

- Introductions of the boundary integral method and level set method
- Introductions of the level set method
- Immersed interface method
- Kernel-free boundary integral method
- Grid-based particle method for moving interface problems
- Problems with incompressible interfaces
- Fluid mixed model of tissue deformations
- Modified bilinear interpolation and FEM for an elliptic interface problem

The working group worked on moving interface free boundary problems. Different ideas and approaches, such as boundary integral method, level set method, immersed interface method, immersed boundary method, and other related topics were thoroughly examined and assessed. The weekly group meetings were very interactive and candid. New collaborations and new ideas were generated. For examples, new methods based on combining different approaches to complement each other's strengths and weaknesses have been proposed and are going to be implemented. The current project including the grid based particle method for moving interface free boundary problems; numerical methods and models for incompressible membranes with bending.

Much of the research focused on analysis of fundamental questions of fluid motion and design and analysis of numerical methods for fluids, and especially methods for problems with interfaces. The Working Group on Interface Problems connected directly with several research interests. Because several of participants had overlapping expertise, we had a great deal to discuss in detail in understanding the advantages and limitations of existing methods and how to push them further toward more realistic problems. However, Dr. T. Beale's expertise is weighted more toward analysis, as opposed to practical computational methods, in comparison with other active participants. It has been valuable to the participants to learn better what is currently being done, what works well in practice, and what does not. Conversely, some analytical point of view contributed to qualitative understanding of behavior of numerical methods, especially the qualitative nature of errors.

Publications:

1. Jun Wang, Qin Cai, Zhilin Li, Hong-Kai Zhao, and Ray Luo, Achieving energy conservation in PoissonBoltzmann molecular dynamics: Accuracy and precision with finite-difference algorithms, *Chemical Physics Letters*, Volume 468, Issues 4-6, 22 January 2009, Pages 112-118.
2. K. Ito, M. Lai, and Zhilin Li, A well-conditioned augmented system for solving Navier-Stokes equations in irregular domains, *J. Comput. Phys.* (2009), doi:10.1016/j.jcp.2008.12.028.
3. X. Wan, Z. Li, and S. Lubkin, Mechanics of mesenchymal contribution to clefting force in branching morphogenesis, *Biomechanics and Modeling in Mechanobiology*, Vol. 7, 417-426, 2008.
4. H. Xie, K. Ito, Z. Li, J. Toivanen, A finite element method for interface problems with locally modified triangulations, *AMS Contemporary Mathematics*, Vol. 466, 2008, 179-190.
5. Q. Jiang, Z. Li, and S. Lubkin, Theoretical & numerical analysis for a fluid mixture model of tissue deformation, *Comm. in Comput. Phy.* Vol. 3, 620-634, 2009.

4.4 Waves and Imaging Working Group

This group held regular meetings over the Fall of 2007 with activities summarized on the webpage <http://www.samsi.info/200708/ranmedia/wg/imaging-random/imaging-index.html>

In the group on Waves and Imaging, a range of collaborations were established. Yvonne Ou from U. Central Florida teamed with Jean-Pierre Fouque and Josselin Garnier to investigate the problem of time-reversal for elastic waves. They started with a review of the literature in the group meeting very soon. Gabriel Peyre from U. Paris-Dauphine teamed up with Laurent Demanet to investigate methods of compressive wave computations, with application to migration. Sava Dediu from NCSU joined with Laurent Demanet to study an optimal-transport approach to the problem of model velocity estimation in one-dimensional seismic inversion. All three collaborations could not have been initiated without the support of SAMSI, and actively benefit from the teleconferencing capabilities that the working room offers. The research produced by the weekly group meeting were the basis for the two discussion sessions in the "Waves and Imaging" workshop.

One of the outcomes of the "Waves and Imaging" working group is the collaboration between Laurent Demanet and Gabriel Peyre on "Compressive wave computation", a novel method for efficiently solving wave equations in the context of inverse problems in seismology. The backdrop for this effort was the group meeting's extensive discussion on nonlinear sampling strategies in imaging, including compressed sensing, during the Fall of 2007. What became apparent is that the ideas of sparsity and undersampling suggest an entirely different strategy for simulating linear wave phenomena on a large computational scale, using nonlinear synthesis from a few eigenfunctions of the Helmholtz operator, chosen at random. The main mathematical question concerned the number of such eigenfunctions needed for a given accuracy guarantee, and was solved during the random media program. Under mild assumptions, the answer is a remarkable $O(\log(N))$ where N is the desired resolution. Gabriel Peyre's visit in November benefited from generous SAMSI funding and was instrumental in establishing

the numerical validity and applicability of this result. In March 2008, the project reached a first milestone with the completion of a preprint treating the one-dimensional case. More collaborators will join our effort as the potential impact of this discovery in reflection seismology is now clear: the compressive viewpoint yields embarrassingly parallel algorithms that promise to help rethink the main computational bottlenecks of adjoint-state methods on large CPU clusters. The inception of this project would not have been possible without the SAMSI Random Media Program and the focus it provided. In fact we would probably not even have thought of starting this project were it not for the opportunity provided by the SAMSI program. Finally, Ray Luo, who is a faculty member in Molecular Biology & Biochemistry at UCI, was invited to the moving interface workshop during the SAMSI program. As part of the program, Luo, Zhilin Li and Hongkai Zhao initiated a project on protein folding mechanism and structure prediction. This work is ongoing and covers the application of efficient numerical methods to study biomolecular structures, functions, and intermolecular interactions at atomic detail and as well as the application of the methods under construction to understand and predict the relations between the sequences, structures and functions of these molecules.

Publications:

1. Laurent Demanet and Gabriel Peyre, *Compressive Wave Computation*, submitted, 2008.
2. Semyon Tsynkov, *On SAR imaging through the Earth Ionosphere*, SIAM Journal on Imaging Sciences, 2 (2009) No. 1, pp. 140–182.
3. Shingyu Leung and Hongkai Zhao, *A New Grid-Based Particle Method for Interface Problems*, Journal of Computational Physics, Volume 228, Issue 8, 2009.
4. Shingyu Leung and Hongkai Zhao, *A Grid Based Particle Method for Evolution of Open Curves and Surfaces*, UCLA-CAM 08-72. Submitted.
5. Jun Wang, Qing Cai, Zhilin Li, Hongkai Zhao, and Ray Luo, *Achieving Energy Conservation in Poisson-Boltzmann Molecular Dynamics: Accuracy and Precision with Finite-Difference Algorithms*, to appear in Chemical Physics Letters.
6. Qin Cai, Jun Wang, Hongkai Zhao, and Ray Luo, *On Removal of Charge Singularity in Poisson-Boltzmann Equation*, to appear in Journal of Chemical Physics.

5 Workshops

5.1 Opening Workshop

The Opening Workshop for the SAMSI program on Random Media was held Sunday-Wednesday, September 23-26, 2007, at the Radisson Hotel RTP in Research Triangle Park, NC. It was preceded, on Sunday, September 23, with tutorials by Eric Vanden-Eijnden (NYU), and Jack Xin (UC- Irvine).

The goal of the opening workshop was focused on the formulation of challenges and directions to be pursued during the Random Media Program. Focus areas during the program included the following topics: time reversal, interface problems, imaging problems, scattering

theory, heterogeneity in biological media, and porous media. During the workshop, several working groups for the program were formed to promote engagement (via web or teleconference) of those who will not be in residence at SAMSI during the program. The workshop engaged a broadly representative segment of the mathematical, statistical and disciplinary sciences.

The workshop was organized by Maarten De Hoop (Purdue University), Zhilin Li (North Carolina State University), Ralph Smith (SAMSI, Directorate Liaison), Hongkai Zhao (UC Irvine).

The workshop included a number of distinguished speakers and young researchers: John Cushman (Purdue University), Weinan E (Princeton University), Bjorn Enquist (Univ. of Texas-Austin), Lisa Fauci (Tulane University), Jean-Pierre Fouque (Univ. of California, Santa Barbara), Tom Hou (California Institute of Technology), Sam Kou (Harvard University), Karl Kunisch (University of Graz), Randy LeVeque (University of Washington), John Lowen-grub (Univ. of California-Irvine), Stanislav Molchanov (UNC Charlotte), Gretar Tryggvason (Worcester Polytechnic Institute), Gunther Uhlmann (University of Washington), Wojbor Woyczynski (Case Western Reserve University); and young researchers: Karen Daniels (NC State), John Fricks (Penn State), Lucy Zhang (Rensselaer Polytechnic Institute), Lucy Zhang (Rensselaer Polytechnic Institute).

During the opening workshop, two panel discussions were conducted. The first one was on interface problems chaired by Gretar Tryggvason (WPI) and Bjorn Enquist (UT Austin). The second one was on time reversal, Stochastic PDEs, and imaging, chaired by Jean-Pierre Fouque (UC Santa Barbara) and Maarten De Hoop (Purdue). A first iteration on the working groups was made. After discussions before the end of the workshop, a list of working groups was formed and the participants signed up for groups of interest. There was an extraordinary response to the working group call, with almost all of the workshop participants remaining for the working group formation.

5.2 Interface Workshop

The interface workshop was held on November 15-16, 2007 at the Radisson Hotel RTP in Research Triangle Park, NC.

The theme of the workshop focused on interface problems. In many science and engineering problems, multiphase systems that involve moving interface and free boundary are quite challenging for both mathematical analysis and numerical simulations. One of the main difficulties is the coupling of the evolution and geometry of the interface with the global dynamics of the bulk. The coupling is often nonlinear and non-local. Singularities, such as discontinuity of material properties and physical quantities across the interface, and topological changes, such as merging and pinch-off, occur during the evolution. Further complications, such as surface diffusion, random media, and multiple scales, can make the problem even more challenging.

Recently there has been significant progress in both theory and numerical methods for moving interface problems. In this workshop, experts from different backgrounds will address aspects of modeling, theory, numeric and applications and their integration. The emphasis in the workshop is to foster discussions, collaborations, identification of new problems in a cross-disciplinary setting, concentrating on numerical methods, analysis, modeling, and applications of interface problems. The workshop was organized by Zhilin Li (NCSU), Ralph

Smith (SAMSI, Directorate Liaison), and Hongkai Zhao (UC-Irvine).

The speakers of the workshop included: Shi Jin (University of Wisconsin-Madison), Jon Wilkening, (University of California-Berkeley), Mark Sussman (Florida State University), Ray Luo (University of California-Irvine), Sigal Gottlieb (University of Massachusetts, Dartmouth), Ping Lin (University of Dundee & National University of Singapore), Patrick Guidotti (University of California-Irvine), J. Thomas Beale (Duke University), Hongkai Zhao (University of California-Irvine), John Strain (University of California-Berkeley), Robert Dillon (Washington State University), Guowei He (Iowa State University), David Chopp (Northwestern University), Richard Tsai (University of Texas-Austin), and Alina Chertock (North Carolina State University).

5.3 Waves and Imaging Workshop

The Waves and Imaging workshop was held on January 31 and February 1, 2008, at the Radisson hotel in Research Triangle Park, NC. A few new approaches have been recently proposed to solve the challenging problems of imaging and inversion from wave measurements, most notably in geophysics and optics. A first example is time reversal, where flipped waveforms sent back into a random medium refocus an order of magnitude better than they would in a uniform medium. A second example is cross-correlation of seismic noise, a procedure that produces the entire Green's function of surface waves from passive receivers. A third example is compressive reverse-time migration where ideas from compressive sampling bring the computational complexity of migration down to the information level of seismic wave fields. The explanation and prediction of all these phenomena stem from some surprising results of statistical stability and probability concentration, which are currently being researched by several groups. The main objectives of this workshop are to: (1), review the extent to which these imaging methods have been developed and understood; (2) expose the progress made in the working group, and (3) discuss open problems and future directions. The workshop was organized by Laurent Demanet (Stanford), Maarten de Hoop (Purdue), Kazufumi Ito and Zhilin Li (NCSU).

The speakers of the workshop included: Margaret Cheney (Rensselaer Polytechnic Institute), Gang Bao (Michigan State University), Yu Chen (New York University), Richard Weaver (University of Illinois at Urbana-Champaign), Luis Tenorio (Colorado School of Mines), Lenya Ryzhik (University of Chicago), Josselin Garnier (Université de Paris VI), Knut Solna (University of California-Irvine), Liliana Borcea (Rice University), William Symes (Rice University), Henri Calandra (Total Corporation), John Schotland (University of Pennsylvania).

5.4 Transition workshop:

This will be held on May 1-2, 2008, at the Radisson hotel in Research Triangle Park, NC. The workshop was organized by Maarten de Hoop (Purdue University), Zhilin Li (North Carolina State University), Ralph Smith (North Carolina State University, SAMSI Directorate Liaison), Hongkai Zhao (UC-Irvine). The goals of this workshop were to

1. Present results generated by this SAMSI program to the applied mathematics, statistics and engineering communities.

2. Formulate follow-up plans for this SAMSI program to continue research and education in this interdisciplinary area.

Several of the speakers presented overview talks about the projects spawned during the program and the significant challenges that remain. For instance, new numerical methods for heterogeneous biological media that merge the immersed boundary method and fluid solvers were discussed. Exciting novel numerical techniques for interfacial free boundary problems involving viscous fluids were also exposed and discussed. Examples of these include hybrid numerical methods that incorporate a separate analytical reduction of the dynamics within the transition layer into a full numerical solution of the interfacial free boundary problem.

The speakers of the workshop included: Greg Forest (University of North Carolina, Chapel Hill), Kazufumi Ito (North Carolina State University), Min Kang (North Carolina State University), Chiu-Yen Kao (Ohio State University), Taufiqar Khan (Clemson University), Isaac Klapper (Montana State University), Anita Layton (Duke University), John Lowengrub (University of California, Irvine), Li-Shi Luo (Old Dominion University), Michael Siegel (New Jersey Institute of Technology), Jason Wilson (Duke University).

6 Education and Outreach

6.1 Credit Courses

The Program offered one 3 credit course in the 2007 Fall semester. The title of the course was “Numerical Methods for Free Boundary and Moving Interface Problems” and the instructors were Kazufumi Ito (NCSU), Zhilin Li (NCSU), and Hongkai Zhao (UC Irvine). Nine students registered in this class including four females. There were about four additional postdocs from SAMSI and NCSU who audited the class.

6.2 SAMSI Two-Day Undergraduate Workshops

February 29-March 1, 2008 at SAMSI. Twenty four undergraduate students from undergraduate colleges and universities across the nation participated in this workshop.

In the workshop, K. Ito (NCSU) presented two lectures on respectively “Level Set Method and Applications” and “Central Voronoi Tessellation and Applications.” H. Zhao (UCI) gave two lectures as well on “Wave Propagation” and “Imaging Using Waves.”

The workshop exposed the students to the idea of mathematical models and their numerical computer implementation, in a wide variety of scenarios and at a level adequate for the wide range of students present. Hands-on computer tutorials helped students grasp the basics of the level set method, wave propagation in random media, and imaging process. Significant emphasis was put on open and often spirited discussions. The workshop was very well attended with students from all over United States. The Workshop accomplished the goals of exposing and interesting a wide diversity of bright students to the area of Random Media, their development, assessment and utilization.

6.3 Graduate students.

The Program contributed to the achievements, education, and Ph.D. projects of many graduate students.

Brandon Lindley (UNC) Brandon Lindley was introduced to biofilms through his participation to the SAMSI program. Brandon graduated in May 08 and took a position at U. South Carolina to work on that topic.

Jason Wilson (Duke) was involved in the interface working group. Jason Wilson is a graduate student at Duke working toward his Ph.D. He was supported by SAMSI for the fall semester, 2007. He took the course in the fall on free boundaries and moving interfaces. His thesis focuses on the construction of overlapping coordinate grids with low distortion on a given, smooth, closed surface in three dimensions. His work has applications to boundary integral methods. While presenting some similarities with the work of Shing-Yu Leung and Hongkai Zhao, Jason's method uses a more detailed representation of the surface which may be of advantage depending on the application.

Ke Xu (UNC) was involved in the heterogeneity in biological media working group. Ke graduates in August, 2009 and she worked with Isaac Klapper (Montana State) while he visited SAMSI. She took the SAMSI course MA581 in the fall 2007 on free boundaries and moving interfaces. She spoke in the working group several time about her research and relation with the SAMSI program.

Hui Xie (NCSU) was involved in the interface working group. Hui Xie is a graduate student at NCSU working toward his Ph.D. He took the SAMSI course MA581 in the fall, 2007 on numerical methods for free boundaries and moving interfaces. He presented a talk in the interface working group. His talk was about the finite element method with a locally modified triangulation for the elliptic interface problems.

Qin Zhang (NCSU) was active in two working groups at SAMSI. One is the working group on Interface Problems. He took the SAMSI course MA581 in the fall 2007 on free boundaries and moving interfaces. He is also one of participants of the SAMSI working group on stochastic PDE. He gave a presentation, titled "Optimal Bilinear Control on Quantum Systems," in the SAMSI Postdoc/Graduate Students Seminar. He also presented a talk on the quantum probability theory and quantum filtering problems in the working group. His thesis topic concerns finding a stable feedback solution for quantum control problem arisen in quantum spin systems under continuous measurement which is closed related to the SAMSI program.

6.4 Efforts Made toward Achieving Diversity

There was a significant percentage of women, minority and new faculty throughout the year long program, which can be seen from the list of speakers and participants.

The invited speakers in the Opening Workshop included Lisa Fauci, Karen Daniels and Luci Zhang where the latter two are new faculty. The invited speakers in the Interface Workshop included Sigal Gottlieb and Alina Chertock.

The core participants in the “Waves and Imaging” group meeting included one minority (Daniel Alfaro) and one woman (Yvonne Ou). The “Waves and Imaging” workshop on Jan 31 and Feb 1 featured two women speakers (Margaret Cheney and Liliana Borcea), one minority speaker (Luis Tenorio), and one speaker from industry (Henri Calandra from Total, France). The attendance of the workshop also included a few more minorities, women, and industry researchers.

Appendix C: Final Report of the Program on Environmental Sensor Networks

1 Introduction

The core purpose of the SAMSI Program on Environmental Sensor Networks is to identify research challenges and opportunities in the use of wireless environmental sensor networks to address critical contemporary problems. They include understanding the effects of global climate change, human activity, and invasive species on ecosystem function, and drive the our need to understand the dynamics of diverse environmental phenomena and their causes. This problem domain offers unique interdisciplinary research challenges. First, the labor cost of deploying and maintaining these networks is very high, which is limiting adoption of the technology. Secondly, uncertainty is dominant, with noise, numerous failure modes, and over/under-sampling issues driven by conflicts between the needs of network connectivity and spatial design for processes of interest. These problems are compounded by inherent issues of dimensionality and scale: datasets for the biological and physical problems of interest consist of sampled multivariate spatio-temporal process with natural scales ranging from minutes to decades and meters to hundreds or thousands of kilometers.

2 Program Organization

A remarkable characteristic of this program is the diversity of disciplines represented in the participants of the Opening Workshop and both Working Groups. Researchers in ecology, computer science, mathematics, electrical and computer engineering, and environmental engineering are working with statisticians specializing in, among other fields, experiment design, sampling techniques, linear models, spatial statistics and hierarchical Bayesian methods.

The program was led by Paul Flikkema (Northern Arizona University) who was in residence at SAMSI during January - May 2008. Two *Working Groups* were formed, whose principal functions were to identify, organize, and nurture collaborative research initiatives. The majority of participants were from outside the Triangle area.

The **Sensor Networks Datasets** working group led by Paul Flikkema (Northern Arizona University) included: Ankit Agarwal (University of Kansas), David Bell (Duke University), Michela Cameletti (SAMSI/Bergamo University), Zoe Cardon (Ecosystems Center, Marine Biological Laboratory), Jim Clark (Duke University), Alan Gelfand (Duke University), Scott Holan (University of Missouri), Rosaria Ignaccolo (SAMSI/Universita' di Torino), Natallia Katenka (University of Michigan), Yongku Kim (SAMSI/Duke), Ernst Linder (University of New Hampshire), Kristian Lum (Duke University), John McGee (UNC Chapel Hill, Renaissance Computing Institute), Yajun Mei (Georgia Institute of Technology), George Michailidis (University of Michigan), Long Nguyen (SAMSI/Duke), Michael Porter (SAMSI/NCSU), Ilka Reis (National Institute for Space Research, Brazil), Karl Rohe (University of California at Berkeley), Sande Satoskar (RENCI), Lance Waller (Emory University), Kim Weems (NCSU), and Bin Yu (UC Berkeley).

The **Sensor Design** Working Group, led by James S. Clark (Duke University) and Jun

Yang (Duke University), included: Ankit Agarwal (University of Kansas), David Bell (Duke University), Michael Breen (EPA), Michela Cameletti (SAMSI/Bergamo University), Zoe Cardon (Ecosystems Center Marine Biological Laboratory), Jorge Cortes (UC San Diego), Jessica Croft (University of Utah), Todd Dawson (UC Berkeley), Carla Ellis (Duke University), Marco Ferreira (University of Missouri), Paul Flikkema (Northern Arizona University), Jeff Frolik (University of Vermont), Alan Gelfand (Duke University), Joe Fred Gonzalez, Jr. (Center for Disease Control), Scott Holan (University of Missouri), Sheryl Howard (Northern Arizona University), Rosaria Ignaccolo (SAMSI/Universita' di Torino), Chris Jones (University of North Carolina) Yongku Kim (SAMSI/Duke University), Hamid Krim (North Carolina State University), Soumen Lahiri (Texas A&M University), David Leslie (Bristol University), Kristian Lum (Duke University), Yajun Mei (Georgia Institute of Technology), George Michailidis (University of Michigan), Long Nguyen (SAMSI/Duke University), Neal Patwari (University of Utah), Michael Porter (North Carolina State University), Ilka Reis (National Institute for Space Research, Brazil), Christine Shoemaker (Cornell University), Bin Yu (UC Berkeley), Yi Zhang (Duke University), and Zhengyuan Zhu (University of North Carolina).

3 Achieving Diversity

The program has had strong participation by female faculty, post-doctoral researchers, and students. Zoe Cardon (Marine Biological Laboratory) and Deborah Estrin (UCLA) serve on the Program Leaders Committee. Estrin, Jennifer Hoeting (Colorado State University), and Kiona Ogle (University of Wyoming) contributed invited presentations at the Opening Workshop. Carla Ellis (Duke University) organized the Fall 2007 SAMSI graduate course on Environmental Sensor Networks.

Participating faculty and researchers include Michela Camelleti, Sheryl Howard, Rosaria Ignaccolo, Cari Kaufman, Christine Shoemaker, Kimberly Weems, and G. Beate Zimmer.

Zoe Cardon (Ecosystems Center, Marine Biological Laboratory) was a leader of the Sensor Networks Datasets Working Group, and provided a critical dataset that will be used in papers now in preparation as well as crucial support in both the interpretation of metadata and development of models.

Participant **Rosaria Ignaccolo**, a SAMSI New Researcher and an Assistant Professor in the Department of Statistics and Applied Mathematics at the Universita' degli Studi di Torino, was an active participant in the Sensor Data working group, and was a key contributor to cleaning and exploratory data analysis of the group's datasets. She has been interested in exploring functional data analysis methods for ecological data and also presented a talk "Functional Analysis and Clustering with Spline Libraries" at the Transition Workshop.

Participant **Sheryl Howard**, an assistant professor in electrical engineering at Northern Arizona University, attended both the Opening and Transition Workshops, and presented an invited talk at the Transition Workshop entitled "Coded Compressive Estimation in Environmental Sensor Networks." Her work led to an NSF grant award for her proposal "BRIGE: Energy-Efficient Communication with Combined Decoding/ Inference". It has also resulted in a Science Foundation Arizona graduate fellowship award for her student Rui Chen, and she is now supporting two undergraduate students (Forrest Schwynn and Hristo Taralov) and one female graduate student (Fauzia Ahmed). She also presented a talk "Combined Source-Channel

Decoding and Transmission Censoring for Power Reduction in a Wireless Sensor Network” at the 2008 International Analog Decoding Workshop (Logan, UT, July 12, 2008). Her student Rui Chen collaborated with Flikkema’s student Saiyi Wang (also awarded a SFAz graduate fellowship) on a poster ”Energy Efficiency in Environmental Sensing Networks: Cross-Layer Approaches for Transmission Censoring” at the Science Foundation Arizona Graduate Research Fellows Grand Challenge Summit, March 27-29, 2009.

Graduate students include Christina Bentrup (Northern Arizona University), Jessica Croft (University of Utah) and Natallia Katenka (University of Michigan), Kristian Lum (Duke University), and Ilka Reis (National Institute for Space Research, Brazil).

Participant **Kristian Lum** is a Ph.D. student in statistics at Duke University. She participated in the Fall 2008 Sensor Networks for Environmental Modeling Course as well as the Opening and Transition workshops. For her preliminary exam in April 2008, she studied how inference is degraded with decreased transmission rates for various models of the sensed data and inference schemes. More recently, she has been analyzing transmission suppression schemes based on approximating the dynamics of the sensed data with linear temporal models, as well as stochastic differential equation models using Ornstein-Uhlenbeck processes.

4 Research Progress

Both Working Groups conducted weekly distributed meetings throughout the program period. Meeting schedule information, notes, presentation slides, reading lists, and participant directories are all available on-line on the SAMSI website. Both groups have focused on cross-disciplinary challenges involving statistics, applied mathematics, engineering, and computer science that are driven by important ecological questions and the characteristics of wireless sensor networks. Each Working Group developed and pursued a detailed research agenda, as outlined in the following.

4.1 Sensor Networks Datasets

Because this group brought together researchers from a very broad set of disciplinary perspectives, our weekly telemeetings were initially dedicated to two types of discussions: (i) short, informal talks and discussions by all working group members about their backgrounds and research interests, and (ii) discussions on research perspectives of the fields represented by the working group members.

Since the group felt strongly that research should be grounded in knowledge of issues that occur in actual experiments and real datasets, these discussions ran concurrently with the process of acquisition and evaluation of example datasets. Unlike in other disciplines, such as bioinformatics, datasets from sensor arrays or wireless networks are very new and extremely rare. They are also very unwieldy, with diverse variables, different sampling intervals, and numerous faults of varying types and severity. Furthermore, necessary physical conversions require interaction with other sensed variables, coupling and propagating uncertainty.

Early on, the group identified three prerequisites for a successful research agenda that is grounded in real data: the dataset should be sufficiently rich in terms of its statistical properties and association with relevant research in ecology or environmental sciences; we

should have sufficient knowledge of the data collection process used; and we should be able to closely interact with a scientist familiar with the experiment and dataset. With these in mind, we studied two datasets:

- Zoe Cardon (Ecosystems Center, Marine Biological Laboratory) presented a dataset for an experiment based on measurement of *water potential* at multiple sites around sagebrush plants in Utah. This experiment was designed to further understanding of soil microbial activity as a function of water in the soil. This dataset is of interest in part because it is from a wired sensor array, and thus is an excellent vehicle for exploring the effects of posited wireless networks. One of the important lessons of this dataset is that, in wireless sensor networking, a rich spectrum of errors and faults will occur regardless of, and in addition to, whatever effects wireless networking may have on the data.
- The group has also assessed a *root structure/soil respiration* dataset from the UCLA Center for Embedded Networked Sensing that used a wireless sensor network with the goal of characterizing the spatio-temporal properties and regulation of soil moisture. The network monitored dynamics of soil respiration, soil moisture, and fine root and rhizomorph (fungi) structure using mini-rhizotrons, with the objectives of understanding ecological processes related to the coupling between soil moisture and fine root and rhizomorph dynamics.

In broad terms, the goal of the working group was to answer the question: *How, and how well, can we answer important and inherently statistical questions in the ecological sciences with data from wireless sensor networks, and how do networks affect our ability to answer those questions?* The specific research questions that fall under this umbrella are challenging; they typically involve multiple, coupled dynamical processes with latent variables, as well as issues of scaling and dimensionality. With respect to wireless sensor networks, the group is working to develop approaches to the crucial open question of modeling energy consumption and efficiency in wireless data gathering networks.

The working group's research plan coalesced around two thematic areas. The first is *Data Analysis*, motivated by the following questions that are relevant to the water potential dataset, but represent challenges found in a wide spectrum of research questions in ecology and environmental science:

- Can we assume that the rate of water loss from deeper soil layers (via transpiration) is proportional to conductance across the soil-root interface?
- Is there a change in the relationship between root-soil conductance for water and soil water potential, e.g., as the season progresses?
- Is the amplitude of daily oscillations in soil water content a driver of soil biogeochemical processes that affect plant root function and growth?

A second theme that we developed is *System-Data-Network Interaction* to explore statistical tools that can be used to address questions such as:

- How do faults and errors interact with various network algorithms to affect our ability to answer ecological questions across time scales?

- How does the energy cost of computation at sensor nodes factor into decisions about network signal processing, inference, coding, and transmission algorithms, given the panoply of errors and faults that may occur?
- How can the model-mediated gathering of data be tuned, based on its explanatory power, given the energy requirements of sampling, computation, and data transmission processes?

The Sensor Datasets Working Group self-organized into two subgroups along the Data Analysis and System-Data-Network Interaction themes, and, in telecon meetings, zeroed in on two problem areas.

The Data Analysis subgroup tackled the above questions in the context of aridland soil-plant-air systems. A unique aspect of this work is inference of hydraulic redistribution and its drivers, the latter including variations of plant-air conductivity in the summer monsoon season. A paper in preparation will compare Bayesian and classical inference techniques for capturing the relative effects of water potential gradients among plants, the atmosphere, and soil at different depths.

The Data Analysis subgroup also addressed automated detection of anomalous data from sensor networks, where Ernst Linder has led the development of algorithms based on the median polish. As part of this work, he advised Jared Murray (Undergraduate Student, Department of Mathematics & Statistics, U. of New Hampshire), on an undergraduate research project (Fall 2008-Spring 2009) to develop an interactive graphical tool called MP-TUNER for automated anomaly detection for multiple time series data from environmental sensor network. Software is currently under development and almost finished, and can be accessed at <http://pubpages.unh.edu/jsb28/>.

There are two specific thrusts in the In System-Data-Network Interaction area. In the first (Howard and Flikkema), we are studying how to couple network-aware source coding, channel coding, transmission control, and Bayesian source-channel inference (a generalization of MAP decoding) at the destination node. This work focuses on the trade-off of uncertainty reduction and energy consumption rather than focusing on information rate, since channel capacity is not limiting in this environmental sensing application. Published results address both global inference at the information sink and a form of in-network cooperative communication wherein nodes use local information to make communication decisions based on prediction of the consequences of candidate decisions using a Bayesian framework. The second thrust (Flikkema with Undergraduate Student Kenji Yamamoto, EE Dept., Northern Arizona U.) is addressing the gap in understanding between theoretical results and practical implementation of in-network algorithms on energy-limited sensor nodes. For example, if the energy cost (stemming from computational complexity) of inference algorithms is too high, it may exceed the reduction in communication energy cost enabled by those algorithms. A real-time power/energy measurement system has been designed, developed, and is under test that will provide accurate and precise estimates of the energy cost of algorithms running on sensor nodes, where electrical current demands vary over five orders of magnitude in both magnitude and time scales.

4.2 Sensor Design

During the first two meetings (Jan. 24 and 31), the working group focused on studying specific ecological applications of wireless sensor networks, in order to better understand the needs of the ecology researchers. The two applications studied were the redwood tree monitoring project of Todd E. Dawson (Integrative Biology, UC Berkeley), and the Duke Forest monitoring project of James S. Clark. Neal Patwari (Electrical Engineering, U. Utah), David Bell (Environmental Science, Duke U.), and Yongku Kim (Statistics, SAMSI) led the discussions.

In the third meeting (Feb. 7), XuanLong Nguyen (Statistics and Computer Science, SAMSI) gave an overview and survey of *suppression* and related techniques in distributed systems and sensor networks. By exploiting redundancy that naturally arise in sensor data, these techniques reduce the amount of data that needs to be communicated to a gateway or base station for collection, thereby conserving energy and prolonging the lifetime of the sensor network deployment.

The next three meetings (Feb. 14, 21, and 28) were devoted to a series of roundtable discussions, wherein each participant prepared a couple of ideas of potential interest to the working group and led the group discussion on these ideas. Paul Flikkema (Electrical Engineering, Northern Arizona U.) talked about joint coding, estimation, and transmission censoring. Marco A. R. Ferreira (Statistics, U. Missouri, Columbia) presented a Bayesian decision-theoretic setup for tackling sensor design problems. Scott Holan (Statistics, U. Missouri, Columbia) proposed looking at adaptive sampling and design problem, and studying models of network failure. Jun Yang (Computer Science, Duke U.) argued for reducing the total maintenance cost of the network instead of total energy consumption, and applying model-driven techniques to the system health monitoring of the network itself. Zhengyuan Zhu (Statistics, UNC Chapel Hill) discussed opportunities of improving data collection efficiency using spatio-temporal sampling design. Neal Patwari and Jessica Croft (Electrical Engineering, U. Utah) proposed considering adaptive deployment and survival strategies for the network. Yongku Kim talked about challenges in statistical analysis, suppression scheme design, and dynamic models. Michela Cameletti and Rosaria Ignaccolo (Statistics, SAMSI) discussed their experience with the Piedmonte PM10 monitoring network and problems in adaptive sampling and model- and entropy-based network design. XuanLong Nguyen presented two specific problems: study of data reduction vs. statistical efficiency of suppression schemes, and sensor selection driven by a spatial model. Ilka A. Reis (Statistics, SAMSI) talked about design of better temporal suppression schemes. Christine Shoemaker (Environmental Engineering, Cornell U.) presented her project on monitoring Cannonsville Reservoir Basin, and challenges in efficient simulation and assessing uncertainty in simulation models. James S. Clark elaborated on the idea of model-based data suppression using soil moisture data as an example.

In the meeting on Mar. 6, Jun Yang summarized the main threads among the problems of interest presented during the roundtable discussions. The group decided to focus on two specific design problems:

- *Design and analysis of data collection schemes.* Given a time series of raw readings, the system can employ a variety of techniques to save communication (and hence energy): a) randomly transmit a reading with some probability; b) transmit only readings that differ by more than ϵ from their predicted values; c) quantize each reading and

only transmit the quantized value if it is different from the last transmitted reading; and d) compress the readings and then transmit the compressed data. Although there has been a lot of work based on these ideas, more rigorous analysis is needed in order to quantify the cost/benefit tradeoff among them and to better understand their relationships and differences. Once formal definitions of cost and benefit are chosen, the design problem involves choosing the best data collection scheme and the optimal parameter setting for it (e.g., probability of transmission, value of ϵ , level of quantization, or compression method).

- *Spatial sensor deployment design.* Given a spatio-temporal model that one wishes to learn using a collection of sensors, where should the sensors be placed to achieve the desired cost/benefit target? While the general experimental design problem has been studied extensively in statistics, the traditional cost models and design constraints are probably inappropriate in the sensor network setting. With cost models and constraints unique to sensor networks, the problem will be a novel and interesting one.

In preparation for tackling the above two problems, the working group reviewed the necessary background knowledge. On Mar. 13, Neal Patwari, representing the electrical engineering perspective, presented models for path loss, interference, and battery power in wireless sensor networks. On Mar. 20, Zhengyuan Zhu, representing the statistical perspective, gave an overview of known results on spatial sampling design.

After these meetings and presentations, subgroups of the Sensor Design Working Group were formed to focus on specific research problems. These subgroups include *sampling/routing design subgroup*, *suppression design subgroup*, and *review article subgroup*.

The *sampling/routing design subgroup* consists of XuanLong Nguyen (statistics and computer science, SAMSI), Jun Yang (computer science, Duke U.), Yi Zhang (computer science, Duke U.), and Zhengyuan Zhu (statistics, U. North Carolina at Chapel Hill). This group is working on jointly designing data sampling and network routing strategies for environmental wireless sensor networks. Traditionally, these two aspects of the design problem have been tackled separately: sampling design (to achieve specific modeling goals) mostly has been the concern of the statistics community, while energy-efficient routing is the focus of the computer science community. Obviously, a truly optimal design must tackle both these aspects, because both sampling and routing have large impact on the consumption of energy, often the most precious resource on battery-powered sensor nodes. The sampling/routing design subgroup is now tackling the design problem by jointly considering these two aspects.

The sampling/routing design subgroup has been meeting regularly since April 2008, and has made considerable progress. Zhengyuan Zhu is now leading the effort in writing an article to summarize the subgroup's findings.

A sizable subset of the Sensor Design Working Group participated in the transitional workshop in October 2008. On behalf of the sampling/routing design subgroup, Zhengyuan Zhu (statistics, U. North Carolina) summarized the findings on the optimal joint design of data sampling and message routing and in wireless networks. Members of the Sensor Design Working Group present at the workshop also discussed possible next steps for the working group.

The sampling/routing design subgroup also sought feedback from and collaboration with the DDDAS team at Duke U. and Northern Arizona, who has been working on deploying

sensors in the Duke Forest to study forest growth. The DDDAS team already includes many active members of the SAMSI working groups: James S. Clark, Paul Flikkema, Alan Gelfand, Kristian Lum, XuanLong Nguyen, Jun Yang, and Yi Zhang. Zhengyuan Zhu joined one of the DDDAS project meeting in October 2008 and discussed the possibility of applying the results of the sampling/routing design subgroup in the practical context of the DDDAS project.

The *suppression design subgroup* consists of Kristian Lum (statistics, Duke U.), Jun Yang, and Yi Zhang. This subgroup is interested in the design of *suppression* schemes, which is a way to reduce communication (and therefore save energy) in sensor networks by using predictive models to suppress reporting of predictable data. However, in the presence of communication failures, missing data is difficult to interpret because it could have been either suppressed or lost in transmission. To date, there has been no solution for handling failures for general, spatio-temporal suppression that uses *cascading*, where a node can use suppression in reporting its readings to another node, which can then use suppression again in further reporting this reading together with other readings to a third node, etc. While cascading further reduces communication, it makes failure handling very difficult, because nodes can act on incomplete and incorrect information and in turn affect other nodes. The subgroup has developed a cascaded suppression framework that fully exploits both temporal and spatial data correlation to reduce communication, and applies coding theory and Bayesian inference to recover missing data resulted from suppression and communication failures. A paper on this subject is currently under submission.

The *review article subgroup* of the Sensor Design Working Group is led by Soumen Lahiri (statistics, Texas A&M U.), and includes XuanLong Nguyen, Jun Yang, and Zhengyuan Zhu. This subgroup is working on a survey article to be submitted to the journal *Statistical Science*. This article will provide the background on wireless sensor networks, and highlight the probabilistic and statistical challenges.

5 Graduate Student Involvement

Three local graduate students actively participated in this program:

David Bell (Duke University) is the SAMSI Graduate Fellow associated with the Sensor Network Dataset working group (Spring 2008) and also the Environmental Risk working group (Fall 2007). He has been active in attending meetings as well as acquisition and exploration of data. He has also been involved in modeling soil moisture data from a local wireless sensor network in Duke Forest with James Clark, Paul Flikkema, Alan Gelfand, Yongku Kim, and XuanLong Nguyen. With his advisor, James Clark, he is developing his dissertation research plan which will involve the use of environmental sensor network data in examining plant-insect interactions in a mixed-hardwood forest. His experience in the SAMSI program prepared him for dealing with data analysis of often faulty sensor data. During his graduate fellowship, he presented a poster at SAMSI's Environmental Sensor Network Workshop (January 2008) concerning modeling of battery data from a wireless sensor network to identify the effects of transmission and data collection on sensor node longevity. He has also given a presentation as an introduction to ecological modeling with sap flux data during SAMSI's PostDoc/Graduate Student Seminar (November 2007), a presentation regarding the use of mathematics and statistics in ecology and environmental sciences at the SAMSI Undergraduate Workshop (March

2008), and gave another presentation during SAMSI's PostDoc/Graduate Student Seminar (April 2008).

Kristian Lum (Duke University) – please see entry under Section 3.

Ilka A. Reis (National Institute for Space Research [INPE], Brazil) has a background in statistics and, currently is pursuing a doctorate in Remote Sensing at INPE. She is interested in developing methods for data collection in sensor networks, especially using data suppression. She attended the kickoff workshop of the SAMSI Environmental Sensor Networks program (January, 13-16, 2008), where she presented the work "Temporal suppression by outlier detection for data collection in sensor networks". She spent the following 8 weeks visiting SAMSI, where she attended the initial meetings of the Working Groups formed as a result of the workshop. While she has since returned to Brazil, she is continuing her involvement in the telemeetings. During her visit to SAMSI, she interacted with researchers to explore the statistical issues involved in the environmental data collection using sensor networks. As a result, she started extending her previous work on temporal data suppression to a more general spatial-temporal suppression scheme. This work, in addition to previous efforts, is expected to form her dissertation.

Another graduate student, Jessica Croft (University of Utah), participated remotely in the telemeetings.

6 Publications and Presentations

Lance Waller organized a session "Monitoring Sensor Networks in Ecology" at the American Statistical Association Environmental Statistics Section's Workshop on Environmetrics (NCAR, Boulder, CO, October 22-24, 2008). Program participants contributed talks as follows:

- Paul Flikkema (collaborative work with Sheryl Howard) - The Roles of Compression and Coding in Inference on Wireless Sensor Networks
- Ernst Linder (collaborative work with Yongku Kim, Zoe Cardon, Scott Holan, Ernst Linder, and Paul Flikkema) - Median Polish Algorithms for Automated Anomaly Detection in Environmental Sensor Networks
- Yongku Kim (collaborative work with Long Nguyen and Scott Holan) - A Correlation Process Prior For Anomaly Detection of Functional Data
- Zhengyuan Zhu (collaborative work with Long Nguyen and Jun Yang) - Optimal Design of Sensor Networks under Energy Budget Constraints

List of Other Presentations (Presentations at SAMSI workshops not included)

- Flikkema, P. The Roles of Compression and Coding in Inference on Wireless Sensor Networks. American Statistical Association, Environmental Statistics Section, Workshop on Environmetrics, NCAR, Boulder, CO, October 22 - 24, 2008.
- Kim, Y. A Correlation Process Prior For Anomaly Detection of Functional Data American Statistical Association, Environmental Statistics Section, Workshop on Environmetrics, NCAR, Boulder, CO, October 22 - 24, 2008.

- Linder, E. Median polish algorithms for automated anomaly detection in Environmental Sensor Networks. American Statistical Association, Environmental Statistics Section, Workshop on Environmetrics, NCAR, Boulder, CO, October 22 - 24, 2008.
- Linder, E. Median polish algorithms for automated anomaly detection in environmental sensor networks. ENAR: Eastern North American Regional Meetings of the International Biometrics Society, San Antonio, TX, March 15 - 18, 2009.
- Zhu, Z. Optimal Design of the Sensor Network under Energy Budget Constraints American Statistical Association, Environmental Statistics Section, Workshop on Environmetrics, NCAR, Boulder, CO, October 22 - 24, 2008.

List of Publications

- Cardon, Z.G., Flikkema, P., Herron, P.M., Holan, S., Kim, Y., Linder, E., and Stark, J.M. A new view of hydraulic redistribution of soil water during rainstorms. To be submitted to Ecology.
- Cardon, Z.G., Stark, J. M., Herron, P.M. (2009) Hydraulic redistribution and the fate of root-derived carbon in soil. Abstract submitted for Ecological Society of America meetings, August 2009, Albuquerque, NM.
- Gelfand, A.E. and Puggioni, G. Analyzing Space-time Sensor Network Data under Suppression and Failure in Transmission, Statistics and Computing (forthcoming).
- He, Y. and Flikkema, P. System-Level Characterization of Single-Chip Radios for Wireless Sensor Network Applications. IEEE WAMICON 2009, April 20-21, 2009, Clearwater, FL USA.
- Howard, S. and Flikkema, P. Integrated Source-Channel Decoding for Correlated Data-Gathering Sensor Networks. IEEE Wireless Communications and Networking Conference (WCNC 2008), March-April 2008
- Howard, S. and Flikkema, P. Progressive Joint Coding, Estimation and Transmission Censoring in Energy-Centric Wireless Data Gathering Networks. Fifth IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS 2008), Sept-Oct 2008.
- Linder, E., Cardon, Z., Murray, J., Holan, S., Flikkema, P., Ignaccolo R., Kim, Y. A Sequential Median Polish for Automated Data Cleaning and Anomaly Detection in Environmental Sensor Networks. Paper in preparation.
- Murray, J. Median polish algorithm for automated anomaly detection in sensor networks (MP-Tuner). Entry to 2009 Student Computing Competition by the American Statistical Association (Section on Computing and Graphical Statistics).
- Murray, J. Median polish algorithm for automated anomaly detection in sensor networks (MP-Tuner). Entry for the 2009 U. of New Hampshire Undergraduate Research Conference. Interactive presentations to be given April 22 and April 24, 2009 (U. of New Hampshire).

- Nguyen, X., Bell, D., Clark, J., Gelfand, A. and Kim, Y. Modeling and computation of wireless sensor network data for environmental monitoring. In preparation.
- Nguyen, X., Yang, J., Yang, Y., and Zhu, Z. Optimal sensor network design under budget constraints. In preparation.
- Nguyen, X., Holand, S., and Kim, Y. A correlation process prior for anomaly detection with functional data. In preparation.
- Nguyen, X., Huang, L. and Joseph, A. (2008). Support vector machines, data reduction and approximate kernel matrices. Proceedings of the 19th European Conference on Machine Learning (ECML), September, Antwerp, Belgium.
- Rajagopal, R., Nguyen, X., Ergen, S. and Varaiya, P. Theory of multiple sequential change-point detection. To be submitted to IEEE Trans. on Signal Processing.
- Rajagopal, R., Nguyen, X., Ergen, S. and Varaiya, P. (2008). Distributed online simultaneous fault detection for multiple sensors. International Conference on Information Processing in Sensor Networks (IPSN), St. Louis, MO.
- Rajagopal, R., Nguyen, X., Coleri-Ergen, S., and Varaiya, P. (2009). Theory of simultaneous fault detection for multiple sensors. Second International Workshop on Sequential Methodologies (IWSM), Troyes, France (invited extended abstract).
- Silberstein, A., Puggioni, G., Gelfand, A., Munagala, K., and Yang, J. Suppression and Failures in Sensor Networks: A Bayesian Approach, Proceedings of the 2007 International Conference on Very Large Data Bases (VLDB '07), Vienna, Austria 2007; 842–853.
- Silberstein, A., Braynard, R., Filpus, G., Puggioni, G., Gelfand, A., Munagala, K., and Yang, J. Data-Driven Processing in Sensor Networks. Proceedings of the 3rd Biennial Conference on Innovative Data Systems Research (CIDR), 2007, Asilomar, California; 10–21.
- Yamamoto, K. and Flikkema, P. Prospector: Multiscale Energy Measurement of Embedded Systems with Wideband Power Supply Signals. In preparation.

7 Workshops

7.1 Planning meeting

The Program Leaders Committee was able to define the program and organize the Opening Workshop via email and teleconference, so a formal planning meeting was not required.

7.2 Opening workshop

The opening workshop for the program was held on January 13-16, 2008, attracted 78 attendees from diverse fields, and met the goal of establishing the composition and activities of the Working Groups. Details of the workshop program are at <http://www.samsi.info/workshops/2007sensor-opening200801.shtml>, and all the presentations at the opening workshop are available at the SAMSI web site.

7.3 Transition workshop

The transition workshop was held October 20-21, 2008, and featured talks by eight program participants (including three female researchers and one graduate student) as well as time for extensive discussions within and between the two Working Groups. Details of the workshop program, including all presentations, are at <http://www.samsi.info/workshops/2008sensor-transition200810.shtml>.

8 Education and Outreach

The Opening Workshop for the Program was preceded by a day of Tutorial Overviews with the following speakers:

- Paul Flikkema, Northern Arizona University: *Ecosystem Inferential Models to Control Data Acquisition and Assimilation*
- Bill Kaiser, Univ. of California-Los Angeles: *Sensor Network Platforms for Rapidly Deployable, Configurable, and Sustainable Observatories*
- Jennifer Hoeting, Colorado State University: *Hierarchical Modeling*
- Kiona Ogle, University of Wyoming: *Data-Model Integration: Examples from Below-ground Ecosystem Ecology*

All the tutorial presentations are available on-line on the SAMSI website.

Paul Flikkema organized a session on Environmental Sensor Networks at a SAMSI Undergraduate Workshop, Feb. 29 - Mar. 1. Speakers were Kenji Yamamoto (undergraduate student, Northern Arizona University), Dave Bell (graduate student, SAMSI/Duke University), XuanLong Nguyen (SAMSI postdoctoral fellow), Yongku Kim (SAMSI postdoctoral fellow), and Michael Porter (SAMSI postdoctoral associate). Paul Flikkema moderated the session.

9 Industrial and Governmental Participation

Yuliy Baryshnikov (Bell Laboratories) and Mike Godin (Monterey Bay Aquarium Research Institute) were invited speakers for the Opening Workshop. There was also participation in the opening workshop and working groups from government agencies, laboratories, and industry, including EPA, Centers for Disease Control and Prevention (CDC), Marine Biological

Laboratory, the IBM T. J. Watson Research Center, the Center for Wireless Communications, University of Oulu (Finland), and the National Institute for Space Research (Brazil).

10 External Support

This program did not have external support.

11 Affiliates Participation

There were working group participants from each of the following university affiliates: University of California - Berkeley, Duke University, University of Michigan, North Carolina State University, and University of North Carolina at Chapel Hill.

APPENDIX D – Workshop Participants Lists

For most of the SAMSI workshops, the participants will be summarized in three tables below. The first table is a summary of all participants by gender, status, field of work/study, affiliation, and location. The second table lists only the participants who received support. The third table lists all workshop participants. The minority status of each participant is available, but we do not include the information here because of privacy issues; the summaries in Section H: Diversity Efforts were compiled from this data.

The key top **Status** entry is as follows:

NRG – New Researcher or Graduate Student S – Students (Education & Outreach)
 FP – Faculty or Professional A – Faculty (Education & Outreach)

2007-08 PROGRAM EVENTS AFTER APRIL 2008

❖ Random Media Program

Random Media Transition Workshop

Participant Summary

May 1-2, 2008

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	4	2	0	3	3	0	6	0	6	6
Unsuppted	8	3	0	5	6	0	11	0	4	2
SAMSI	3	1	0	1	3	0	4	0		

Random Media Transition Workshop

Workshop Participants

May 1-2, 2008

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Beale	J. Thomas	Male	Duke University	Mathematics	FP
Forest	Greg	Male	UNC Chapel Hill	Mathematics & Biomedical Engineering	FP
Heller	Martin	Male	SAMSI	Mathematics	NRG
Ito	Kazi	Male	NCSU	Mathematics	FP

Jiang	Qunlei	Female	North Carolina State University	Mathematics	NRG
Kang	Min	Female	North Carolina State University	Mathematics	FP
Kao	Chiu-Yen	Female	Ohio State	Department of Mathematics	NRG
Khan	Taufiqar	Male	Clemson University	Mathematical Sciences	NRG
Klapper	Isaac	Male	Montana State University	Department of Mathematical Sciences	FP
Layton	Anita	Female	Duke University	Mathematics	NRG
Li	Zhilin	Male	North Carolina State University	Mathematics	FP
Lowengrub	John	Male	U California, Irvine	Mathematics	FP
Luo	Li-Shi	Male	Old Dominion University	Department of Mathematics & Statistics	FP
McAdoo	Bonnie	Female	Clemson University	Mathematical Sciences	NRG
Siegel	Michael	Male	New Jersey Institute of Technology	Mathematical Sciences	NRG
Smith	Ralph	Male	North Carolina State University & SAMSI	Mathematics	FP
Spiller	Elaine	Female	SAMSI	Mathematics	NRG
Wang	Cheng	Male	University of Tennessee	Mathematics	NRG
Wilson	Jason	Male	Duke University	Mathematics	NRG
Xie	Hui	Male	NC State Univ	Mathematics	NRG
Zhong	Weigang	Male	SAMSI	Mathematics	NRG

❖ **Education and Outreach Program**

SAMSI/CRSC Interdisciplinary Workshop for Undergraduates

Participant Summary

May 19-23, 2008

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Institutions Represented	Number of States Represented
Supported	4	11	0	0	15	15	0	13	10
Unsuppted	4	4	0	5	3	8	0	3	1
SAMSI	3	1	0	4	0	4	0		

SAMSI/CRSC Interdisciplinary Workshop for Undergraduates

Workshop Participants

May 19-23, 2008

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Abdalla	Widad	Female	University of Puerto Rico, Cayey	Mathematics	S
Armentrout	Megan	Female	Whitworth University	Mathematics	S
Canseco	Veronica	Female	University of Illinois at Urbana-Champaign	Mathematics	S
Cheng	Guang	Male	SAMSI	Statistics	A
Cole-Manning	Camme	Female	NCSU	Mathematics	A
Costanzo	Kate	Female	Drew University	Mathematics/Sociology	S
Enstrom	Betsy	Female	Duke University	Statistics	A
Gao	Yajing	Male	Duke University	Biomedical Eng, Mathematics	S
Israel	Alicia	Female	Texas A&M University - College Station	Appl Mathematical Sciences	S
Johnson	Terri	Female	Meredith College	Mathematics	S
Konopacka	Roza	Female	City College of New York	Mathematics	S
Madar	Vared	Female	SAMSI	Statistics	A

Manning	Jim	Male	University of South Carolina	Mathematics and Statistics	S
Minges	Erik	Male	University of North Carolina in Wilmington	Physics / Applied Mathematics	S
Myers	Ashley	Female	North Carolina State University	Statistics	S
Pal	Jayanta	Male	SAMSI	Statistics	A
Porter	Michael	Male	NCSU / SAMSI	Statistics	A
Robles Vega	Evelyn	Female	University of Puerto Rico at Cayey	Mathematics-Physics	S
Sapp	Stephanie	Female	Johns Hopkins University	Appl Mathematics & Statistics	S
Sherman	Toby	Male	Virginia Tech	Mathematics and Chemical Eng	S
Silva	Sanjeeka	Female	Meredith College	Mathematics and Computer Science	S
Smith	Erickson	Male	North Carolina State University	Applied Mathematics	S
Stitzinger	Ernie	Male	NCSU	Mathematics	A
Tan	Khoon Yu	Male	University of Michigan, Ann Arbor	Statistics and Act Mathematics	S
Weems	Kim	Female	NCSU	Statistics	A
Weiss	Madeline	Female	California State University, Chico	Mathematics & Statistics	S
White	Gentry	Male	NCSU	Statistics	A

❖ Risk Program

Risk Revisited: Progress and Challenges Workshop

Participant Summary

May 21, 2008

Participants	Male	Female	Unspec-ified	Faculty/Professional	New Researcher/Student	Stat	Math	Other	# of Institutions Represented	# of States Represented
Supported	3	1	0	1	3	4	0	0	2	2
Unsuppted	10	6	0	9	7	16	0	0	11	1
SAMSI	2	2	0	2	2	4	0	0		

Risk Revisited: Progress and Challenges Workshop

Workshop Participants

May 21, 2008

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Berger	Jim	Male	SAMSI	Statistics	FP
Bobashev	Georgiy	Male	RTI	Department of Statistics	FP
Cheng	Guang	Male	Duke U	Department of Statistics	NRG
Cooley	Dan	Male	Colorado State U	Department of Statistics	NRG
Das	Sourish	Male	U of Connecticut	Department of Statistics	NRG
Dey	Dipak	Male	U of Connecticut	Department of Statistics	FP
Enstrom	Betsy	Female	Duke U	Department of Statistics	NRG
Evangelou	Evangelos	Male	University of North Carolina	Department of Statistics	NRG
Fricker	Ron	Male	Naval Postgraduate School	Department of Statistics	FP
Gaioni	Elijah	Male	U of Connecticut	Department of Statistics	NRG
Heffernan	Janet	Female	Lancaster University / J. Heffernan Consulting	Department of Statistics	FP

Ignaccolo	Rosalba	Female	Universita' degli Studi di Torino	Department of Statistics	NRG
Katzoff	Myron	Male	National Center for Health Statistics	Department of Statistics	FP
Kim	Yongku	Male	SAMSI	Department of Statistics	NRG
Madar	Vered	Female	SAMSI	Department of Statistics	NRG
Munoz	Pilar	Female	Technical University of Catalonia	Department of Statistics	FP
Nail	Amy	Female	NCSU	Department of Statistics	NRG
Qin	Xiao	Female	Beihang University and UNC	Department of Statistics	NRG
Rios Insua	David	Male	University Rey Juan Carlos	Department of Statistics	FP
Sedransk	Nell	Female	NISS and SAMSI	Statistics	FP
Smith	Richard	Male	University of North Carolina	Department of Statistics	FP
Shen	Haipeng	Male	University of North Carolina	Department of Statistics	NRG
Wang	Jenting	Female	SUNY-Oneonta	Department of Statistics	FP
Wolpert	Robert	Male	Duke U	Department of Statistics	FP

❖ **2008 Summer Program**

Meta Analysis
Participant Summary
June 2-13, 2008

Participants	Male	Female	Unspecified	Faculty/Professional	New Researcher/Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	19	7	0	6	20	23	1	3	21	12
Unsupported	23	17	0	15	25	39	0	0	18	3
SAMSI	0	0	0	0	0	0	0	0		

Meta Analysis
Workshop Participants
June 2-13, 2008

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Ahmad	Faiz	Male	GlaxoSmithKline Inc.	STAT	NRG
Barrett	Jessica	Female	Medical Research Council Biostatistics Unit	STAT	NRG
Basu	Sanjib	Male	Northern Illinois University	STAT	FP
Bayarri	Susie	Female	University of Valencia	STAT	FP
Berger	James	Male	SAMSI	STAT	FP
Bortz	David	Male	University of Colorado	MATH	NRG
Casella	George	Male	U of Florida	STAT	FP
Demidenko	Eugene	Male	Dartmouth Medical School	STAT	FP
Deng	Chunqin	Male	Talecris Biotherapeutics	STAT	NRG
Dukic	Vanja	Female	University of Chicago	STAT	NRG
Dunson	David	Male	Duke University	STAT	FP
Gatsonis	Constantine	Male	Brown University	STAT	FP
Harrell	Leigh	Female	Virginia Tech	EDUC	NRG
He	Qianchuan	Male	UNC-Chapel Hill	STAT	NRG
Hedges	Larry	Male	Northwestern U	STAT	FP
Higgins	Julian	Male	Medical Research Council	STAT	FP
Hua	Zhaowei	Female	UNC-Chapel Hill	STAT	NRG
Jackson	Daniel	Male	MRC Biostatistics Unit	STAT	NRG

Johnson	Nels	Male	Virginia Tech	STAT	NRG
Kaizar	Eloise	Female	Ohio State U	STAT	NRG
Kim	Yongku	Male	SAMSI	STAT	NRG
Kinney	Satkartar	Female	NISS	STAT	NRG
Kounali	Daphne	Female	Centre of Multilevel Modelling	STAT	NRG
Krishen	Alok	Male	Glaxo Smith Kline Inc.	STAT	FP
Lin	Danyu	Male	UNC	Biostats	NRG
Liu	Fei	Female	University of Missouri	STAT	NRG
Madar	Vared	Female	SAMSI	STAT	NRG
Mak	Timothy	Male	Imperial College London	STAT	NRG
McCandless	Lawrence	Male	Imperial College London	STAT	NRG
Moreno	Elias	Male	University of Granada	STAT	FP
Morton	Sally	Female	RTI International	STAT	FP
Olkin	Ingram	Male	Stanford University	STAT	NRG
O'Rourke	Keith	Male	O'Rourke Consulting	STAT	NRG
Petricka	Jalean	Female	Duke University	LIFE	NRG
Plante	Jean-Francois	Male	University of Toronto	STAT	NRG
Platt	Robert	Male	McGill University	STAT	FP
Pungpapong	Vitara	Female	Purdue University	STAT	NRG
Rice	Kenneth	Male	University of Washington	STAT	NRG
Sedransk	Nell	Female	SAMSI/NISS	STAT	FP

Sherif	Bintu	Female	RTI International	STAT	NRG
Shrier	Ian	Male	McGill University	STAT	FP
Stangl	Dalene	Female	Duke University	STAT	FP
Stevens	John	Male	Utah State University	STAT	NRG
Stuart	Elizabeth	Female	Johns Hopkins Bloomberg School of Public Health	STAT	NRG
Sun	Junfeng	Male	U of Nebraska Medical Center	STAT	NRG
Thorlund	Kristian	Male	Copenhagen Trial Unit	STAT	NRG
Tiwari	Ram	Male	Food and Drug Administration	STAT	NRG
Trikalinos	Thomas	Male	Tufts Medical Center	LIFE	NRG
Tzeng	Jung-Ying	Female	NC State University	STAT	NRG
Umbach	David	Male	National Institute of Environmental Health Sciences, NIH	STAT	NRG
Unal	Cemal	Male	Pozen, Inc.	STAT	NRG
Wang	Jen-Ting	Female	SUNY-Oneonta	STAT	FP
Warren	Liling	Female	GSK	STAT	NRG
Williams	Matthew	Male	Department of Statistics at Virginia Tech	STAT	NRG
Wolpert	Robert	Male	Duke U	STAT	FP
Wouhib	Abera	Male	CDC	STAT	FP
Xia	Jessie	Female	NISS	STAT	NRG
Young	Stan	Male	NISS	STAT	FP
Zhang	Lingsong	Male	Harvard School of Public Health	STAT	NRG
Zhang	Ying	Female	POZEN, Inc	STAT	FP

Zhao	Yue	Female	UNC-CH	STAT	NRG
Zhou	Jasmin	Male	National Institute of Statistical Sciences	STAT	NRG
Zou	Fei	Female	UNC	STAT	NRG

❖ **Environmental Sensor Networks**

Sensor Networks Transition Workshop

Participant Summary

October 20-21, 2008

Participants	Male	Female	Unspecified	Faculty/Professional	New Researcher/Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	2	4	0	3	3	2	0	4	5	4
Unsuppted	7	0	0	2	5	4	0	3	5	3
SAMSI	1	0	0	1	0	1	0	0		

Sensor Networks Transition Workshop

Workshop Participants

October 20-21, 2008

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Berger	James	Male	SAMSI	STAT	FP
Cardon	Zoe	Female	Marine Biological Laboratory	LIFE	FP
Clark	Jim	Male	Duke U	BIO	FP
Flikkema	Paul	Male	Northern Arizona U	ENG	FP
Holan	Scott	Male	U Missouri	STAT	NRG
Howard	Sheryl	Female	Northern Arizona U	ENG	NRG
Ignaccolo	Rosalba	Female	Universit� degli Studi di Torino	STAT	NRG
Lahiri	Soumendra	Male	Texas A&M	STAT	FP

Linder	Ernst	Male	U of New Hampshire	STAT	NRG
Nguyen	Long	Male	Duke U	STAT	NRG
Shoemaker	Christine	Female	Cornell	ENG	FP
Yang	Jun	Male	Duke U	COMP	NRG
Zhang	Yi	Male	Duke U	COMP	NRG
Zhu	Zhengyuan	Male	UNC	STAT	NRG

2008-09 PROGRAM EVENTS THROUGH JULY 2009

❖ **Sequential Monte Carlo Methods**

Sequential Monte Carlo Methods Opening Workshop

Participant Summary

September 7-10, 2008

Participants	Male	Female	Unspecified	Faculty/Professional	New Researcher/Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	38	8	0	16	30	15	8	23	34	16
Unsuported	65	19	0	30	54	69	27	38	31	2
SAMSI	3	1	0	1	3	2	1	0		

Sequential Monte Carlo Methods Opening Workshop

Workshop Participants

September 7-10, 2008

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Airoldi	Edo	Male	Princeton University	STAT	NRG
Alcala	Jose	Male	New York University	MATH	NRG
Argon	Nilay	Female	University of North Carolina at Chapel Hill	ENGG	NRG

Armagan	Artin	Male	Duke University	STAT	NRG
Bain	Melanie	Female	University of North Carolina	OTHR	NRG
Belmonte	Miguel	Male	University of Warwick	STAT	NRG
Berger	James	Male	SAMSI	STAT	FP
Berrocal	Veronica	Female	Department of Statistical Science - Duke University	STAT	NRG
Bhadra	Anindya	Male	University of Michigan	STAT	NRG
Bhatnagar	Nayantara	Female	UC Berkeley	COMP	NRG
Bickel	Peter	Male	Princeton	STAT	FP
Bishwal	Jaya	Male	University of North Carolina at Charlotte	STAT	NRG
Boomer	Karen "KB"	Female	Bucknell University	STAT	NRG
Bornn	Luke	Male	University of British Columbia	STAT	NRG
Briers	Mark	Male	QinetiQ Ltd	STAT	NRG
Briggs	Jonathan	Male	University of Auckland	STAT	FP
Bugallo	Monica	Female	Stony Brook University	ENGG	NRG
Butala	Mark	Male	University of Illinois at Urbana-Champaign	ENGG	NRG
Carvalho	Carlos	Male	The University of Chicago Graduate School of Business	STAT	NRG
Chen	Fang	Male	Rutgers University	STAT	FP
Chen	Hao	Male	SAS Institute	STAT	NRG
Chen	Rong	Male	Duke University – FSB	STAT	NRG
Chopin	Nicolas	Male	CREST-ENSAE	STAT	NRG

Clark	Daniel	Male	Heriot-Watt University	ENGG	NRG
Clyde	Merlise	Female	Duke University	STAT	FP
Colvin	Jacob	Male	University of California, Santa Cruz	STAT	NRG
Corberan	Ana	Female	University of Valencia	STAT	NRG
Cornebise	Julien	Male	Telecom Paristech	STAT	NRG
Crisan	Dan	Male	Imperial College London	MATH	FP
Dance (Bradley)	Sarah	Female	University of Reading	MATH	NRG
Das	Sourish	Male	SAMSI	STAT	NRG
DeJong	David	Male	University of Pittsburgh	SOCL	FP
Del Moral	Pierre	Male	INRIA	COMP	FP
Deng	Shaozhong	Male	University of North Carolina at Charlotte	MATH	NRG
Djuric	Petar	Male	Stony Brook	ENGG	FP
Doucet	Arnaud	Male	Department of Statistics University of British Columbia	STAT	FP
Dunson	David	Male	Duke University	STAT	FP
Falin	Lee	Male	VA Bioinfo	STAT	NRG
Fan	Kai	Male	North Carolina State University	MATH	NRG
Fearnhead	Paul	Male	Lancaster University	STAT	FP
Fokoue	Ernest	Male	Kettering University	STAT	NRG
Ghosh	Sujit	Male	NC State University	STAT	FP
Godsill	Simon	Male	University of Cambridge	ENGG	FP

Goel	Prem	Male	The Ohio State University	STAT	FP
Green	Nathan	Male	Dstl	MATH	NRG
Griffiths	Robert	Male	University of Oxford	STAT	FP
Guerron	Pablo	Male	North Carolina State University	SOCL	NRG
Hannig	Jan	Male	Academic	PHYS	NRG
He	Qianchuan	Male	University of North Carolina at Chapel Hill	MATH	NRG
Holenstein	Roman	Male	University of British Columbia	STAT	NRG
Huber	Mark	Male	Duke University	MATH	FP
Ikoma	Norikazu	Male	Kyushu Institute of Technology	ENGG	FP
Ionides	Edward	Male	University of Michigan	STAT	NRG
Ji	Chunlin	Male	Duke University	STAT	NRG
Johannes	Michael	Male	Columbia University, Graduate School of Business	SOCL	NRG
Kim	Songhan	Male	Portland State University	ENGG	NRG
Kimura	Tomoaki	Male	Waseda Univ. Dpt. of Science and Engineering Matsumoto lab.	COMP	NRG
Koutsourelakis	Steve	Male	Cornell University	ENGG	NRG
Law	Wai	Female	Duke University	MATH	NRG
Leman	Scotland	Male	Virginia tech	STAT	NRG
Lin	Ming	Female	UNC	COMP	FP
Liu	Fei	Female	Iowa	STAT	NRG
Liu	Jun	Male	University of Chicago	STAT	NRG

Lopes	Hedibert	Male	Cornell University	PHYS	FP
Loredo	Thomas	Male	University of South Carolina	STAT	FP
Lynch	James	Male	University of Georgia	SOCL	FP
Lyubimov	Konstantin	Male	Duke University	STAT	NRG
Macaro	Christian	Male	Duke University	MATH	NRG
Manolopoulou	Ioanna	Female	North Carolina State University	STAT	NRG
McLain	Alex	Male	Duke University	STAT	NRG
Merl	Daniel	Male	SUNY at Stony Brook	ENGG	NRG
Mernick	Kevin	Male	New Jersey Institute of Technology	MATH	FP
Mihaylova	Lyudmila	Female	Lancaster University	ENGG	FP
Moore	Matthew	Male	University of North Carolina - Chapel Hill	MATH	NRG
Morales	Mario	Male	Hunter College, CUNY	STAT	NRG
Moulines	Eric	Male	Ecole Nationale Supérieure	MATH	FP
Mukherjee	Chiranjit	Male	Duke University	STAT	NRG
Mulder	Joris	Male	Utrecht University	STAT	NRG
Munoz	Maria Pilar	Female	Ohio State University	SOCL	FP
Myung	Jay	Male	Duke University	STAT	NRG
Niemi	Jarad	Male	Imperial College London	MATH	NRG
Okten	Giray	Male	Florida State University	MATH	FP
Olasunkanmi	Obanubi	Male	Imperial College London	MATH	NRG
Owen	Megan	Female	SAMSI and NCSU	MATH	NRG

Papaspiliopoulos	Omiros	Male	Barcelona GSE	STAT	FP
Pelletier	Denis	Male	North Carolina State University	STAT	NRG
Pena	Edsel	Male	University of South Carolina	STAT	FP
Peterson	Chris	Male	Colorado State	MATH	FP
Petralia	Francesca	Female	Duke University	STAT	NRG
Petris	Giovanni	Male	U of Arkansas	STAT	FP
Polson	Nick	Male	University of Chicago	STAT	FP
Porter	Michael	Male	University of California Santa Cruz	STAT	FP
Prado	Raquel	Female	University of California	MATH	FP
Redelings	Benjamin	Male	University of Southern California	COMP	NRG
Robert	Christian	Male	Universite Paris Dauphine	STAT	FP
Rodriguez	Abel	Male	University of California	STAT	NRG
Rogers	Chris	Male	U of Cambridge	MATH	FP
Roos	Jason	Male	Duke University	SOCL	NRG
Roy	Deb	Female	Pennsylvania State	MATH	NRG
Rozgic	Viktor	Male	University of Southern California	ENGG	NRG
Rubenthaler	Sylvain	Male	Universit� de Nice-Sophia Antipolis	STAT	NRG
Rubio-Ramirez	Juan	Male	Duke University	SOCL	NRG
Schoolfield	Clyde	Male	University of Florida	MATH	FP
Schott	Sarah	Female	Duke University	MATH	NRG
Septier	Francois	Male	Cambridge University	ENGG	NRG

Shen	Bingxin	Female	Stony Brook University	ENGG	NRG
Shi	Minghui	Female	Duke U	STAT	NRG
Stark	Christopher	Male	NSF	MATH	FP
Stroud	Jonathan	Male	George Washington U	STAT	FP
Sun	Dongchu	Male	U Missouri	STAT	FP
Tadesse	Mahlet	Female	Georgetown University	STAT	NRG
ter Braak	Cajo	Male	Wageningen University and Research Centre	STAT	FP
Thomas	Andrew	Male	University of St Andrews	STAT	FP
Thomas	Len	Male	U St. Andrews	STAT	FP
Ueno	Genta	Male	The Institute of Statistical Mathematics	PHYS	NRG
Vaswani	Namrata	Female	Iowa State University	ENGG	NRG
Verma	Vandi	Female	California Institute of Technology - Jet Propulsion laboratory	OTHR	NRG
Vo	Ba-Ngu	Male	The University of Melbourne	ENGG	FP
Vogelstein	Joshua	Male	Johns Hopkins	BIOSCI	NRG
Voss	Jochen	Male	University of Warwick	MATH	NRG
Wang	Hao	Male	Duke University	STAT	NRG
Wang	Kai	Male	Duke University	STAT	NRG
Weare	Jonathan	Male	NY University	MATH	FP
West	Mike	Male	Duke University	STAT	FP
White	Gentry	Male	NCSU	STAT	NRG

Wolpert	Robert	Male	Duke University	STAT	FP
Xu	Zhenli	Male	UNC-Charlotte	MATH	NRG
Yang	Hongxia	Female	Duke University	STAT	NRG
Yasamin	Ahmad	Male	SAMSI	STAT	NRG
Yin	Junming	Male	UC Berkeley	COMP	NRG
Yoshida	Ryo	Male	Institute of Statistical Mathematics	BIOSTAT	NRG
Zhang	Baqun	Male	NC State University	STAT	NRG
Zhou	Enlu	Female	University of Maryland, College Park	ENGG	NRG
Zou	Fei	Female	UNC	BIOSTAT	NRG

SMC Mid-Program
Participant Summary
February 19-20, 2009

Participants	Male	Female	Unspecified	Faculty/Professional	New Researcher/Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	7	1	0	3	5	1	0	7	7	5
Unsuppted	17	4	0	5	16	13	3	5	11	3
SAMSI	4	1	0	0	5	5	0	0		

SMC Mid-Program
Workshop Participants
February 19-20, 2009

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Argon	Nilay	Female	University of North Carolina	Statistics and Operations Research	NRG
Bain	Melanie	Female	University of North Carolina	Statistics and Operations Research	NRG
Briers	Mark	Male	QinetiQ Ltd	Statistics	NRG

Carvalho	Carlos	Male	University of Chicago	Booth School of Business	NRG
Clark	Daniel	Male	Heriot-Watt University	Eng	NRG
Coates	Mark	Male	McGill University	Engineering/Operations Research	FP
Das	Sourish	Male	SAMSI and Duke University	Statistics	NRG
Djuric	Petar	Male	Stony Brook University	Electrical and Computer Engineering	FP
Dunson	David	Male	Duke University	Statistical Science	FP
Fearnhead	Paul	Male	Lancaster University	Statistics	FP
Fokoue	Ernest	Male	Kettering University	Mathematics	NRG
Godsill	Simon	Male	University of Cambridge	Engineering	FP
Green	Nathan	Male	DSTL	Mathematics	NRG
Ji	Chunlin	Male	Duke University	Department of Statistical Science	NRG
Liu	Bin	Male	SAMSI	Statistics	NRG
Lopes	Hedibert	Male	University of Chicago	Booth School of Business	FP
Lynch	Jim	Male	University of South Carolina	Statistics	FP
Lyubimov	Konstantin	Male	University of Georgia	Social Sciences	NRG
Macaro	Christian	Male	SAMSI and Duke	Statistics	NRG
Manolopoulou	Ioanna	Female	SAMSI	Statistics	NRG
Mukherjee	Chiranjit	Male	Duke University	Statistics	NRG
Rozgic	Viktor	Male	U Southern California	Electrical Engineering – Systems	NRG
Schott	Sarah	Female	Duke University	Mathematics	NRG
Septier	Francois	Male	University of Cambridge	Signal Processing Lab.	NRG

Shi	Minghui	Female	Duke University	Statistical Science	NRG
Taddy	Matt	Male	University of Chicago	Booth School of Business	NRG
Vaswani	Namrata	Female	Iowa State University	Electrical and Computer Engineering	NRG
Vidyashankar	Anand	Male	Cornell University	Statistical Science and Social Statistics	NRG
Wang	Hao	Male	Duke University	Statistical Science	NRG
West	Mike	Male	Duke University	Statistics	FP
White	Gentry	Male	SAMSI NC State U	Statistics	NRG
Yardim	Caglar	Male	UCSD	SIO	NRG
Yoshida	Ryo	Male	SAMSI	Statistics	NRG
Zhang	Baqun	Male	NCSU	Statistics	NRG

Adaptive Design, Sequential Monte Carlo, and Computer Modeling
Participant Summary
April 15-17, 2009

Participants	Male	Female	Unspecified	Faculty/Professional	New Researcher/Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	6	0	0	2	4	4	0	2	5	3
Unsuported	23	8	0	13	19	21	5	5	17	10
SAMSI	5	1	0	1	4	6	0	0		

Adaptive Design, Sequential Monte Carlo, and Computer Modeling
Workshop Participants
April 15-17, 2009

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Argon	Nilay	Female	University of North Carolina	ENG	NRG
Bain	Melanie	Female	University of North Carolina, Chapel Hill	Statistics and Operations Research	NRG

Bayarri	Susie	Female	University of Valencia	Statistics and Operations Research	FP
Berger	James	Male	SAMSI	Statistics	FP
Bhat	K Sham	Male	Pennsylvania State University	Statistics	NRG
Bingham	Derek	Male	Simon Fraser University	Statistics and Actuarial Science	NRG
Colvin	Jacob	Male	University of California, Santa Cruz	Applied Math & Statistics	NRG
Cornebise	Julien	Male	SAMSI	Statistics	NRG
Dalbey	Keith	Male	University at Buffalo	Mechanical and Aerospace Engineering	NRG
Das	Sourish	Male	SAMSI and Duke University	Statistical Science	NRG
de Villiers	Johan	Male	University of Pretoria	Electrical, Electronic and Computer Engineering	NRG
Feddag	Mohand	Male	University of Southampton	Statistical Sciences Research Institute	NRG
Flournoy	Nancy	Female	University of Missouri	Statistics	FP
Gattiker	James	Male	Los Alamos National Laboratory	Statistics	FP
Godsill	Simon	Male	University of Cambridge	Engineering	FP
Higdon	Dave	Male	LANL	Statistical Sciences Group	FP
Ji	Chunlin	Male	Duke University	Department of Statistics Science	NRG
Johannesson	Gardar	Male	Lawrence Livermore National Laboratory	Statistics	NRG
Lee	Herbert	Male	University of California, Santa Cruz	Applied Math & Statistics	FP
Liu	Xuyuan	Male	Georgia Institute of Technology	Industrial Engineering	NRG
Liu	Fei	Female	University of Missouri-Columbia	Statistics	NRG
Liu	Bin	Male	SAMSI	Statistics	NRG

Lopes	Danilo	Male	Duke University	Statistical Science	NRG
Lopes	Hedibert	Male	University of Chicago	Booth School of Business	FP
Loredo	Thomas	Male	Cornell University	Department of Astronomy	FP
Lynch	James	Male	U South Carolina	Statistics	FP
Manolopoulou	Ioanna	Female	SAMSI	Statistics	NRG
Ncube	Moeti	Male	Florida State University	Statistics	NRG
Notz	William	Male	Ohio State University	Statistics	FP
Patterson	Angela	Female	General Electric	Statistics	FP
Pitman	Bruce	Male	University at Buffalo	Mathematics	FP
Rodriguez	Abel	Male	University of California, Santa Cruz	Applied Mathematics and Statistics	NRG
Sain	Steve	Male	NCAR	Statistics	FP
Spiller	Elaine	Female	Marquette University	Math, Stat, and CS	NRG
Storlie	Curtis	Male	U New Mexico	Statistical Science and Social Statistics	NRG
Vidyashankar	Anand	Male	Cornell University	Statistical Science	NRG
Wang	Jianyu	Male	Duke University	Statistical Science	NRG
Wang	Hao	Male	Duke University	Statistics	NRG
West	Mike	Male	Duke U	Statistics	FP
White	Gentry	Male	SAMSI	Statistics	NRG
Williams	Brian	Male	LANL	Statistical Science	NRG
Wolpert	Robert	Male	Duke U	Statistical Sciences Research Institute	FP

Woods	Dave	Male	University of Southampton	Statistics	NRG
-------	------	------	---------------------------	------------	-----

❖ **Algebraic Methods in Systems Biology and Statistics**

Algebraic Methods Opening Workshop

Participant Summary

September 14-17, 2008

Participants	Male	Female	Unspecified	Faculty/Professional	New Researcher/Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	43	15	1	20	39	14	33	12	38	15
Unsuppted	38	18	0	28	29	21	27	9	19	5
SAMSI	2	2	0	1	2	2	1	0		

Algebraic Methods Opening Workshop

Workshop Participants

September 14-17, 2008

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Allman	Elizabeth	Female	University of Alaska Fairbanks	MATH	FP
Bakalov	Bojko	Male	North Carolina State University	MATH	NRG
Barker	Brandon	Male	Cornell University	STAT	NRG
Bayarri	Susie	Female	U Valencia	STAT	FP
Beerenwinkel	Niko	Male	ETH Zurich	LIFE	NRG
Beerli	Peter	Male	Florida State University	LIFE	FP
Berger	James	Male	SAMSI	STAT	FP
Buczynska	Weronika	Female	Texas A&M	MATH	NRG
Cartwright	Dustin	Male	University of California, Berkeley	MATH	NRG
Chen	Hegang	Male	University of Maryland School of Medicine	STAT	FP
Chen	Teng	N/A	University of Central Florida	MATH	NRG

Chen	Wenjie	Female	UNC-Chapel Hill	STAT	NRG
Chifman	Julia	Female	University of Kentucky	MATH	NRG
Chuang	Jer-Chin	Male	Duke University	MATH	NRG
Coleman	Deidra	Female	North Carolina State University	STAT	NRG
Conradi	Carsten	Male	Max Planck Institute Dynamics of Complex Technical Systems	MATH	NRG
Cox	Lawrence	Male	National Center for Health Statistics/CDC	MATH	FP
Craciun	Gheorghe	Male	University of Wisconsin	MATH	NRG
Deems	Thomas	Male	North Carolina State University	MATH	NRG
Dickenstein	Alicia	Female	Universidad de Buenos Aires	MATH	FP
Dimitrova	Elena	Female	Clemson University	MATH	NRG
Dinwoodie	Ian	Male	Duke University	STAT	FP
Drton	Mathias	Male	University of Chicago	STAT	NRG
Fienberg	Stephen	Male	Carnegie Mellon University	STAT	FP
Fleming	Ronan	Male	UCSD	LIFE	NRG
Francis	Andrew	Male	University of Western Sydney	MATH	FP
Friedrich	Thomas	Male	Freie Universitaet Berlin	MATH	NRG
Garfield	David	Male	Duke University	LIFE	NRG
Ginestet	Cedric	Male	Imperial College	BIOSTAT	NRG
Gnacadjia	Gilles	Male	Amgen	MATH	FP
Gopalkrishnan	Manoj	Male	University of Southern California	COMP	NRG

Gunawardena	Jeremy	Male	Harvard Medical School	LIFE	FP
Haney	Richard	Male	Cellular Statistics	STAT	FP
Hara	Hisayuki	Male	University of Tokyo	STAT	FP
Hartemink	Alexander	Male	Duke University	COMP	NRG
Heitsch	Christine	Female	Georgia Institute of Technology	MATH	NRG
Hinkelmann	Franziska	Female	Virginia Bioinformatics Institute	MATH	NRG
Hoeschele	Ina	Female	Virginia Tech	STAT	FP
Horn	Mary Ann	Female	National Science Foundation	MATH	FP
Hosten	Serkan	Male	SF State U	MATH	FP
Hower	Valerie	Female	Georgia Institute of Technology	MATH	NRG
Huber	Mark	Male	Duke University	MATH	FP
Jaromczyk	Jerzy	Male	University of Kentucky	COMP	FP
Jarrah	Abdul Salam	Male	Virginia Tech	MATH	NRG
Jing	Naihuan	Male	North Carolina State Univ	MATH	FP
Johannsen	David	Male	Naval Surface warfare Center	MATH	FP
Kahle	Thomas	Male	Max Planck Institute for Mathematics in the Sciences	MATH	NRG
Kaltofen	Erich	Male	North Carolina State U	MATH	NRG
Kogan	Irina	Female	North Carolina State University	MATH	NRG
Kondor	Imre	Male	Gatsby Unit, University College London	COMP	NRG
Kubatko	Laura	Female	Ohio State University	STAT	FP

Kuo	Lynn	Female	University of Connecticut	STAT	FP
Laubenbacher	Reinhard	Male	Virginia Polytechnic Institute and State University	STAT	NRG
Layne	Lori	Female	Clemson University	MATH	NRG
Lee	Tong	Male	Hunter College of City University of New York	MATH	NRG
Lewis	Robert	Male	Fordham U	MATH	FP
Lin	Shaowei	Male	University of California, Berkeley	MATH	NRG
Lunagomez	Simon	Male	Duke University	STAT	NRG
Magwene	Paul	Male	Duke University	BIOSCI	FP
Manon	Christopher	Male	University of Maryland	MATH	NRG
Marchette	David	Male	Naval Surface Warfare Center	STAT	FP
Maruri Aguilar	Hugo	Male	London School of Economics	STAT	NRG
Matias	Catherine	Female	CNRS, Laboratoire Statistique & GÃ©nome	STAT	NRG
McCandlish	David	Male	Duke University	LIFE	NRG
Minimair	Manfred	Male	Seton Hall University	COMP	NRG
Mishra	Bud	Male	Courant Institute	MATH	FP
Mortveit	Henning	Male	Virginia Tech	MATH	NRG
Nagel	Uwe	Male	U of Kentucky	MATH	FP
Ohler	Uwe	Male	Duke University	LIFE	NRG
Owen	Megan	Female	SAMSI	MATH	NRG
Pachter	Lior	Male	UC Berkeley	MATH	FP

Pantea	Casian	Male	University of Wisconsin – Madison	MATH	NRG
Perduca	Vittorio	Male	Universita' degli Studi di Torino	MATH	NRG
Perez Millan	Mercedes	Female	Universidad de Buenos Aires	MATH	NRG
Petrovic	Sonja	Female	University of Illinois at Chicago	MATH	NRG
Pistone	Giovanni	Male	Politecnico di Torino	MATH	FP
Provan	Scott	Male	University of North Carolina	MATH	FP
Qu	Xianggui	Male	Oakland University	STAT	NRG
Reishus	Justin	Male	USC	COMP	NRG
Rempala	Greg	Male	Medical College of GA	STAT	FP
Rhodes	John	Male	University of Alaska Fairbanks	MATH	FP
Riccomagno	Eva	Female	Universita di Genova	STAT	FP
Rong	Yongwu	Male	George Washington University	MATH	FP
Savageau	Michael	Male	University of California	LIFE	FP
Schardl	Christopher	Male	University of Kentucky	LIFE	FP
Shen	Jian	Male	Texas State University	MATH	FP
Shiu	Anne	Female	University of California, Berkeley	MATH	NRG
Siebert	Heike	Female	DFG Research Center Matheon/ Free University Berlin	MATH	NRG
Sitharam	Meeram	Female	U Florida	COMP	FP
Slavkovic	Aleksandra	Female	Penn State University	STAT	NRG
Solhjo	Soroosh	Male	Johns Hopkins University School of Medicine	LIFE	NRG

Stigler	Brandilyn	Female	Mathematical Biosciences Institute	MATH	NRG
Stone	Eric	Male	North Carolina State University	STAT	NRG
Sturmfels	Bernd	Male	University of California	MATH	FP
Sullivant	Seth	Male	North Carolina State University	MATH	NRG
Szanto	Agnes	Female	NCSU	MATH	FP
Takemura	Akimichi	Male	University of Tokyo	STAT	FP
Thomas	Rene	Male	Université libre de Bruxelles	LIFE	FP
Tyler	Brett	Male	Virginia Polytechnic Institute and State University	LIFE	FP
Uhler	Caroline	Female	UC Berkeley	STAT	NRG
Veliz-Cuba	Alan	Male	Virginia Tech	MATH	NRG
Vera-Licona	Paola	Female	Rutgers University	MATH	NRG
Vince	Andrew	Male	University of Florida	MATH	FP
Volny	Frank	Male	Clemson University	MATH	NRG
Wang	Guanyu	Male	George Washington University	LIFE	FP
Watanabe	Sumio	Male	Tokyo Institute of Technology	MATH	FP
Wells	Benjamin	Male	North Carolina State University	STAT	NRG
Wolpert	Robert	Male	Duke U	STAT	FP
Wynn	Henry	Male	London School of Economics	STAT	NRG
Yamada	Richard	Male	University of Michigan	STAT	NRG
Yarahmadian	Shantia	Male	Indiana University, Molecular Biology Institute	MATH	NRG

Yasamin	Ahmad	Male	SAMSI	STAT	NRG
Yoshida	Ruriko	Female	University of Kentucky	STAT	NRG
Yoshida	RYo	Male	Institute of Statistical Mathematics	STAT	NRG
Yuster	Debbie	Female	DIMACS	MATH	NRG
Zhu	Mingfu	Male	Clemson University	MATH	NRG
Zou	Yi Ming	Female	U Wisconsin	MATH	FP
Zuk	Or	Male	Broad Institute of MIT and Harvard	MATH	NRG
Zwiernik	Piotr	Male	University of Warwick	STAT	NRG

Discrete Models in Systems Biology
Participant Summary
December 3-5, 2008

Participants	Male	Female	Unspecified	Faculty/Professional	New Researcher/Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	14	9	0	3	20	5	13	5	17	10
Unsuppted	12	7	0	7	12	9	7	3	7	2
SAMSI	1	1	0	0	2	1	1	0		

Discrete Models in Systems Biology
Workshop Participants
December 3-5, 2008

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Anderson	David	Male	University of Wisconsin - Madison	Mathematics	NRG
Chen	Wenjie	Female	UNC-CH	STAT	NRG
Demitrius	Lloyd	Male	Harvard	Statistics	FP
Dimitrova	Elena	Female	Clemson University	Mathematical Sciences	NRG

Elaydi	Saber	Male	Trinity University	Mathematics	FP
Friedrich	Thomas	Male	FU Berlin, Germany	Mathematics and Computer Science	NRG
Gao	Shuhong	Male	Clemson University	Mathematical Sciences	NRG
Hinkelmann	Franziska	Female	Virginia Bioinformatics Institute	Mathematics	NRG
Hosten	Serkan	Male	San Francisco State University and SAMSI	Mathematics	FP
Jarrah	Abdul Salam	Male	Virginia Tech	Virginia Bioinformatics Institute	NRG
Jenista	Michael	Male	Duke University	Math Department	NRG
Kondor	Imre Risi	Male	University College London	Gatsby Computational Neuroscience Unit	NRG
Lan	Ling	Female	Medical College of Georgia	Department of Biostatistics	NRG
Laubenbacher	Reinhard	Male	Virginia Tech	Virginia Bioinformatics Institute	FP
Lipan	Ovidiu	Male	University of Richmond	Physics and Mathematics	FP
Lu	Huitian	Male	South Dakota State University	STAT	FP
Macauley	Matthew	Male	Clemson University	Mathematical Sciences	NRG
Megraw	Molly	Female	Duke University	Bio Sci	NRG
Mitra	Indranil	Male	Clemson University	Mathematical Sciences	NRG
Mortveit	Henning	Male	Virginia Tech	Mathematics & Virginia Bioinformatics Institute	NRG
Owen	Megan	Female	SAMSI	STAT	NRG
Pawlikowska	Iwona	Female	Medical College of Georgia	Biostatistics	NRG
Piazza	Carla	Female	University of Udine	Mathematics and Computer Science	NRG
Provan	Scott	Male	Univ. North Carolina	Statistics and Op Research	FP

Rempala	Grzegorz A	Male	Medical College of Georgia	Biostatistics	FP
Sevim	Volkan	Male	Duke University	Physics	NRG
Shiu	Anne	Female	University of California Berkeley	Mathematics	NRG
Smith	James	Male	University of Warwick	Statistics	FP
Solhjo	Soroosh	Male	Johns Hopkins U School of Medicine	Biomedical Engineering	NRG
Stallmann	Tim	Male	Duke University	Mathematics	NRG
Stigler	Brandilyn	Female	Southern Methodist University	Mathematics	NRG
Stone	Eric	Male	North Carolina State University	Statistics	NRG
Sullivant	Seth	Male	North Carolina State University	Mathematics	NRG
Thakar	Juilee	Female	Pennsylvania State University	Physics	NRG
Thomas	Rachel	Female	Duke University	Mathematics	NRG
Tzeng	Jung-Ying	Female	NC State University	Statistics	FP
Ucar	Duygu	Female	Ohio State University	Computer Science and Engineering	NRG
Vera-Licona	Martha Paola	Female	Rutgers University	DIMACS and the Mathematics Department	NRG
Xu	Hongyan	Male	Medical College of Georgia	Biostatistics	NRG
Yamada	Richard	Male	University of Michigan	Applied Mathematics - Computational Biology	NRG
Yang	Hongxia	Male	Duke University	Statistics	NRG
Yarahmadian	Shantia	Male	Indiana University	Molecular Biology	NRG
Yasamin	Saeid	Male	SAMSI	STAT	NRG
Ye	Tianjun	Female	Georgia Tech	Mathematics	NRG

Algebraic Statistical Models

Participant Summary

January 15-17, 2009

Participants	Male	Female	Unspec- ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	10	3	0	3	10	8	4	1	12	7
Unsuppted	12	5	0	9	8	5	7	5	10	5
SAMSI	3	1	0	0	4	3	1	0		

Algebraic Statistical Models

Workshop Participants

January 15-17, 2009

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Allman	Elizabeth	Female	University of Alaska Fairbanks	Mathematics and Statistics	FP
Chen	Wenjie	Female	UNC-Chapel Hill	Statistics	NRG
Cox	Lawrence	Male	National Center for Health Statistics	Office of Research and Methodology	FP
Das	Sourish	Male	SAMSI, Duke University	Statistics	NRG
Dinwoodie	Ian	Male	Duke University	DSS	FP
Friedrich	Thomas	Male	FU Berlin, Germany	Mathematics and Computer Science	NRG
Garcia-Puente	Luis David	Male	Sam Houston State University	Mathematics and Statistics	NRG
Gupta	Shuva	Male	Florida State University	Statistics	NRG
Hara	Hisayuki	Male	University of Tokyo	Technology Management for Innovation	FP
Henmi	Masayuki	Male	The Institute of Statistical Mathematics	Mathematical Analysis and Statistical Inference	NRG
Hosten	Serkan	Male	San Francisco State University	Mathematics	NRG
Ke	Weiming	Male	South Dakota State University	Mathematics and Statistics	NRG

Lauritzen	Steffen	Male	University of Oxford	Statistics	FP
Maruri-Aguilar	Hugo	Male	London School of Economics	Statistics	NRG
Morton	Jason	Male	Stanford University	Mathematics	NRG
Owen	Megan	Female	SAMSI	Mathematics	NRG
Petrovic	Sonja	Female	University of Illinois at Chicago	Mathematics Statistics and Computer Science	NRG
Pistone	Giovanni	Male	Politecnico di Torino	DIMAT (Mathematics)	FP
Rhodes	John	Male	University of Alaska Fairbanks	Mathematics and Statistics	FP
Riccomagno	Eva	Female	University of Genova	Statistics	FP
Richards	Donald	Male	Penn State University	Statistics	FP
Richardson	Thomas	Male	University of Washington	Statistics	FP
Romer	Megan	Female	Penn State University	Statistics	NRG
Sheridan	Paul	Male	Tokyo Institute of Technology	Graduate School of Information and Computing Sciences	NRG
Stokes	Erik	Male	Michigan Technological University	Mathematical Sciences	NRG
Sullivant	Seth	Male	North Carolina State University	Mathematics	NRG
Takemura	Akimichi	Male	University of Tokyo	Graduate School of Information Science and Technology	FP
Tian	Jin	Male	Iowa State University	Department of Computer Science	NRG
Tzeng	Jung-Ying	Female	NC State University	Statistics	FP
Xiao	Han	Male	University of Chicago	Department of Statistics	NRG
Xing	Chuanhua	Female	Duke University	Biology	NRG

Yasamin	Saeid	Male	SAMSI	Statistics	NRG
Yoshida	Ruriko	Female	University of Kentucky	Statistics	NRG
Yoshida	Ryo	Male	SAMSI	Statistics	NRG

Molecular Evolution and Phylogenetics

Participant Summary

April 2-3, 2009

Participants	Male	Female	Unspecified	Faculty/Professional	New Researcher/Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	10	8	0	4	14	4	7	7	17	11
Unsuppted	10	9	0	8	11	6	7	6	7	4
SAMSI	3	1	0	1	3	2	2	0		

Molecular Evolution and Phylogenetics

Workshop Participants

April 2-3, 2009

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Ane	Cecile	Female	University of Wisconsin - Madison	Departments of Statistics and of Botany	NRG
Arias	Tatiana	Female	U Missouri	Biological Sci	NRG
Bloomquist	Erik	Male	U California, Los Angeles	Biostatistics	NRG
Chifman	Julia	Female	University of Kentucky	Mathematics	NRG
Dickenstein	Alicia	Female	Universidad de Buenos Aires	Dto. de Matematica, FCEN	FP
Durak	M. Zeki	Male	Cornell University	Department of Food Science and Technology	NRG
Elissaveta	Arnaoudova	Female	University of Kentucky	Computer Science	FP
Fernandez-Sanchez	Jesus	Male	Universitat Politecnica de Catalunya	Matematica Aplicada I	NRG
Gremaud	Pierre	Male	SAMSI	Math	FP

Gross	Kevin	Male	North Carolina State University	Statistics	NRG
Hinkelmann	Franziska	Female	Virginia Bioinformatics Institute	Mathematics	NRG
Hodge	Terrell	Female	W Michigan U	Math	FP
Huggins	Peter	Male	Carnegie Mellon University	Computational Biology	NRG
Jaromczyk	Jerzy	Male	University of Kentucky	Computer Science	FP
Kim	Junhyong	Male	University of Pennsylvania	Penn Genome Frontiers Institute	FP
Koelle	Katia	Female	Duke University	Biology	NRG
Kubatko	Laura	Female	Ohio State University	Statistics and Evolution, Ecology, & Organismal Biology	FP
Kuo	Lynn	Female	University of Connecticut	Statistics	FP
Lam	Fumei	Female	UC Davis	Math	NRG
Laubenbacher	Reinhard	Male	Virginia Tech	Statistics	FP
Matsen	Frederick	Male	University of California, Berkeley	Life	NRG
Owen	Megan	Female	SAMSI	Math	NRG
Perez Millan	Mercedes	Female	Universidad de Buenos Aires	Math	NRG
Petrovic	Sonja	Female	University of Illinois at Chicago	Mathematics Statistics and Computer Science	NRG
Provan	Scott	Male	University of North Carolina	Statistics and Operations Research	FP
Reishus	Dustin	Male	University of Southern California	Computer Science	NRG
Schardl	Christopher	Male	U Kentucky	Life	FP
Shiu	Anne	Female	UC Berkeley	Math	NRG
Stone	Eric	Male	North Carolina State University	Statistics	NRG

Sullivant	Seth	Male	North Carolina State University	Mathematics	NRG
Sumner	Jeremy	Male	University of Tasmania	School of Maths and Physics	NRG
Tannor	David	Male	W Michigan University, Kalamazoo	Mathematics	NRG
Thorne	Jeffrey	Male	North Carolina State University	Genetics and Statistics	FP
Warnow	Tandy	Female	University of Texas	Math	FP
Xing	Julia Chuanhua	Female	Duke University	Department of Biology	NRG
Yarahmadian	Shantia	Male	Indiana Molecular Biology Institute	Biology	NRG
Yasamin	Saeid	Male	SAMSI	Statistics	NRG
Yellick	Jason	Male	SAMSI	Statistics	NRG
Yoshida	Ruriko	Female	University of Kentucky	Statistics	NRG
Yuan	Hsiang-yu	Male	Duke University	Computational Biology and Bioinformatics	NRG
Zhao	Yichuan	Male	Georgia State University	Mathematics and Statistics	NRG

Algebra Transition Workshop
Participant Summary
June 18-20, 2009

Participants	Male	Female	Unspecified	Faculty/Professional	New Researcher/Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	11	8	0	7	12	8	11	0	14	9
Unsuppted	8	4	0	4	8	5	6	1	8	4
SAMSI	1	1	0	0	2	1	1	0		

Algebra Transition Workshop

Workshop Participants

June 18-20, 2009

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Arnaoudova	Elissaveta	Female	University of KY	STAT	NRG
Dimitrova	Elena	Female	Clemson University	MATH	NRG
Dinwoodie	Ian	Male	Duke University	STAT	FP
Evans	Christina	Female	George Washington University	MATH	NRG
Fontana	Roberto	Male	Politecnico di Torino	STAT	FP
Friedrich	Thomas	Male	Freie Universitaet Berlin	MATH	NRG
Garcia-Puente	Luis David	Male	Sam Houston State University	STAT	NRG
Gnacadjia	Gilles	Male	Amgen	MATH	FP
Hara	Hisayuki	Male	University of Tokyo	STAT	FP
Hodge	Terrell	Female	Western Michigan University	MATH	FP
Huggins	Peter	Male	Carnegie Mellon University	STAT	NRG
Kidwell	Paul	Male	Purdue University	STAT	NRG
Laubenbacher	Reinhard	Male	Virginia Tech	STAT	FP
Mao	Yue	Male	Clemson University	MATH	NRG
Murrugarra Tomairo	David	Male	Virginia Tech	MATH	NRG
Owen	Megan	Female	SAMSI and NCSU	MATH	NRG
Petrovic	Sonja	Female	University of Illinois at Chicago	MATH	NRG
Pistone	Giovanni	Male	Politecnico di Torino	MATH	FP

Provan	Scott	Male	University of North Carolina	MATH	FP
Rong	Yongwu	Male	George Washington University	MATH	FP
Shiu	Anne	Female	University of California, Berkeley	MATH	NRG
Siebert	Heike	Female	DFG Research Center Matheon/ Free University Berlin	MATH	NRG
St. John	Katherine	Female	UCLA	MATH	NRG
Stigler	Brandy	Female	Southern Methodist University	MATH	NRG
Sullivant	Seth	Male	North Carolina State University	MATH	NRG
Sun	Xiaoqian	Male	Clemson University	STAT	NRG
Uyenoyama	Marcy	Female	Duke University	LIFE	NRG
Wynn	Henry	Male	London School of Economics	STAT	FP
Yamada	Richard	Male	University of Michigan	STAT	NRG
Yasamin	Ahmad	Male	SAMSI	STAT	NRG
Yoshida	Ruriko	Female	University of Kentucky	STAT	FP

❖ **Summer Program on Psychometrics**

Summer Program on Psychometrics

Participant Summary

July 7-17, 2009

Participants	Male	Female	Unspecified	Faculty	New Researcher/ Student	Stat/Math	Other/Unspecified	Number of Institutions Represented	Number of States Represented
Supported	10	10	0	6	14	8	12	15	12
Unsuppted	37	13	0	28	22	20	30	25	7
SAMSI	1	0	0	1	0	1	0		

Summer Program on Psychometrics

Workshop Participants

July 7-17, 2009

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Alonzo	Alicia	Female	University of Iowa	Teaching & Learning	FP
Atkinson	Thomas	Male	Memorial Sloan Kettering Cancer	Statistics	NRG
Banks	David	Male	Duke University	Statistical Science	FP
Basch	Ethan	Male	Memorial Sloan Kettering Cancer	Other	FP
Benness	George Anthony	Male	Fordham University	Psychology	NRG
Bollen	Ken	Male	University of North Carolina	Sociology	FP
Burdick	Donald	Male	MetaMetrics, Inc.	Statistics	FP
Cai	Li	Male	University of California, L.A.	GSE&IS and Psychology	NRG
Cao	Jing	Female	Southern Methodist U	Statistics	NRG
Chahine	Saad	Male	University of Toronto	Human Development and Applied Psychology	NRG
Cheng	Ying	Female	University of Notre Dame	Psychology	NRG
Cho	Sun-Joo	Male	University of California, Berkeley	Statistics	NRG
Cleeland	Charlie	Male	University of M. D. Anderson Cancer Center	Life	FP
Cooke	Ben	Male	Duke University	Academic Resource Center	NRG
Cui	Ying	Female	University of Alberta	Educational psychology	NRG
Das	Sourish	Male	SAMSI, Duke University	Statistics	NRG
de la Torre	Jimmy	Male	Rutgers University	Education	NRG
Fairclough	Diane	Female	University of Colorado Denver, School of Public Health	Biostatistics and Informatics	FP

Feldman	Betsy	Female	University of California, Berkeley	Graduate School of Education	NRG
Finkelman	Matthew	Male	Tufts University School of Dental Medicine	Statistics	NRG
Fuentes	Jose	Male	San Diego State University	Mathematics and Statistics	NRG
Gilligan	Theresa	Female	RTI Health Solutions	Patient Reported Outcomes	NRG
Harrell	Leigh	Female	Virginia Tech	Statistics	NRG
Hartigan	Brian	Male	University of North Carolina Wilmington	Psychology	NRG
Henson	Robert	Male	University of North Carolina, Greensboro	Statistics	NRG
Hill	Cheryl	Female	RTI Health Solutions	Patient Reported Outcomes	NRG
Huff	Kristen	Female	College Board	R&D	FP
Jang	Eunice	Female	Ontario Institute	Education	NRG
Johnson	Matthew	Male	Columbia U	Statistics	FP
Johnson	Valen	Male	U Texas	Statistics	FP
Karelitz	Tzur	Male	Education Development Center, Inc.	Center for Science Education	FP
Lam	Tsz Cheung	Male	Rutgers University	Educational Psychology	NRG
Levy	Roy	Male	Arizona State University	Education	NRG
Loye	Nathalie	Female	University of Montreal	Administration et fondements de l'Éducation	NRG
Lu	Jun	Male	American U	Statistics	FP
Madden	James	Male	Louisiana State University	Mathematics	FP
McGill	Mike	Male	Virginia Tech	Education	NRG
McGowan	Herle	Female	North Carolina State University	Statistics	NRG
McLeod	Lori	Female	RTI Health Solutions	Patient Reported Outcomes	FP

Morales	Knashawn	Female	University of Pennsylvania	Biostatistics and Epidemiology	FP
Nelson	Lauren	Female	RTI Health Solutions	Patient Reported Outcomes	FP
Nugent	Rebecca	Female	Carnegie Mellon University	Statistics	NRG
Peruggia	Mario	Male	Ohio State University	Statistics	FP
Price	Mark	Male	RTI Health Solutions	Patient Reported Outcomes	NRG
Rapkin	Bruce	Male	Albert Einstein College of Medicine of Yeshiva University	Div of Community Collaboration & Implementation	FP
Rijmen	Frank	Male	Educational Testing Service	Psychology	FP
Rivera-Medina	Carmen	Female	University of Puerto Rico	Institute of Psychological Research	FP
Rouder	Jeff	Male	U Missouri	Psychology	FP
Rupp	Andre	Male	University of Maryland	EDMS	FP
Schwartz	Carolyn	Female	Tufts University, School of Medicine	Medicine and Orthopaedic Surgery	FP
Sheng	Yanyan	Female	Southern Illinois University	Other	NRG
Sinharay	Sandip	Male	Educational Testing Service	Sociology	FP
Speckman	Paul	Male	U Missouri	Statistics	FP
Stenner	Jack	Male	MetaMetrics	Education	FP
Sun	Dongchu	Male	Missouri	Statistics	FP
Swartz	Richard	Male	U Texas	Other	FP
Tatsouka	Curtis	Male	Case Western	Statistics	FP
Thissen	David	Male	UNC	Statistics	FP
Tractenberg	Rochelle	Female	Georgetown University Medical Center	Neurology	NRG

Uenlue	Ali	Male	University of Augsburg	Institute of Mathematics	FP
Van Zandt	Trish	Female	Ohio State University	Sociology	FP
von Davier	Matthias	Male	Educational Testing Service	Statistics	NRG
Wang	Jun	Male	North Carolina State University	Statistics Department	NRG
Wang	Xiaojing	Male	Duke University	Statistical Science	NRG
Williams	Valerie	Female	RTI Health Solutions	Patient Reported Outcomes	FP
Wilson	Mark	Male	UC Berkeley	Education	FP
Wu	Hao	Male	Ohio State University	Psychology	NRG
Yue	Yu	Male	Baruch College, City University of New York	Statistics and CIS	NRG
Zhang	Song	Male	U Texas	Comp Sci	NRG
Zhang	Jingshun	Male	University of Toronto	Education	NRG

❖ **2009-10 PROGRAM EVENTS THROUGH JULY 2009**

❖ **Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change: Summer School on Spatial Statistics**

Spatial Summer School

Participant Summary

July 28 – August 1, 2009

Participants	Male	Female	Unspecified	Faculty/Professional	New Researcher/Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	13	11	1	4	21	22	1	2	20	13
Unsuported	8	9	0	1	16	12	1	4	9	4
SAMSI	1	0	0	0	1	1	0	0		

Spatial Summer School

Workshop Participants
July 28 – August 1, 2009

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Banerjee	Sudipto	Male	University of Minnesota	Biostats	FP
Bornn	Luke	Male	University of British Columbia	Statistics	NRG
Chang	Xiaohui		University of Chicago	Statistics	NRG
Chen	Lisha	Famale	Yale University	Assistant Professor of Statistics	NRG
Chen	Jiehua	Female	Columbia University	Statistics	NRG
Das	Sourish	Male	SAMSI -- Duke University	Statistics	NRG
Furrer	Reinhard	Male	University of Zurich	Math	FP
Gunning	Patricia	Female	NISS	Statistics	NRG
Guo	Ruixin	Male	University of Missouri, Columbia	Statistics	NRG
Hammerling	Dorit	Female	University of Michigan	Environmental Engineering, Phd	NRG
Herring	Amy	Female	University of North Carolina	Biostatistics	NRG
Holt	Nathan	Male	University of Florida	Statistics	NRG
Homrighausen	Darren	Male	Carnegie Mellon University	Statistics	NRG
Hughes	John	Male	Pennsylvania State University	Statistics, PhD	NRG
Hurtado Rua	Sandra M	Female	University of Connecticut	Statistics	NRG
Joo	Eun	Male	Duke University	Statistics	NRG
Katzoff	Myron	Male	George Washington University	Mathematical Statistics	FP
Kim	Harry	Male	University of California, Berkeley	Statistics	NRG

Kolovos	Alexander	Male	SAS	N/A	NRG
Liang	Ye	Male	University of Missouri	Statistics	NRG
Liu	Yajun	Female	University of Missouri	Statistics	NRG
Lopiano	Kenneth	Male	University of Florida	Statistics	NRG
Nychka	Doug	Male	NCAR	Statistics	FP
Rister	Krista	Female	Texas A&M University	Statistics	NRG
Rosenberg	David	Male	University of California, Berkeley	Statistics, PhD	NRG
Sain	Stephen	Male	NCAR	Statistics	FP
Schmaltz	Chester	Male	University of Missouri	Statistics, MA	NRG
Sharma	Bhawna	Female	North Carolina State University	Statistics	NRG
Shen	Ling	Female	University of Colorado, Boulder	Geography	NRG
Stark	Glenn	Male	University of New Mexico	Statistics	NRG
Sun	Ying	Female	Texas A&M University	Statistics	NRG
Torres	Pedro	Male	North Carolina State University	Statistics	NRG
Toto	Criselda	Female	NISS	Statistics	NRG
Wang	Fangpo	Female	Duke University	Statistics	NRG
Wang	Jianqiang	Female	NISS	Statistics	NRG
Wang	Ziwei	Female	University of California, Santa Cruz	Statistics	NRG
Wang	Xia	Female	University of Connecticut	Statistics	NRG
Wei	Rong	Male	University of Wisconsin, Madison	Animal Sciences	NRG
Wilson	James	Male	Clemson University	Statistics	NRG

Xue	Yun	Female	Michigan State University	Statistics	NRG
Yang	Hongxia	Female	Duke University	Statistics	NRG
Zhao	Yingqi	Female	University of North Carolina	Biostatistics	NRG
Zhuang	Lili	Female	Ohio State University	Statistics	NRG

❖ **Education and Outreach Program**

SAMSI/CRSC Industrial Mathematical & Statistical Workshop for Graduates

Participant Summary

July 21-29, 2008

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Institutions Represented	Number of States Represented
Supported	20	13	0	0	33	33	0	23	18
Unsuppted	2	2	0	0	4	2	2	2	2
SAMSI	0	0	0	0	0	0	0		

SAMSI/CRSC Industrial Mathematical & Statistical Workshop for Graduates

Workshop Participants

July 21-29, 2008

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Akapame	Sydney	Male	Montana State University	STAT	NRG
Beavers	Daniel	Male	Baylor University	STAT	NRG
Bhattacharya	Abhishek	Male	U Arizona	STAT	NRG
Blanton	Jacob	Male	Louisiana State University	MATH	NRG
Cargill	Daniel	Male	New Jersey Int. of Tech.	MATH	NRG
Causley	Matthew	Male	New Jersey Institute of Technology	MATH	NRG
Chai	Juanjuan	Female	Indiana U	MATH	NRG
Chalmers	Nancy	Female	University of South Carolina	STAT	NRG

Chen	Wei	Male	Johns Hopkins University	STAT	NRG
Cui	Jintao	Male	Louisiana State University	MATH	NRG
Daley	Caitlin	Female	NCSU	STAT	NRG
Gewecke	Nicholas	Male	University of Tennessee	MATH	NRG
Giulio	Genovese	Male	Dartmouth College	MATH	NRG
Hofer	Marian	Female	U of California	STAT	NRG
Holm	Kathleen	Female	North Carolina State University	STAT	NRG
Jacob	Jobby	Male	Clemson University	MATH	NRG
Joshi	Adarsh	Male	Texas A&M University	STAT	NRG
Joshi	Yogesh	Male	New Jersey Institute of Technology	MATH	NRG
Kaur	Manmeet	Female	New Jersey Insitute Of Technology	MATH	NRG
Klein	Viviane	Female	Oregon State University	MATH	NRG
Markova	Denka	Female	Baylor University	STAT	NRG
Njoh	Linda	Female	Baylor U	STAT	NRG
Peh	Lu Ee	Female	University of Dayton	MATH	NRG
Qi	Peng	Male	University of Wisconsin-Madison	MATH	NRG
Qiu	Yu	Female	Iowa State University	ENGG	NRG
Robledo	Lucinda	Female	University of California, Santa Cruz	MATH	NRG
Singh	Shashi	Male	U of Hawaii	MATH	NRG
Soloveva	Svetlana	Female	Moscow State University	MATH	NRG
Stanley	Jeffrey	Male	Texas A&M University	STAT	NRG

Tan	Shuguang	Male	University of Florida	MATH	NRG
Thompson	Clay	Male	NCSU	BMA	NRG
Torabi	Solmaz	Female	University of California, Irvine	ENGG	NRG
Walters	Mark	Male	University of South Carolina	MATH	NRG
Wanner	Nathan	Male	North Carolina State University	MATH	NRG
Wright	Justin	Male	NCSU	MATH	NRG
Yan	Bokai	Male	U Wisconsin-Madison	MATH	NRG
Yin	Shuxin	Female	Auburn University	STAT	NRG

Undergraduate Two-Day Workshop

Participant Summary

October 31-November 1, 2008

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspe-cified	Number of Institutions Represented	Number of States Represented
Supported	7	7	0	0	14	14	0	14	10
Unsuppted	7	6	0	0	13	11	2	4	2
SAMSI	9	5	0	14	0	14	0		

Undergraduate Two-Day Workshop

Workshop Participants

October 31-November 1, 2008

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Agbai	Didi	Male	Benedict College	Finance	S
Ayesh	Afeefa	Male	Meredith College	Mathematics (& Statistics minor)	S
Bain	Melanie	Female	SAMSI & UNC	STAT	A

Bishwal	Jaya	Male	SAMSI	STAT	A
Cao	Yue	Male	U Tennessee	Electrical Engineering and Mathematics	S
Cornebise	Julien	Male	SAMSI	STAT	A
Daley	Lynsie	Female	Utah State University	Statistics	S
Das	Sourish	Male	SAMSI & Duke	STAT	A
Dreiding	Rebecca	Female	Virginia Commonwealth U	Mathematical Sciences	S
Ezerioha	Nnadozie	Male	Benedict College	Physics/Eng	S
Petralia	Francesca	Female	SAMSI & Duke	STAT	A
Gerke	Travis	Male	University of Florida	Statistics and Mathematics	S
Gordon-Wright	Rachael	Female	NCSU	Mathematics, Statistics	S
Green	Nathan	Male	SAMSI	STAT	A
Gremaud	Pierre	Male	NCSU & SAMSI	MATH	A
Ji	Chunlin	Female	SAMSI	STAT	A
Kosel	Alison	Female	Pomona College	mathematics, neuroscience	S
Kwan	Kevin	Male	Carnegie Mellon University	Business Administration (Finance), Statistics	S
Liu	Zack	Male	University of Texas	Mathematics (Dean's Scholars) and Economics	S
Macaro	Christian	Male	SAMSI	STAT	A
Manolopoulou	Ioanna	Female	SAMSI	STAT	A
Moore	Russell	Male	U Wisconsin	Physics and Math	S
Myers	Ashley	Female	NCSU	Statistics	S

Payne	Rebecca	Female	Pomona College	Mathematics and Psychology (Double Major)	S
Piarulli	Kevin	Male	Ithaca College	Mathematics, CS Minor	S
Popovic	Natalija	Female	University of Illinois	Mathematics	S
Reeson	Craig	Male	Duke University	STAT	S
Schott	Sarah	Female	SAMSI	Mathematics	A
Shi	Ce	Male	Duke University	Statistical Science, Economics, and Mathematics	S
Shi	Minghui	Female	University of Illinois	Mathematics and Economics	S
Tillett	Shannon	Female	Clemson University	Mathematical Sciences	S
Tronetti	Alexandra	Female	Carnegie Mellon University	Economics-Statistics	S
Tutt	Andrew	Male	Duke University	Mathematics	S
Voss	Jochen	Male	SAMSI	STAT	A
White	Gentry	Male	SAMSI/ NCSU	STAT	A
Wilson	Anna	Female	UNC	Atmospheric Science (Climatology) & Math (Applied)	S
Worley	Mitchell	Male	Wofford College	Double Major in Chemistry and Mathematics	S
Zhang	Rui	Female	University of Michigan	Statistics; Cellular and Molecular Biology	S
Zhang	Baqun	Male	SAMSI / NCSU	STATS	A
Zheng	Perry	Male	Duke University	Mathematics / Economics	S
Zimmer	Stephanie	Female	NCSU	Statistics	S

Blackwell-Tapia Conference
Participant Summary
November 15-16, 2008

Participants	Male	Female	Unspec- ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	25	21	1	17	30	13	34	0	29	17
Unsuppted	21	10	0	17	14	8	18	5	16	3
SAMSI	1	0	0	1	0	1	0	0		

Blackwell-Tapia Conference
Workshop Participants
November 15-16, 2008

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Alexander	Clemontina	Female	NCSU	STAT	NRG
Arellano	John	Male	Rice U	MATH	NRG
Austin	Joshua	Male	U Maryland	MATH	NRG
Basu	Kanadpriya	Male	U South Carolina	MATH	NRG
Bayarri	Susie	Female	U Valencia	STAT	FP
Berger	James	Male	SAMSI	STAT	FP
Bridges	Clifford	Male	U Maryland	MATH	NRG
Brizzotti	Murilo	Male	NCSU	STAT	NRG
Buckmire	Ron	Male	Occidental College	MATH	FP
Carden	Russell	Male	Rice U	MATH	NRG
Carraminana	Rodrigo	Male	U Illinois	MATH	FP
Castillo-Chavez	Carlos	Male	Arizona State	MATH	FP
Catepillan	Ximena	Female	Millersville U	MATH	FP

Ceniceros	Hector	Male	U California	MATH	FP
Chism	Lyrial	Female	U Mississippi	MATH	NRG
Cintron	Ariel	Male	NCSU	MATH	NRG
Cline	Jon	Male	Case Western	LIFE	NRG
Colbert-Kelly	Sean	Male	Purdue	MATH	NRG
Coleman	Deidra	Female	NCSU	STAT	NRG
Davies	Kalatu	Female	Rice U	STAT	NRG
Enriquez	Marco	Male	Rice U	MATH	NRG
Gallegos	Angela	Female	Tulane U	MATH	NRG
Goins	Edray	Male	Purdue U	MATH	FP
Golubitsky	Martin	Male	Ohio State	MATH	FP
Gonzalez	Oscar	Male	U Texas	MATH	FP
Guevara	Alvaro	Male	Louisiana State	MATH	NRG
Harris	Leona	Female	College of NJ	MATH	FP
Hernandez	Troy	Male	U Illinois	STAT	NRG
Hicks	Illya	Male	Rice U	MATH	NRG
Horne	Rudy	Male	Florida State	MATH	NRG
Houston	Johnny	Male	Elizabeth City State U	MATH	FP
Huerta	Gabriel	Male	U New Mexico	STAT	FP
Hughes-Oliver	Jacqueline	Female	NCSU	STAT	FP
Jackson	Monica	Female	American U	STAT	NRG

Jennings	Otis	Male	Duke U	BUS	NRG
Jimenez	Silvia	Female	Louisiana State U	MATH	NRG
Kemajou	Elisabeth	Female	Southern Illinois U	MATH	NRG
Konate	Souleymane	Male	U Central Florida	MATH	NRG
Laubenbacher	Reinhard	Male	Virginia Tech	MATH	FP
Light	Emily	Female	U Michigan	STAT	NRG
Martinez	Josue	Male	Texas A&M	STAT	NRG
Massey	William	Male	Princeton	ENG	FP
Megginson	Robert	Male	U Michigan	MATH	FP
Melara	Luis	Male	Shippensburg U	MATH	FP
Meza	Juan	Male	LBNL	MATH	FP
Moore	Tanya	Female	Building Diversity in Science	OTHER	FP
Morgan	Carolyn	Female	Hampton U	STAT	FP
Morgan	Morris	Male	Hampton U	ENG	FP
Munoz Maldonado	Yolanda	Female	Michigan Tech	STAT	FP
Murillo	David	Male	Arizona State	MATH	NRG
Nkengla	Mechie	N/A	U of Illinois	MATH	NRG
Oluyede	Broderick	Male	Georgia Southern	STAT	FP
Ortega	Omayra	Female	Arizona State	MATH	FP
Pantula	Sastry	Male	NCSU	STAT	FP
Papakonstantinou	Joanna	Female	Rice U	MATH	NRG

Pararai	Mavis	Female	Indiana U of PA	STAT	FP
Patterson	Sam	Male	Carleton College	MATH	FP
Ramos	Jaime	Male	Rice U	STAT	NRG
Reyna Jr.	Nabor	Male	Rice U	MATH	NRG
Rezaei	Mahmoud	Male	Clemson U	STAT	NRG
Rios-Doria	Daniel	Male	Arizona State	MATH	NRG
Robbins	Danielle	Female	NCSU	MATH	NRG
Sancier-Barbosa	Flavia	Female	Southern Illinois	MATH	NRG
Sellers	Kimberly	Female	Georgetown U	STAT	FP
Shakiban	Cheri	Female	U Minnesota	MATH	FP
Sifuentes	Josef	Male	Rice U	MATH	NRG
Simms	Anthony	Male	Meyerhoff Scholar	MATH	NRG
Sircar	Treena	Female	U South Carolina	MATH	NRG
Somersille	Stephanie	Female	U California, Berkeley	MATH	NRG
Tapia	Richard	Male	Rice U	MATH	FP
Teguia	Alberto	Male	Duke U	MATH	NRG
Thornton	Timothy	Male	U California, SF	STAT	FP
Tullius	Toni	Female	Rice U	MATH	NRG
Turner	Jesse	Male	Rice U	STAT	NRG
Valdez-Jasso	Daniela	Female	NCSU	MATH	NRG
Villalobos	Cristina	Female	U Texas	MATH	FP

Washington	Talitha	Female	U Evansville	MATH	FP
Wilson	Ulrica	Female	Morehouse College	MATH	FP
Woldegebreal	Eyerusalem	Female	U St. Thomas	MATH	NRG

Undergraduate Two-Day Workshop

Participant Summary

February 27-28, 2009

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Institutions Represented	Number of States Represented
Supported	19	8	0	0	27	23	4	19	13
Unsuppted	6	1	0	7	0	7	0	3	2
SAMSI	3	1	0	4	0	4	0		

Undergraduate Two-Day Workshop

Workshop Participants

February 27-28, 2009

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Adams	John	Male	Virginia Commonwealth U	Statistics	S
Bostwick	Michael	Male	University of Connecticut	Statistics	S
Brouillette	Stephen	Male	Louisiana State University	Mathematics	S
Colbert	Cory	Male	Virginia Commonwealth University	Mathematics	S
Culiuc	Amalia	Female	Mount Holyoke College	Mathematics	S
Diaz	Alexander	Male	Sam Houston State University	Math & Stats	S
Dillon	Matthew	Male	University of North Carolina Wilmington	Mathematics	S
Dinwoodie	Ian	Male	Duke University	Statistics	A
Galgon	Geoff	Male	California Institute of Technology	Mathematics	S

Garcia-Puente	Luis	Male	Sam Houston State University	Math & Stats	A
Gliner	Genna	Female	University of California, Davis	Applied Mathematics	S
Gopalan	Giri	Male	California Institute of Technology	Applied Mathematics	S
Gremaud	Pierre	Male	SAMSI	Mathematics	A
Hopkins	David	Male	Colorado State University	Math General and Applied	S
Howard	Andrew	Male	Sam Houston State University	Computer Science & Sociology; minor in Mathematics	S
Ilic	Ognjen	Male	Harvard University	Physics and Mathematics	S
Ivan	Radu-Andrei	Male	University of Massachusetts Amherst	Electrical Engineering	S
Keys	Kevin	Male	University of Arizona	Mathematics	S
Kottmeyer	Alexa	Female	Mount Holyoke College	Mathematics	S
Lauer	Abigail	Female	Elon University	Mathematics	S
Lee	Jinwoo	Male	California Institute of Technology	Biology	S
Nielsen	Mark	Male	Utah State University	Mathematics and Statistics	S
Owen	Megan	Female	SAMSI	Mathematics	A
Pankow	Anne	Female	University of Washington	Statistics, Mathematics, and Economics	S
Pistone	Giovanni	Male	Politecnico di Torino	Statistics	A
Rush	Cynthia	Female	University of North Carolina	Mathematics, Statistics and Operations Research	S
Sadowski	Peter	Male	California Institute of Technology	Computer Science	S
Spielvogel	Sarah	Female	Sam Houston State University	Math, Spanish	S
Stigler	Brandy	Female	Southern Methodist University	Mathematics	A
Sullivant	Seth	Male	North Carolina State University	Mathematics	A

Thorne	Jefferey	Male	North Carolina State University	Statistics	A
Vishniakou	Siarhei	Male	Cornell University	Engineering Physics	S
Wells	Ben	Male	North Carolina State University	Statistics	A
White	Gentry	Male	SAMSI and North Carolina State University	Statistics	A
Yasamin	Saeid	Male	SAMSI	Statistics	A
Young	Andrew	Male	Appalachian	Applied Mathematics	S
Zagardo	Michelle	Female	Mount Holyoke College	Mathematics	S
Zheng	Perry	Male	Duke University	Math/Econ/Computer Science	S

Graduate Student Probability Workshop

Participant Summary

May 1-3, 2009

Participants	Male	Female	Unspecified	Faculty	Student	Number of Institutions Represented	Number of States Represented
All	84	31	0	6	109	37	22
SAMSI	0	0	0	0	0		

Graduate Student Probability Workshop

Workshop Participants

May 1-3, 2009

Last Name	First Name	Gender	Affiliation	Status
Aldous	David	Male	University of California – Berkeley	S
Almada	Sergio	Male	Georgia Institute of Technology	S
Al-sharadqah	Ali	Male	University of Alabama - Birmingham	S
Babatunde	Ayilara Ibrahim	Male	Obafemi Awolowo University	S
Baek	Changryong	Male	UNC - Chapel Hill	S
Balachandran	Prakash	Male	Duke University	S

Bichuch	Maxim	Male	Carnegie Mellon University	S
Blair-Stahn	Nathaniel	Male	University of Washington	S
Bloemendal	Alex	Male	University of Toronto	S
Borysov	Petro	Male	UNC - Chapel Hill	S
Budhiraja	Amarjit	Male	UNC - Chapel Hill	A
Burr	Meredith	Female	Tufts University	S
Cabanski	Chris	Male	UNC - Chapel Hill	S
Canepa	Elena Cristina	Female	Carnegie Mellon University	S
Cecil	Matt	Male	University of Connecticut	S
Chakrabarty	Arijit	Male	Cornell University	S
Chavez	Esteban	Male	Duke University	S
Chen	Hua	Male	North Carolina State University	S
Chen	Li	Female	Oregon State University	S
Chen	Ao	Male	University of Illinois - Urbana-Champaign	S
Chen	Jiang	Male	UNC - Chapel Hill	S
Chen	Xia	Male	University of Tennessee - Knoxville	S
Chronopoulou	Alexandra	Female	Purdue University	S
Cisewski	Jessi	Female	UNC - Chapel Hill	S
Corwin	Ivan	Male	Courant Institute, New York University	S
Crosskey	Miles	Male	Duke University	S
Deshpande	Amogh	Male	North Carolina State University	S

Djordjevic	Jasmina	Female	University of Niš	S
Esunge	Julius	Male	Louisiana State University	S
Fang	Ming	Male	University of Minnesota	S
Fellouris	Georgios	Male	Columbia University	S
Feng	Yaqin	Female	UNC - Charlotte	S
Ganguly	Arnab	Male	University of Wisconsin - Madison	S
Georgiou	Nicos	Male	University of Wisconsin - Madison	S
Gong	Ruoting	Male	Georgia Institute of Technology	S
Grieves	Justin	Male	University of Tennessee - Knoxville	S
Guettes	Sabrina	Female	University of Wisconsin - Madison	S
Guo	Xiaoqin	Male	University of Minnesota	S
Gupta	Ankit	Male	University of Wisconsin - Madison	S
Haidari	Arman	Male	Massachusetts Institute of Technology	S
Hannig	Jan	Male	UNC - Chapel Hill	A
Hao	Xuemiao	Male	University of Iowa	S
Hoffmeyer	Allen	Male	Georgia Institute of Technology	S
Hu	Ken	Male	Massachusetts Institute of Technology	S
Hu	Xueying	Female	University of Michigan	S
Jackson	Aaron	Male	Duke University	S
Ji	Chuanshu	Male	UNC - Chapel Hill	A
Jiang	Yunjiang	Male	Stanford University	S

Karabash	Dmytro	Male	Courant Institute, New York University	S
Kaupila	Helena	Female	Columbia University	S
Kilanowski	Philip	Male	Ohio State University	S
Kim	Kunwoo	Male	University of Illinois - Urbana-Champaign	S
Klimova	Alexandra	Female	Canegie Mellon University	S
Kobayashi	Kei	Female	Tufts University	S
Kolba	Tiffany	Female	Duke University	S
Leadbetter	Ross	Male	UNC - Chapel Hill	A
Lee	Chia Ying	Female	Brown University	S
Lee	Seonjoo	Female	UNC - Chapel Hill	S
Lee	Mihee	Female	UNC - Chapel Hill	S
Lei	Pedro	Male	University of Kansas	S
Li	Zhongyang	Female	Brown University	S
Li	Zhiqiang	Male	University of Tennessee - Knoxville	S
Lin	Hao	Male	University of Wisconsin - Madison	S
Little	Anna	Female	Duke University	S
Liu	Xin	Female	UNC - Chapel Hill	S
Luo	Shishi	Female	Duke University	S
Lyons	Russell	Male	Indiana University	S
Ma	Jinyong	Male	Georgia Institute of Technology	S
Mattingly	Jonathan	Male	Duke University	A

McKinley	Scott	Male	Duke University	S
Mester	Peter	Male	Indiana University	S
Miller	Jason	Male	Stanford University	S
Mostafael	Hamidreza	Male	Islamic Azad University North Tehran Branch	S
Ni	Kai	Male	Georgia Institute of Technology	S
Oprisan	Adina	Female	University of Texas - Arlington	S
Pasour	Virginia	Female	Duke University	S
Raman	Balaji	Male	University of Connecticut	S
Reiner	Bobby	Male	University of Michigan	S
Reinhold	Dominik	Male	UNC - Chapel Hill	S
Restrepo	Ricardo	Male	Georgia Institute of Technology	S
Rezaei	Mahmoud	Male	Clemson University	S
Ruf	Johannes	Male	Columbia University	S
Samara	Marko	Male	Ohio State University	S
Sang	Hailin	Male	University of Cincinnati	S
Schott	Sarah	Female	Duke University	S
Serrano	Rafael	Male	University of York	S
Shabalin	Andrey	Male	UNC - Chapel Hill	S
Shkolnikov	Mykhaylo	Male	Stanford University	S
Smith	Aaron	Male	Stanford University	S
Song	Jian	Male	University of Kansas	S

Srinivasan	Ravi	Male	Brown University	S
Stroock	Daniel	Male	Massachusetts Institute of Technology	S
Thomas	Rachel	Female	Duke University	S
Thompson	Russ	Male	Cornell University	S
Tokle	Joshua	Male	University of Washington	S
Tone	Cristina	Female	Indiana University	S
Turner	Matthew	Male	University of Tennessee - Knoxville	S
Varatharajan	Sarvesh Kumar	Male	University of Kansas	S
Varkey	Paul	Male	University of Illinois - Chicago	S
Veillette	Mark	Male	Boston University	S
Viquez	Juan	Male	Purdue University	S
Wang	Ting	Male	University of Michigan	S
Wang	Fangfang	Female	UNC - Chapel Hill	S
Watkins	Andrea	Female	Duke University	S
Whitmeyer	Joseph	Male	UNC - Charlotte	S
Wu	Wei-Ying	Male	Michigan State University	S
Xin	Linwei	Male	Georgia Institute of Technology	S
Xing	Fei	Female	University of Tennessee - Knoxville	S
Xu	Weijun	Male	Harvard University	S
Xu	Fangjun	Male	University of Connecticut	S
Xue	Yun	Female	Michigan State University	S

Yang	Hongxia	Female	Duke University	S
Zhang	Hongzhong	Male	City University of New York	S
Zhang	Hao	Female	UNC - Charlotte	A
Zhu	Lingjiong	Male	Courant Institute, New York University	S

SAMSI/CRSC Undergraduate Workshop

Participant Summary

May 18-22, 2009

Participants	Male	Female	Unspecified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Institutions Represented	Number of States Represented
Supported	7	9	0	0	16	12	4	13	11
Unsuppted	5	5	1	9	2	11	0	4	1
SAMSI	5	4	0	9	0	9	0		

SAMSI/CRSC Undergraduate Workshop

Workshop Participants

May 18-22, 2009

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Ahmed	Munadir	Male	Macalester College	STAT	S
Attarian	Adam	Male	NCSU	MATH	A
Bain	Melanie	Female	SAMSI/UNC	STAT	A
Balicki	Robert	Male	University of California, Berkeley	STAT	S
Chen	Wenjie	Female	SAMSI/UNC	STAT	A
Choi	Erica	Female	Carnegie Mellon University	STAT	S
Conces	Carola	Female	National Center for Education Research	EDUC	S
Cook	Nicholas	Male	University of North Carolina at Chapel Hill	MATH	S

Das	Sourish	Male	SAMSI	STAT	A
Dickey	Kristen	Female	Loyola University Chicago	MATH	S
Falls	William	Male	University Buffalo	PHYS	S
Gehring	Ryan	Male	self employed	STAT	S
Gordon-Wright	Rachael	Decline	North Carolina State University	MATH	S
Gupta	Himani	Female	Pennsylvania State University	MATH	S
Gupta	Nikhil	Male	Macalester College	OTHR	S
Gremaud	Pierre	Male	SAMSI / NCSU	MATH	A
Hancock	Amy	Female	Washington State University	MATH	S
Ji	Chunlin	Male	Duke U	STAT	A
Keegan	Lindsay	Female	University of Florida	MATH	S
Kepler	Grace	Female	NCSU	MATH	A
Macaro	Christian	Male	SAMSI/Duke U.	MATH	A
Manning	Cammey	Female	Meredith College	MATH	A
Manolopoulou	Ioanna	Female	SAMSI	STAT	A
Murray	Jared	Male	Duke U	STAT	S
Owen	Megan	Female	SAMSI	MATH	A
Rush	Cynthia	Female	University of North Carolina	STAT	S
Schott	Sarah	Female	Duke U.	MATH	A
Shi	Minghui	Female	Duke U	STAT	A
Skowron	Robert	Male	University of Illinois, Urbana Champaign	MATH	S

Snell	Margaret	Female	New Mexico Institute of Mining and Technology	MATH	S
Stitzinger	Ernie	Male	NCSU	MATH	A
Stryjewski	Lisa	Female	Rice University	STAT	S
Tran	Hien	Male	NCSU	STAT	A
Weems	Kim	Female	NCSU	STAT	A
Yasamin	Saed	Male	SAMSI	STAT	A
Yellick	Jason	Male	SAMSI	STAT	A
Zhang	Baqun	Male	NCSU	STAT	A

Industrial Math/Stat Graduate Workshop

Participant Summary
July 20-28, 2009

Participants	Male	Female	Unspecified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Institutions Represented	Number of States Represented
Supported	26	11	0	0	37	31	6	32	17
Unsuported	2	1	0	3	0	3	0	1	1
SAMSI	0	0	0	0	0	0	0		

Industrial Math/Stat Graduate Workshop

Workshop Participants
July 20-28, 2009

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Brown	Aaron	Male	Tufts University	MATH	S
Byrne	Erin	Female	University of Colorado Boulder	MATH	S
Choi	Heejun	Male	Purdue University	MATH	S
Ding	Lili	Female	University of Cincinnati	STAT	S

Gabrys	Robertas	Male	Utah State University	STAT	S
Gremaud	Pierre	Male	NCSU	MATH	A
Griep	Chad	Male	University of Rhode Island	MATH	S
Heaton	Matthew	Male	Duke University	STAT	S
Ipsen	Ilse	Female	NCSU	MATH	A
Jayaram	Magathi	Female	Utah State University	ENGG	S
Katzfuss	Matthias	Male	Ohio State University	STAT	S
Kim	Noory	Male	Towson U	STAT	S
Kirshtein	Jenya	Female	University of Denver	MATH	S
Kumar	Nitesh	Male	University of California Merced	APP MATH	S
Laungrungrong	Busaba	Female	Arizona State University	STAT	S
Li	Yi	Male	Duke University	MATH	S
Lomuscio	Michael	Male	Western Carolina University	MATH	S
Morales	Mario	Male	Hunter College	MATH	S
Pearson	Dale	Male	Texas Tech University	PHYS	S
Pedings	Kathryn	Female	College of Charleston	MATH	S
Porter	Jacob	Male	University of California Davis	MATH	S
Proctor	William	Male	North Carolina State University	ENGG	S
Raghuram	Karthik	Male	University of California Santa Barbara	ENGG	S
Ramachandar	Shahla	Female	University of Texas	STAT	S
Richards	Gregory	Male	Kent State University	STAT	S

Samarakoon	Nishantha	Male	Kansas State University	STAT	S
Shafahi	Maryam	Female	University of California Riverside	ENGG	S
Shen	Chongyi	Male	University of Iowa	STAT	S
Skorczewski	Tyler	Male	University of California Davis	MATH	S
Smith	Ralph	Male	NCSU	MATH	A
Soodhalter	Kirk	Male	Temple University	MATH	S
Sun	Jie (Rena)	Female	University of Michigan	STAT	S
Vu	Duy	Male	Penn State University	STAT	S
Wang	Min	Male	Northern Illinois University	MATH	S
Wiltshire	Jelani	Male	FSU	STAT	S
Yang	Hongxia	Female	Duke University	STAT	S
Zhang	Jingyan	Female	Penn State University	MATH	S
Zhong	Peng	Male	University of Tennessee	MATH	S
Zhou	Kun	Male	Penn State University	MATH	S
Zou	James	Male	Harvard	ENGG	S

APPENDIX E – Workshop Programs and Abstracts

1. Random Media Transition Workshop

Schedule

Thursday, May 1, 2008

Radisson RTP, (Room F/G, 3rd Floor)

- 8:45-9:15 Registration and Continental Breakfast
- 9:15-9:30 Welcome
Ralph Smith, North Carolina State University
- 9:30-10:15 *Heterogeneity in Biological Materials*
Greg Forest, University of North Carolina
- 10:15-10:30 Break
- 10:30-11:15 *Fluctuations for the Tagged Particle in Exclusion Process with Particle Disorder*
Min Kang, North Carolina State University
- 11:15-Noon *A Decomposition Approach for the Immersed Interface Problem*
Anita Layton, Duke University
- Noon-1:00 Lunch (Room F/G, 3rd Floor)
- 1:00-1:45 *Modeling, Analysis, and Computations of the Influence of Surfactant on the Breakup of Bubbles and Drops in a Viscous Fluid*
Michael Siegel, New Jersey Institute of Technology
- 1:45-2:30 *Controlling the Morphology of Viscous Fingering Patterns: A Surprising Discovery*
John Lowengrub, University of California, Irvine
- 2:30-2:45 Break
- 2:45-3:30 *Shape Optimization for Elliptic Eigenvalue Problems*
Chiu-Yen Kao, Ohio State University
- 3:30-4:15 *Lattice Boltzmann Simulation of Flow through Porous Media*
L-S Luo, Old Dominion University

Friday, May 2, 2008

Radisson RTP, (Room F/G, 3rd Floor)

- 8:30-9:00 Registration and Continental Breakfast
- 9:00-9:45 *Image Reconstruction in Diffuse Optical Tomography*
Taufiqar Khan, Clemson University
- 9:45-10:30 *Material Properties of Heterogeneous Viscous and Viscoelastic Fluids*
Isaac Klapper, Montana State University
- 10:30-10:45 Break
- 10:45-11:30 *Adaptive Tikhonov Regularization for Inverse Problems*
Kazufumi Ito, North Carolina State University
- 11:30-12:15 *An Algorithm for Generating Overlapping Grids and Partitions of Unity for Integrating on Implicitly*
Jason Wilson, Duke University
- 12:15-1:15 Lunch

SPEAKER ABSTRACTS

Greg Forest

University of North Carolina
forest@amath.unc.edu

“Heterogeneity in Biological Materials”

In this talk I will give an overview of the projects that were spawned during the Fall semester in our working group on Heterogeneity in Biological Media at SAMSI, the progress that has been made up til now, and the significant challenges that remain.

Kazufumi Ito

North Carolina State University
kito@ncsu.edu

“Adaptive Tikhonov Regularization for Inverse Problems”

Tikhonov regularization method plays a critical role in ill-posed inverse problems, arising in industrial applications including computerized tomography, inverse scattering and image processing. The goodness of the inverse solution heavily depends on selection of the regularization parameter. Commonly used methods rely on a priori knowledge of the noise level. A method that automatically estimates the noise level and selects the regularization parameter automatically is presented.

Min Kang

North Carolina State University
mkang2@ncsu.edu

“Fluctuations for the Tagged Particle in Exclusion Process with Particle Disorder”

We study the asymptotic distribution of the fluctuations from the mean of the velocity of a tagged particle performing totally asymmetric simple exclusion on the integer lattice with random disorder. The fluctuations are investigated in the subcritical case on a non-Gaussian scale.

Chiu-Yen Kao

Ohio State University
kao@math.ohio-state.edu

“Shape Optimization for Elliptic Eigenvalue Problems”

Identification or optimization of shapes arises in many science and engineering applications. In this talk, we focus on the optimal shape design related to elliptic eigenvalue problems. Specific applications for identifying structures of photonic crystal, optimization of quality factor of an acoustic resonator, and determining the optimal spatial arrangement of favorable and unfavorable regions for a species to survive will be discussed.

Taufiqar Khan

Clemson University
khan@CLEMSON.EDU

“Image Reconstruction in Diffuse Optical Tomography”

In this talk, an overview of the basics of image reconstruction in diffuse optical tomography (DOT), a typical computational framework to solve the deterministic inverse problem, and some results involving an iteratively regularized Gauss-Newton method will be presented. The question of how to reformulate an ill-posed inverse problem, such as DOT, in order to convert it into a well-posed one in a deterministic and/or in a statistical setting will be raised. Furthermore, a particular statistical formulation of the computational inverse problem using Bayes' formula will be mentioned to generate discussions among participants of the transition workshop.

Isaac Klapper

Montana State University
klapper@math.montana.edu

“Material Properties of Heterogeneous Viscous and Viscoelastic Fluids”

Anita Layton

Duke University
alayton@math.duke.edu

“A Decomposition Approach for the Immersed Interface Problem”

We consider the immersed boundary problem in which the fluid, described by Navier-Stokes flows, is separated into two regions by an elastic boundary. The moving elastic boundary exerts a singular force on the local fluid. The model solution is obtained using a decomposition approach, which splits the solution into a “Stokes” part and a “regular” part. The Stokes solution is given by the Stokes equations and the singular boundary force; that solution is obtained using the immersed interface method, which computes second-order accurate approximations by incorporating known jumps in the solution or its derivatives into a finite difference method. The regular solution is given by the Navier-Stokes equations and a body force; that solution is obtained using a time-stepping method that combines the semi-Lagrangian discretization and the Backward Difference Formula. Because the body force is continuous, jump conditions are not necessary in the computations associated with the regular solution. For problems with stiff boundary forces, the decomposition approach can be combined with fractional time-stepping, in which the boundary is advanced quickly using boundary integrals, and using a smaller time step to maintain numerical stability, and the overall solution is updated using a larger time step to reduce computational cost.

John Lowengrub

UC Irvine

lowengrb@math.uci.edu

“Controlling the Morphology of Viscous Fingering Patterns: A Surprising Discovery”

A variety of pattern forming phenomena, ranging from the growth of bacterial colonies to snowflake formation, share similar underlying physical mechanisms and mathematical structure. Viscous fingering, considered here, is a paradigm for such phenomena. Prediction and control of the shape of emergent patterns is difficult due to the non-locality and nonlinearity of the system. Here, we report the discovery of a remarkable strategy to precisely control the pattern shape and the evolving interfacial instabilities over some ten orders of magnitude in length. There exist denumerable attracting, self-similarly evolving symmetric, universal patterns. Experiments confirm the feasibility of the strategy, which is summarized in a morphology diagram.

Li-Shi Luo

Old Dominion University

lluo@odu.edu

“Lattice Boltzmann Simulation of Flow through Porous Media”

The lattice Boltzmann equation (LBE) is a numerical method for computational fluid dynamics (CFD). As opposed to conventional CFD method based on direct discretizations of the Navier-Stokes equations, the LBE method is derived from kinetic theory and the Boltzmann equation. Due to its kinetic origin, the LBE method has some features different from conventional CFD methods. In this talk, I will first present the

derivation of the LBE method from kinetic equation so some of its features can be clearly seen. I will then show some applications of the LBE for flow through porous media and interfacial flows to demonstrate the capability of the LBE method.

Michael Siegel

New Jersey Institute of Technology
misieg@oak.njit.edu

“Modeling, Analysis, and Computations of the Influence of Surfactant on the Breakup of Bubbles and Drops in a Viscous Fluid”

We present an overview of experiments, numerical simulations, and mathematical analysis of the breakup of a low viscosity drop in a viscous fluid, and consider the role of surface contaminants, or surfactants, on the dynamics near breakup. As part of our study, we address a significant difficulty in the numerical computation of fluid interfaces with soluble surfactant that occurs in the important limit of very large values of bulk Peclet number Pe . At the high values of Pe in typical fluid-surfactant systems, there is a narrow transition layer near the drop surface or interface in which the surfactant concentration varies rapidly, and its gradient at the interface must be determined accurately to find the drop's dynamics. Accurately resolving the layer is a challenge for traditional numerical methods. We present recent work that uses the narrowness of the layer to develop fast and accurate 'hybrid' numerical methods that incorporate a separate analytical reduction of the dynamics within the transition layer into a full numerical solution of the interfacial free boundary problem.

Jason Wilson

Duke University
jwilson@math.duke.edu

“An Algorithm for Generating Overlapping Grids and Partitions of Unity for Integrating on Implicitly Defined Curves and Surfaces”

It is well known that the trapezoid rule achieves super-algebraic convergence when used to integrate smooth integrands with compact support. Algorithms based on the trapezoid rule have also been developed to integrate singular and nearly singular integrands for use in Boundary Integral Methods. In order to apply these methods on a closed smooth surface, one needs a set of overlapping patches covering the surface as well as an associated smooth partition of unity. This talk discusses an algorithm that automatically generates such a set of patches and partitions given an implicitly defined smooth closed curve or surface. The focus of the talk will be on the curve case which is easy to visualize. The algorithm easily generalizes to handle surfaces.

RISK REVISITED: PROGRESS AND CHALLENGES

Wednesday May 21, 2008

8:50 - 9:00	Welcome
9:00 - 10:00	Bayesian GLMs Dipak Dey , University of Connecticut "Flexible Skewed Link Function for the Dichotomous Response Data: Generalized Extreme Value Link" Sourish Das , University of Connecticut "Analyzing Extreme Drinking Behavior of Patients suffering Alcohol Dependence Disorder Using Pareto Regression"
10:00 - 10:20	Break
10:20 - 11:20	Environmental Risk Rosalba Ignaccolo , Universita' degli Studi di Torino and SAMSI "Impact Evaluation of Changing Ozone Standards on Mortality" Evangelos Evangelou , University of North Carolina and SAMSI "Multivariate Generalized Linear ARMA Processes: An Application to Hurricane Activity"
11:20 - 12:20	Risk in the Service Sector Pilar Munoz , Technical University of Catalonia "Impact of Electricity: Financial, Macroeconomic and Environmental" Haipeng Shen , University of North Carolina "Classification of Services with Application in Service Risk Management: Progress and Challenges"
12:20 - 1:10	Lunch
1:10 - 2:40:	Biosurveillance and Epidemic Modeling (joint with QMDNS)

Georgiy Bobashev, RTI
"Local and Global Epidemic Models. Can they be Practical and Useful?"

Ron Fricker, Naval Postgraduate School
"Optimizing Biosurveillance Systems"

Myron Katzoff, National Center for Health Statistics
"A Further Consideration of Two Problems Related to Biosurveillance "

2:40-3:40 Multivariate and Spatial Extreme Value Theory

Dan Cooley, Colorado State University
"Hierarchical Spatial Modeling for Extremes"

Xiao Qin, Beihang University and UNC
"New Classes of Multivariate Survival Functions"

3:40 - 4:00 Break

4:00 - 4:30 Adversarial Risk

David Rios Insua, University Rey Juan Carlos
"Advances in Adversarial Risk Analysis"

4:30 - 5:30 SAMSI New Researcher Session

Vered Madar, SAMSI
"Bayesian Model Selection for The Farlie-Gumbel-Morgenstern Copula for Describing Two Generalized Extreme Value Variables"

Guang Cheng, Duke University and SAMSI
"Semiparametric Additive Isotonic Regression"

8:00-10:00 Evening Reception

Risk Revisited: Progress and Challenges

May 21, 2008

SPEAKER ABSTRACTS

Georgiy Bobashev

RTI

bobashev@rti.org

"Local and Global Epidemic Models. Can they be Practical and Useful?"

A large body of evidence shows that the emergence of highly transmissible influenza strain is very likely in future. Public health officials and policy makers are turning to modelers for suggestions about the practical steps to recognize and contain future epidemic. One of the challenges is to produce a model that would both scientific and practical values. I will present an approach that uses a system of models, each most appropriate at its own temporal and spatial scale. One of such models is a stochastic equation-based epidemic model describing the global transmission of pandemic flu. Using simulation analysis, we show that interventions should not be considered independently of each other. When the epidemic starts in Asia, travel restrictions can delay the arrival of flu to the US and allow public health to better prepare for the pandemic. If, in the time afforded, control measures such as administration of antiviral medication and self-isolation are instituted, the result is a significant reduction in cases worldwide and in the U.S. We show that accounting for seasonality in the transmission rate is critical for making the decision on the optimal combination of the interventions at the global scale. At the same time, local models are much more useful in describing how particular interventions are implemented in practice; and how a particular public policy translates into a change of a parameter value. Surveillance analysis tools, such as TranStat become critical for the early estimation of immediate risk.

Guang Cheng

Duke University and SAMSI

chengg@stat.duke.edu

"Semiparametric Additive Isotonic Regression"

This paper is about the efficient estimation of semiparametric additive isotonic regression model, i.e. $Y = X\beta + \sum h_j(W_j) + \epsilon$. Each additive component h_j is assumed to be a monotone function. It is shown that the least square estimator of the parametric component is asymptotically normal. Moreover, the isotonic estimator for each additive functional component is proved to have the oracle property, which means it can be estimated with the highest asymptotic accuracy, equivalently, as if the other components were known.

Dan Cooley

Colorado State University

cooleyd@stat.colostate.edu

"Spatial Hierarchical Modeling of Precipitation Extremes from a Regional

Climate Model”

Regional climate models (RCMs) are tools which allow scientists to begin to understand how different forcings may affect climate. There has been some statistical work done to characterize the difference in mean behavior between control and future scenarios as predicted by RCMs. The goal of this work is to characterize the extremes as produced by a RCM and to additionally examine the difference in extremes between a control and future scenario.

To characterize the spatial behavior of extreme precipitation we construct a hierarchical model. The data level is formed by the point process representation of extremes, and the process level is based on a conditional autoregressive (CAR) model since our data are on a regular lattice. Because we are interested in modeling not only how much the extremes change but also how they appear to be changing, we spatially model all three (location, shape, and scale) of the extreme value distribution's parameters.

Sourish Das

University of Connecticut
sourish.das@gmail.com

“Analyzing Extreme Drinking Behavior of Patients suffering Alcohol Dependence Disorder using Pareto”

In this paper, we examine two issues of importance in study of Alcohol Dependence Disorder (ADD). First, we examine the association between extreme alcohol ingestion and single nucleotide polymorphism (SNPs) within GABRA2 gene and second we examine the efficacy of three types of psychosocial treatment for alcoholism: Cognitive Behavioral Therapy (CBT), Motivational Enhancement Therapy (MET), and twelve-step facilitation (TSF). European-American subjects (n=812, 73.4% male) provided DNA samples for the analysis. All were participants in Project Matching Alcoholism Treatment to Client Heterogeneity (MATCH), a multi-center randomized clinical trial. The study length consists of 3 month treatment and 12 month post treatment periods. We develop a novel Pareto regression model with unknown shape parameter, for analyzing extreme drinking behavior of these patients suffering ADD. We consider a Generalized Linear Model (GLM) framework, using log-link between the shape parameter of random component and systematic component. In order to incorporate the longitudinal component of the study, we add in the time and interaction between time and SNPs into systematic component with other covariate information like age. We present a Monte Carlo based Bayesian method to implement the analysis.

Dipak Dey

University of Connecticut
dey.dipak@gmail.com

“Flexible Skewed Link Function for the Dichotomous Response Data: Generalized Extreme Value Link”

The choice of the links is one of most critical issues involved in modeling binary data as substantial bias in the mean response estimates can be yielded if the link could be

misspecified. The objective of this study is to introduce a flexible skewed link function for modeling categorical data. The commonly used complementary log-log (Cloglog) link is prone to link misspecification because of its positive and fixed skewness. We propose a new link function based on the generalized extreme value (GEV) distribution. The GEV link has a very wide range of skewness, which is purely decided by its shape parameter. Using Bayesian methodology, we can automatically detect the skewness in the data along with the model fitting by the GEV link. Various theoretical properties are examined and explored in details. We compare the logit, the probit, the Cloglog and the GEV links under different scenarios. The possibility applying this link to the large p , small n cases is also discussed. The deviance information criterion measure is used for guiding model selection when comparing different links.

Key Words: Latent variable; Complementary log-log link; Generalized extreme values distribution; Prior Elicitation; Markov chain Monte Carlo

Evangelos Evangelou

University of North Carolina, Chapel Hill and SAMSI
vangelis@email.unc.edu

“Multivariate Generalized Linear ARMA Processes: An Application to Hurricane Activity”

In this paper we propose a multivariate framework for investigating the relationship between hurricane activity and global warming. Papers such as Saunders and Lea (Nature, 2008) find evidence of correlation between the number of US landfalling hurricanes and local sea surface temperatures. We propose a modelling strategy involving a bivariate process where one component is Poisson and the other is Gaussian. Since standard time series analysis shows significant auto-correlations, we use a multivariate generalized linear ARMA model. Our analysis can be viewed as an extension of the methodology by Davis, Dunsmuir and Streett (2003, 2005) to multiple dimensions. Our maximum likelihood analysis shows that a multivariate framework can be a powerful tool for simultaneously analyzing hurricane activity and global warming in the presence of correlation between the two.

Ronald Fricker

Naval Postgraduate School
rdfricke@nps.edu

“Optimizing Biosurveillance Systems”

Motivated by the threat of bioterrorism, biosurveillance systems are being developed and implemented throughout the United States. Biosurveillance is the regular collection, analysis, and interpretation of real-time and near-real-time indicators of possible disease outbreaks and bioterrorism events by public health organizations. Little is known about how effective these systems will be at quickly detecting a bioterrorism attack, but there is some evidence in the form of excessive false alarm rates that they are being suboptimally employed. This talk will provide an overview of the problem and describe an approach for managing the trade-off between the aggregate "system" false alarm rates and the power to detect a localized bioterrorism attack.

Rosalba Ignaccolo

Universita' degli Studi di Torino and SAMSI
rosalba.ignaccolo@gmail.com

“Impact Evaluation of Changing Ozone Standards on Mortality”

We present a risk assessment analysis of the potential effect that various regulatory standards for ozone may have on the incidence of non-accidental mortality. The analysis uses roll-back functions as models for the potential effect of regulatory standards. The statistical methods are based on the hierarchical Bayesian models. The objective is to obtain estimates of the effects of various regulatory standards, estimates of their variability, and the effects of various modeling assumptions on those estimates.

Myron Katzoff

National Center for Health Statistics
mjk5@cdc.gov

"A Further Consideration of Two Problems Related to Biosurveillance"

For the detection of a catastrophic public health event and the subsequent collection of information to monitor progress in addressing its consequences, we may expect that the statistical methods employed for those purposes will draw upon experience acquired in other applications. The first part of this talk will consider the application of ideas from extreme value theory to the detection of “outbreaks” and the estimation of the probabilities of occurrence of values for disease incidence rates that might be viewed as extreme. Since my study of these ideas is at a very early stage, there will be more “vigor” than “rigor” in this part of my talk. In the remaining time, I will visit the application of some adaptive sampling techniques that I believe will have promise in obtaining information about the health status of individuals affected by the types of public health events of interest. With the occurrence of such events, the affected individuals may be “hidden” or hard-to-locate because it is unlikely that there will be a sampling frame of them available for our immediate use and they may be well-disperse throughout other populations. The adaptive sampling techniques were originally developed as probability sampling design alternatives for collecting information on populations at risk for AIDS/HIV.

Vered Madar

SAMSI
madar@post.tau.ac.il

“Bayesian Model Selection for The Farlie-Gumbel-Morgenstern Copula for Describing Two Generalized Extreme Value Variables”

The Bivariate Farlie-Gumbel-Morgenstern Copula (Johnson and Kotz 1975) has a bad reputation for providing restricted range of dependence (Joe 1997). Yet the simple linear-like structure of this copula is very appealing, and with slight changes one can increase further its range of dependence (Güven and Kotz 2008). We suggest a Bayesian model

selection procedure for this copula for describing the joint distribution of two generalized extreme value variables.

Reference:

Guven B. Kotz S. (2008). Test of independence for generalized Farlie-Gumbel-Morgenstern distributions. *Journal of Computational and Applied Mathematics*. 212, 102-111.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. CRC Press

Johnson, N.L. and Kotz, S. (1975). A vector multivariate hazard rate. *Journal of Multivariate Analysis*. 5, 53-66

Maria Pilar Munoz

Technical University of Catalonia
pilar.munoz@upc.edu

“Impact of Electricity: Financial, Macroeconomic and Environmental”

To analyze and predict the electricity consumption and price are a problem of real interest at present. Estimation of these variables and the relations with other ones are necessary to better understand several risk aspects. This talk is focused on three different perspectives of this topic and summarizes the work accomplished at SAMSI risk program in collaboration with other authors.

- Financial aspect (with Nik Tuzov, Purdue University):
This work studies US Energy Market with the objective to check what variables influence the power price, in particular at extreme levels. The research in this topic is ongoing.
- Macroeconomic aspect (with Dave Dickey, NCSU):
The relations with the Spanish electricity and the USDollar/Euro Exchange rate are investigated in this point. The conclusion is that both variables are cointegrated, getting an estimation of the long run equilibrium relationship. In addition, estimated volatilities for both series are also related.
- Environmental aspect (with Jen-Ting Wang, SUNY-Oneonta). It is commonly known that the carbon dioxides (CO₂) emissions are one of the major causes of global warning. This motivated us to estimate and predict the CO₂ emissions produced by the electricity generation in US. Estimations and forecasts of the fossil fuels used in the generation process have allowed us to compare the trends in the fossil fuels consumption and to make predictions of CO₂ emissions.

Xiao Qin

University of North Carolina, Chapel Hill
qinx@email.unc.edu

“New Classes of Multivariate Survival Functions”

Ledford and Tawn (1997, JRSSB) and later Ramos and Ledford (2007a, 2007b) extended the traditional approach based on bivariate extreme value theory to model the joint tail distribution, by incorporating positively and negatively associated asymptotical independence into asymptotical dependence and exact independence. However, to the best of our knowledge, their model is limited to the bivariate case. This paper generalizes their models into multivariate cases under certain assumptions. A technique to construct the angular measure constrained by their models is proposed. A rich parametric class suitable to model the multivariate joint tails is derived.

David Rios Insua

University Rey Juan Carlos

david.rios@urjc.es

“Advances in Adversarial Risk Analysis”

In the talk, I shall summarise key findings and ongoing problems in relation with adversarial risk analysis. After stating the ARA problem, I shall critically assess some previous approaches and provide an alternative Bayesian solution. I shall then outline applied problems that we are facing in the areas of auctions, cybersecurity and counterterrorism.

Haipeng Shen

University of North Carolina, Chapel Hill

shenhaipeng@gmail.com

“Classification of Services with Application in Service Risk Management: Progress and Challenges”

Classification of services is a strong tool for gaining insights into different types of services and has been particularly used for benchmarking as well as strategic positioning of services. There have been numerous works, mainly started from 70s, on classifying services in general or within specific contexts. However, most of them are not based on any empirical study. We apply statistical methods to an empirical pilot study on classification of services for risk management purposes. The pilot study has revealed some of the challenges of applying service classification in the area of risk management. We discuss these challenges and the next stages of the research.

**Summer 2008 Program on Meta-analysis: Synthesis and
Appraisal of Multiple Sources of Empirical Evidence
June 2-13, 2008
SCHEDULE**

Monday, June 2, 2008

[Radisson Hotel RTP](#)

Tutorials

8:00-8:45 Registration and Continental Breakfast

8:45-9:00 Welcome

9:00-10:15 Overview of Meta-analysis

Statistical Methods for Combining the Results of Independent Studies
Ingram Olkin, Stanford University

10:15-10:45 Coffee Break

10:45-Noon *Statistical Methods for Combining the Results of Independent Studies*, continued

Noon-1:00 Lunch

1:00-3:00 *Statistical Methods for Combining the Results of Independent Studies*, continued

3:00-3:30 Coffee Break

3:30-5:00 *Likelihood Basis for Multiple Data Sources*
Keith O'Rourke, Duke University

Tuesday, June 3, 2008

[Radisson Hotel RTP](#)

8:00-9:00 Registration and Continental Breakfast

9:00-10:15 *Likelihood Basis given Sparse Evidence and Common Parameter Focus*
Keith O'Rourke, Duke University

10:15-10:45 Coffee Break

10:45-11:15 *Likelihood Basis given Sparse Evidence and Common Parameter Focus*,
continued

11:15-Noon *Integrated Likelihood for Common Parameter Focus*

Vanja Dukic, University of Chicago

Noon-1:00 Lunch

1:00-3:00 *Conditional Likelihood for Common Parameter Focus, Exchangeability and links between paradigms.*
Ken Rice, University of Washington

3:00 – 3:30 Coffee Break

3:30 – 5:00 *Likelihood or “pre-Posterior” Data Analysis Session*
Keith O’Rourke, Duke University

Wednesday, June 4, 2008

[Radisson Hotel RTP](#)

8:00-9:00 Registration and Continental Breakfast

9:00-10:15 *Bayesian MA*
Vanja Dukic, University of Chicago
Ken Rice, University of Washington

10:15-10:45 Coffee Break

10:45-Noon *Bayesian MA*, continued

Noon-1:00 Lunch

1:00-3:00 *Practical Obstacles in Meta-analysis*
Julian Higgins, Cambridge University

3:00 – 3:30 Coffee Break

3:30 – 5:00 *Data Analysis Session*
Ken Rice, University of Washington

5:00-5:30 Poster Advertisement Session: 2 minute ads by each poster presenter

6:30–8:30 Poster Session and Reception
SAMSI will provide poster presentation boards and tape. The board dimensions are 4 ft. wide by 3 ft. high. They are tri-fold with each side being 1 ft. wide and the center 2 ft. wide. Please make sure your poster fits the board. The boards can accommodate up to 16 pages of paper measuring 8.5 inches by 11 inches.

Thursday, June 5, 2008

[Radisson Hotel RTP](#)

Opening Workshop

- 8:00-8:45 Registration and Continental Breakfast
- 8:45-9:00 Welcome
- 9:00-10:00 *Overcoming the Scope and Limitations of the Literature: Some Examples of Complex Evidence Synthesis.*
Julian Higgins, Cambridge University
- 10:00–10:30 Coffee Break
- 10:30–11:45 *MA-Sports Medicine*
Ian Shrier, McGill University
- 11:45-1:00 Lunch
{Program Leaders' Lunch with Keith Crank and Sara Murphy, ASA}
- 1:00–3:15 *Bayesian Meta-analysis of Diagnostic Test Accuracy Studies*
Constantine Gatsonis, Brown University
- Empirical Insights from Genetic Meta-analysis: Challenges, Biases, and Unique Considerations*
Tom Trikalinos, Tufts University
- 3:15-3:45 Coffee Break
- 3:45-5:00 New Researcher Session I:
Combining Information from Randomized and Observational Data: A Simulation Study
Eloise Kaizar, Ohio State University
- Generalizing Results from a Randomized Trial to a Broader Population: Bridging Observational and Experimental Data*
Elizabeth Stuart, Johns Hopkins University

Friday, June 6, 2008

[Radisson Hotel RTP](#)

- 8:00-9:00 Registration and Continental Breakfast
- 9:00-10:00 *Recent Advances: Robust and Multidimensional Meta-analysis Models*
Eugene Demidenko, Dartmouth Medical School
- 10:00–10:30 Coffee Break

- 10:30–11:45 *The Exact Distributions of Test Statistics Resulting from the Random Effects Model for Meta-Analysis*
Dan Jackson, Cambridge University
- 11:45-1:00 Lunch
- 1:00–3:15 *Issues in Hierarchical and Non Hierarchical Combining of Information*
Susie Bayarri, University of Valencia
- Nonparametric Bayes Data Fusion*
David Dunson, NIEHS
- 3:15-3:45 Coffee Break
- 3:45-5:00 New Researcher Session II:
- Meta-analysis of Diagnostic Test Accuracy Assessment Studies with Varying Number of Thresholds*
Vanja Dukic, University of Chicago
- Hierarchical Dependence in Meta-Analysis*
John Stevens, Utah State University

Monday, June 9 – Friday, June 15, 2008

[SAMSJ RTP](#)

- 12:00-1:30 Working Week Lunch Forum
- Monday: **Rafael Irizarry**, Johns Hopkins University
- Tuesday: **Robert Platt**, McGill University
- Wednesday: **Dan Jackson**, MRC Cambridge
- Thursday: **Sally Morton**, RTI

**Program on Meta-analysis: Synthesis and Appraisal of Multiple Sources
of Empirical Evidence
June 2-13, 2008**

SPEAKER ABSTRACTS

Susie Bayarri

University of Valencia
susie.bayarri@uv.es

“Issues in Hierarchical and Non Hierarchical Combining of Information”

We first consider important issues that arise specially when hierarchically combining several sources of (exchangeable) information. One such issues refers to uncertainty about the likelihood, whose effect can get dramatically magnified as the number of combined sources increases. This issue gets special relevance when combining published experiments, as there is considerable uncertainty in the selection mechanisms. A possible solution is to resort to robust Bayes analysis. Another issue in hierarchical combinations refers to the uncertainty about the (prior) relating the different sources of information. As the number of combined sources gets large, inadequacy in the specification of the second level can severely affect in unanticipated ways inferences about parameters in the data level, even if they are not subject to the meta-analytic combination. We present a "quick" (and dirty) fix which alleviate the problem, and also an appropriate checking of those latent layers models. Quite different but important issues arise when combining very disparate sources of information, as it is the case when one has to combine computer simulators data with field data; we'll briefly consider this emerging and increasingly important form of data-merging.

Eugene Demidenko

Dartmouth Medical School
Eugene.Demidenko@Dartmouth.EDU

“Recent Advances: Robust and Multidimensional Meta-analysis Models”

While the sample size of individual studies is typically large the number of studies is usually small. This fact may contradict the normal assumption usually used when the meta-analysis model is estimated. We suggest a robust version of the meta-analysis model when the distribution of the random effects is not normal and tails are heavy. An example of a robust model is when the weighted median is used instead of the mean. This suggests a new distribution for the meta-analysis model as a convolution of the double-exponential and normal densities.

Multivariate meta-analysis model is another extension when several auxiliary variables, besides the variable of interest/exposure, are available. We show that an addition of a new variable to the meta-analysis may improve the efficiency, especially when it correlates with the variable of interest. The discussion follows chapter 5 of the book on

the mixed models recently published by the author as well a recent paper on the multivariate meta-analysis model published in the Journal of Statistical Planning and Inference. The theory is illustrated with a classic example on efficacy of the BCG vaccine (13 studies) and a recent meta-analysis with preventive carotenoids for lung cancer (7 studies).

Vanja Dukic

University of Chicago
vdukic@health.bsd.uchicago.edu

“Meta-analysis of Diagnostic Test Accuracy Assessment Studies with Varying Number of Thresholds”

Current meta-analytic methods for diagnostic test accuracy are generally applicable to a selection of studies reporting only estimates of sensitivity and specificity, or at most, to studies whose results are reported using an equal number of ordered categories. In this article, we propose a new meta-analytic method to evaluate test accuracy and arrive at a summary receiver operating characteristic (ROC) curve for a collection of studies evaluating diagnostic tests, even when test results are reported in an unequal number of non-nested ordered categories. We discuss both non-Bayesian and Bayesian formulations of the approach. In the Bayesian setting, we propose several ways to construct summary ROC curves and their credible bands. We illustrate our approach with data from a recently published meta-analysis evaluating a single serum progesterone test for diagnosing pregnancy failure.

David Dunson

NIEHS
dunson1@niehs.nih.gov

“Nonparametric Bayes Data Fusion”

There is increasing interest in borrowing strength and learning of commonalities in data from multiple sources. Some classical examples include meta analysis, multi-center studies and longitudinal data analysis, while emerging areas include multi-task learning, functional data analysis, and joint modeling of data having fundamentally different measurement scales. Often, data are high-dimensional in such settings, and it is necessary to discover a sparse latent structure in the data and exploit this structure in building joint models. Widely used methods, such as parametric random effects models, are clearly insufficiently flexible in such cases. This talk will focus on nonparametric Bayes methods that rely on partitioning to build flexible and sparse dependence structures across disparate data. I provide a brief review and illustration of approaches based on the Dirichlet process and extensions, such as the nested Dirichlet process and the hierarchical Dirichlet process. In addition, new approaches based on local partition processes are described. The methods will be illustrated through applications to multi-center studies, functional data analysis, compressive sensing and image analysis.

Constantine Gatsonis
Brown University
gatsonis@stat.brown.edu

“Bayesian Meta-analysis of Diagnostic Test Accuracy Studies”

Interest in evidence based diagnosis has grown rapidly in recent years and has highlighted the need for systematic reviews in this area. We will discuss statistical methods for diagnostic accuracy studies, with a focus on studies reporting estimates of sensitivity and specificity. The need to account for between-study differences in the threshold for test positivity is a fundamental aspect of systematic reviews of test accuracy has led to Summary Receiver Operating Characteristic curve analysis. The reviews also need to account for other sources of within-and between-study heterogeneity and to address issues such as errors in the reference standard, the use of multiple cutpoints, and verification bias. In this presentation, I will discuss hierarchical and mixed model methods for research synthesis in this area of meta-analysis and will also discuss open problems requiring new methodologic development.

Julian Higgins
University of Cambridge
julian.higgins@mrc-bsu.cam.ac.uk

Wednesday: “Practical Obstacles in Meta-analysis”

I will discuss some of the fundamental practical obstacles that most meta-analyses face ‘in the field’. I will draw on my experiences of working with review authors in The Cochrane Collaboration, and make use of several ‘problematic’ data sets. The main problems that I will address are (i) deciding which studies to include; (ii) addressing the ‘quality’ of the studies identified; (iii) dealing with variation across the studies included; and (iv) addressing publication bias.

Thursday: “Overcoming the Scope and Limitations of the Literature: Some Examples of Complex Evidence Synthesis”

“Complex evidence synthesis” has been used to describe methods that go beyond multiple sources of similar evidence to synthesize studies of different questions, often using different study designs, to address a question wider than any of the individual studies. I will present some examples from medical research. First, I will discuss the broadening of eligibility criteria in meta-analyses of randomized trials to incorporate studies of lesser quality and lesser relevance. A special case of the latter is including studies that compare interventions other than those of primary interest but that allow indirect evaluation of the main question (for example, A and B can be compared by combining studies of A vs C with studies of B vs C). Such ‘multiple treatment meta-analyses’ also allow new questions to be addressed, in particular the question of which intervention is ‘best’. Second, I will discuss the extension of meta-analyses in human genome epidemiology to include studies that partially address the question of interest, but

that, when combined with other studies and reasonable assumptions, contribute information to the synthesis. The complex evidence syntheses can be implemented conveniently within a Bayesian framework, for example using WinBUGS.

Dan Jackson

MRC Biostatistics Unit at the University of Cambridge, Institute of Public Health,
daniel.jackson@mrc-bsu.cam.ac.uk

"The Exact Distributions of Test Statistics Resulting from the Random Effects Model for Meta-Analysis"

The random effects model is routinely used in meta-analysis and can incorporate both covariate effects and multivariate study outcomes. Standard procedures for implementing this type of model typically estimate the between-study (heterogeneity) variance and then effectively regard this as fixed and known when pooling the studies' results. Although justifiable asymptotically, this type of procedure requires sufficiently large numbers of studies. The exact distributions of a variety of typical test statistics are therefore examined in order to assess the suitability of this type of approach. Initially the exact distribution of Cochran's heterogeneity statistic is derived, in order to assess the degree of uncertainty in the heterogeneity parameter. The exact distributions of the usual estimates of treatment effect, for simple special case, are then derived, in order to give an indication of the number of studies that are needed in practice when adopting the standard procedures; in particular the implications of multiple testing in the context of meta-regression will be examined. The talk will conclude with a discussion of 'less asymptotic' approaches when implementing the random effects model, and the other types of issues and concerns when using this model.

Eloise Kaizar

Ohio State University
ekaizar@stat.osu.edu

"Combining Information from Randomized and Observational Data: A Simulation Study"

Randomized controlled trials have become the gold standard of evidence in medicine. They earned this status because they offer strong internal validity. However, subject recruitment may introduce selection bias that limits trials' external validity. To mitigate the selection bias some turn to meta-analysis to widen the recruitment pool; in practice, this method is not likely to eliminate all selection bias. Observational studies are also commonly used in medical and epidemiological research. Complementary to randomized trials, these studies tend to have strong external validity or broad generalizability, but because of treatment self-selection often have severely limited internal validity. We propose a response surface framework for combining both randomized and observational data in a single overarching probability model that models the selection bias of the randomized studies and the self-selection bias of the observational studies. Simulations

show that our framework may produce a single estimate with less bias than estimates derived using current methods.

Ingram Olkin

Stanford University
iolkin@stat.stanford.edu

“Meta-analysis: Statistical Methods for Combining the Results of Independent Studies”

Meta-analysis enables researchers to synthesize the results of a number of independent studies designed to determine the effect of an experimental protocol such as an intervention, so that the combined weight of evidence can be considered and applied. Increasingly meta-analysis is being used in the health sciences, education and economics to augment traditional methods of narrative research by systematically aggregating and quantifying research literature. A Google scholar search on meta-analysis plus different fields of research uncovered close to 200,000 hits in the social sciences (psychology, sociology, education), and a like number in medicine. The range of applications is surprisingly broad. Two meta-analytic examples are the effectiveness of mammography in the detection of breast cancer, and an evaluation of gender differences in mathematics education.

The information explosion in almost every field coupled with the movement towards evidence based analyses and decision making, and cost-effective analysis has served as a catalyst for the development of procedures to synthesize the results of independent studies

In this workshop we provide an historical perspective of meta-analysis, discuss some of the issues such as various types of bias and the effects of heterogeneity. The statistical methodology will include discussions of nonparametric and parametric models; effect sizes for proportions, fixed versus random effects, regression and anova models. New material on multivariate models will also be presented.

Robert Platt

McGill University
robert.platt@mcgill.ca

“Defining Causal Effects in RCTs and Observational Studies and Considerations for Inclusion in Meta-analysis”

This session outlines the definition of causal effects using counterfactual random variables. The intent-to-treat (ITT) causal effect typically reported in randomized trials and used in meta-analyses has a well-defined meaning as the comparison between expectations of counterfactual random variables. However, as is well-known, this effect may underestimate the biological causal effect of an intervention due to non-compliance. Observational studies, on the other hand, may allow estimation of the biological causal effect, but suffer inherently from the potential for unmeasured confounding. We use counterfactuals and directed acyclic graphs to make links between the ITT effect and the

biological causal effect, and discuss the implications for combining information from randomized trials and observational studies in the same analysis.

Ian Shrier

McGill University
ian.shrier@mcgill.ca

“Meta-Analysis – Sports Medicine”

Clinical sports medicine is a relatively young field and most of the evidence comes from non-randomized trials, and extrapolations from basic and applied exercise physiology. The first part of this presentation will demonstrate some of the issues related to meta-analyses through an “interactive” systematic review of the literature on whether stretching prevents injury. The second part of the presentation discusses some ideas on whether randomized trials are always estimating the parameter of most interest to clinicians, and always the most helpful in making causal inferences. The final part of the presentation proposes that a structural approach to bias (all epidemiological bias is either due to absence of conditioning on a common cause, or the presence of conditioning on a common effect) may provide a transparent framework for meta-analysts to decide whether or not it is appropriate to combine studies using different 1) designs and/or 2) regression models.

John Stevens

Utah State University
john.r.stevens@usu.edu

“Hierarchical Dependence in Meta-Analysis: Methods”

Studies to be combined in a meta-analysis may have sampling and/or hierarchical dependence. The former can be accounted for at the sampling level to avoid overlapping information. We review methods to estimate this sampling dependence. We also present a novel approach to account for dependence at the hierarchical level also, effectively down-weighting extreme effect size estimates that are hierarchically dependent. This hierarchical dependence is estimated using both random effects and Bayesian models. Implementation, comparison, and interpretation of results are discussed.

Elizabeth Stuart

Johns Hopkins University
estuart@jhsph.edu

“Generalizing Results from a Randomized Trial to a Broader Population:
Bridging Observational and Experimental Data”

While the immediate question in any randomized trial is the efficacy of the program among the study participants, the broader question is generally one of effectiveness: What are the effects of the program in real-world settings, among a broader population?

Little work has been done in thinking about how to make these kinds of generalizations from randomized trials to broader populations. We explore this issue by using a unique combination of data: a group randomized trial of Positive Behavior Interventions and Supports (PBIS), a school-wide violence prevention program, embedded within the broader statewide implementation of the PBIS program in schools across Maryland. The trial involved the random assignment of 37 Maryland elementary schools to PBIS or a control condition. We address the question of how the randomized trial of PBIS can inform policymakers about the broader effectiveness of the program statewide. Extensive data is available on the schools in the trial, as is information on schools statewide, including school characteristics, student suspensions, and achievement test results. Using the rich set of school characteristics available, we use propensity scores to examine how similar the randomized trial schools are to schools statewide and then weight the trial schools to represent the full set of schools in the state. We lay out the assumptions underlying this approach, being particularly clear about the types of schools to which we can and cannot generalize the findings from the randomized trial. In addition to assisting policymakers in assessing the broader effectiveness of the PBIS program, this work helps to provide a framework for considering the role of randomized trials within questions of broader program effectiveness.

Thomas A. Trikalinos

Tufts University

TTrikalinos@tufts-nemc.org

“Empirical Insights from Genetic Meta-analysis: Challenges, Biases, and Unique Considerations”

Meta-analysis, the quantitative synthesis of information from different studies, is used extensively to describe the genetic epidemiology of complex diseases. It summarizes quantitative information on genetic risks and provides the framework to explore and explain between-study diversity. For these reasons, meta-analytic techniques and evidence based medicine concepts have a key role in distinguishing genuine genetic associations of disease from spurious ones. We will discuss challenges in the conduct and interpretation of meta-analyses in genetic epidemiology through the presentation of empirical evidence and multiple examples.

SAMSI/CRSC Undergraduate Workshop

May 18 -May 23, 2008

<http://www.ncsu.edu/crsc/events/ugw08/index.php>

(All sessions are in Harrelson G100 unless otherwise noted.)

Sunday, May 18

7:00 Welcoming Reception in Honors Village Commons Multipurpose Room

Monday, May 19

8:30 Meet participants at Becton Hall. Transport to SAMSI.

9:10 Introduction to SAMSI, followed by presentations from current SAMSI programs.
 9:15 Environmental Sensor Networks (David Bell)
 10:00 Risk Analysis (Sourish Das)
 10:45 Break
 11:00 Random Media (Dr. Elaine Spiller)
 11:45 Lunch at SAMSI
 12:30 Vans transport participants to Harrelson Hall
 1:15 Introduction and Background (Dr. Ralph Smith)
 1:45 Introduction to the Forward Problem: Solving the Harmonic Oscillator System (Dr. Elaine Spiller)
 2:45 Break { Refreshments/Drinks available in HA 326
 3:00 Brief Introduction to the Computing System and MATLAB (Dr. Wiegang Zhong)
 4:30 Vans take participants to Lake Crabtree
 5:00 Dinner at Lake Crabtree

Tuesday, May 20

9:00 Linear Inverse Problems: A MATLAB Tutorial. (Qin Zhang)
 10:30 Break - Refreshments/Drinks available in HA 326
 10:45 Introduction to Basic Statistics and Probability (Justin Shows and Betsy Enstrom)
 12:15 Lunch
 1:15 Introduction to Statistical Inference (Dr. Vered Madar, Dr. Guang Cheng, Evangelos Evangelou, and Jaeun Choi)
 2:45 Break – Refreshments/Drinks available in HA 326
 3:15 Regression and Least Squares: A MATLAB Tutorial (Dr. Michael Porter)

Wednesday, May 21

9:00 Rotating Sessions (Cox Hall)

Vibrating Beam Data Collection at CRSC Laboratory

- Adam Attarian
- Dr. Ralph Smith

Graduate School Panel

- Dr. Ernie Stitzinger, Mathematics Department, NCSU
- Dr. Kim Weems, Statistics Department, NCSU

Career Panel (Facilitator: Dr. Cammey Cole Manning)

- Dr. Karen Chiswell, GlaxoSmtihKline
- Dr. Emily Lada, SAS

12:00 Box Lunches

1:15 Re°ection on Modeling and Data Collection (David Bell and Dr. Elaine Spiller)

- 2:15 Break – Refreshments/Drinks available in HA 326
- 2:30 Nonlinear Optimization and its Relationship to Statistical Inverse Problem (Martin Heller)
- 4:00 Teams Work on Inverse Problem (All)

Thursday, May 22

- 9:00 Statistical Analysis of Vibrating Beam Data (Dr. Gentry White)
- 10:00 Break – Refreshments/Drinks available in HA 326
- 10:15 Alternative Beam Model (Dr. Ralph Smith)
- 11:15 Teams Work on Inverse Problem (All)
- 12:30 Lunch
- 1:30 What could we do better? Alternative Statistical Models (Dr. Jayanta Pal, Qin Zhang, and Martin Heller)
- 2:30 Break – Refreshments/Drinks available in HA 326
- 3:00 Teams Work on Inverse Problem; Begin to Prepare Reports (All)
- 5:00 Dinner Break
- 6:30 Bowling (Meet under Harrelson Hall)

Friday, May 23

- 9:00 Presentations and Discussion (All)
- 10:30 Break – Refreshments/Drinks available in HA 326
- 10:45 Presentations and Discussion (All)
- 11:45 Closing Remarks & Workshop Evaluation (Drs. Cammey Cole Manning and Ralph Smith)
- 12:00 Lunch
- 1:00 Participants Depart for Home

SAMSI Graduate Fellow Seminar Day

Wednesday, May 7, 2008

NISS-SAMSI Building, Room 104

- 8:55-9:00 Opening Remarks
Ralph Smith, SAMSI Associate Director and Professor of Mathematics
North Carolina State University
- 9:00-9:20 “Texture Classification by Local Vector Autoregressive Models” Martin Heller, North Carolina State University
- 9:20-9:40 “Multivariate Generalized Linear ARMA Processes: An Application to Hurricane Activity”
Evangelos Evangelou, University of North Carolina
- 9:40-10:00 “Examining Wireless Sensor Network Function and the Environmental Processes Being Monitored”
David Bell, Duke University
- 10:00-10:20 “A Finite Element Method for Interface Problems with Locally Modified Triangulations”
Hui Xie, North Carolina State University

- 10:20-11:00 Coffee Break
- 11:00-11:20 “Stress Communication and Filtering of Viscoelastic Layers in Oscillatory Strain” Brandon Lindley, University of North Carolina
- 11:20-11:40 “Parameter Inference in Situations of Reduced Number of Transmissions” Kristian Lum, Duke University
- 11:40-12:00 “Objective Bayesian Analysis in One-way and Two-way Mixed Models for the Binary Response Data” Iris Lin, University of Missouri at Columbia
- 12:00-1:00 Lunch
- 1:00-1:20 “Quantum Probability Theory, Quantum Filtering and Control” Qin Zhang, North Carolina State University
- 1:20-1:40 “Performance Evaluation of Statistical Methods for Data Mining in Pharmacovigilance” Jaeun Choi, University of North Carolina
- 1:40-2:00 “Consistent Estimation and Variable Selection for Right-Censored Survival Data” Justin Shows, North Carolina State University
- 2:00-2:20 “Modeling Heterogeneity in Biological Materials by a Modified Immersed Boundary Method” Ke Xu, University of North Carolina
- 2:20 Closing Remarks
Ralph Smith, SAMSI and North Carolina State University

Tutorials and Workshop on Sequential Monte Carlo Methods
Opening Workshop
September 7-10, 2008

SCHEDULE

Sunday, September 7, 2008
[Radisson Hotel RTP](#)

Overview Tutorials

- 8:00-8:55 Registration and Continental Breakfast
- 8:55-9:00 Welcome
- 9:00-10:30 *On the Convergence and the Applications of Sequential Monte Carlo Methods*
Pierre Del Moral, INRIA Bordeaux
- 10:30-11:00 Coffee Break
- 11:00-12:30 *Sequential Monte Carlo and Related Methods for Analysing Complex Stochastic Systems*
Paul Fearnhead, Lancaster University

- 12:30 –1:45 Lunch (2nd Floor Room ABC)
- 1:45- 3:15 *An Introduction to Sequential Monte Carlo Schemes*
Hedibert Lopes, University of Chicago
- 3:15-3:45 Coffee Break
- 3:45-5:15 *Sequential Monte Carlo: General Frameworks and Applications*
Jun Liu, Harvard University

Monday, September 8, 2008

[Radisson Hotel RTP](#)

- 8:15-9:00 Registration and Continental Breakfast
- 9:00-9:15 Welcome
- 9:15-12:00 *Theory of Sequential Monte Carlo:*
- Uniform Approximations of Discrete Time Filters*
Dan Crisan, Imperial College
- Theory of Sequential Monte Carlo*
Eric Moulines, Ecole Nationale Supérieure des Télécommunications
- Coffee Break
- Discussion Session:*
Peter Bickel, University of California at Berkeley
Sylvain Rubenthaler, Université de Nice Sophia Antipolis
Nicholas Chopin, ENSAE-CREST
- 12:00-1:15 Lunch (2nd Floor Room ABC)
- 1:15-3:30 *Tracking and Large Scale Dynamic Systems:*
- Particle Filtering for Large Dimensional State Spaces with Multimodal Likelihoods*
Namrata Vaswani, Iowa State University
- Random Set/Point Process in Multi-target Tracking*
Ba-Ngu Vo, University of Melbourne
- Discussion Session:*
Monica Bugallo, Stony Brook University
Daniel Clark, Harriot-Watt University
Simon Godsill, University of Cambridge

- 3:30 – 4:00 Coffee Break
- 4:00-5:00 *New Researcher Session:*
- The Computational Complexity of Estimating Convergence Times*
Nayantara Bhatnagar, University of California, Berkeley
- SMC Methods for NASA Applications*
Vandi Verma, NASA
- Variance Reduction for Particle Filters of Systems with Time Scale Separation*
Jonathan Weare, Courant Institute
- 5:00-5:45 *Poster Advertisement Session (2 minute ads each)*
- 6:30–8:30 *Poster Session and Reception (2nd Floor Room ABC)*
 SAMSI will provide poster presentation boards and tape. The board dimensions are 4 ft. wide by 3 ft. high. They are tri-fold with each side being 1 ft. wide and the center 2 ft. wide. Please make sure your poster fits the board. The boards can accommodate up to 16 pages of paper measuring 8.5 inches by 11 inches.

Tuesday, September 9, 2008

[Radisson Hotel RTP](#)

- 9:00-12:00 *Applications in Economics and Decision Making:*
- Particle Learning and Smoothing*
Carlos Carvalho, University of Chicago
- The New Macroeconometrics: An Introductory Review*
Juan Rubio-Ramirez, Duke University
- Coffee Break
- Discussion Session Leaders:*
Michael Johannes, Columbia University
David de Jong, University of Pittsburgh
Hedibert Lopes, University of Chicago
- 12:00-1:15 Lunch (2nd Floor Room ABC)
- 1:15-3:45 *Continuous Time and Financial Applications:*
- Inference and Filtering for Diffusion Processes using Monte Carlo in the Path Space*
Omiros Papaspiliopoulos, Universitat Pompeu Fabra

Uses of Particle Filtering in Finance
Chris Rogers, Cambridge University

Discussion Session Leaders:
Ed Ionides, University of Michigan
Nick Polson, University of Chicago
Jonathan Stroud, University of Pennsylvania

3:45 – 4:15 Coffee Break

4:15-5:15 *New Researcher Session:*

The Ensemble Kalman Filter: a State Estimation Method for Hazardous Weather Prediction

Sarah Dance, University of Reading

Particle Methods for High-Dimensional Traffic Estimation Problems

Ludmila Mihaylova, Lancaster University

State-space Smoothing Using Sequential Monte Carlo

Mark Briers, QinetiQ Limited

Wednesday, September 10, 2008

[Radisson Hotel RTP](#)

9:00-12:00 *Population Methods and Other Aspects of Methodology:*

Adaptive Importance Sampling in General Mixture Classes

Christian Robert, Ceremade - Université Paris-Dauphine

Particle Markov Chain Monte Carlo

Arnaud Doucet, University of British Columbia

Coffee Break

Discussion Session Leaders:
Rong Chen, Rutgers University
Merlise Clyde, Duke University

12:00-1:15 Lunch (2nd Floor Room ABC)

1:15–2:45 *Working Group Formation and Initial Meeting*

2:45-3:30 *Working Group Reports*

Thursday and Friday: Initial working group meetings at SAMSI.

**Program on Sequential Monte Carlo Methods
Opening Workshop
September 7-10, 2008**

SPEAKER ABSTRACTS

Christophe Andrieu

University of Bristol
c.andrieu@bris.ac.uk

“Particle Markov Chain Monte Carlo”

Markov chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) methods have emerged as the two main tools to sample from high-dimensional probability distributions. Although asymptotic convergence of MCMC algorithms is ensured under weak assumptions, the performance of these latter is unreliable when the proposal distributions used to explore the space are poorly chosen and/or if highly correlated variables are updated independently. We show here how it is possible to build efficient high-dimensional proposal distributions using SMC methods. This allows us not only to improve over standard MCMC schemes but also to make Bayesian inference feasible for a large class of statistical models where this was not previously the case. We demonstrate these algorithms on various non-linear non-Gaussian state-space models, a stochastic kinetic model and Dirichlet process mixtures.

(joint with A. Doucet and R. Holenstein)

Nayantara Bhatnagar

University of California, Berkeley
nayan@stat.berkeley.edu

“The Computational Complexity of Estimating Convergence Times”

In practice, there are many diagnostics used to test convergence and our aim is to formally analyze the complexity of the computational problem. We present some results on the computational complexity of estimating the convergence time of a Markov chain. This is joint work with Andrej Bogdanov, Elchanan Mossel and Salil Vadhan.

Carlos Carvalho

University of Chicago
carlos.carvalho@chicagogsb.edu

“Particle Learning and Smoothing”

This paper provides novel particle learning (PL) methods for sequential parameter learning and smoothing in state space models with non-normal errors, non-linear observation equations, and non-linear state evolutions. The methods extend existing particle methods by incorporating unknown parameters, utilizing sufficient statistics, for the parameters and/or the states, and allowing for nonlinearities in the state and/or observation equation. We also show how to solve the state smoothing problem, integrating out parameter uncertainty. Previously, the only approach available for this marginal smoothing problem is MCMC. We show that our algorithms outperform MCMC, as well existing particle filtering algorithms such as the mixture Kalman filter.

Dan Crisan

Imperial College London
d.crisan@imperial.ac.uk

“Uniform Approximations of Discrete Time Filters”

Throughout recent years, various sequential Monte Carlo methods have been widely applied to various applications involving the evaluation of the generally intractable stochastic discrete time filter. Although convergence results exist for finite time intervals, a stronger form of convergence, namely, uniform convergence, is required for bounding the error on an infinite time interval. I will present a number of results containing easily verifiable conditions for the filter applications that are sufficient for the uniform convergence of certain particle filters. Essentially, the conditions require the observations to be accurate enough. No mixing or ergodicity conditions are imposed on the signal process. This is joint work with Kari Heine.

Sarah Dance

University of Reading
s.l.dance@reading.ac.uk

“The Ensemble Kalman Filter: a State Estimation Method for Hazardous Weather Prediction”

Numerical weather prediction models require an estimate of the current state of the atmosphere as an initial condition. Observations only provide partial information, so they are usually combined with prior information, in a process called data assimilation. The dynamics of hazardous weather such as storms is very nonlinear, with only a short predictability timescale, thus it is important to use a nonlinear, probabilistic filtering method to provide the initial conditions. Unfortunately, the state space is very large (about 10^7 variables) so approximations have to be made.

The Ensemble Kalman filter (EnKF) is a quasi-linear filter that has recently been proposed in the meteorological and oceanographic literature to solve this problem. The filter uses a forecast ensemble (a Monte Carlo sample) to estimate the prior statistics. While such filters look promising, a number of issues have arisen in the development and application of ensemble-based data assimilation techniques. In this talk we will consider

some of the fundamental problems associated with sampling errors due to small ensemble sizes, and discuss the merits of some of the various implementation schemes.

David DeJong

University of Pittsburgh
dejong@pitt.edu

“An Efficient Approach to Analyzing State-Space Representations”

We develop a numerical procedure that facilitates efficient likelihood evaluation and filtering in applications involving non-linear and non-Gaussian state-space models. The procedure approximates necessary integrals using continuous approximations of target densities. Construction is achieved via efficient importance sampling, and approximating densities are adapted to fully incorporate current information.

Pierre Del Moral

INRIA
Pierre.Del-Moral@inria.fr

“On the Convergence and the Applications of Sequential Monte Carlo Methods”

This lecture is concerned with the convergence analysis and the applications of sequential Monte Carlo methods. We provide some recent stochastic models including Feynman-Kac distributions flows and their statistical interpretations in terms of interacting particle systems and genealogical tree based models. We discuss a variety of application model areas including stochastic engineering (signal processing, rare event simulation), particle physics, computational chemistry (directed polymers, Schroedinger ground state energies calculations) and biology (population dynamics, genetic algorithms). In the second part of the lecture, we provide a series of convergence results including multivariate and functional central limit theorems and uniform exponential concentration estimates w.r.t. the time parameter.

Paul Fearnhead

Lancaster University
p.fearnhead@lancs.ac.uk

“Sequential Monte Carlo and Related Methods for Analysing Complex Stochastic Systems”

We consider analysis of complex stochastic models based upon partial information. MCMC and reversible jump MCMC are often the methods of choice for such problems, but in some situations they can be difficult to implement; and suffer from problems such as poor mixing, and the difficulty of diagnosing convergence. Here we review three alternatives to MCMC methods: importance sampling, the forward-backward algorithm, and sequential Monte Carlo (SMC). We discuss how to design good proposal densities for importance sampling, show some of the range of models for which the forward-

backward algorithm can be applied, and show how resampling ideas from SMC can be used to improve the efficiency of the other two methods.

Jun Liu

Harvard University
jliu@stat.harvard.edu

“Sequential Monte Carlo: General Frameworks and Applications”

Sequential Monte Carlo is built on the importance sampling principle and utilizes resampling and Markov chain iterations to improve efficiencies. Its basic building block, sequential importance sampling (SIS), can be understood as a generic strategy to sequentially/recursively construct an importance sampling distribution for high-dimensional problems and it produces weighted multiple samples as its end result. With these multiple samples, new information can be easily “learnt” by adjusting the associated importance weights. The recursive nature of state-space models make it ideal to develop nonlinear filters based on the SIS strategy. Since the importance weights tend to be more and more skewed as the system evolves, ideas of resampling, rejection sampling, kernel density estimation, MCMC iterations are necessary for the control of Monte Carlo variations in SIS. We show how these basic ideas can be implemented by examples ranging from energy minimization for polymer folding to target tracking and contingency table analysis.

Hedibert Lopes

University of Chicago
hlopes@chicagogsb.edu

“An Introduction to Sequential Monte Carlo Schemes”

The tutorial starts reviewing Monte Carlo sampling via importance function and its natural role into drawing from unconventional posteriors. Then sequential importance sampling is introduced to deal with online estimation of state vectors in, potentially nonnormal and/or nonlinear, dynamic models. Sequential particle impoverishment is discussed and auxiliary particle filters are introduced to replenish the particles. Next, we present particle filters that deal with sequential parameter learning, smoothing and take advantage of (the possible existence of) sufficient statistics for states and/or parameters.

Lyudmila Mihaylova

Lancaster University
mila.mihaylova@lancaster.ac.uk

“Particle Methods for High-Dimensional Traffic Estimation Problems”

Traffic flow on motorways is a nonlinear, many-particle phenomenon, with complex interactions between vehicles such as traffic jams, stop-and-go-waves. To manage urban and freeway road traffic, traffic data is collected in traffic control centers in many countries. This data is often used for traffic monitoring, control, and information dissemination. Direct traffic measurements from sensors are corrupted by noises, or some

data may be missing, and additionally data may be aggregated over a longer time period. This talk presents a formulation of the traffic estimation problem within Bayesian framework and particle filters aimed at on-line traffic flow prediction in a centralised and in a parallelised manner. The filters' performance and suitability to large networks will be discussed.

Eric Moulines

Ecole Nationale Supérieure des Télécommunications
eric.moulines@orange.fr

“Theory of Sequential Monte Carlo”

Despite many theoretical advances, the large-sample theory of SMC remains a question of central interest. In this talk, we establish a law of large numbers and a central limit theorem as the number of particles gets large. We introduce the concepts of "weighted sample" consistency and asymptotic normality, and derive conditions under which the transformations of the weighted sample used in the SMC algorithm preserve these properties. To illustrate our findings, we analyze SMC algorithms to approximate the filtering distribution in state-space models. We show how our techniques allow to relax restrictive technical conditions used in previously reported works and provide grounds to analyze more sophisticated sequential sampling strategies, including branching, resampling at randomly selected times, auxiliary sampling, etc.

Omiros Papaspiliopoulos

Universitat Pompeu Fabra
omiros.papaspiliopoulos@upf.edu

“Inference and Filtering for Diffusion Processes using Monte Carlo in the Path Space”

Diffusion processes is a large family of time-series models with a wide and increasing range of applications. They can be used either to model directly observed data, or as a (partially observed or latent) component of more complex hierarchical models. From a statistical point of view interest lies in the estimation of unknown parameters of such models and of the process itself when it is only partially observed.

However, inference for partially observed diffusions involves marginal laws (e.g. the transition kernel of the process) which are typically intractable. This raises serious theoretical and computational challenges.

The talk focuses on the computational challenge and develops appropriate Monte Carlo methodology for parameter and process estimation. The Monte Carlo methods we consider include rejection sampling (RS), importance sampling (IS) and sequential versions of it, and Markov chain Monte Carlo (MCMC). We show that it is natural to derive theoretical algorithms in the infinite-dimensional space of the diffusion paths, that is the path space. Practical implementation of these algorithms can then be achieved either approximately, by finite-dimensional projections (discretizations) or by exact retrospective methods.

The infinite-dimensional setup sheds light and gives solutions to problems which are masked in alternative (and popular) methods which first project to finite-dimensions and then design the Monte Carlo algorithm. The limiting behaviour (as the approximation gets finer) of such finite-dimensional algorithms often has serious deficiencies, which include infinite variance of IS weights, poor mixing of MCMC algorithms for simulation of paths, reducibility of MCMC algorithms which update unobserved paths and parameters. We will also demonstrate how the infinite-dimensional setup justifies certain ad-hoc finite-dimensional algorithms which have proved successful in this context.

Christian Robert

Université Paris Dauphine
xian@ceremade.dauphine.fr

“Adaptive Importance Sampling in General Mixture Classes”

In this work, joint with O. Cappé, R. Douc, A. Guillin and J.M. Marin, we propose an adaptive algorithm that iteratively updates both the weights and component parameters of a mixture importance sampling density so as to optimise the performance of importance sampling, as measured by an entropy criterion. The method is shown to be applicable to a wide class of importance sampling densities, which includes in particular mixtures of multivariate Student t distributions. The performance of the proposed scheme is studied on both artificial and real examples, highlighting in particular the benefit of a novel Rao-Blackwellisation device which can be easily incorporated in the updating scheme.

Chris Rogers

University of Cambridge
l.c.g.rogers@statslab.cam.ac.uk

“Uses of Particle Filtering in Finance”

This talk will be a free-form discussion of a number of examples from finance, where particle filtering offers itself as a natural way to fit models of varying degrees of complexity. The successes and limitations, frustrations and fixes, will be discussed, more to highlight what we would like to be able to do than to make exorbitant claims.

Sylvain Rubenthaler

Université de Nice-Sophia Antipolis
rubentha@unice.fr

“Propagation of Chaos for Various Particle Systems, Coalescence Trees and Applications”

For various particle systems (genetic in discrete and continuous time, Bird and Nanbu systems), we write a coalescent-tree based functional representation of the q -th tensor product of the empirical measure associated to a particle system. This representation uses

combinatorics on trees and allows for an extension of the Wick formula. As a consequence, we prove the convergence of U-statistics of such systems (a.s. and with a CLT).

Juan Rubio-Ramirez,
Duke University
juan.rubio-ramirez@duke.edu

“The New Macroeconometrics: An Introductory Review”

Namrata Vaswani
Iowa State University
namrata@iastate.edu

“Particle Filtering for Large Dimensional State Spaces with Multimodal Likelihoods”

We study efficient importance sampling techniques for particle filtering (PF) when either (a) the observation likelihood is frequently multimodal or heavy-tailed, or (b) the state space dimension is large or both. When the likelihood is multimodal, but the state transition prior is narrow enough, the optimal importance density is usually unimodal. Under this assumption, many techniques have been proposed. But when the prior is broad, this assumption does not hold. We study how existing techniques can be generalized to situations where the optimal importance density is multimodal, but is unimodal conditioned on a part of the state vector. Sufficient conditions to test for the unimodality of this conditional posterior are derived. Our result is directly extendable to testing for unimodality of any posterior.

The number of particles, N , to accurately track using a PF increases with state space dimension, thus making any regular PF impractical for large dimensional tracking problems. But in most such problems, most of the state change occurs in only a few dimensions, while the change in the rest of the dimensions is small. Using this property, we propose to replace importance sampling from a large part of the state space (whose conditional posterior is narrow enough) by posterior mode tracking.

Applications in sequentially estimating spatially varying physical quantities such as temperature or pressure in a large area using a network of sensors which may be nonlinear and/or may have non-negligible failure probabilities and in dynamic computer vision problems such as deformable contour tracking or landmark shape tracking have been studied and improved performance demonstrated with respect to existing work.

Vandi Verma
NASA Jet Propulsion Laboratory
vandi@jpl.nasa.gov

“SMC Methods for NASA Applications”

Ba-Ngu Vo
University of Melbourne

bv@ee.unimelb.edu.au

“Random Set/Point Process in Multi-target Tracking”

Driven primarily by aerospace applications, multi-target tracking has been an intensive research area since the early 1970s. Today multi-target filtering has found its way into a range of diverse applications. Mahler's Finite set statistics (FISST) provides a general systematic foundation for multi-target filtering based on the theory of random finite set (RFS). The theory of RFS, or point process, is a rigorous mathematical discipline for dealing with random spatial patterns that has long been used by statisticians in many diverse applications including agriculture, geology, seismology, and epidemiology. The RFS framework has led to the development of novel and efficient multi-target filters, which attracted substantial interests. This talk outlines recent developments of RFS theory in multi-target filtering.

Jonathan Weare

New York University, Courant Institute
weare@cims.nyu.edu

“Variance Reduction for Particle Filters of Systems with Time Scale Separation”

I present a particle filter construction for a system that exhibits time scale separation. The separation of time scales allows two simplifications: i) the use of the averaging principle for the dimensional reduction of the dynamics for each particle during the prediction step and ii) the factorization of the transition probability for the Rao-Blackwellization of the update step. The resulting particle filter is faster and has smaller variance than the particle filter based on the original system. I present the results of numerical tests on a multiscale stochastic differential equation and on a multiscale pure jump diffusion motivated by chemical reactions.

Program on Algebraic Methods in Systems Biology and Statistics
Opening Workshop
September 14-17, 2008

SCHEDULE

Sunday, September 14, 2008
[Radisson Hotel RTP](#)

Overview Tutorials

11:15-Noon Registration

Noon –1:00 Lunch

- 1:00- 2:15 *Algebraic Statistics*
Bernd Sturmfels, University of California, Berkeley
- 2:15-2:30 Break
- 2:30-3:45 *An Introduction to Systems Biology*
Reinhard Laubenbacher, Virginia Bioinformatics Institute
- 3:45-4:00 Break
- 4:00-5:15 *Phylogenetics*
Elizabeth Allman, University of Alaska

Monday, September 15, 2008

[Radisson Hotel RTP](#)

- 8:00-8:45 Registration and Continental Breakfast
- 8:45-9:00 Welcome
- 9:00-9:45 *Algebraic Statistical Models*
Mathias Drton, University of Chicago
- 9:45-10:00 Questions and Discussion
- 10:00-10:45 *The Geometry of Multisite Phosphorylation*
Jeremy Gunawardena, Harvard University
- 10:45-11:00 Questions and Discussion
- 11:00-11:15 Coffee Break
- 11:15-Noon *Combinatorial Insights into RNA Folding*
Christine Heitsch, Georgia Institute of Technology
- Noon-12:15 Questions and Discussion
- 12:15-1:30 Lunch
- 1:30-2:15 *Algebra, Automata, Algorithms, Biology and Beyond*
Bud Mishra, Courant Institute, New York University
- 2:15-2:30 Questions and Discussion
- 2:30-3:15 *Reverse Engineering Nested Canalizing Boolean Networks*
Abdul Jarrah, Virginia Bioinformatics Institute

- 3:15-3:30 Questions and Discussion
- 3:30-3:45 Break
- 3:45-4:30 *Species and Genomes: Lessons from my Favorite Symbionts*
Chris Schardl, University of Kentucky
- 4:30-4:45 Questions and Discussion
- 4:45-6:00 *Panel Discussion:*
Jeremy Gunawardena, Harvard University
Ina Hoeschele, Virginia Polytechnic Institute
Alexander Hartemink, Duke University
Greg Rempala, Medical College of Georgia
Brett Tyler, Virginia Polytechnic Institute
- 6:00-6:30 *Poster Advertisement Session* (2 minute ads each)
- 6:30-8:30 *Poster Session and Reception*
SAMSI will provide poster presentation boards and tape. The board dimensions are 4 ft. wide by 3 ft. high. They are tri-fold with each side being 1 ft. wide and the center 2 ft. wide. Please make sure your poster fits the board. The boards can accommodate up to 16 pages of paper measuring 8.5 inches by 11 inches.

Tuesday, September 16, 2008

[Radisson Hotel RTP](#)

- 8:15-9:00 Registration and Continental Breakfast
- 9:00-9:45 *Algebraic Structure of System Design Space for Metabolic Pathways and Gene Circuits*
Michael Savageau, University of California, Davis
- 9:45-10:00 Questions and Discussion
- 10:00-10:45 *Polynomial Equations and Instabilities in Biochemical Reaction Networks*
Gheorghe Craciun, University of Wisconsin
- 10:45-11:00 Questions and Discussion
- 11:00-11:15 Coffee Break
- 11:15-Noon *Algebraic Geometry, Empirical Process, and Singular Model Evaluation*
Sumio Watanabe, Tokyo Institute of Technology
- Noon-12:15 Questions and Discussion

- 12:15-1:30 Lunch
- 1:30-2:15 *Using Groebner Bases to Reconstruct Regulatory Modules in C. elegans*
Brandy Stigler, Mathematical Biosciences Institute
- 2:15-2:30 Questions and Discussion
- 2:30-3:15 *Algebraic Combinatorics for Predicting Virus Assembly Pathways*
Meera Sitharam, University of Florida
- 3:15-3:30 Questions and Discussion
- 3:30-3:45 Break
- 3:45-4:30 *The Algebra and Statistics of Biological Sequence Alignment*
Lior Pachter, University of California, Berkeley
- 4:30-4:45 Questions and Discussion
- 4:45-6:00 *Panel Discussion:*
Bernd Sturmfels, University of California, Berkeley
Seth Sullivant, North Carolina State University
Rudy Yoshida, University of Kentucky
Peter Beerli, Florida State University
Elizabeth Allman, University of Alaska

Wednesday, September 17, 2008

[Radisson Hotel RTP](#)

- 8:15-9:00 Registration and Continental Breakfast
- 9:00-9:45 *Inferring Genetic Regulatory Networks in Host-Pathogen Interactions*
Brett Tyler, Virginia Polytechnic Institute
- 9:45-10:00 Questions and Discussion
- 10:00-10:45 *Mathematical Models of Evolutionary Escape*
Niko Beerenwinkel, ETH Zurich
- 10:45-11:00 Questions and Discussion
- 11:00-11:15 Coffee Break
- 11:15-Noon *Between Algebraic Statistics and Information Geometry*
Eva Riccomagno, Università di Genova
- Noon-12:15 Questions and Discussion

- 12:15-1:30 Lunch
- 1:30-2:15 *Algebraic Statistics for $p1$ Random Graph Models: Markov Bases and Their Uses*
Stephen Fienberg, Carnegie Mellon University
- 2:15-2:30 Questions and Discussion
- 2:30–2:45 Break
- 2:45-5:00 *Break-out Sessions for Working Groups*

(**Thursday and Friday:** Possible working group meetings at SAMSI)

Program on Algebraic Methods in Systems Biology and Statistics
Opening Workshop
September 14-17, 2008

SPEAKER ABSTRACTS

Elizabeth Allman
University of Alaska Fairbanks
e.allman@uaf.edu

“Phylogenetics”

This tutorial will give an introduction to the mathematics and statistics of phylogenetics, the branch of biology which seeks to infer evolutionary relationships between organisms.

There are several approaches to the inference of phylogenetic trees, from DNA or protein sequences, including the statistical Maximum Likelihood and Bayesian methods. These methods depend upon a probabilistic model describing the evolution of sequences from a common ancestor.

Many of the models used in data analysis are algebraic: the joint distribution of patterns in the sequences is the image of a polynomial parameterization. This allows the application of viewpoints and techniques from algebraic geometry.

This talk will give an overview of mathematical phylogenetics, emphasizing the places where algebraic methods have been and will likely continue to be useful in enhancing our understanding.

Niko Beerenwinkel
ETH Zurich

niko.beerenwinkel@bsse.ethz.ch

“Mathematical Models of Evolutionary Escape”

We introduce and analyze a class of waiting time models for the accumulation of genetic changes. Conjunctive Bayesian networks are defined by a partially ordered set of mutations and by the rate of fixation of each mutation, or the conditional probability of its fixation. The partial order encodes constraints on the order in which mutations can fixate in the population, shedding light on the mutational pathways underlying the evolutionary process. We present solutions to maximum likelihood parameter estimation and to likelihood-based model selection. These models can be used to compute the probability of a pathogen escaping from selective pressure by accumulating mutations. Similarly, we discuss applications to the evolution of cancer.

Gheorghe Craciun

University of Wisconsin
craciun@math.wisc.edu

“Polynomial Equations and Instabilities in Biochemical Reaction Networks”

Biochemical reaction network models give rise to polynomial dynamical systems that are usually high dimensional, nonlinear, and have many unknown parameters. Due to the presence of these unknown parameters (such as reaction rate constants) direct numerical simulation of the chemical dynamics is practically impossible. On the other hand, we will show that important properties of these systems are determined only by the network structure, and do not depend on the unknown parameters. Also, we will show how some of these results can be generalized to systems of polynomial equations that are not necessarily derived from chemical kinetics. In particular, we will point out connections with classical problems in algebraic geometry, such as the real Jacobian conjecture.

Mathias Drton

University of Chicago
drton@uchicago.edu

“Algebraic Statistical Models”

Many statistical models are defined in terms of polynomial constraints, or in terms of polynomial or rational parametrizations. In such algebraic statistical models, there is often an intimate connection between the geometry of parameter spaces and the behavior of statistical procedures. This talk will exemplify such connections for classical methods of likelihood inference such as likelihood ratio and Wald tests.

Stephen Fienberg

Carnegie Mellon University
fienberg@stat.cmu.edu

“Algebraic Statistics for p_1 Random Graph Models: Markov Bases and Their Uses”

In a seminal 1981 paper, Holland and Leinhardt described what they referred to as the p_1 model for describing dyadic interactions in a social network summarized in the form of a directed graph. Their model which is log-linear in form, allows for effects due to differential attraction (popularity) and expansiveness, as well as an additional effect due to reciprocation. Fienberg and Wasserman re-represented the p_1 model in contingency table form and gave it a log-linear representation in that setting. In this paper we reconsider the Holland-Leinhardt p_1 model using the tools of algebraic geometry now embodied in the area of research now referred to as algebraic statistics. In particular, we derive Markov bases for p_1 and we link these to the results on Markov bases for log-linear models for contingency tables. We briefly describe some of potential uses of the Markov bases, including the problem of goodness-of-fit, and we discuss some possible generalizations to the class of $p^?$ models.

Stephen E. Fienberg, Sonja Petrovi C, and Alessandro Rinaldo

Jeremy Gunawardena

Harvard Medical School
jeremy@hms.harvard.edu

“The Geometry of Multisite Phosphorylation”

With the emergence of systems biology, ordinary differential equation models are often used to study the dynamics of biomolecular networks within cells. Such studies are hampered by intractable nonlinearities in the equations and a lack of knowledge of the relevant parameters. Simulation is usually the only option. However, when such models are derived from the principle of mass-action, their steady states necessarily form an algebraic variety over $R(a)$ - the field of rational functions in the parameters with real coefficients. This suggests that algebraic geometry may provide a framework for making assertions about the steady state behaviour of such systems in a parameter-independent manner. In this talk I will discuss multisite protein phosphorylation in which a kinase and a phosphatase act on a substrate with n phosphorylation sites. In this case the substrate phospho-forms, at steady state, form a rational, projective algebraic curve over $R(a)$, from which several insights into the systems properties of multisite phosphorylation can be deduced. These predictions are currently being experimentally tested in our laboratory.

Christine Heitsch

Georgia Tech
heitsch@math.gatech.edu

“Combinatorial Insights into RNA Folding”

An RNA molecule is a linear biochemical chain which folds into a three dimensional structure via a set of 2D base pairings known as a nested secondary structure. Reliably determining a secondary structure for large RNA molecules, such as the genomes of most viruses, is an important open problem in computational molecular biology. We give

combinatorial results which yield insights into the interaction of local and global constraints in RNA secondary structures and suggest new directions in understanding the folding of RNA viral genomes.

Abdul Salam Jarrah

Virginia Tech
ajarrah@vbi.vt.edu

“Reverse Engineering Nested Canalizing Boolean Networks”

Inferring a biochemical network from experimental data is one of the main research areas in systems biology. Data such as transcripts are used to infer either the structure (topology) or the function (dynamics) of a gene regulatory network. Although there usually are many models that fit the given data, the desired models are biologically meaningful and have some favorable properties such as canalization. Boolean nested canalizing networks have been recently shown to have robust and stable dynamics and have been suggested as appropriate models of gene regulatory networks. In this talk, we present a method for inferring gene regulatory networks as Boolean nested canalizing networks. This method based on the framework of polynomial dynamical systems and uses tools from computational algebraic geometry.

Reinhard Laubenbacher

Virginia Bioinformatics Institute
reinhard@vbi.vt.edu

“An Introduction to Systems Biology”

This tutorial will provide an introduction to the key concepts and central problems of systems biology. No advanced biological background is required.

Bud Mishra

Courant Institute, NYU
mishra@nyu.edu

“Algebra, Automata, Algorithms, Biology and Beyond”

In this talk, I will introduce a new approach to modeling dynamics of biological systems and its relations to certain problems in differential algebra, automata theory and algorithmics. The questions, addressed here, are central to the success of the emerging field of systems biology and relate to questions in decidability theory, algorithmic algebra, hybrid automata models, etc. A particular focus in this talk is on approaches embedded in an embryonic program, dubbed “Algorithmic Algebraic Model Checking,” and its power and limitations.

Lior Pachter

University of California, Berkeley

lpachter@math.berkeley.edu

“The Algebra and Statistics of Biological Sequence Alignment”

We will explain the biological sequence alignment problem, and discuss its connections to algebraic statistics. In particular, we will overview recent theoretical and practical developments including a counterexample to the "square root of n" conjecture by Cynthia Vinzant, and an algorithm for exact statistics of BLAST by Kevin McLoughlin. Finally, we will discuss a new approach to "statistical alignment" that we are developing.

Eva Riccomagno

Universita` di Genova
riccomagno@dima.unige.it

“Between Algebraic Statistics and Information Geometry”

The interaction of two established mathematical theories, algebraic geometry and differential geometry, with mathematical statistics and probability has lead to Algebraic Statistics and Information Geometry, respectively.

The awareness that important probabilistic and statistical notions and models have an algebraic and/or geometrical nature prompted research at a fundamental level. Both algebraic statistics and information geometry are showing how mathematical statistics is located at the frontier of current research in mathematical science. In algebraic statistics statistical models, especially exponential models, are studied as algebraic varieties. Whilst information geometry is pinned upon differential geometry and was started by the observation that Fisher information can be seen as a Riemannian metric on a statistical model.

A purpose of this talk is to contribute in the direction of a closer interaction between algebraic statistics and information geometry. We will do this by presenting some examples from the introductory chapter and the final chapter in [Gibilisco, P., Riccomagno, E., Rogantin, M-P. and Wynn, H.P. (eds) Algebraic and geometric methods in statistics. CUP, Cambridge].

This presentation is in collaboration with G. Pistone and H.P. Wynn.

Michael Savageau

University of California, Davis
masavageau@ucdavis.edu

“Algebraic Structure of System Design Space for Metabolic Pathways and Gene Circuits”

Determining quality of performance for a biological system is critical to identifying and elucidation its design principles. This important task is greatly facilitated by enumeration

of regions within the system's design space that exhibit qualitatively distinct functions. First, I will review a few examples of design spaces that have proved useful in revealing design principles for elementary gene circuits. Second, I will present an approach to the generic construction of design spaces. This approach is grounded in the power-law equations that characterize traditional chemical kinetics and, by transformation, the rational functions that characterize biochemical kinetics. In steady state, the analysis of these equations can be reduced to that of linear algebraic equations. Third, these methods will be illustrated with applications from common classes of biochemical network motifs, including unbranched pathways, branched pathways, moiety-transfer cycles, and elementary gene circuits. Finally, in the case of moiety-transfer cycles, predictions will be tested with experimental data from human erythrocytes.

Chris Schardl

University of Kentucky
schardl@uky.edu

“Species and Genomes: Lessons from my Favorite Symbionts”

Despite the large and rapidly growing number of reports of species with sequenced genomes, no species have actually been sequenced. In fact, only one or a few individuals within each species has been sequenced. The difference between an individual genome and the population of genomes in a species is profound, and both genomics and phylogenetics need to take greater account of whole species. To illustrate this, I will present relevant findings and outstanding questions from the past 20 years of work on the epichloae, a group of fungi that are symbiotic with grasses and are well known for producing suites of bioprotective metabolites. For tens of millions of years, variation in their host interactions and beneficial characteristics has made these symbionts key factors in evolutionary adaptability of the grasses.

Meera Sitharam

University of Florida
sitharam@ufl.edu

“Algebraic Combinatorics for Predicting Virus Assembly Pathways”

Viruses and other macromolecular assemblies are outstanding examples of spontaneous, rapid, nanoscale self-assembly processes in nature. Yet this assembly process is poorly understood. Better understanding can help arrest assembly for controlling infections and can help encourage assembly for gene therapy with viral vectors as well as for engineering robust nanoscale self-assembly processes. While the final X-ray crystallography structure is often available, what is lacking is snapshot data that would illuminate the process of assembly. We use algebraic geometry, combinatorial rigidity consistent with biophysical principles for modeling the nanoscale molecular interaction. From this we extract microscale assembly rules, and use these rules in conjunction with the action of symmetry groups to model assembly pathways at the microscale.

We avoid both expensive dynamical simulation as well as to avoid blackbox models

obtained purely by automated, data-intensive, statistical inference. We instead develop intuitive, static mathematical theories consistent with existing biophysical principles, i.e., we develop new biophysical theories. The resulting multiscale models of assembly pathways in turn yield efficient algorithms to predict assembly pathway probabilities when the final assembled structure is input.

The advantages of this type of modeling are the following.

(a) The developed models are tractable, static, modular, transparent: i.e., its parts are forward and backward analyzable and hence better tunable and testable.

(b) The developed theory is consistent with and based on existing biophysical principles and can bring to bear considerable mathematical muscle. As a result, the models lend themselves to intuitive reasoning as opposed to only simulation. This helps to intelligently cut down experimental possibilities and guide decisions in the design of further laboratory experiments, vastly improving efficiency.

(c) They can be combined with other theories and models of the same system or of other systems that interact with them.

(d) While the developed model is reduced down to the simplest static principles, dynamical simulation can be incorporated if necessary; similarly, while the developed model is transparent, it can incorporate blackbox models that are obtained by pure automated statistical inference.

We will motivate our model by giving some success stories of its predictions on real viruses.

This work was supported in part by: an NSF-QUBIC grant(2002-2006), a NSF-NER grant (2004-2006) and an NSF-DMS/NIGMS grant (current). Current collaborators: Mavis Agbandje-Mckenna, Director center of structural biology, and Miklos Bo'na, Mathematics, both at the University of Florida.

Brandy Stigler

Mathematical Biosciences Institute

bstigler@mail.smu.edu

“Using Groebner Bases to Reconstruct Regulatory Modules in *C. elegans*”

Since the completion of the cell lineage of the nematode *Caenorhabditis elegans*, key genes have been identified in cell fate specification. In particular, the gene *pal-1* is required for development of muscle and ectoderm cells during embryogenesis. Of biological importance is the description of the network of interactions among these so-called tissue identity genes.

In this study we utilized the systems-biology approach of reverse engineering, that is, the construction of mathematical models based on system-wide observations, to model the network of the tissue identity genes specified by *pal-1*. We developed an algorithm using tools from computational algebraic geometry to construct polynomial dynamical systems (PDSs), which are polynomial functions over a finite field, from experimental data. The

algorithm encodes all PDSs that fit a given data set in a zero-dimensional ideal and selects a minimal PDS by computing a Groebner basis for the ideal. This encoding allows for the construction of the entire discrete model space and the computation of model distribution via the Groebner fan.

We have applied the algorithm to microarray time series data for a collection of *pal-1*-dependent genes. We present the results of the method, which includes a small number of most likely PDSs, as well as predicted regulatory modules for muscles and ectoderm development.

Bernd Sturmfels

University of California
bernd@math.berkeley.edu

“Algebraic Statistics”

This tutorial offers an introduction to Algebraic Statistics for non-experts.

Brett Tyler

Virginia Polytechnic Institute and State University
bmt Tyler@vt.edu

“Inferring Genetic Regulatory Networks in Host-Pathogen Interactions”

The outcome of a host-pathogen interaction may be considered to be governed by a genetic regulatory network that encompasses both organisms. High throughput functional genomics data can be generated with describes the concentrations of mRNAs, proteins and metabolites during the interaction. However, deconvoluting this information into a computational network model that has useful predictive value remains a major challenge. One of the severest challenges is that functional genomics data typically contain drastically fewer samples (e.g. time points) than variables (e.g. genes). I will report progress in two approaches we are using to address this challenge. In the first, quantitative disease resistance in soybean against the oomycete pathogen *Phytophthora sojae*, we are using genetical genomics to infer genetic regulatory networks that are associated with disease resistance. We have assayed 297 recombinant inbred lines of soybean segregating for *P. sojae* resistance, using 2600 Affymetrix GeneChips that contain probes for both host and pathogen genes. Using methods refined using yeast data we are using our data to identify networks of expression QTLs associated with the disease resistance QTLs. In the second project we are using transcriptional profiles of the oxidative stress responses of yeast, of *Arabidopsis* plant tissue and of *P. sojae* to evaluate the use of summary variables, such as those obtained using Principal Components Analysis, to create sequential dynamical models of the responses. We have created an approach called biologically plausible interpolation to infer families of models consistent with the data and to predict additional experiments that most cost-effectively refine the models.

Sumio Watanabe

Tokyo Institute of Technology
swatanab@pi.titech.ac.jp

“Algebraic Geometry, Empirical Process, and Singular Model Evaluation”

A statistical model which has hierarchical structure or hidden variables is nonidentifiable and singular. In singular statistical models, it has been difficult to estimate its generalization error from random samples. In this presentation, I show that there exist two universal equations among four errors, Bayes and Gibbs, generalization and training. By using these universal equations, we can predict Bayes and Gibbs generalization errors from Bayes and Gibbs training errors without any knowledge of the true distribution. This result is mathematically equal to a generalization of AIC to singular statistical models, which is proved by resolution of singularities and empirical process theory on an algebraic variety.



Blackwell-Tapia Conference November 14-15, 2008

SCHEDULE

Friday, November 14, 2008

[Radisson Hotel RTP](#)

- | | |
|------------|---|
| 12:30-1:30 | Registration and Coffee/Refreshments |
| 1:30- 1:50 | Welcome and Introduction |
| 1:50-2:30 | Lecture: Jacqueline M. Hughes-Oliver , North Carolina State University
<i>Analysis of High-Dimensional Structure-Activity Screening Datasets Using the Optimal Bit String Tree</i> |
| 2:30-3:20 | Panel Discussion: <i>Getting Undergraduates Involved in Research</i>
Carlos Castillo-Chavez , Arizona State University (Chair)
Reinhard Laubenbacher , Virginia Tech
Juan Meza , Lawrence Berkeley National Lab
Peter Mucha , UNC – Chapel Hill
Michael Shearer , NC State University |
| 3:20-3:45 | Coffee Break |
| 3:45-4:30 | Short Talks I: Tim Thornton , University of California, San Francisco
<i>Statistical Methods for Genetic Association Studies in Structured Populations</i> |

Angela Gallegos, Tulane University
Crocodylia, Sex Determination and Delay Differential Equations

4:30-5:10 Lecture: **Freda Porter**, Porter Scientific
Technologies for Addressing Environmental Challenges

5:10-6:30 Poster Set-up

6:30-8:30 Poster Session and Reception
SAMSI will provide poster presentation boards and tape. The board dimensions are 4 ft. wide by 3 ft. high. They are tri-fold with each side being 1 ft. wide and the center 2 ft. wide. Please make sure your poster fits the board. The boards can accommodate up to 16 pages of paper measuring 8.5 inches by 11 inches.

Saturday, November 15, 2008

[Radisson Hotel RTP](#)

8:30-9:00 Registration and Continental Breakfast

9:00-9:30 Lecture: **Oscar Gonzalez**, University of Texas, Austin
Predicting Geometric Properties of DNA from Hydrodynamic Diffusion Data

9:30-10:15 Short talks II: **Rudy Horne**, Florida State University
Solitary Waves in Discrete Media in the Presence of Four-Wave Mixing Products

Yolanda Munoz Maldonado, Michigan Technological University
Testing the Equality of Mean Functions for Continuous Time Stochastic Processes

10:15-10:45 Coffee Break

10:45-11:15 Lecture: **Gabriel Huerta**, University of New Mexico
Statistical Approaches for Parameter Estimation in Climate Models

11:15-12:00 Short talks III: **Ulrica Wilson**, Morehouse College
A Criterion for Finding Cyclic $k_p((t))$ -division Algebras

Tanya Moore, Building Diversity in Science
Using Mathematics to Transform Communities

12:00-1:40 Lunch (Galleria Restaurant, First Floor)

1:40-2:00 Presentation: *Opportunities at SAMSI, the Math Institutes, and NSF*
Jim Berger, SAMSI
Cheri Shakiban, IMA
Peter March, NSF

2:00-3:00 Panel Discussion: *Career Opportunities in the Mathematical Sciences*

Carolyn Morgan, Hampton University (Chair)
Tanya Moore, Building Diversity in Science
Bob Rodriguez, SAS
Nell Sedransk, NISS
Janet Spoonamore, Army Research Office

- 3:00-3:20 Coffee Break
- 3:20-4:00 Lecture: **Richard Tapia**, Rice University
Optimization: The Cradle of Contemporary Mathematics
- 4:00-5:00 Blackwell-Tapia Lecture:
Juan Meza, Lawrence Berkeley National Laboratory
Optimization: The Difference Between Theory and Practice
- 5:00-6:15 Break
- 6:15-6:30 Conference Group Photos
- 6:30-9:00 Conference Reception and Banquet
6:30 Reception
7:00 Dinner is Served
7:45 Juan Meza Receives Award



**Blackwell-Tapia Conference
November 14-15, 2008**

SPEAKER TITLES/ABSTRACTS

Angela Gallegos
Tulane University
agalleg@tulane.edu

“Crocodilia, Sex Determination and Delay Differential Equations”

The crocodilia have multiple interesting characteristics that affect their population dynamics. They are among several reptile species which exhibit temperature-dependent sex determination (TSD) in which the temperature of egg incubation determines the sex of the hatchlings. Their life parameters, specifically birth and death rates, exhibit strong age-dependence. We develop delay-differential equation (DDE) models describing the evolution of a crocodilian population. In using the delay formulation, we are able to account for both the TSD and the age-dependence of the life parameters while

maintaining some analytical tractability. In our single-delay model we also find an equilibrium point and prove its local asymptotic stability. We numerically solve the different models and investigate the effects of multiple delays on the age structure of the population as well as the sex ratio of the population. For all models we obtain very strong agreement with the age structure of crocodilian population data as reported in Smith and Webb (Aust. Wild. Res. **12**, 541–554, 1985). We also obtain reasonable values for the sex ratio of the simulated population. This is joint work with Tenecia Plummer, David Uminsky, Cinthia Vega, Clare Wickman and Michael Zawoiski.

Oscar Gonzalez

University of Texas, Austin

og@math.utexas.edu

“Predicting Geometric Properties of DNA from Hydrodynamic Diffusion Data”

The sequence-dependent curvature and flexibility of DNA is critical for its packaging into the cell, recognition by other molecules, and conformational changes during biochemical processes. However, few methods are available for directly probing these properties at the basepair level. In this talk, a model for estimating sequence-dependent curvature and other geometric properties of DNA from hydrodynamic data on short sequences is described. The model is based on a generalized diffusion equation for DNA in dilute solution, with a coefficient matrix determined by the Stokes equations in the spatial domain around a single molecule. By comparing experimental measurements of this matrix with predictions based on direct numerical solution of the Stokes equations around sequence-dependent geometries, various structural features of DNA can be studied. In a preliminary application, we use the model to predict the hydrated radius of DNA under different assumptions on DNA curvature. Our results indicate that previous estimates of the radius, which were based on an assumption of zero curvature, are likely to be underestimates.

Rudy Horne

Florida State University

horne@math.fsu.edu

“Solitary Waves in Discrete Media in the Presence of Four-Wave Mixing Products”

In this talk, I will discuss solutions that arise in a vector discrete model of the Nonlinear Schrödinger equation where nonlinear inter-component coupling and four-wave mixing are taken into account. We show that the solutions to this model give rise to two single mode branch solutions as well as two mixed mode branch solutions. These solutions are obtained explicitly and their stability is analyzed in the so-called anti-continuum limit. Also, we connect this analysis to recent experiments that motivated this work.

Gabriel Huerta

University of New Mexico

ghuerta@stat.unm.edu

“Statistical Approaches for Parameter Estimation in Climate Models”

To quantify the uncertainties arising in climate prediction it is necessary to estimate a multidimensional probability distribution. This is known as the calibration problem. The computational cost of evaluating such a probability distribution for a climate model is impractical using traditional methods such as Gibbs/Metropolis algorithms. This talk will describe an optimization based method that has been applied for non-linear problems in geophysics and that is currently in use to calibrate parameters of an atmospheric general circulation model (AGCM).

Furthermore, we will also consider adaptive Monte Carlo based methods in the context of a climate model that is able to approximate the noise and response behavior of the AGCM. Comparisons and efficiency evaluations between approaches will be made. Another aspect of this talk is to overview the current role of spatial methods in providing emulators to climate model output and reducing computational burden. In particular we will discuss the use of Gaussian process (GP) in this context and on potential limitations and challenges for these methods.

Jacqueline Hughes-Oliver

North Carolina State University
hughesol@ncsu.edu

“Analysis of High-Dimensional Structure-Activity Screening Datasets Using the Optimal Bit String Tree”

A new classification method called the Optimal Bit String Tree (OBSTree) is proposed to identify quantitative structure-activity relationships (QSARs) in high-throughput screening studies. This recursive partitioning method introduces the concept of a chromosome to describe the simultaneous presence or absence of a combination of molecular features within a compound. Chromosomes are combined with a subset of descriptors (or predictor variables) to create a splitting variable, and these splitting variables form the search space for recursively splitting a compound collection in order to identify those compounds having both similar molecular structure and similar biological activity. Because of the resulting explosion in size of the search space, care is needed when exploring this space. We use a new stochastic searching scheme that consists of a weighted sampling scheme, simulated annealing, and a trimming procedure. Simulation studies and application to screening for monoamine oxidase (MAO) inhibitors show that OBSTree is advantageous in accurately and effectively identifying QSAR rules and finding different classes of active compounds.

Juan Meza

Lawrence Berkeley National Laboratory
JCMeza@lbl.gov

“Optimization: The Difference Between Theory and Practice”

There's an old saying, "In theory, there's no difference between theory and practice, but in practice there is". In this talk, I will discuss some of the challenges one faces when trying to solve optimization problems arising in real-world applications and what roles theory and practice play in developing new optimization algorithms. Today, scientists are working on problems such as designing nanostructures with specific properties, predicting the structure of proteins, finding new supernovae, and determining vulnerabilities in the electric power grid. In part, this is due to an increased ability to mathematically model new physical and engineering processes and the rapid rise of computational modeling and simulation. The resulting simulation-based optimization problems, however, have very different characteristics than classical problems and usually do not fit within the standard theoretical assumptions. In many cases, for example, there is noise associated with the evaluation of the objective function, usually through numerical errors in the solution of the equations. In other cases, no derivative information is available or the function may not be sufficiently smooth for standard methods. I will discuss several optimization techniques for the solution of these types of problems and some lessons learned in applying theory to practical problems.

Yolanda Munoz Maldonado

Michigan Technological University
ymunoz@mtu.edu

“Testing the Equality of Mean Functions for Continuous Time Stochastic Processes”

One of the most common activities in Statistics is the comparison of means for two or more groups. This task is usually carried out by the method called Analysis of Variance (ANOVA). When the analysis is done on functional data, the implementation of this technique becomes complicated due to the dimensionality of the problem. In this talk, we modify the test statistic of a permutation test used to compare the similarity between two sets of curves. The modified statistic is shown to be a U-statistic, and using its asymptotic distribution and following classical ANOVA reasoning, it allows for comparison of two or more groups of functions. A small Monte-Carlo simulation shows comparable power between the permutation test and our proposed approach when the number of groups analyzed is two. It also provides evidence that the U-statistic performs well for three sets of curves. We apply the U-statistic test to a ganglioside profile data set.

Tanya Moore

Building Diversity in Science
tanya.a.moore@gmail.com

“Using Mathematics to Transform Communities”

Can mathematics be used to empower a community? How does a biostatistician transfer math skills to work in the government and non-profit sectors? How is statistics really used in the field of public health? During this talk I will share highlights of my journey

from studying mathematics to working in a city health department and for a non-profit that is committed to supporting and encouraging emerging scientist and mathematicians.

Freda Porter

Porter Scientific

fporter@porterscientific.com

“Technologies for Addressing Environmental Challenges”

The protection of water resources is vital in today’s environment. A number of environmental issues are presented along with the latest technologies including 1) corrosion control coatings and processes, industrial water recovery, and monitoring solutions; 2) EPA Brownfields properties and remediation technologies; 3) Leaky landfills and groundwater monitoring; and 4) UST removal and remediation, where EPA guidelines’ function is to reduce leaking USTs that contaminate water supplies. Risk-based modeling of natural bioattenuation for groundwater contamination along with monitoring is suggested for measuring the extent of contamination. The mathematical underpinning of estimating the rate of natural bioattenuation is discussed.

Richard Tapia

Rice University

rat@rice.edu

“Optimization: The Cradle of Contemporary Mathematics”

In contrast to other disciplines in mathematics, problems in optimization are usually quite easy to state and to understand—even for those with limited mathematical sophistication. As such, important optimization problems embedded in some controversy have played a major role in motivating and promoting mathematical activity.

Writing circa 200 BC, the Greek mathematician Zenodorus considered the so-called isoperimetric problem: Determine, from all simple closed planar curves of the same perimeter, the one that encloses the greatest area.

In this talk the speaker will argue that the isoperimetric problem has been the most influential mathematics problem of all time. It played a major role in motivating the calculus of variations activity credited to the Bernoullis, Newton, Euler, and Lagrange in the late 1600’s and early 1700’s. In turn the early calculus of variations led to the golden era of mathematics that we recognize as the 18th and 19th centuries. Yet a complete proof of the isoperimetric problem eluded these early pioneers. Indeed, it was Weierstrass who first gave a complete proof more than a century later. In this talk the speaker will demonstrate that Euler and later Lagrange in the derivation of their , now well-known, Euler-Lagrange equation necessary condition were one direct observation away from deriving a sufficiency condition that would have given a straightforward resolution of the isoperimetric problem. Finally the derivation of the Euler-Lagrange equation presented by Euler and Lagrange is well known to be flawed. A correct derivation was

given by du Bois-Raymond some 150 years later. We argue quite surprisingly that the du Bois-Raymond's derivation can be viewed as presenting the Euler-Lagrange equation as a Lagrange multiplier rule. As such, it would be the world's first Lagrange multiplier rule and would precede the very notion of Lagrange multiplier rules.

Timothy Thornton

University of California, San Francisco
timothy_thornton@sbcglobal.net

“Statistical Methods for Genetic Association Studies in Structured Populations”

Genetic association testing has proven to be a valuable tool for the mapping of complex diseases. Technological advances have made it feasible to perform case-control association studies on a genome-wide basis. Some of the characteristics of the data include missing information, and the need to analyze hundreds of thousands or millions of genetic markers in a single study, which puts a premium on computational speed of the methods. The observations in these studies can have several sources of dependence, including population structure and relatedness among the sampled individuals, where some of this structure may be unknown. We describe a new approach to this problem.

Ulrica Wilson

Morehouse College
uwilson@morehouse.edu

“A Criterion for Finding Cyclic $k_p((t))$ -division Algebras”

What are all of the different types of division algebras? This question is far from being answered, but there is much that can be said. One strategy is to identify all the possible constructions of division algebras over a particular field. For example, thanks to Frobenius, we know that there are exactly two \mathbb{R} -division algebras, \mathbb{R} itself, and Hamilton's quaternions. This kind of classification is optimal because we have an explicit list of \mathbb{R} -division algebras (up to isomorphism). Classifying division algebras over other fields has proven to be much more difficult. Cyclic division algebras form a particularly nice class of division algebras. In this talk along with describing this special class of division algebras I will give a criterion for determining the cyclicity of division algebras over the Laurent series field $k_p((t))$.



Program on Algebraic Methods in Systems Biology and Statistics
Discrete Models in Systems Biology Workshop
December 3-5, 2008

SCHEDULE

Wednesday, December 3, 2008

- 8:15-8:45 Registration and Continental Breakfast
- 8:45-9:00 Welcome
- 9:00-10:00 **Lloyd Demetrius**, Harvard University
“Statistical Mechanics and Evolutionary Theory”
- 10:00-10:30 Break
- 10:30-11:15 **Abdul Jarrah**, Virginia Tech
“Polynomial Dynamical Systems as Discrete Models of Biological Networks”
- 11:15-11:30 Break
- 11:30-12:30 *Discussion: Goals and wishes (of the workshop)*
- 12:30-2:30 Lunch
- 2:30-3:15 **Anne Shiu**, UC Berkeley
“Siphons, Primary Decomposition, and the Global Attractor Conjecture”
- 3:15-3:30 Break
- 3:30-4:15 **David Anderson**, University of Wisconsin
“Persistence and Stationary Distributions of Biochemical Reaction Networks”
- 4:15-4:30 Break
- 4:30-5:00 *Poster Advertisements (2 minute ads each)*
- 5:00–7:00 *Poster Session and Reception*
SAMSII will provide poster presentation boards and tape. The board dimensions are 4 ft. wide by 3 ft. high. They are tri-fold with each side

being 1 ft. wide and the center 2 ft. wide. Please make sure your poster fits the board. The boards can accommodate up to 16 pages of paper measuring 8.5 inches by 11 inches.

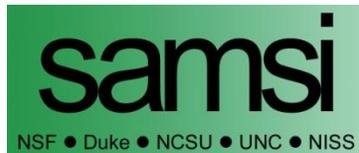
Thursday, December 4, 2008

- 8:30-9:00 Registration and Continental Breakfast
- 9:00-10:00 **Joshua Socolar**, Duke University
“Continuous Dynamics and Boolean Approximations in Complex Networks”
- 10:00-10:30 Break
- 10:30-11:15 **Juilee Thakar**, Pennsylvania State University
“Systems-Level Regulation of Pathogen-Immune System Interactions”
- 11:15-11:30 Break
- 11:30-12:30 *Discussion: Open problems*
- 12:30-2:30 Lunch
- 2:30-3:15 **Duygu Ucar**, Ohio State University
“Data mining Techniques for Functional Protein Clustering”
- 3:15-3:30 Break
- 3:30-4:15 **Ovidiu Lipan**, University of Richmond
“A Discrete Stochastic Model for Stress Response in CHO Mammalian Cells”
- 4:15-4:30 Break
- 4:30-5:15 **Jim Smith**, University of Warwick
“Discrete Modeling using Chain Event Graphs”
- 5:15-6:30 *Discussion: Second chances (to ask the “dumb” questions)*

Friday, December 5, 2008

- 8:00-8:30 Registration and Continental Breakfast
- 8:30-9:30 **Carla Piazza**, University of Udine
“Hybrid Automata and Systems Biology”
- 9:30-9:45 Break

- 9:45-10:30 **Henning Mortveit**, Virginia Tech
“Graph Dynamical Systems - A Mathematical Framework for Interaction-Based Systems, Their Analysis and Simulations”
- 10:30-10:45 Break
- 10:45-11:30 **Greg Rempala**, Medical College of Georgia
“Algebraic Methods for Inferring Biochemical Networks: a Maximum Likelihood Approach”
- 11:30-12:30 *Discussion: Building bridges and closing*
- 12:30-1:30 Lunch



**Discrete Models in Systems Biology
December 3-5, 2008**

SPEAKER TITLES/ABSTRACTS

David Anderson

University of Wisconsin
anderson@math.wisc.edu

“Persistence and Stationary Distributions of Biochemical Reaction Networks”

The dynamics of biochemical reaction systems can be modeled either deterministically or stochastically. Typically, the equations governing the dynamics of these models are quite complex. Further, there is oftentimes little knowledge about the exact values of the different system parameters, and, worse still, these system parameter values may vary from cell to cell. However, the network structure of a given system induces the corresponding equations (up to parameter values) governing its dynamics. I will show in this talk how this fact may be exploited to infer qualitative properties of large classes of biochemical systems and, most importantly, to learn which properties are independent of the details of the system parameters. I will give results for both stochastically and deterministically modeled systems. For deterministically modeled systems I will focus on persistence of trajectories, which in some important cases is sufficient to guarantee global asymptotic stability of equilibria. For stochastically modeled systems I will focus on the existence, and form, of stationary distributions.

Lloyd Demetrius

Harvard University

ldemetr@fas.harvard.edu

“Statistical Mechanics and Evolutionary Theory”

The statistical parameter evolutionary entropy, a measure of the uncertainty in age of the mother of a randomly chosen newborn, provides a framework for explaining the large diversity in life span, body size, and metabolic rate that describes natural populations. I will describe the analytical basis for this claim and discuss the relation between thermodynamic processes and evolutionary theory.

Abdul Jarrah

Virginia Tech
ajarrah@vbi.vt.edu

“Polynomial Dynamical Systems as Discrete Models of Biological Networks”

Mathematical models are an essential part of the new field of systems biology as they are the only way to formalize and analyze models that capture the dynamics and provide insights at the system level. Recently polynomial dynamical systems over finite fields have been introduced as a new framework for modeling and analyzing biological networks as multi-states finite dynamical systems, generalizing Boolean networks and logical models. Within this algebraic framework, using tools from computational algebra and algebraic geometry, the whole model space is presented and different algebraic methods are proposed for identifying a particular model from the model space.

Furthermore, methods for analyzing the dynamics of classes of polynomial systems have been developed. In this talk I will present methods for the development of polynomial dynamical systems models as well as methods for the analysis of their dynamics.

Ovidiu Lipan

University of Richmond
olipan@richmond.edu

“A Discrete Stochastic Model for Stress Response in CHO Mammalian Cells”

In many biological systems the interactions that describe the coupling between different units in a genetic network are nonlinear and stochastic. We study the interplay between stochasticity and nonlinearity using the responses of Chinese-hamster ovary (CHO) mammalian cells to different temperature shocks. The experimental data show that the mean value response of a cell population can be described by a mathematical expression which is valid for a large range of heat shocks conditions. A nonlinear model was developed to explain the mean value response. Moreover, the theoretical model predicts a specific biological probability distribution of responses for a cell population. The prediction was experimentally confirmed by measurements at single cell level. The computational approach can be used to study other nonlinear stochastic biological phenomena. The mathematical formalism is based on the discrete stochastic master equation built on a set of transition probabilities. The transition probabilities are directly connected with the biological phenomena. The mathematical formalism uses the factorial cumulants as dynamic variables.

Henning Mortveit

Virginia Tech
hmortvei@vbi.vt.edu

“Graph Dynamical Systems - A Mathematical Framework for Interaction-Based Systems, Their Analysis and Simulations”

This talk will be on Graph Dynamical Systems (GDS). These are dynamical systems constructed from (i) a graph where each vertex has a state, (ii) a sequence of vertex functions, and (iii) an update scheme. Here the update scheme specifies how the vertex functions are assembled to form the dynamical system map that governs the discrete time evolution. For example, applying the vertex functions in parallel corresponds to generalized cellular automata. If the vertex functions are applied according to a fixed vertex sequence we obtain the class of sequential dynamical systems.

The framework of graph dynamical systems is natural for representing distributed, interaction-based systems. Such systems are often referred to as complex systems, and examples range from socio-technical systems to biological systems. The GDS representation allows for accurate system descriptions that are amenable to mathematical analysis and that also maps well to implementations and hardware.

This talk will be an introduction to GDS with examples of theory and applications. The theory part will include graph based characterizations, comparisons and enumerations of phase space properties. The application examples will be taken from transportation and epidemiology - this part of the talk will focus on aspects of modeling and implementation.

Cara Piazza

University of Udine
carla.piazza@dimi.uniud.it

“Hybrid Automata and Systems Biology”

Most of the observable natural phenomena exhibit a mixed discrete-continuous behavior characterized by laws changing according to a phase cycle. Such behaviors can be modeled in a very natural way by a class of automata called hybrid automata. In this class, the evolution of measurable quantities, such as concentrations of reactants, is represented according to both dynamical system evolutions on dense domains and rules phases through a discrete transition structure.

The double nature, both discrete and continuous, of hybrid automata make them particularly suitable in the modeling of systems exhibiting a mixed behavior which cannot be characterized in a proper way using either discrete or continuous formalisms. For such reasons, since their introduction, hybrid automata have initiated a new tradition, promising powerful tools for modeling and reasoning about complex engineered or natural systems.

In this context, one of the basic problems is the reachability one which requires to decide whether it is possible to move from an automaton state to another.

Unfortunately, the flexibility and expressive power of hybrid automata soon lead to undecidability and complexity results which cast doubts on their suitability as a general tool that can be algorithmized and efficiently implemented.

In order to control both undecidability and complexity, one can either impose syntactic conditions and concentrate on classes of hybrid automata or define semantic approximation and discretization techniques.

After a brief introduction on hybrid automata, this talk will present such results and will show some application of hybrid techniques in systems biology.

Greg Rempala

Medical College of Georgia
grem Pala@mcg.edu

“Algebraic Methods for Inferring Biochemical Networks: a Maximum Likelihood Approach”

We present a novel method for identifying a biochemical reaction network based on multiple sets of estimated reaction rates in the corresponding equations arriving from various (possibly different) experiments. The current method, unlike some of the graphical approaches proposed in the literature, uses the values of the experimental measurements only relative to the geometry of the biochemical reactions under the assumption that the underlying reaction network is the same for all the experiments. The method is illustrated with a numerical example of a hypothetical network arising from a “mass transfer”-type model. Jointed work with Gheorghe Craciun and Casian Pantea.

Anne Shiu

University of California, Berkeley
annejls@gmail.com

“Siphons, Primary Decomposition, and the Global Attractor Conjecture”

In a biochemical reaction network, the concentrations of chemical species evolve in time, governed by the differential equations of mass-action kinetics. The nicest networks are the toric dynamical systems, which are those whose steady states are a special kind, called complex balancing steady states. Algebraically, the steady state loci and moduli spaces form toric varieties. One might ask whether we can characterize the limiting behavior of such systems. The assertion that a trajectory of such a system converges to a point on the toric variety (rather than a boundary point of the positive orthant) is the content of the Global Attractor Conjecture, which has been open for thirty years. The concept of a “siphon” (in the work of D. Angeli, P. De Leenheer, and E. Sontag), or equivalently a “semi-locking set” (in the work of D. Anderson), describes the possible zero-coordinates of boundary steady states; understanding their structure has been an important goal in pursuing the conjecture. An algebraic approach to this family of ideas will be presented; in particular, primary decomposition plays a prominent role.

No prior knowledge of chemical reaction network theory or toric geometry will be assumed.

Jim Smith

University of Warwick
j.q.smith@warwick.ac.uk

“Discrete Modeling using Chain Event Graphs”

Chain Event Graphs encode a new class of finite discrete models that strictly contains discrete Bayesian Network models and their context specific generalizations as a very special case. They provide a particularly powerful graphical framework for eliciting, querying, encoding, performing inferences and estimating highly asymmetric models in an efficient and transparent way. Such model classes arise naturally in both biological and social contexts. The class exhibits many of the advantages of Bayesian Networks. There are direct analogues of graphical conditional independence querying techniques. The framework supports conjugate inference with complete data and hence efficient exact search algorithms over the model class. Furthermore, like the Bayesian Network, the class encodes algebraic constraints on a class of polynomials and so it can be mapped into its own associated albeit typically inhomogeneous algebraic parametrization. Finally, being closely linked to an event tree Chain Event Graphs admit an excellent framework for expressing causal extensions of this model class. The talk will demonstrate these properties using a number of examples.

Joshua Socolar

Duke University
socolar@phy.duke.edu

“Continuous Dynamics and Boolean Approximations in Complex Networks”

Complex systems are often modeled as Boolean networks in attempts to capture their logical structure and reveal its dynamical consequences. Approximating the dynamics of continuous variables by discrete values and Boolean logic gates may, however, introduce dynamical possibilities that are not accessible to the original system. We study a class of systems motivated by modeling of transcriptional regulatory networks. In small networks, details of the switching characteristics and pulse propagation select stable attractors that are not captured by Boolean approximations. In large random networks, continuous systems often fail to exhibit the complex dynamics of corresponding Boolean models in the disordered (chaotic) regime, even when each element appears to be a good candidate for Boolean idealization.

Juilee Thaker

Pennsylvania State University
jthakar@phys.psu.edu

“Systems-Level Regulation of Pathogen-Immune System Interactions”

Pathogenic bacteria can modulate host immune responses to enable their establishment and persistence. We have examined a respiratory infection model system in which the immune response is generally successful in clearing the pathogen. We study the interactions between host’s immune components and pathogen’s virulence factors by synthesizing a network based on existing experimental information and integrating it in a Boolean and piecewise linear model. Our Boolean model offers predictions regarding cytokine regulation, key immune components and clearance of primary and secondary infections; we experimentally validate two of these predictions. The piecewise linear model extends our Boolean model by making predictions about the timescales of each process, the activity thresholds of each component and about novel regulatory interactions. Some of these predictions are supported by the literature, and many can serve as targets of future experiments.

Duygu Ucar

Ohio State University

ucar@cse.ohio-state.edu

“Data mining Techniques for Functional Protein Clustering”

Complex relations among biological entities can be efficiently represented in the form of interaction networks. However, this representation, by itself, does not reveal the useful information about the underlying system. Computational methods need to be studied to be able to extract this information from noisy and scale-free biological interactions networks. We studied Data Mining techniques to deduce functional protein clusters from a Protein-Protein Interactions (PPI) network of *Saccharomyces cerevisiae*. Major problems we attacked in this study are knowledge discovery from noisy and scale-free interactions networks by taking into consideration the necessity for multiple cluster membership.

Program on Algebraic Methods in Systems Biology and Statistics
Algebraic Statistical Models Workshop
January 15-17, 2009

SCHEDULE

Thursday, January 15, 2009

- | | |
|-------------|--|
| 8:15-8:45 | Registration and Continental Breakfast |
| 8:45-9:00 | Welcome |
| 9:00-10:00 | Steffen Lauritzen , University of Oxford
<i>“Combining Statistical Models - Towards Structural Meta-Analysis”</i> |
| 10:00-10:30 | Break |
| 10:30-11:30 | Jin Tian , Iowa State University
<i>“Causal Inference and Algebraic Methods”</i> |
| 11:30-12:30 | Discussion |
| 12:30-2:00 | Lunch |
| 2:00-3:00 | Elizabeth Allman , University of Alaska, Fairbanks
<i>“Applications of Kruskal's Theorem to the Identifiability of Algebraic Statistical Models”</i> |
| 3:00-3:30 | Break and Poster Set-up |
| 3:30-4:30 | Sonja Petrovic , University of Illinois
<i>“Markov Bases of p_1 Random Graph Models”</i> |

- 4:30-5:00 *Poster Advertisements* (2 minute ads each)
- 5:00-7:00 *Poster Session and Reception*
SAMSI will provide poster presentation boards and tape. The board dimensions are 4 ft. wide by 3 ft. high. They are tri-fold with each side being 1 ft. wide and the center 2 ft. wide. Please make sure your poster fits the board. The boards can accommodate up to 16 pages of paper measuring 8.5 inches by 11 inches.

Friday, January 16, 2009

- 8:30-9:00 Registration and Continental Breakfast
- 9:00-10:00 **Thomas Richardson**, University of Washington
"Analysis of the Binary Instrumental Variable Model"
- 10:00-10:30 Break
- 10:30-11:30 **Donald Richards**, Pennsylvania State University
"Finite-Sample Inference with Incomplete Multivariate Normal Data"
- 11:30-12:30 Open Problem Discussion
- 12:30-2:00 Lunch
- 2:00-3:00 **Ruriko Yoshida**, University of Kentucky
"Geometry of Cophylogeny and its Applications to Genome Evolution"
- 3:00-3:30 Break
- 3:30-4:00 Discussion
- 4:00-5:00 **Akimichi Takemura**, University of Tokyo
"Minimality Properties of Markov Bases and Normality of Semigroups"

Saturday, January 17, 2009

- 8:00-8:30 Registration and Continental Breakfast
- 8:30-9:30 **Hugo Maruri-Aguilar**, London School of Economics
"Design Fan, Term Orders and Zonotopes"
- 9:30-10:30 **Jason Morton**, Stanford University
"Algebraic Models for Multilinear Dependence"
- 10:30-11:00 Break

- 11:00-12:00 **Serkan Hosten**, San Francisco State University
“Algebra, Geometry, and Algorithms for Maximum Likelihood Estimation”
- 12:00-12:30 Discussion and Closing
- 12:30-1:30 Lunch
- 1:30 Adjournment

**Algebraic Statistical Models
January 15-17, 2009**

SPEAKER TITLES/ABSTRACTS

Elizabeth Allman

University of Alaska, Fairbanks
e.allman@uaf.edu

“Applications of Kruskal's Theorem to the Identifiability of Algebraic Statistical Models”

A statistical model with n observed discrete random variables and one hidden discrete random variable is a simple example of a 'conditional independence model' when the observations are independent given a fixed state for the hidden variable.

In the 1970s, J. Kruskal proved that such models with 3 observed variables are identifiable provided the state space for the observed variables is large enough, and the parameters are sufficiently generic. In this talk, we review Kruskal's result and show that it can be applied to prove the identifiability of a diverse collection of models with more observed variables and more complex hidden structure, including phylogenetic models, random graph models, and hidden Markov models.

Serkan Hosten

San Francisco State University
serkan@math.sfsu.edu

“Algebra, Geometry, and Algorithms for Maximum Likelihood Estimation”

The talk will be a review of the role of algebraic geometry has played in ML estimation and an invitation to open problems in this direction.

Steffan Lauritzen

University of Oxford
steffen@stats.ox.ac.uk

“Combining Statistical Models - Towards Structural Meta-Analysis”

Graphical models have proved their value for the modelling and analysis of complex stochastic systems, not least because of their fundamental use of conditional independence to establish modularity and enable local specification and computation.

This lecture is concerned with formalizing a calculus for combination of structural information in the form of statistical models for separate systems which in part are concerned with the behaviour of identical quantities. The work follows up on the notion of a meta-Markov model as discussed by Dawid and Lauritzen (1993) [Annals of Statistics].

The lecture represents joint work with Sofia Massa, University of Padova.

Hugo Maruri-Aguilar

London School of Economics
H.Maruri-Aguilar@lse.ac.uk

“Design Fan, Term Orders and Zonotopes”

Computing the algebraic fan of an experiment is closely related to the computation of the Universal Grobner basis for the design ideal. The crucial object required for doing this is a collection of partially ordering vectors. I intend to present general results on determining this set of vectors. I use previously known results from geometry together with special polytopes called zonotopes.

Jason Morton

Stanford University
jason@math.stanford.edu

“Algebraic Models for Multilinear Dependence”

We discuss a new statistical technique inspired by research in tensor geometry and making use of cumulants, the higher order tensor analogs of the covariance matrix. For non-Gaussian data not derived from independent factors, tensor decomposition techniques for factor analysis such as Principal Component Analysis and Independent Component Analysis are inadequate. Seeking a small, closed space of models which is computable and captures higher-order dependence leads to a proposed extension of PCA and ICA, Principal Cumulant Component Analysis (PCCA). Estimation is performed by maximization over a Grassmannian. Joint work with L.-H. Lim.

Sonja Petrovic

University of Illinois, Chicago
petrovic@math.uic.edu

“Markov Bases of p_1 Random Graph Models”

The p1 model describes dyadic interactions in a social network, which is summarized in the form of a directed graph. The model is log-linear in form, and it allows for effects due to differential attraction (popularity) and expansiveness, as well as an additional effect due to reciprocation. Fienberg has given an introductory talk to these models in the Opening Workshop. Since then, we have been able to understand better the Markov bases for these models. However, their complex structure can not always be described explicitly (for a whole family of models). This talk will explain some problems that remain unsolved.

This talk builds on joint work with Stephen Fienberg and Alessandro Rinaldo.

Donald Richards

Pennsylvania State University
richards@stat.psu.edu

“Finite-Sample Inference with Incomplete Multivariate Normal Data”

We review results obtained recently for a class of problems in finite-sample inference with two-step, monotone incomplete, data from a multivariate normal population. We present a stochastic representation for the exact distribution of the maximum likelihood estimator of the population mean vector; ellipsoidal confidence regions for the mean through a generalization of Hotelling’s statistic; and Stein-rule shrinkage estimators for the mean. We also discuss the algebraic difficulties inherent in extensions of these results to three-step monotone, or to non-monotone, incomplete multivariate normal data.

Thomas Richardson

University of Washington
thomasr@u.washington.edu

“Analysis of the Binary Instrumental Variable Model”

The instrumental variable model comprises a randomly assigned treatment (Z), an exposure variable (X) and a response variable (Y). It is well known that when all three of these variables are binary, the potential outcomes model is not identified by the joint distribution $p(x,y,z)$. Consequently many statistical analyses impose additional assumptions, or change the causal estimand of interest in order to achieve identification.

Here we take a different approach, directly characterizing and displaying the set of distributions compatible with the observed data. This provides insights into the variation dependence between average causal effects for various compliance groups, that are partially identified. The analysis also leads directly to a re-parameterization that may be used for Bayesian inference and the development of models that incorporate baseline covariates.

(Joint work with James Robins, Harvard)

Akimichi Takemura

University of Tokyo
takemura@stat.t.u-tokyo.ac.jp

“Minimality Properties of Markov Bases and Normality of Semigroups”

We discuss various notions of minimality of Markov bases, such as indispensable moves, indispensable monomials and distance reduction by a Markov basis. These notions seem to be related to the normality of the semigroups associated with a configuration defining a toric ideal, although there are only a few results relating the notion on the minimality of Markov basis to the normality of the semigroup.

Jin Tian

Iowa State University
jtian@cs.iastate.edu

“Causal Inference and Algebraic Methods”

In this talk I will provide an introduction to causal inference problems in causal Bayesian Networks (CBNs) that may potentially be addressed by algebraic methods. I will discuss recent work in identifying causal effects in CBNs with hidden variables. I will discuss recent developments and open problems in identifying constraints on the probability distributions induced by CBNs.

Ruriko Yoshida

University of Kentucky
ruriko.yoshida@uky.edu

“Geometry of Cophylogeny and its Applications to Genome Evolution”

The diversity of species is related to the separation of gene pools over evolutionary time. In this process two or more lineages often stay closely associated with one another: genes with species and hosts with symbionts (parasites or mutualists). The concept of codivergence, the divergence of one lineage (species or gene) as a result of the divergence of another, has fascinated researchers for a long time. However, researchers assume that the host tree and the parasite tree (or gene trees) are reconstructed independently or assume that the true trees are given. In practice, since phylogenetic trees are reconstructed independently, this means they assume implicitly that the host tree and the parasite tree have developed independently, i.e., that the hosts and the parasites do not exhibit codivergence. The starting point of our approach is to relax this assumption and to study the joint probabilities for the host-parasite trees or the gene trees without assuming their independent development. In this paper we focus on its underlying algebraic and polyhedral geometric structures. Specifically, we define a notion of the spaces of cophylogenetic trees as well as some preliminary results using kernels defined on the cross product of the space of dissimilarity maps to analyzing codivergence on plants-endophytes phylogenetic trees and also on gene trees. We end this talk with several open

problems related to gene codivergence and coevolutions in terms of polyhedral geometry and algebra.

Program on Sequential Monte Carlo Methods
Mid-Program Workshop
February 19-20, 2009

SCHEDULE

Thursday, February 19, 2009

- 8:15-8:45 Registration and Continental Breakfast
- 8:45-9:00 Welcome
- 9:00-9:30 **Hedibert Lopes**, University of Chicago
Particle Learning: a semester later
- 9:30-10:00 **Paul Fearnhead**, Lancaster University
A Particle Smoother with Linear Computational Cost
- 10:00-10:15 Break
- 10:15-10:45 **Francois Septier**, Signal Processing Laboratory, Cambridge University
Multi-target Tracking using MCMC-Based Particle Algorithm
- 10:45-11:15 **Nathan Green**, Defence Science and Technology Laboratories (Webex)
- 11:15-11:45 **Mark Briers**, QinetiQ (Webex)
An Application of ABC Using SMC to Multiple Source Term Estimation
- 11:45-12:15 **Daniel Clark**, Heriot Watt University (Webex)
Joint Target-Detection and Tracking Smoothers
- 12:15-2:00 Lunch
- 2:00-2:30 **Chunlin Ji**, Duke University
Dynamic Spatial Mixture Modelling and its Application in Cell Tracking
- 2:30-3:00 **Viktor Rozjic**, University of Southern California
Performance of the Resample-move Algorithm on the Simulated Multi-target Tracking Dataset
- 3:00-3:30 **Gentry White**, North Carolina State University
A Kalman Filter Based Emulator for Source Term Estimation

- 3:30-4:00 **Ernest Fokoue**, Kettering University
Variational Mean Field Approach to Efficient Multitarget Tracking
- 4:00-4:30 **David Dunson**, Duke University
Sourish Das, Duke University
Bayesian Distribution Regression via Augmented Particle Filtering
- 4:30-4:45 Break and Poster Set-up
- 4:45-5:00 *Poster Advertisements* (2 minute ads each)
- 5:00-7:00 *Poster Session and Reception*
SAMSI will provide poster presentation boards and tape. The board dimensions are 4 ft. wide by 3 ft. high. They are tri-fold with each side being 1 ft. wide and the center 2 ft. wide. Please make sure your poster fits the board. The boards can accommodate up to 16 pages of paper measuring 8.5 inches by 11 inches.

Friday, February 20, 2009

- 8:30-9:00 Registration and Continental Breakfast
- 9:00-9:30 **Petar Djuric**, Stony Brook University
Tracking Multiple Targets with Multiple Particle Filters
- 9.30-10:00 **Namrata Vaswani**, Iowa State University
PF-EIS and PF-MT: Particle Filtering (PF) with Efficient Importance Sampling (EIS) and with Mode Tracking (MT) and Applications in Deformable Contour Tracking
- 10:00-10:15 Break
- 10:15-10:45 **Carlos Carvalho**, University of Chicago
Model Assessment and Adaptive Design
- 10:45-11:15 **Matt Taddy**, University of Chicago
Particle Learning for General Mixtures
- 11:15-11:45 **Ioanna Manolopoulou**, SAMSI
Targeted Sequential Resampling from Large Data Sets in Mixture Modeling
- 11:45-12:15 **Mark Coates**, McGill University
Weak L_p Bounds on the Performance of the Leader Node Particle Filter
- 12:15-2:00 Lunch
- 2:00-5:00 Discussion Session for Working Groups

**Sequential Monte Carlo Methods
Mid-Program Workshop
February 19-20, 2009**

SPEAKER TITLES/ABSTRACTS

Mark Briers

QinetiQ

mark.briers@gmail.com

“An Application of ABC Using SMC to Multiple Source Term Estimation”

In this talk we will discuss and demonstrate recent advances within Approximate Bayesian Computation using SMC-based approximations in the context of estimating the number and location of multiple simultaneous CBRN releases.

Authors: Mark Briers and Keith Copsey

Carlos Carvalho

University of Chicago

carlos.carvalho@chicagobooth.edu

“Model Assessment and Adaptive Design”

Daniel Clark

Heriot Watt University

D.E.Clark@hw.ac.uk

“Joint Target-Detection and Tracking Smoothers”

A multi-object Bayes filter analogous to the single-object Bayes filter can be derived using Finite Set Statistics for the estimation of an unknown and randomly varying number of target states from random sets of observations. The joint target-detection and tracking (JoTT) filter is a truncated version of the multi-object Bayes filter for the single target detection and tracking problem. Despite the success of Finite-Set Statistics for multi-object Bayesian filtering, the problem of multi-object smoothing with Finite Set Statistics has yet to be addressed. I propose multi-object Bayes versions of the forward-backward and two-filter smoothers and derive optimal nonlinear forward-backward and two-filter smoothers for jointly detecting, estimating and tracking a single target in cluttered environments. I also derive optimal Probability Hypothesis Density (PHD) smoothers, restricted to a maximum of one target and show that these are equivalent to their Bayes filter counterparts.

Mark Coates

McGill University

markjcoates@gmail.com

“Weak Lp Bounds on the Performance of the Leader Node Particle Filter”

The leader node particle filter finds application in sensor networks that strive to track a moving target. Each node in the sensor network is capable of sensing and computation, and the leader node is the node responsible for tracking the target using a particle filter. In order to keep communication local, the leader node is changed periodically to keep it close to the target. During this changeover, the particle representation must be exchanged, and this generally involves additional approximation, either through a reduced number of particles or parametric approximation. In this paper, I will present some error bounds for the leader node particle filter, which indicate how the approximation step impacts performance.

Petar Djuric

Stony Brook University
djuric@ece.sunysb.edu

“Tracking Multiple Targets with Multiple Particle Filters”

In this presentation, we build on our previous work for tracking multiple targets with multiple particle filters, where each particle filter tracks its own target. We avoid the collapse of traditional particle filtering by considering an interconnected network of such particle filters where each of them works on a relatively low dimensional space. We assume that our interest is in finding the marginal posterior distributions of the state vectors describing the different targets and not in the joint posterior of all the targets. We test the method on the problem of multiple target tracking based on sensor data which represent a superposition of contributions of all the targets in the field. The computer simulations demonstrate the performance of the newly proposed method and compare it with other implementations of particle filtering.

David Dunson

Duke University
dunson@stat.duke.edu

Sourish Das

Duke University
sourish.das@stat.duke.edu

“Bayesian Distribution Regression via Augmented Particle Filtering”

To limit assumptions in modeling of conditional response distributions, hierarchical mixtures-of-experts models allow the mixing weights in a regression model to vary flexibly with predictors. Nonparametric Bayes methods can be used to incorporate infinitely-many components, allowing effective model dimension to increase with sample size. However, MCMC algorithms for posterior computation often encounter mixing problems due to multimodality of the posterior. Focusing on a broad class of probit stick-breaking process priors for conditional response distributions indexed by time, space or predictors, we propose an efficient augmented particle filter for posterior computation and approximation of marginal likelihoods. The algorithm sequentially updates random length latent normal vectors within each particle as subjects are added, avoiding

truncation of the infinite collection of random measures. Through marginalization after data augmentation, the approach bypasses the need to update parameters, dramatically improving efficiency while avoiding degeneracies. The method can be applied broadly for continuous, count or categorical response variables. The methods are illustrated using simulated examples and an epidemiologic application.

Paul Fearnhead

Lancaster University
p.fearnhead@lancaster.ac.uk

We consider methods for smoothing: estimating past values of a state given observations to date. We describe methods based on sequential Monte Carlo, and develop a novel approach that is computationally more efficient than common existing approaches: the new method has a computational cost that is linear in the number of particles, rather than a quadratic cost.

This method is motivated and applied to athletics data.
Paul Fearnhead, David Wyncoll and Jon Tawn

Ernst Fokoue

Kettering University
efokoue@kettering.edu

“Variational Mean Field Approach to Efficient Multitarget Tracking”

We present various aspects of a variational mean field alternative to MCMC-based particle algorithms for multitarget tracking. Our proposed method is motivated by both clarity and efficiency with an emphasis on the derivation of updating schemes that are fast. In the spirit of traditional variational mean field inference, our intractable posteriors of interest are approximated by more tractable counterparts with the immediate advantage being the rapid generation of the desired proposals. Our work is compared to Francois Septier’s MCMC-based particle algorithm from which most of the building blocks of our methods are borrowed.

Chunlin Ji

Duke University
chunlin.ji@duke.edu

“Dynamic Spatial Mixture Modelling and its Application in Cell Tracking”

We discuss dynamic spatial mixture modelling for inhomogeneous point processes. A time varying spatial Dirichlet process Gaussian mixture model is proposed to characterize the underlying dynamic of intensity of the spatial inhomogeneous point process. Consequently, the components in the mixture model are able to represent the positions of targets. A Poisson measurement model is presented for the spatial point process observations, where we assume that a single target may generate a set of spatial point

observations. Furthermore, a consequence of the Poisson model is that the measurement likelihood may be evaluated without explicit data association. Bayesian inference for the intensity of a dynamic spatial inhomogeneous point process is presented in detail and we also provide the particle filter implementation of the proposed Bayesian filtering framework. Illustrative simulation examples of extended target tracking and cell fluorescent microscopic imaging tracking will be presented.

Hedibert Lopes

University of Chicago

Hedibert.Lopes@chicagobooth.edu

“Particle Learning: a semester later”

The main developments and ideas generated during the Particle Learning (PL) working group Thursday meetings will be summarized. After briefly reviewing PL itself, I describe the current research agendas of the various PL subgroups:

- a) Particle learning in autoregressive models with structured priors (Prado and Lopes);
- b) Expanding the particle learning framework to models without conditional sufficient statistic structure (Niemi, Mukherjee, Carvalho and Lopes);
- c) Sequential Monte Carlo methods for long memory stochastic volatility models (Macaro and Lopes);
- d) Particle learning DSGE models (Petralia, Chen, Carvalho and Lopes);
- e) The role of options, stochastic volatility and jumps in the interest rate risk premia dynamics (Lund and Lopes).

Ioanna Manolopoulou

SAMSI

ioannamanolopoulou@gmail.com

“Targeted Sequential Resampling from Large Data Sets in Mixture Modeling”

One of the challenges of Markov Chain Monte Carlo in large datasets is the need to scan through the whole data at each iteration of the sampler, which can be computationally prohibitive. Several approaches have been developed to address this, typically drawing computationally manageable samples of the data. Here we consider the specific case when most of the data provides no information about the parameters of interest. The motivating application arises in flow cytometry, where interest lies in identifying specific rare cell subtypes and characterizing them according to their corresponding markers. We present an MCMC approach where an initial sample of the full data is used to draw a further set of datapoints which contains more information about rare events, and extend it to a Sequential Monte Carlo framework whereby the selected sample is augmented sequentially as estimates improve.

Viktor Rozjic

University of Southern California

rozgic@usc.edu

“Performance of the Resample-move Algorithm on the Simulated Multi-target Tracking Dataset”

In this talk I am going to present implementation details for the application of the resample-move algorithm on the simulated multi-target tracking dataset.

This work is a part of the bigger initiative within the tracking workgroup, led by Francois Septier and Prof. Simon Godsill, where the goal was to compare performance of various multi-target tracking algorithms on the simulated dataset.

Francois Septier

University of Cambridge
fjms2@cam.ac.uk

“Multi-target Tracking using MCMC-Based Particle Algorithm”

Detection and tracking of multiple targets are essential components of modern sensor systems. The purpose of multiple target tracking algorithms is to determine the number of targets and their respective kinematic parameters from sequences of noisy observations. The difficulty of this problem has increased as sensor systems in the modern battlefield are required to detect and track targets in very low probability of detection and in environments with heavy clutter.

With the parallel advances in modern computational power and the developments in optimal non-linear techniques such as particle methods, it is now possible to solve complex state space models efficiently, potentially achieving significant performance gains.

In this talk, we will address the problem of detection and tracking of independent targets. We will present the MCMC-Based Particle algorithm used to perform the sequential inference. Some results will also be provided to illustrate the ability of this algorithm to detect and track multiple targets in hostile environments with high noise and low detection probabilities.

Matthew Taddy

University of Chicago
matt.taddy@chicagobooth.edu

“Particle Learning for General Mixtures”

We consider the use of efficient particle filtering methods in the estimation of general mixture models. More specifically, we develop a set of filtering recursions for the analysis of finite mixture models with known number of components (MM) as well as Dirichlet process (DP) mixture models. Our approach exactly samples from a particle

approximation to the joint distribution of parameters and hidden states (or mixture indicators) avoiding the usual problems associated with sequential importance sampling and providing a Monte Carlo alternative to "hard to converge" MCMC methods. Central to our strategy is the use of conditional sufficient statistics for learning about parameters (more here). We illustrate the proposed methodology first via a finite mixture of Poisson followed by multivariate density estimation problems.

Namrata Vaswani

Iowa State University
namrata@iastate.edu

“PF-EIS and PF-MT: Particle Filtering (PF) with Efficient Importance Sampling (EIS) and with Mode Tracking (MT) and Applications in Deformable Contour Tracking”

Our key contribution in the design of PF-EIS and of PF-MT (and of PF-EIS-MT) is in the importance sampling step. We will use static importance sampling to first explain the two ideas. The extension to sequential importance sampling or particle filtering (sequential importance sampling + resampling) is simple. The aim of both EIS and MT is to achieve accurate tracking with a smaller number of particles by improving the effective particle size. EIS does this by trying to find the maximum number of dimensions on which a Gaussian approximation to the optimal importance density can be used. MT addresses large dimensional problems and replaces importance sampling on the "compressible" part of the state space by conditional posterior mode tracking. We have successfully used PF-MT for deformable boundary contour tracking from image sequences. The main ideas will be discussed and our results shown.

Gentry White

North Carolina State University and SAMSI
gawhite@ncsu.edu

“A Kalman Filter Based Emulator for Source Term Estimation”

Deterministic models for dynamic systems can be computationally expensive. In order to gain information on how a dynamic system behaves over a range of inputs can require a large number of simulator runs, which can be prohibitive. One solution is to construct a statistical emulator using a reduced number of simulator runs. The resulting emulator is based on a computationally simpler model and allows for the estimation of simulator output at new input values along with a measure of uncertainty. In the case of the source term estimation problem we can use an emulator based on the Kalman Filter/Sootheer as a solution to a dynamic linear model. The construction begins with a linear approximation to the non-linear deterministic model plus a Gaussian process prior on the input space. The resulting emulator can be easily evaluated using existing software and techniques.

SAMSI UNDERGRAD WORKSHOP SCHEDULE
Algebraic Methods in Systems Biology and Statistics
February 27 – 28, 2009

Friday

- 8:15 Shuttle from Radisson to SAMSI (Group #1)
8:35 Shuttle from Radisson to SAMSI (Group #2)
8:55 Shuttle from Radisson to SAMSI (Group #3)
- 9:15-9:30 Welcome and Introductions
- 9:30-10:20 **Brandy Stigler**, Southern Methodist University
Introduction to System Biology
- 10:20-10:50 **Gentry White**, North Carolina State University and SAMSI
Introduction to R
- 10:50-11:10 Break
- 11:10-12:00 **Seth Sullivant**, North Carolina State University
Introduction to Algebraic Statistics
- 12:00-1:00 Lunch
- 1:00-1:40 **Luis Garci-Puente**, Sam Houston State University and SAMSI
Algebraic Statistical Models
- 1:40-2:20 **Ian Dinwoodie**, Duke University and SAMSI
Comparing Binary Dynamics
- 2:20-3:00 **Giovanni Pistone**, Politecnico di Torino and SAMSI
Statistical Design of Experiments and Algebra
- 3:00-3:20 Break
- 3:20-3:30 Summary of Experience
- 3:30-4:30 **Ian Dinwoodie**, Duke University
Giovanni Pistone, Politecnico di Torino and SAMSI
Ben Wells, North Carolina State University
Saied Yasamin, SAMSI
Interactive Session
- 4:30-5:00 **Pierre Gremaud** North Carolina State University and SAMSI
Discussion of Graduate Schools and Career Options
- 5:00 Shuttle to Radisson (Group #1)
5:20 Shuttle to Radisson (Group #2)

5:40 Shuttle to Radisson (Group #3)

6:00 Dinner at the Radisson Hotel

Saturday

8:20 Shuttle from Radisson to SAMSI (Group #1)

8:40 Shuttle from Radisson to SAMSI (Group #2)

9:00 – 9:50 **Jeffrey Thorne**, North Carolina State University
Evolutionary Biology and Phylogenetics

9:50 – 10:40 **Megan Owen**, SAMSI
Tree Metrics / Dissimilarity Measures / Tree Space

10:40 – 11:00 BREAK

11:00 - 12:00 **Megan Owen**, SAMSI
Jason Yellick, North Carolina State University
Interactive Session on Phylogenetic Trees

Noon Adjournment and Departure

Program on Algebraic Methods in Systems Biology and Statistics Molecular Evolution and Phylogenetics Workshop April 2-3, 2009

SCHEDULE

Thursday, April 2, 2009

8:15-8:45 Registration and Continental Breakfast

8:45-9:00 Welcome

9:00-10:00 **Jeffrey Thorne**, North Carolina State University
Making Inferences About the Impact of Phenotype on Genotype from the Ancestral Lineage

10:00-10:15 Break

10:15-11:00 **Cecile Ane**, University of Wisconsin, Madison
Identifiability of Trait Evolution Models

11:00-11:45 **Laura Kubatko**, Ohio State University
Distributions Arising on Gene Trees Under the Coalescent Model

- 11:45-12:15 Discussion with Speakers
- 12:15-2:15 Lunch
- 2:15-3:15 **Junhyong Kim**, University of Pennsylvania
Known Unknowns and Unknown Unknowns in Phylogeny Reconstruction
- 3:15-4:00 **Eric Stone**, North Carolina State University
Something Old, Something New: A phylogenetic application of the combinatorial graph Laplacian
- 4:00-4:15 Break and Poster Set-up
- 4:15-4:45 Discussion with Speakers
- 4:45-5:00 Poster Advertisements (2 minute ads each)
- 5:00–7:00 Poster Session and Reception
SAMSI will provide poster presentation boards and tape. The board dimensions are 4 ft. wide by 3 ft. high. They are tri-fold with each side being 1 ft. wide and the center 2 ft. wide. Please make sure your poster fits the board. The boards can accommodate up to 16 pages of paper measuring 8.5 inches by 11 inches.

Friday, April 3, 2009

- 8:30-9:00 Registration and Continental Breakfast
- 9:00-10:00 **Seth Sullivant**, North Carolina State University
The Geometry of Phylogenetic Mixtures
- 10:00-10:15 Break
- 10:15-11:00 **Jeremy Sumner**, University of Tasmania
Markov Invariants in Phylogenetics: the Quartet Case Done to Death
- 11:00-11:45 **Sonja Petrovic**, University of Illinois, Chicago
Group-based Models in Phylogenetics and Related Problems
- 11:45-12:15 Discussion with Speakers
- 12:15-2:15 Lunch
- 2:15-3:15 **Tandy Warnow**, University of Texas, Austin
SATe: A New Method for Simultaneous Estimation of Alignments and Trees
- 3:15-4:00 **Jesus Fernandez-Sanchez**, UPC Barcelona Tech
Phylogenetic Invariants of Equivariant Evolutionary Models
- 4:00-4:45 **Fumei Lam**, University of California, Davis
Generalizing the Four Gamete Test

- 4:45-5:15 Discussion with Speakers
5:15-6:30 Open Problems and Wrap-up

Molecular Evolution and Phylogenetics
April 2-3, 2009

SPEAKER TITLES/ABSTRACTS

Cecile Ane

University of Wisconsin, Madison
ane@stat.wisc.edu

“Identifiability of Trait Evolution Models”

Most often, biologists build phylogenetic trees in order to use them and analyze evolution of traits. Many so-called 'comparative methods' have been proposed for the analysis of trait evolution on trees. These methods provides ways to accommodate the dependence that arises from shared ancestry and that might obscure correlation among traits. I will draw analogies with models in spatial statistics: where dependence arises from the spatial structure of sampling units. Comparative methods have mostly been assumed to share the same properties as standard statistical approaches, that rely on independent random samples. I will show that some basic properties do not necessarily hold, such as consistency of estimates, and the BIC approximation.

Jesus Fernandez-Sanchez

Universitat Politecnica de Catalunya
jesus.fernandez.sanchez@ucp.edu

“Phylogenetic Invariants of Equivariant Evolutionary Models”

Since a number of statistical evolutionary models can be viewed as algebraic varieties, tools and results coming from algebraic geometry can be applied to the study of problems related with phylogenetic reconstruction. Indeed, the generators of the ideals associated to these varieties should allow to determine the topology of the phylogenetic tree associated to a given set of taxa. Due to the recent work of Draisma and Kuttler, most of the wide used evolutionary models can be described by the action of a finite group on an algebraic variety (/equivariant models/). We will review this approach on the study of evolutionary models and we will discuss how a deep understanding of their geometry may improve some phylogenetic reconstruction methods. We will also make use of some facts of representation theory to show that for reconstruction purposes, it is enough to take into account invariants coming from the edges of phylogenetic trees.

Junhyong Kim

University of Pennsylvania
junhyong@sas.upenn.edu

“Known Unknowns and Unknown Unknowns in Phylogeny Reconstruction”

Phylogenies are tree graphs of genealogical relationship between organisms or bio-molecules. Over the past 50 years, various algorithms and statistical methods have been devised to estimate such tree graphs from biological data. More recently it has been standard to assume a Markov Model of evolution over the edges of the tree graph. Here I show that such models generate a geometric family of probability distributions for the joint appearance of character states on the leaves of the tree. Different estimation methods can also be analyzed in this geometric context and explicitly geometric estimators can be derived using algebraic invariants. More importantly, different properties of the methods can be studied using geometric reasoning. Using this background I show some recent results for statistical consistency of certain phylogenetic estimators and for identifiability of certain models under relaxed conditions for model mixing.

Laura Kubatko

Ohio State University
lkubatko@stat.osu.edu

“Distributions Arising on Gene Trees Under the Coalescent Model”

In the field of phylogenetics, primary interest is generally on estimation of the species phylogeny, the tree that represents the actual sequence of speciation events that have led to the present configuration of species. However, numerous evolutionary processes can give rise to variation in the true evolutionary histories of individual genes, which are represented by gene trees. In this talk, we examine several distributions related to gene trees that arise when the coalescent process is used to model the relationship between gene trees and species trees. Study of these distributions gives insight into the challenges involved in using multi-locus data to estimate species-level phylogenies.

Fumei Lam

University of California, Davis
flam@cs.ucdavis.edu

“Generalizing the Four Gamete Test”

For binary input, the four gamete test gives a concise necessary and sufficient condition for the existence of a perfect phylogeny, and is the building block for many theoretical results and practical algorithms. In this talk, we discuss recent work to generalize the four gamete test (joint work with Dan Gusfield and Srinath Sridhar).

Sonja Petrovic

University of Illinois, Chicago
petrovic@math.uic.edu

“Group-based Models in Phylogenetics and Related Problems”

This talk will be an overview on phylogenetic invariants for group-based models, what is known so far, and how some of the algebraic conjectures could be used.

Eric Stone

North Carolina State University
eric_stone@ncsu.edu

“Something Old, Something New: A phylogenetic application of the combinatorial graph Laplacian”

Graphs have been used to represent a variety of relationships among biological data. The phylogenetic tree is one such graph whose purpose is to convey the pattern of descent relating a collection of species. On a phylogenetic tree, extant species are positioned as leaves, or pendent vertices in graph theoretical parlance. Crucially, ancestral species populate the interior of this graph and by definition are not observed. It is the province of phylogenetic reconstruction to identify from data where the speciation events away from these common ancestors have occurred. Thus, phylogenetic reconstruction is what creates a tree in the graph theoretic sense. In this talk, we take a step back and consider an application of graph theory to the latent tree encoded by the pairwise relationships between extant species. In particular, we show that a celebrated result of Miroslav Fiedler on spectral graph cutting extends to the latent tree case. We discuss how this extension can be used for phylogenetic reconstruction from distance data. Finally, we connect our new result to a classical application of multidimensional scaling in numerical taxonomy.

Seth Sullivant

North Carolina State University
smsulli2@ncsu.edu

“The Geometry of Phylogenetic Mixtures”

Phylogenetic mixture models are used to model evolutionary histories where possibly different regions in the genome have evolved according to different lineages. The mixtures models are specified by choosing a model of sequence evolution and a collection of trees. A fundamental open question about phylogenetic mixtures is whether or not the collection of trees used in the definition of the model is well-specified by the family of probability distribution in the model. In other words, are the tree parameters identifiable? We will describe results to this effect in the group based models. Our results depend heavily on the use of computer algebra software, and suggest some surprising geometric features of these models.

This is joint work with Elizabeth Allman, Sonja Petrovic, and John Rhodes.

Jeremy Sumner

University of Tasmania
jsumner@utas.edu.au

“Markov Invariants in Phylogenetics: the Quartet Case Done to Death”

It is possible to define "Markov invariants" as polynomials that transform as one-dimensional "representations" of the time evolution of the generic continuous time Markov chain. This is

done by viewing the transition matrices of the process as embedded within a continuous Lie group. In this way, Markov invariants retain some of the complex structure of the process, while greatly reducing the number of free parameters present. This approach has obvious attractive features in the light of the bias/variance tradeoff issue of statistical inference, but, as defined, the invariants are oblivious to the tree structure that underlies phylogenetic models. To employ these invariants in applied studies, we need to be able systematically to find linear combinations that are "tree informative". Algebraically, this means that these combinations need to satisfy the requirements of "phylogenetic invariants". In the quartet case, I will show that this can be done by demanding they also transform as irreducible representations of the group of symmetries of leaf permutations; itself a finite permutation group.

Jeffrey Thorne

North Carolina State University
thorne@statgen.ncsu.edu

“Making Inferences About the Impact of Phenotype on Genotype from the Ancestral Lineage”

There is a rich body of population genetic theory that accompanies the study of intraspecific genetic variation. Although many of the most important evolutionary events in the history of biology can only be studied via interspecific comparisons, it is difficult to apply population genetic theory to the study of interspecific genetic variation. However, some progress is being made in the situation where mutation rates are low. In this talk, we will focus on our attempts to infer the impact of phenotype on genotype in the low mutation rate regime. We will also overview simulation results that we have obtained for the scenario where mutation rates are higher.

This is joint work with Sang Chul Choi (now at Rutgers University) and Reed Cartwright (North Carolina State University).

Tandy Warnow

University of Texas
tandy@cs.utexas.edu

“SATE: A New Method for Simultaneous Estimation of Alignments and Trees”

Inferring an accurate evolutionary Tree of Life requires high quality alignments of molecular sequence datasets from large numbers of species. However, this task is often difficult, very slow and idiosyncratic, especially when the sequences have high rates of insertions and deletions (collectively, "indels") and substitutions. We present SATE, (Simultaneous Alignment and Tree Estimation), the first fully automated method that quickly and accurately estimates both DNA alignments and trees using the maximum likelihood (ML) criterion. It operates on much larger numbers of unaligned nucleotide sequences than other simultaneous methods that use likelihood, and in an extensive simulation study that included datasets of up to 1000 sequences, it dramatically improved tree and alignment accuracy compared to the best two-phase methods currently available.

Schedule for the 3rd Annual Graduate Student Conference in Probability

May 1-3, 2009

Hosted by The Department of Statistics and Operations Research at UNC- Chapel Hill and The Department of Mathematics at Duke University

Friday, May 1st

Due to the large number of speakers, we will have talks run in parallel. The speaker listed first will be in Hanes Room 120 and the speaker listed second will be in Hanes Room 125.

8:00-9:00 am Registration and Breakfast (Hanes 3rd Floor)

9:00-9:25 am Welcome Session (Hanes Room 120)

9:30-9:50 am Two type stochastic model for concentration in yeast cell - **Ankit Gupta**
Variations and Hurst index estimation for a Rosenblatt process using longer filters
Alexandra Chronopoulou

9:55-10:15 am Reaction-diffusion equations with extra parameters - **Yaqin Feng**
CLT's for Hilbert-space valued random fields under a strong mixing condition
Cristina Tone

10:15-10:30 am Coffee Break (Hanes 3rd Floor)

10:30-11:30 am **David Aldous: Keynote Address (Murphey Room 116)**
Spatial random networks

11:30-1:00 pm Lunch (Hanes 3rd Floor)

1:00-1:20 pm Large deviations for additive functionals of Markov processes - **Adina Oprisan**
Brownian motion on manifolds with manifold time-space - **Dmytro Karabash**

1:25-1:45 pm Error analysis of the simulation method for a Jump Type Markov process -
Arnab Ganguly
Formulas for stopped Levy processes at CUSUM stopping times -
Georgios Fellouris

1:50-2:30 pm Survival and limiting configurations in the two-type Richardson model, part 2
Nathaniel Blair-Stahn
First passage times of Levy subordinators: moments and computation -

Stahn Mark Veillette

- 2:30-2:45 pm Coffee Break (Hanes 3rd Floor)
- 2:45-3:05 pm Fluctuations of branching random walks - **Ming Fang**
On evaluation points for stochastic integrals - **Julius Esunge**
- 3:10-3:30 pm Models of dissemination through pairwise contact - **Joseph Whitmeyer**
Feynman-Kac formula for heat equation driven by fractional white noise -
Jian Song
- 3:35-4:15 pm What I believe about what you believe about what I believe, and so on ad
infinitum - **Paul Varkey**
Heat kernel measures on path and loop groups - **Matt Cecil**
- 4:15-4:30 pm Coffee Break (Hanes 3rd Floor)
- 4:30-5:30 pm **Russell Lyons: Keynote Address (Murphey Room 116)**
Asymptotic enumeration of spanning trees via traces and random walks
- 5:30 pm Opening Reception (Hanes 3rd Floor)

Saturday, May 2nd

Due to the large number of speakers, we will have talks run in parallel. For the talks before lunch, the speaker listed first will be in Gardner Room 008 and the speaker listed second will be in Gardner Room 105. For the talks after lunch, the speaker listed first will be in Wilson Room 107 and the speaker listed second will be in Peabody Room 215.

- 8:00-8:45 am Breakfast (Hanes 3rd Floor)
- 8:45-9:05 am Moderate deviation of intersection of ranges of random walks in the stable
case -
Justin Grieves
Volatility of Eurodollar futures and Gaussian HJM term structure models
Balaji Raman
- 9:10-9:30 am A dynamical version of the Kratky-Porod model of semi-flexible polymers
—
Philip Kilanowski/Marko Samara
Statistical analysis of volatility component models - **Fangfang Wang**
- 9:35-10:15 am Traffic jams, polymer growth, and random matrices - **Ivan Corwin**
Optimal trading strategies under arbitrage - **Johannes Ruf**
- 10:15-10:30 am Coffee Break (Hanes 3rd Floor)
- 10:30-11:30 am **Daniel Stroock: Keynote Address (Gardner Room 105)**

Gaussian measures in infinite dimensions

11:30-1:00 pm Lunch (Hanes 3rd Floor)

1:00-1:20 pm Markov chains on left-regular bands - **Aaron Smith**

Asymptotic tail probability of the maximum exceedance over a renewal threshold -

Xuemiao Hao

1:25-1:45 pm Comparison theorems for random walks on quotients of finitely generated groups

Russ Thompson

Optimal consumption with investment in incomplete semimartingale markets -

Helena Kauppila

1:50-2:30 pm Percolation with two robust clusters - **Peter Mester**

An optimal portfolio of correlated futures with small transaction costs -
Maxim Bichuch

2:30-2:45 pm Coffee Break (Hanes 3rd Floor)

2:45-3:05 pm A new total variation distance bound on Kac Random Walk - **Yunjiang Jiang**

Drawdowns and drawups in a finite time horizon - **Hongzhong Zhang**

3:10-3:30 pm Soft edge results for longest increasing paths on the planar lattice –

Nicos Georgiou

The malfunction probability and surplus ruin probability for non-profit organizations - **Li Chen**

3:35-4:15 pm Eigenvalues for Wishart matrices - **Weijun Xu**

Transition densities of symmetric α -stable processes - **Joshua Tokle**

4:15-4:30 pm Coffee Break (Hanes 3rd Floor)

4:30-5:30 pm **David Aldous: Remarks on Teaching (Mitchell Room 005)**

Remarks on teaching an undergraduate "Probability in the Real World" course

5:30 pm Dinner (Hanes 3rd Floor)

Sunday, May 3rd

Due to the large number of speakers, we will have talks run in parallel. The speaker listed first will be in Hanes Room 120 and the speaker listed second will be in Hanes Room 125.

- 8:00-8:45 am Breakfast (Hanes 3rd Floor)
- 8:45-9:05 am Complete integrability in Burgers turbulence - **Ravi Srinivasan**
Metastability in mean field models - **Mykhaylo Shkolnikov**
- 9:10-9:30 am Fitting circles to scattered data: parameter estimates have no moments -
Ali Al-sharadqah
Stochastic integration with respect to stable and tempered stable random
measures
Matthew Turner
- 9:35-9:55 am Variable bandwidth kernel density estimation with clipping procedures –
Hailin Sang
Effect of friction on noise - **Kunwoo Kim**
- 10:00-10:40 am Effect of truncation on heavy-tailed models - **Arijit Chakrabarty**
A view towards heteroclinicity of a dynamical system perturbed by
small noise
Sergio Almada
- 10:40-10:55 am Coffee Break (Hanes 3rd Floor)
- 10:55-11:15 am Weak convergence of stochastic integrals driven by continuous time
random
Walks - **Meredith Burr**
Randomization of forcing in large systems of PDE for improvement of
energy
Estimates - **Chia Ying Lee**
- 11:20-12:00 pm Inference in the presence of Volterra noise - **Bobby Reiner**
Thick points of the Gaussian free field - **Jason Miller**
- 12:05-12:45 pm Linear dependence of binary random vectors of fixed weight - **Ricardo
Restrepo**
Fractal and smoothness properties of space-time Gaussian models -
Yun Xue
- 12:50-1:10 pm Space-time Poisson processes applied to default data - **Cristina Canepa**
Viscosity and Principal-Agent problem - **Ruoting Gong**

Thank you to all of our sponsors

SAMSI/CRSC Undergraduate Workshop May 17 - May 22, 2009

Sunday, May 17

6:00 Welcoming Reception (Multipurpose Room, King Village)

Monday, May 18

8:15 Participants meet outside office at King Village. Transport to SAMSI.

9:00 *Breakfast at SAMSI*

9:30 Announcements and Introduction to SAMSI (Dr. Pierre Gremaud)

10:15 SAMSI Talk: Sequential Monte Carlo (Dr. Christian Macaro)

11:15 *Break*

11:30 Group Pictures

12:00 *Lunch at SAMSI*

1:00 Introduction to the Forward Problem: Solving the Harmonic Oscillator System
(Dr. Megan Owen)

2:30 *Break*

2:45 Brief Introduction to the Computing System and MATLAB (Dr. Ioanna Manolopoulou)

4:15 Vans take participants to Lake Crabtree

5:00 Dinner at Lake Crabtree

Tuesday, May 19

8:15 Participants meet outside office at King Village. Transport to SAMSI.

9:00 *Breakfast at SAMSI*

9:30 Linear Inverse Problems: A MATLAB Tutorial. (Wenjie Chen)

11:00 *Break*

11:15 Basic Probability and Statistics (Sarah Schott)

12:45 *Lunch at SAMSI*

- 1:45 Introduction to Statistical Inference (Dr. Saeid Yasamin)
- 3:15 *Break*
- 3:30 Regression and Least Squares: A MATLAB Tutorial. (Baqun Zhang)
- 5:00 Vans take participants to King Village

Wednesday, May 20 (All sessions are in SAS Hall Room 4101 unless otherwise noted.)

- 8:30 *Breakfast in SAS Hall Room 4101*
- 9:00 Graduate School Panel
- Melanie Bain, Operations Research, UNC
- Dr. Ernie Stitzinger, Mathematics Department, NCSU
- Dr. Kim Weems, Statistics Department, NCSU
- 10:00 Vibrating Beam Data Collection at CRSC Laboratory in Cox 309
- Adam Attarian, CRSC and Mathematics Department, NCSU
- Dr. Grace Kepler, CRSC, NCSU
- Dr. Hien Tran, CRSC and Mathematics Department, NCSU
- 10:45 *Break*
- 11:00 Career Panel
- Dr. Karen Chiswell, GlaxoSmtihKline
- Dr. Scott Pope, SAS
- Dr. Jeff Scroggs, Mathematics Department, NCSU
- 12:00 *Lunch with Panelists (provided)*
- 1:00 Reflection on the Data Collection and Modeling Experiences (Dr. Sourish Das)
- 2:00 Introduction to Optimization (Melanie Bain)
- 3:00 *Break*
- 3:15 Solving the Vibrating Beam: Inverse Problem (Jason Yellick)

Thursday, May 21 (All sessions are in SAS Hall Room 4101.)

- 8:30 *Breakfast in SAS Hall Room 4101*
- 9:00 Statistical Analysis for the Vibrating Beam Inverse Problem (Dr. Sourish Das)
- 10:00 *Break*

- 10:15 Alternative Beam Model (Dr. Pierre Gremaud)
- 11:15 Teams work on Inverse Problem
- 12:30 *Lunch{Participants on their own}*
- 1:30 What could we do better? Alternative Statistical methods
(Francesca Petralia)
- 2:30 Teams work on Inverse Problem; Begin to prepare reports
- 3:30 *Break*
- 3:45 Teams continue work on Inverse Problem and preparation of reports
- 5:00 *Dinner Break{Participants on their own}*
- 6:30 Bowling

Friday, May 22 (All sessions are in SAS Hall Room 4101.)

- 8:30 *Breakfast in SAS Hall Room 4101*
- 9:00 Presentations and Discussion
- 10:30 *Break*
- 10:45 Presentations and Discussion
- 11:45 Closing Remarks & Workshop Evaluation
(Drs. Pierre Gremaud and Cammey Cole Manning)
- 12:00 *Lunch{Participants on their own}*

**Program on Algebraic Methods in Systems Biology and Statistics
Transition Workshop
June 18-20, 2009**

SCHEDULE

Thursday, June 18, 2009

- 8:15-8:45 Registration and Continental Breakfast
- 8:45-9:00 Welcome

- 9:00-10:00 **Reinhard Laubenbacher**, Virginia Bioinformatics Institute
“Stochastic Algebraic Models”
- 10:00-10:20 Break
- 10:20-11:00 **Heike Siebert**, Freie Universität Berlin
“Modularity of Discrete Regulatory Networks”
- 11:00-12:00 **Luis David Garcia-Puente**, Sam Houston State University
“Applications of Toric Varieties in the Sciences”
- 12:00-12:30 Second Chances
- 12:30-2:00 Lunch
- 2:00-3:00 **Katherine St. John**, City University of New York
“Comparing Phylogenetic Trees”
- 3:00-3:40 **Megan Owen**, SAMSI
“Computing the Geodesic Distance in Tree Space in Polynomial Time”
- 3:40-4:00 Break and Poster Set-up
- 4:00-5:00 **Marcy Uyenoema**, Duke University
“Genomic conflict and DNA Sequence Variation”
- 5:00-5:30 Second Chances
- 5:30-6:00 Poster Advertisements (2 minute ads each)
- 6:00-8:00 Poster Session and Reception
SAMSI will provide poster presentation boards and tape. The board dimensions are 4 ft. wide by 3 ft. high. They are tri-fold with each side being 1 ft. wide and the center 2 ft. wide. Please make sure your poster fits the board. The boards can accommodate up to 16 pages of paper measuring 8.5 inches by 11 inches.

Friday, June 19, 2009

- 8:30-9:00 Registration and Continental Breakfast
- 9:00-10:00 **Henry Wynn**, London School of Economics
“Betti Numbers, State Polytopes and the Connectivity of Experimental Design”
- 10:00-10:15 Break
- 10:15-11:15 **Ruriko Yoshida**, University of Kentucky
“Markov Bases and Subbases for Bounded Contingency Tables”
- 11:15-12:15 **Peter Huggins**, Carnegie Mellon University

“Extensions of Parametric Inference”

- 12:15-12:40 Second Chances
- 12:40-2:00 Lunch
- 2:00-3:00 **Gilles Gnacadja**, Amgen
“Selected Problems about the Equilibrium States of Networks of Reversible Binding Reactions”
- 3:00-3:40 **Anne Shiu**, University of California, Berkeley
“Siphons in Biochemical Reaction Networks”
- 3:40-4:00 Break
- 4:00-5:00 **Olga Milenkovic**, University of Illinois, Urbana-Champaign
“Information Theoretic Methods for the Reverse Engineering of the Topology and Dynamics of Gene Regulatory Networks”
- 5:00-5:30 Second Chances

Saturday, June 20, 2009

- 8:30-9:00 Registration and Continental Breakfast
- 9:00-10:00 **Elena Dimitrova**, Clemson University
“Parameter Estimation for Drought Response-related Gene Networks in Rice”
- 10:00-10:20 Break
- 10:20-11:00 **Paul Kidwell**, Purdue University
“Non-Parametric Modeling of Partially Ranked Data with Application to Survey Design”
- 11:00-Noon **Seth Sullivant**, North Carolina State University
“Algebraic Challenges for Gaussian Graphical Models”
- Noon-12:30 Second Chances and Closing
- 12:30-2:00 Lunch

**Algebra Transition Workshop
June 18-20, 2009**

SPEAKER TITLES/ABSTRACTS

Elena Dimitrova

Clemson University
edimit@clemson.edu

“Parameter Estimation for Drought Response-related Gene Networks in Rice”

Rice is a keystone crop in worldwide food supply but like all crops, it is susceptible to yield reduction as a result of water deficit. At the cellular level, water stress induces numerous plant responses including massive rearrangements of gene expression patterns, accumulation of specific hormones, membrane protection, and protein stabilization. The understanding of these mechanisms on genetic level would help create new cultivars that have a high yield potential under normal and water deficit conditions. We present several existing statistical approaches towards this goal and propose a promising stochastic modeling method that is based on computational algebra and combinatorics.

Luis David Garcia-Puente

Sam Houston State University
lgarcia@shsu.edu

“Applications of Toric Varieties in the Sciences”

Geometric modeling builds computer models for industrial design and manufacture from basic units, called patches. Many patches, including Bezier curves and surfaces, are pieces of toric varieties, which are objects from algebraic geometry. Statistical models are families of probability distributions used in statistical inference to study the distribution of observed data. Many statistical models, including the log-linear or discrete exponential models are also pieces of toric varieties. Toric varieties also play an important role in the study of systems of nonlinear ordinary differential equations that derive from chemical reaction networks. In this talk, I will show how toric varieties arise in these diverse fields and the direct connections between these applied subjects. In particular, I will discuss the role of maximum likelihood estimation in geometric modeling and dynamical systems.

Gilles Gnacadja

Amgen
gnacadja@amgen.com

“Selected Problems about the Equilibrium States of Networks of Reversible Binding Reactions”

We recently proposed the class of complete networks of reversible binding reactions in an effort to describe many reaction networks that are studied in pharmacology. An outcome of this effort is a positive polynomial P such that, given a vector b of total (free and bound) concentrations of the so-called elementary species, the vector x of equilibrium concentrations of these species is uniquely given by $P(x) = b$. The polynomial P is parameterized with structural and kinetic information about the network, and the equation $P(x) = b$ admits an auspicious transformation into a fixed-point equation $F(x) = x$ where the function F is positive and order-reversing. We will discuss two outstanding issues relevant to applications of this work:

1. The identifiability of kinetic and structural parameters from the complete or partial and aggregate knowledge of equilibrium state; and
2. The prospect of exploiting the fixed-point equation to calculate the equilibrium state with speed and a priori assurance of success.

Peter Huggins

Carnegie Mellon University
phuggins@andrew.cmu.edu

“Extensions of Parametric Inference”

For many graphical models, MAP inference can be performed for all model parameters simultaneously by using parametric inference. In practice, many real-world applications can benefit from extensions of the original parametric inference framework. Specifically we consider constrained parametric inference and parametric k -best inference. We derive complexity bounds and efficient easy-to-use algorithms which mirror the best-known results for parametric inference. In particular, parametric k -best inference has surprisingly tractable complexity -- polynomial in k -- which also lends new insight into the complexity of standard parametric inference.

Paul Kidwell

Purdue University
kidwellpaul@gmail.com

“Non-Parametric Modeling of Partially Ranked Data with Application to Survey Design”

Statistical models on full and partial rankings of n items are often of limited practical use for large n due to computational consideration. We explore the use of non-parametric models for partially ranked data and derive computationally efficient procedures for their use for large n . The derivations are largely possible through combinatorial and algebraic manipulations based on the lattice of partial rankings. A bias-variance analysis and an experimental study demonstrate the applicability of the proposed method. This estimation procedure finds a ready application to survey question design via selection of the best partial ranking form for eliciting subject preferences. By allowing the question form to vary over partial rankings a smoothing is performed which may reduce both MSE and the

cognitive burden associated with providing full rankings. A decision theoretic formulation is then possible in the space of survey cost and optimal estimator form with respect to MSE.

Reinhard Laubenbacher

Virginia Bioinformatics Institute
reinhard@vbi.vt.edu

“Stochastic Algebraic Models”

This talk will focus on several different types of dynamic algebraic models for biological networks and their relationship. After discussing the deterministic case and some central open problems, they are related to problems about stochastic polynomial models and some results are presented.

Olgica Milenkovic

University of Illinois, Urbana-Champaign
milenkov@ad.uiuc.edu

“Information Theoretic Methods for the Reverse Engineering of the Topology and Dynamics of Gene Regulatory Networks

We consider the problem of reverse engineering the topology and dynamics of gene networks when only small training sample sets are available for the modeling approach. In particular, we propose a combination of methods from the theory of Markov random fields and algebraic list-decoding to accomplish this task, and provide analytical results describing the model complexity and sample-set size trade-offs needed for accurate modeling. Our methods are tested on the E.coli SOS repair system network.

Megan Owen

SAMSI
maowen@ncsu.edu

“Computing the Geodesic Distance in Tree Space in Polynomial Time”

The geodesic distance between two phylogenetic trees is the length of the shortest path between them in tree space, as introduced by Billera, Holmes, and Vogtmann (2001). We present the first known polynomial time algorithm for computing the geodesic distance. We construct a bipartite graph to represent constraints on the geodesic path through tree space. To find the geodesic distance, we repeatedly solve a minimum weight vertex cover problem on this graph.

Ann Shiu

University of California, Berkeley

annejls@berkeley.edu

“Siphons in Biochemical Reaction Networks”

In a biochemical reaction network, the concentrations of chemical species evolve in time, governed by the polynomial differential equations of mass-action kinetics. Siphons in a biochemical reaction system are subsets of the species that have the potential of being absent in a steady state. We present a new method that computes siphons and determines which of them are relevant. This method relies on the primary decomposition of monomial and binomial ideals. The importance of such a procedure is for verifying whether large biochemical reaction systems are persistent; "persistence" is the property that no species concentration tends to zero. As an application, we can compute for a given system, the set of initial conditions for which persistence is easily verified; this set consists of regions of a chamber decomposition. This is joint work with Bernd Sturmfels.

Heike Siebert

Freie Universitat Berlin
siebert@mi.fu-berlin.de

“Modularity of Discrete Regulatory Networks”

Analyzing complex networks is a difficult task, regardless of the chosen modeling framework. For a discrete regulatory network, even if the number of components is in some sense manageable, we have to deal with the problem of analyzing the dynamics in an exponentially large state space. A well known idea to approach this difficulty is to break the network down to smaller building blocks, analyze them in isolation and then draw conclusions concerning the original network. However, this approach faces several difficulties. How do we identify suitable building blocks, what is a sensible way to derive the rules governing their behavior in isolation, and what are the rules to derive information about the networks dynamics from the dynamical properties of its building blocks? In this talk we address these questions, not only applying the notion of motif or module to the network structure but also to the system's dynamics, and illustrating the benefit of understanding the rules relating the structural to the dynamical building blocks.

Katherine St. John

City University of New York
stjohn@lehman.cuny.edu

“Comparing Phylogenetic Trees”

Evolutionary histories, or phylogenies, form an integral part of much work in biology. In addition to the intrinsic interest in the interrelationships between species, phylogenies are used for drug design, multiple sequence alignment, and even as evidence in a recent criminal trial. A simple representation for a phylogeny is a rooted, binary tree, where the leaves represent the species, and internal nodes represent their hypothetical ancestors.

This talk will focus on some of the elegant questions that arise from assembling, summarizing, and visualizing phylogenetic trees.

Seth Sullivant

North Carolina State University
smsulli2@ncsu.edu

“Algebraic Challenges for Gaussian Graphical Models”

Gaussian graphical models have a long history and have been widely used in statistics, economics, and the social sciences, often under many different names (for example, structural equation models). Despite their ubiquity, there remain fundamental open problems about their mathematical structure. Nearly all of these problems are open because of algebraic and combinatorial difficulties. The purpose of this talk will be to highlight some of these challenges including: identifiability, (non)smoothness, maximum likelihood estimation, and constraints.

Marcy Uyenoema

Duke University
marcy@duke.edu

“Genomic Conflict and DNA Sequence Variation”

I will use self-incompatibility (SI) in flowering plants to illustrate the need for the development of a framework for inferring the nature of the evolutionary process from patterns of DNA sequence variation.

Sexual antagonism reflects differences in evolutionary pressures between the sexes. Although most plants are hermaphroditic, sexual antagonism may arise between genetic factors that control the pollen (male) and pistil (female) components of reproduction. Tight genetic linkage between the pollen and pistil components within the S-locus is essential to the operation of SI. Although linkage implies that the evolutionary fates of the male and female components are conjoined, the S-locus region may bear the hallmarks of sexual antagonism.

I will describe the selective pressures to which the S-locus is subject and observed genetic patterns that may reflect those pressures.

Henry Wynn

London School of Economics
H.Wynn@lse.ac.uk

“Betti Numbers, State Polytopes and the Connectivity of Experimental Design”

The now standard method of obtaining a saturated polynomial (regression) basis for an experimental, gives much additional information, for example as we vary the monomial ordering. One example is the state polytope, whose lower boundary is obtained for

special classes of designs. Another is the connectivity structure of the basis as measured by the Betti number of the design ideal and certain simplicial complexes. Roughly speaking, high Betti number are associated with less connectivity and more lower degree monomial terms. Low Betti numbers mean fewer isolated “effects” and more higher order interactions. These features are studied in detail for two level designs (the square-free case) and for both regular fractions and non-standard designs. The relationship between the state polytope, average degrees aberration and the Betti numbers is studied.

Ruriko Yoshida

University of Kentucky
ruriko.yoshida@gmail.com

“Markov Bases and Subbases for Bounded Contingency Tables”

In this talk, we focus on bounded two-way contingency tables under independence model and show that if these bounds on cells are positive, i.e., they are not structural zeros, the set of basic moves of all 2×2 minors connects all tables with given margins. We end this paper with a conjecture that if we know that the given margins are positive, the set of basic moves of all 2×2 minors connects all incomplete contingency tables with given margins. This is joint work with Fabio Rapallo.

**Program on Psychometrics
July 7-17, 2009**

SCHEDULE

Tuesday, July 7, 2009

(Radisson, RTP)

- | | |
|-------------|---|
| 8:15-8:45 | Registration and Continental Breakfast |
| 8:45-9:00 | Welcome |
| 9:00-12:00 | Yanyan Sheng , Southern Illinois University
<i>“Bayesian Analysis of Item Response Theory Models”</i> |
| 10:00-10:30 | Break |
| 12:00-2:00 | Lunch |
| 2:00-3:00 | David Thissen , University of North Carolina
<i>“IRTPRO Demonstration”</i> |
| 3:00-3:15 | Break |

3:15-4:15 **Richard Swartz**, University of Texas, MD Anderson Cancer Center
 “Bayesian and Classical Computerized Adaptive Testing Item Selection Algorithms”

Wednesday, July 8, 2009
(Radisson, RTP)

8:30-9:00 Registration and Continental Breakfast

9:00-12:00 **Sun-Joo Cho**, University of California, Berkeley
 Frank Rijmen, Educational Testing Service
 Mark Wilson, University of California, Berkeley
 “A Nonlinear Mixed Models Approach to IRT”

10:00-10:30 Break

12:00-2:00 Lunch

2:00-4:00 **Mario Peruggia**, Ohio State University
 “Hierarchical Bayes Models for Response Time Data”

Trish Van Zandt, Ohio State University
 “An Overview of Response Time Models in Psychology”

3:00-3:15 Break

4:00-5:00 *Poster Advertisements* (2 minute ads each)

5:00–7:00 *Poster Session and Reception*
 SAMSI will provide poster presentation boards and tape. The board dimensions are 4 ft. wide by 3 ft. high. They are tri-fold with each side being 1 ft. wide and the center 2 ft. wide. Please make sure your poster fits the board. The boards can accommodate up to 16 pages of paper measuring 8.5 inches by 11 inches.

Thursday, July 9, 2009
(Radisson, RTP)

8:30-9:00 Continental Breakfast

9:00-12:00 **Mathias von Davier**, Educational Testing Service
 “Notes on Models for Cognitive Diagnosis”

10:00-10:30 Break

- 12:00-1:30 Lunch
- 1:30-2:30 **Sandip Sinharay**, Educational Testing Service
“A Critical Evaluation of Diagnostic Score Reporting: Some Theory and Applications”
- 2:30-2:45 Break
- 2:45-4:00 **Dongchu Sun**, University of Missouri
“Bayesian Hierarchical Models for Recognition-Memory Experiments”
- Jun Lu**, American University
“A Bayesian Approach for Assessing Human Memory Using Process-Dissociation Procedure”

Friday, July 10, 2009
(Radisson, RTP)

- 8:30-9:00 Continental Breakfast
- 9:00-12:00 **Matthew Johnson**, Columbia University
“An Introduction to Rater Models”
- 10:00-10:30 Break
- 12:00-2:00 Lunch
- 2:00-4:00 **Paul Speckman, Dongchu Sun, and Jeff Rouder**, University of Missouri
“Item-Response Models for Measuring Thresholds in Performance”
- 3:00-3:15 Break

Monday July, 13 – Friday July, 17
(SAMSI)

- 9:00-12:00 Working Group Meetings:
Peer Review Working Group (Room 104)
Cognitive Diagnostic Models Working Group (Room 150)
Longitudinal Assessment of PRO Working Group (Room 203)
- 12:00-1:00 Lunch

1:00-5:00 Working Group Meetings:
Peer Review Working Group (Room 104)
Cognitive Diagnostic Models Working Group (Room 150)
Longitudinal Assessment of PRO Working Group (Room 203)

Program on Psychometrics: Peer Review Working Group

The PRWG will meet spontaneously during the second week of the program. Currently scheduled talks are listed below. Contributed talks will be added to the schedule during the course of the Psychometric Program. A white paper will be prepared by the working group and summarized at a joint meeting at the end of the week. Non-meeting time will be devoted to group collaboration on topics related to peer review.

SCHEDULE

Monday July 13, 2009 – Friday July 17, 2009

Monday July 13, 2009

11:30-12:30 **David Banks**, Duke University
“Judgement in JASA”

12:30-2:00 Lunch

Tuesday July 14, 2009

11:30-12:30 **Valen E. Johnson**, University of Texas, MD Anderson Cancer Center
“An Overview of NIH R01 Peer Review Scores”

12:30-2:00 Lunch

Wednesday July 15, 2009

11:00-12:00 **Jing Cao**, Southern Methodist University
“A Bayesian Approach to Ranking and Rater Evaluation: An Application to Grant Reviews”

12:00-1:30 Lunch

Thursday July 16, 2009

11:00-12:00 **Song Zhang**, University of Texas, Southwestern Medical Center
“A Bayesian Hierarchical Model for Multi-rater Data with Fine Scales”

12:00-1:30 Lunch

Friday July 17, 2009

11:00-12:00 Discuss Draft of White Paper

12:00-1:00 Lunch

1:00 Adjourn

Cognitive Diagnostic Models Working Group (CDMWG)

The CDMWG will meet during the second week of the program. The program will be more structured during the beginning of the week, and more open during the end of the week. Talks currently scheduled during this week are listed below. Additional talks will be added to the schedule during the course of the Psychometric Program. A white paper will be prepared by the working group and summarized at a joint meeting at the end of the week.

SCHEDULE

Monday July 13, 2009 – Friday July 17, 2009

Monday July 13, 2009

9:00-10:00 Introduction

10:00-11:30 **Matthew Finkelman**, Tufts University
Kristin Huff, College Board
Curtis Tatsuoka, Case Western University
“Diagnostic Assessment Approaches”

11:30-12:30 **David Banks**, Duke University

12:30-2:00 Lunch

2:00-3:30 **Tzur Karelitz**, Education Development Center
Jere Confrey, North Carolina State University
Alicia Alonzo, University of Iowa
“Developmental Theories for Diagnostic Assessment”

3:30-5:00 **Andre Rupp**, University of Maryland
Ying Cui, University of Alberta
Nathalie Loyer, University of Montreal
“Task & Q-matrix Construction”

Tuesday July 14, 2009

8:30-10:00 **Andre Rupp**, University of Maryland
Jimmy de la Torre, Rutgers University

Robert Henson, University of North Carolina, Greensboro
“Fully Parametric Models for Classification”

10:00-11:30 **Curtis Tatsuoka**, Case Western University
Ying Cui, University of Alberta
Rebecca Nugent, Carnegie Mellon University
“Non-parametric and Semi-parametric Models for Classification”

11:30-12:30 **Valen Johnson**, University of Texas, MD Anderson Cancer Center
12:30-2:00 Lunch

2:00-3:30 **Roy Levy**, Arizona State University
Robert Henson, University of North Carolina, Greensboro
Jimmy de la Torre, Rutgers University
“Challenges in Estimation, Programming, and Implementation”

3:30-5:00 **Ying Cui**, University of Alberta
Roy Levy, Arizona State University
Jimmy de la Torre, Rutgers University
“Model Fit Assessment & Refinement”

Wednesday July 15, 2009

8:30-10:30 **Curtis Tatsuoka**, Case Western University
Ying Cheng, University of Notre Dame
Matthew Finkelman, Tufts University
“Optimal Test Design and Computerized Adaptive Testing”

10:30-12:00 **Eunice Jang**, University of Toronto
Neil Heffernan, Worcester Polytechnic Institute
Kristin Huff, College Board
“Score Reporting & Subsequent Action”

12:00-1:30 Lunch

1:30-3:00 **Andre Rupp**, University of Maryland
Tiffany Barnes, University of North Carolina, Charlotte
Neil Heffernan, Worcester Polytechnic Institute
“Validation”

3:00-5:00 Moderated Panel Session
“Future Challenges for Diagnostic Assessment”

The morning of the fourth day will be devoted to identifying the cutting-edge methods in cognitive diagnosis modeling, and three to five important gaps that must be filled to advance the field forward. The cutting-edge methods and research gaps will constitute the research agenda of the working group. The topics in these research agenda will be distributed into three

90-minute time blocks that will extend from Thursday afternoon and the first part of Friday morning. Participants will be asked to select a research topic in each time block. The 1.5 hour-block will be used to discuss potential research projects that can be done in specific topics, sign-up participants who can collaborate on these projects, and outline strategies and time frame for completing these projects. Starting the latter part of Friday morning (10:30-12:30), participants will report the summary of the discussion of potential projects to all the participants of the working group. The working group will adjourn after a lunch.

Longitudinal Assessment of PRO Working Group (LAPROWG)

The LAPROWG will meet during the second week of the program. The program will be more structured during the beginning of the week, and with more spontaneous meetings during the end of the week. Talks currently scheduled during this week are listed below. Additional talks may be added to the schedule during the course of the Psychometric Program. A white paper will be prepared by the working group and summarized at a joint meeting at the end of the week.

SCHEDULE

Monday July 13, 2009 – Friday July 17, 2009

Monday July 13, 2009

- 9:00-9:30 **Richard Swartz**, University of Texas M. D. Anderson Cancer Center
Introductions, Purpose of Working Group
- 9:30-10:20 **Carolyn Schwartz**, DeltaQuest
“Importance of Responsiveness to Change and Mediators to the Measurement of Change in Health Outcomes”
- 10:30-Noon **Ken Bollen**, University of North Carolina
“Longitudinal Measurement of Patient Reported Outcomes: Latent Curve Models Using Structural Equation Models”
- 1:00-3:15 **Ethan Basch**, Memorial Sloan Kettering Cancer Center
Charles Cleeland, University of Texas M. D. Anderson Cancer Center
“Practical Needs in Trial Design and Detecting True Change Over Time - Clinicians’ Perspective”
- Diane Fairclough**, University of Colorado, Denver
“Practical Needs in Trial Design and Detecting True Change Over Time - Patient Perspective of the Patient Experience”
- 3:30-5:00 **Carolyn Schwartz**, DeltaQuest
“Methods to Detect Response Shift and Responsiveness to Change ”

Tuesday, July 14

- 9:00-10:30 **Bruce Rapkin**, Albert Einstein College of Medicine
“Cognitive Factors in the Quality of Life Rating Response Scales and How to Include/model This Information”
- 10:45-Noon **Diane Fairclough**, University of Denver, Colorado
“Impact of Missing Data When Evaluating Change Over Time”
- 1:00- 2:30 **Li Cai**, University of California, Los Angeles
“Multidimensional IRT and Potential Applications to Assessing PROs Over Time and Detecting Response Shift”
- 3:00-5:15 **Jeff Sloan**, Mayo Clinic
“Precision, Validity and Sensitivity vs. Response Burden in PRO Endpoints – Facilitating Detection of True Change: Using Single-item vs. Multi-item Scales to Monitor Change”
- Richard Swartz**, The University of Texas M. D. Anderson Cancer Center
“Precision, Validity and Sensitivity vs. Response Burden in PRO Endpoints – Facilitating Detection of True Change: Considering the Precision vs. Burden Tradeoff within CAT”

Wednesday, July 15

- 9:00-10:30 **Jeff Sloan**, Mayo Clinic
“Interpreting Minimally Important Differences While Accounting for Measurement Variability/response Shift”
- 10:45-Noon Brainstorming Next Steps / Outline White Papers
- 1:00-3:00 Outline White Paper / Discuss Datasets

Thursday, July 16

- 12:00-1:00 Lunch
- 3:00-5:00 Updates

Friday, July 17

- 9:00-Noon Revise Outline for White Papers/ Delegate Duties / Develop Timeline to Complete White Paper.
- 12:00-1:00 Lunch

The white paper to be produced in the LAPROWG has two main goals:

- 1) to discuss the state of the art and recommend policy and procedures for analyzing longitudinal PRO data
- 2) to identify areas needing methodological improvement when considering longitudinal PRO data

Summer School: Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change

**July 28 – August 1, 2009
SCHEDULE**

Tuesday July 28, 2009

[SAMSI](#) (Room 150)

- | | |
|-----------|---|
| 7:45-8:45 | Shuttles from Radisson RTP to SAMSI |
| 8:45-9 | Registration |
| 9-12 | Estimating curves and surfaces and splines
Exercises and group activities / Tutorials in R
(Doug Nychka , NCAR) |
| 12-1:15 | Lunch |
| 1:15-4:30 | Spatial process models and Kriging
Exercises and group projects / Tutorials in R
(Doug Nychka , NCAR) |
| 4:30-5:30 | Shuttles to Radisson RTP |
| 6-8 | Banquet at Radisson RTP |

Wednesday July 29, 2009

[SAMSI](#) (Room 150)

- | | |
|-------|---|
| 8-9 | Shuttles from Radisson RTP to SAMSI |
| 9-10 | Estimating covariances and nonGaussian models
(Doug Nychka , NCAR) |
| 10-12 | Multivariate spatial data and models
Exercises and Examples
(Stephen Sain , NCAR) |

- 12-1:15 Lunch
- 1:15- 3:00 Group projects
- 3:00 - 4:30 Sparse matrix methods and Kriging
Examples
(**Reinhard Furrer**, University of Zurich)
- 4:30-5:30 Shuttles to [Radisson RTP](#)

Thursday July 30, 2009

[SAMSI](#) (Room 150)

- 8-9 Shuttles from [Radisson RTP](#) to SAMSI
- 9-10:30 Application to large spatial data sets
Examples
(**Reinhard Furrer**, University of Zurich)
- 10:30 -12 Group projects
- 12-1:15 Lunch
- 1:15 - 4:30 Bayesian methods for spatial data
Examples
(**Sudipto Banerjee**, University of Minnesota)
- 4:30 – 5:30 Evening on your own: Shuttles to the [Radisson RTP](#) or to [The Streets at Southpoint](#) (return on your own, via hotel shuttle (919) 549-8631 or taxi)

Friday July 31, 2009

[SAMSI](#) (Room 150)

- 8-9 Shuttles from [Radisson RTP](#) to SAMSI
- 9 -12 Spatial autoregressive models for epidemiological data
(**Sudipto Banerjee**, University of Minnesota)
- 12- 1:15 Lunch
- 1:15-4:30 Examples and group projects using R packages.
- 4:30-5:30 Shuttles to [Radisson RTP](#)

Saturday August 1, 2009

[SAMSI](#) (Room 150)

- 8-9 Shuttles from [Radisson RTP](#) to SAMSI

(check-out at the hotel and bring your belongings with you to SAMSI)

9- 11:30 Student/group presentations and discussion.

11:30-12:30 Shuttles to airport and [Radisson RTP](#)

APPENDIX F – Workshop Evaluations

F.1 Overview of Workshop Evaluation

At each workshop, the participants are asked to complete the SAMSI Workshop Evaluation, which asks the participants to rank the Workshop in terms of scientific quality, staff, helpfulness, meeting facilities, lodging, and local transportation and then asks a series of questions. A sample evaluation is included for review.

The following results are summaries of the evaluations completed after a total of fourteen Workshops in the previous year broken into two categories: Program Workshops and Education and Outreach (E&O) Workshops. The Program Workshops include the seven workshops held to date for the 2008-09 Program Year together with the three 2007-08 Program Workshops that occurred after the submission date of the 2007-08 Annual Report and were therefore not included in that Annual Report. The E&O Workshops include the two workshops held to date for 2008-09 together with the final two E&O Workshops from the 2007-08 year.

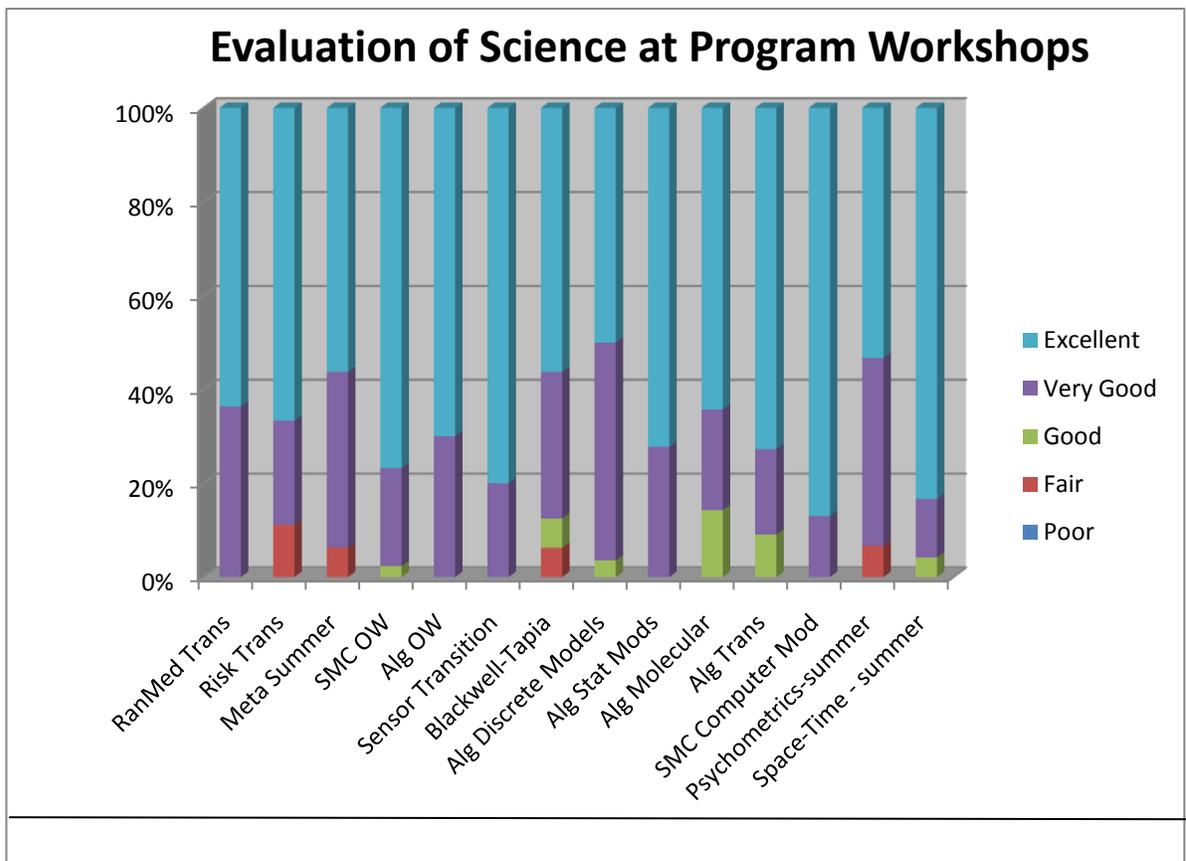
There are three remaining Program Workshops and two E&O Workshop scheduled for 2009 and evaluations of these programs will be included in the 2009-10 Annual Report.

SAMSI Workshop Evaluation Form

Your feedback on this workshop is requested by SAMSI's funding agencies, who view it as important for assessing and improving our performance. Your feedback is also gratefully appreciated by SAMSI's directors, because it will enable us to immediately improve SAMSI activities. Please fill out this form and hand it to a SAMSI Staff Member, or return it by mail.

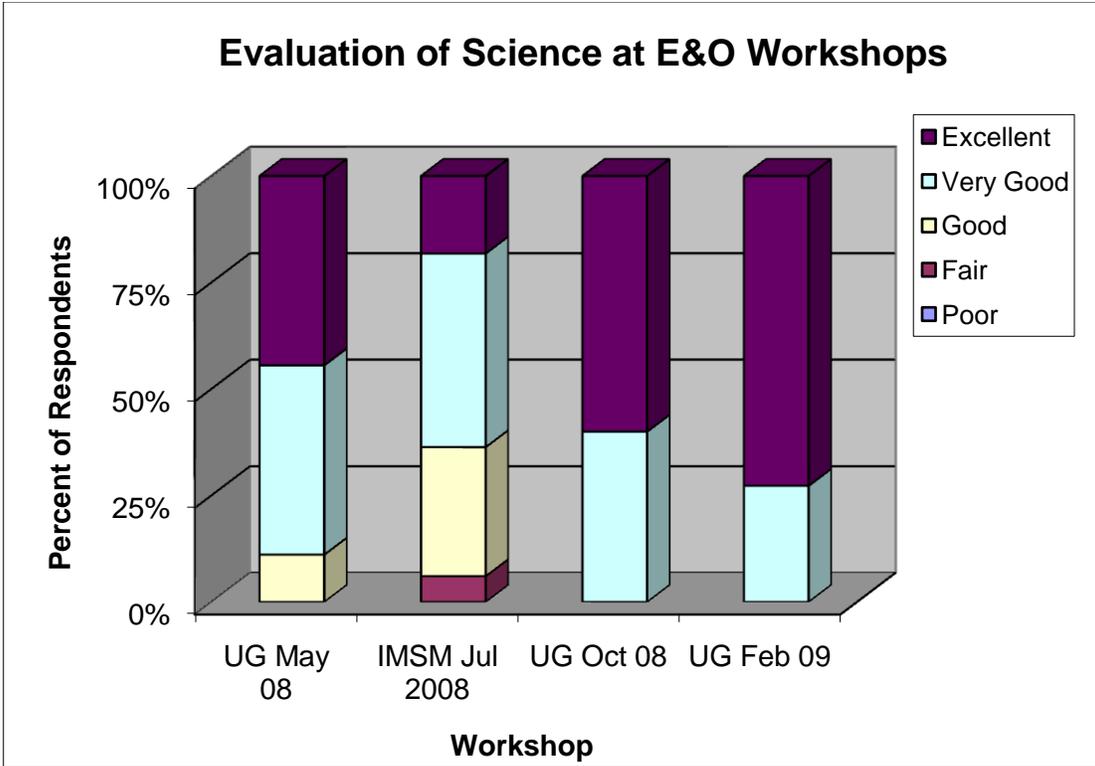
1. Personal Information: We are required by our funding agencies to obtain information – in a standard format – about all participants in SAMSI activities. If you have not already done so, please go to www.samsi.info/PartInfo/200708/participantinformationform0708.html to provide this information. Note that if you have participated in a SAMSI activity since last July 1 and completed this webform, you need not do so again, unless your personal information has changed.

2. General Ratings:	Poor	Fair	Good	Very Good	Excellent
a. Scientific Quality	1	2	3	4	5
b. Staff Helpfulness	1	2	3	4	5
c. Meeting Room/AV Facilities	1	2	3	4	5
d. Lodging	1	2	3	4	5

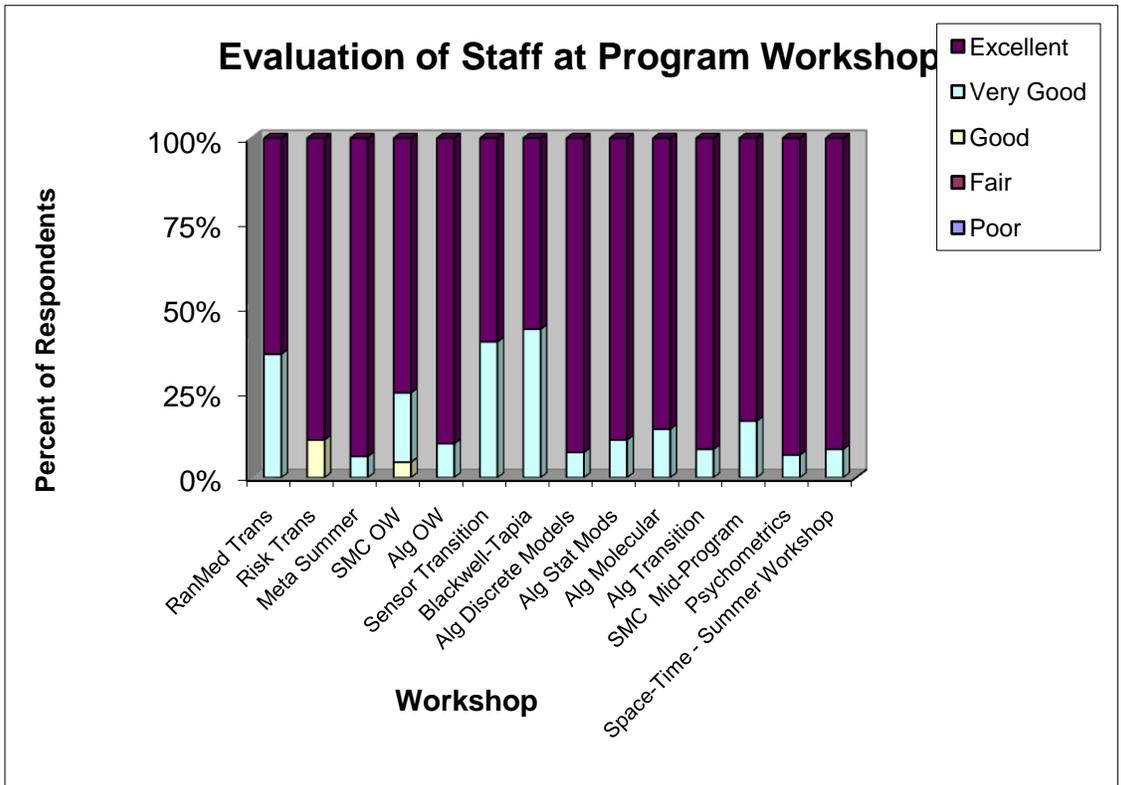


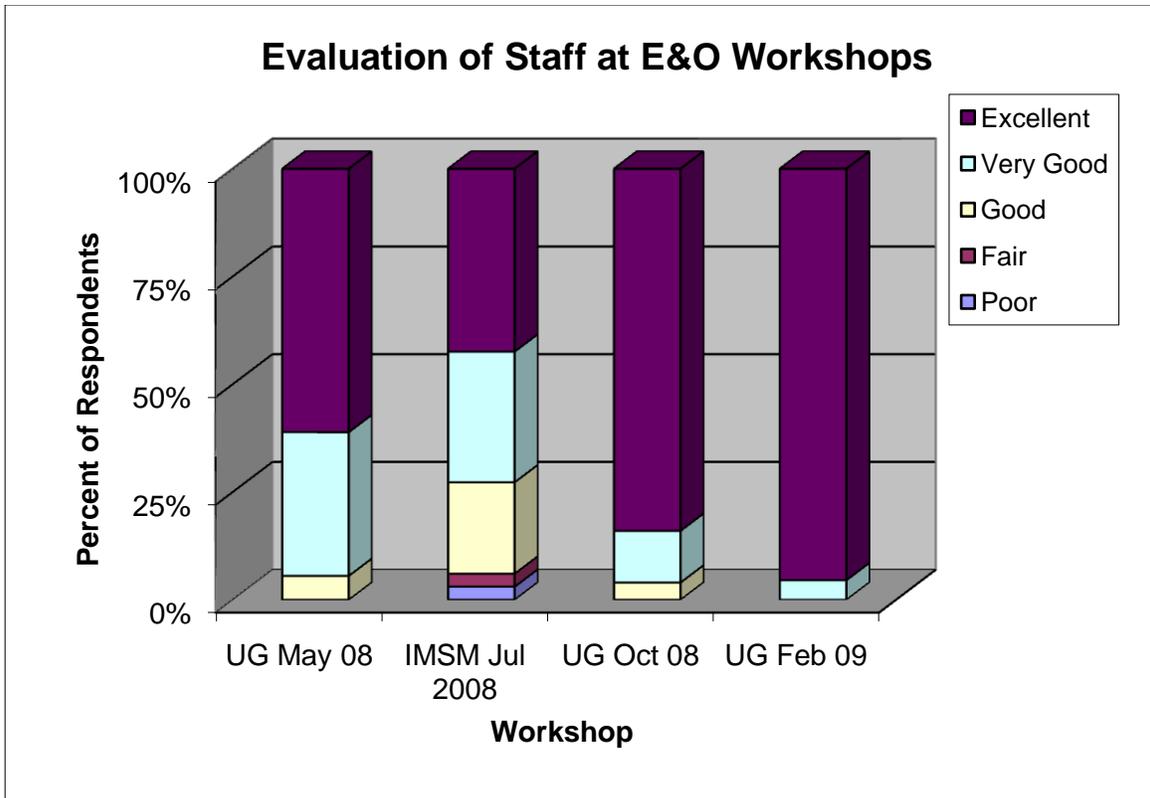
F.2.2 E&O Workshops (4 Events)

- Two-Day Undergraduate Workshop May 2008 (UG May 08)
- Industrial Mathematical, Statistical Modeling Workshop July 2007 (IMSM)
- Two-Day Undergraduate Workshop October 2008 (UG Oct 08)
- Two-Day Undergraduate Workshop February 2009 (UG Feb 09)

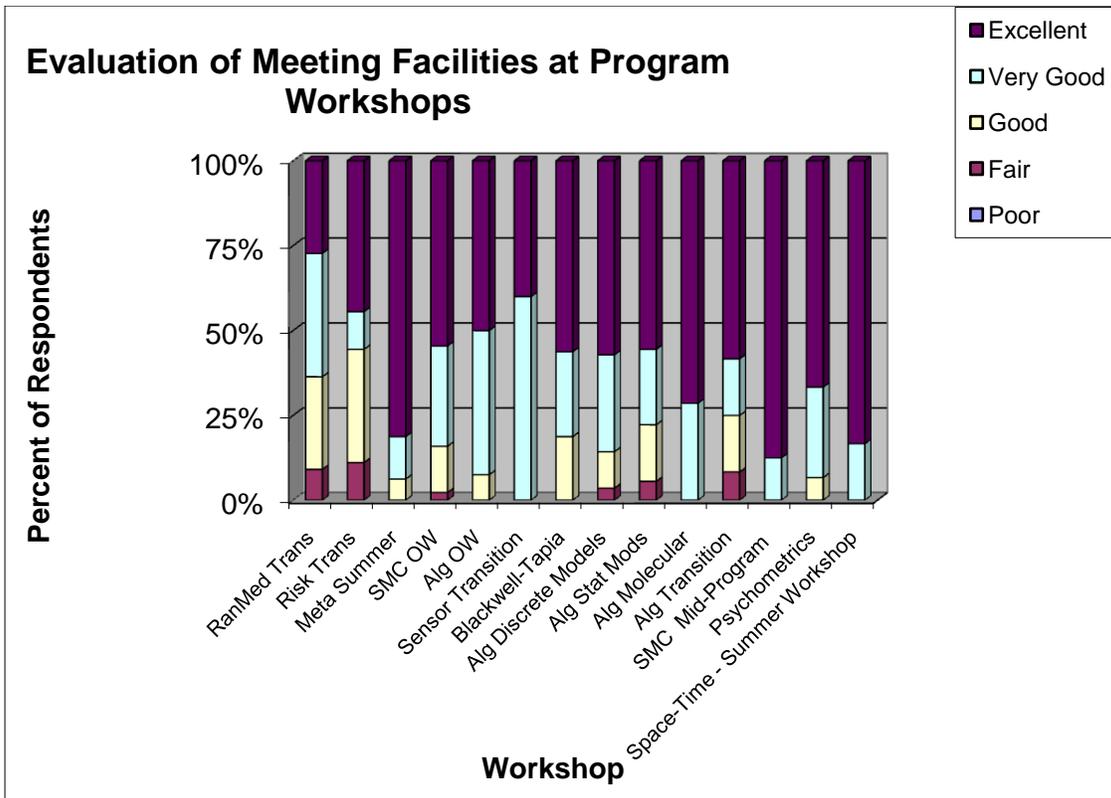


F.3 Evaluation of Staff

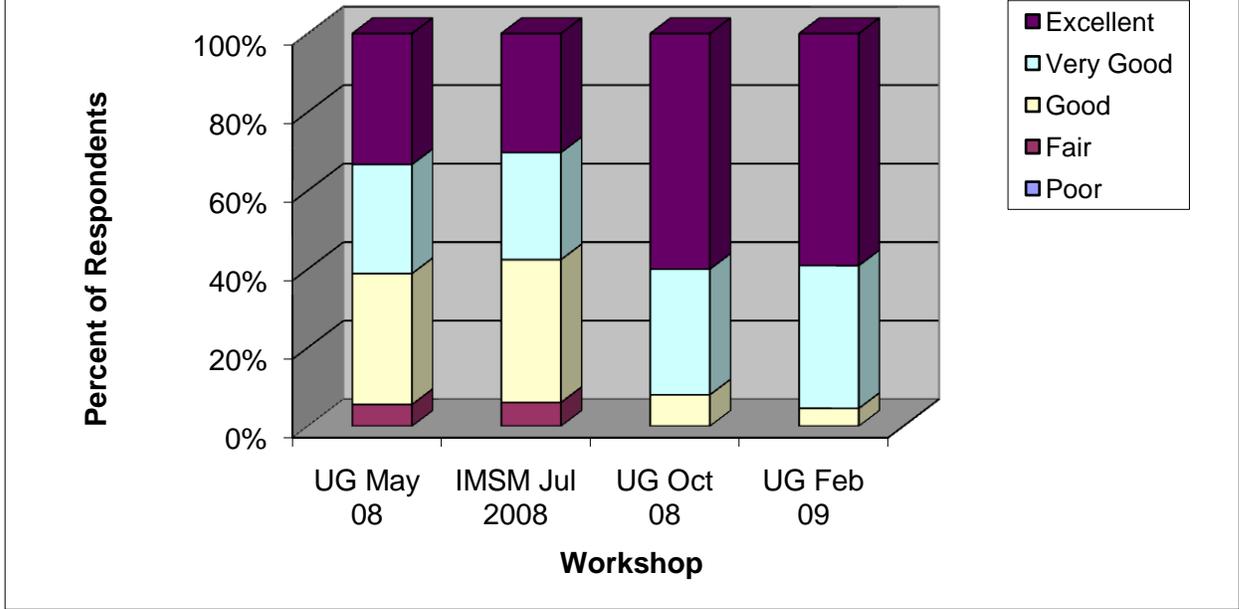




F.4 Evaluation of Meeting Room and Facilities

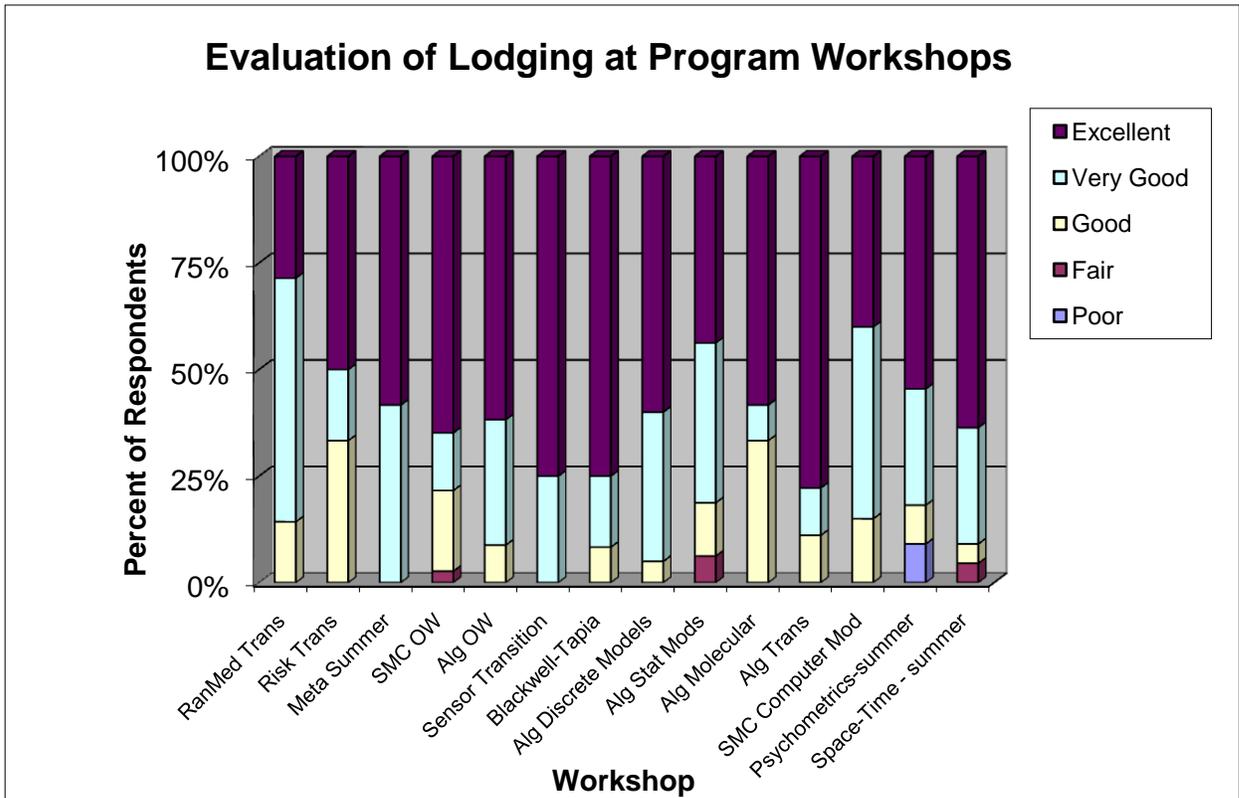


Evaluation of Meeting Facilities at E&O Workshops

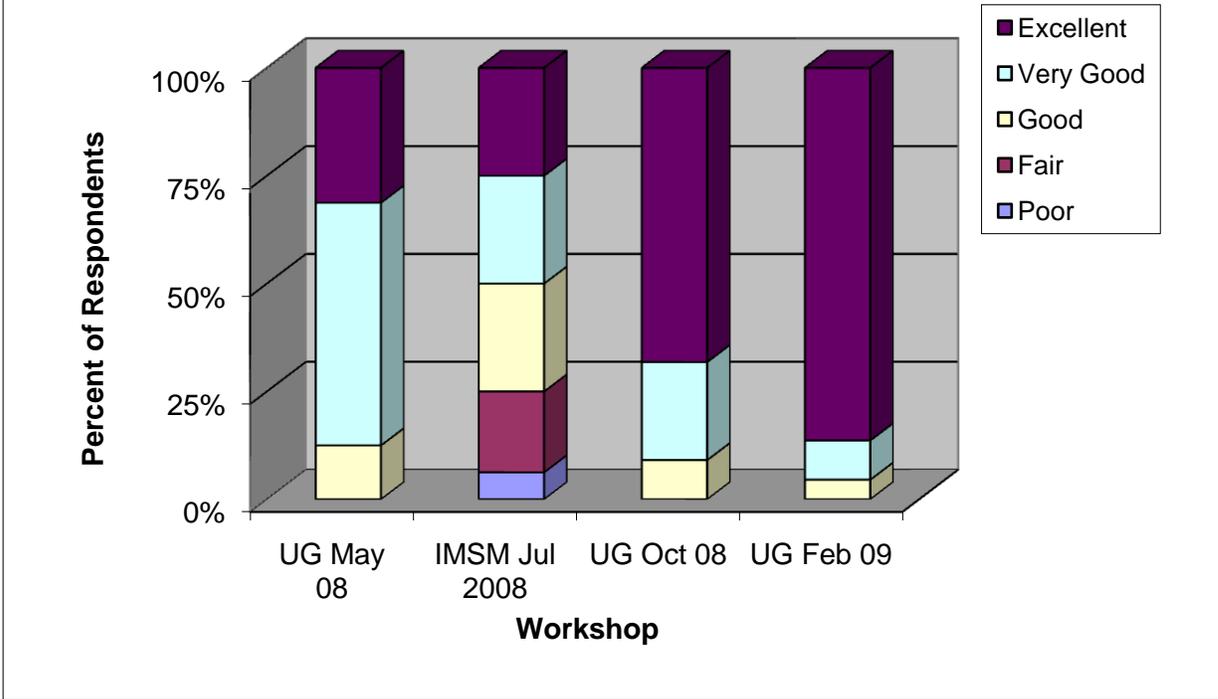


F.5 Evaluation of Lodging

Evaluation of Lodging at Program Workshops

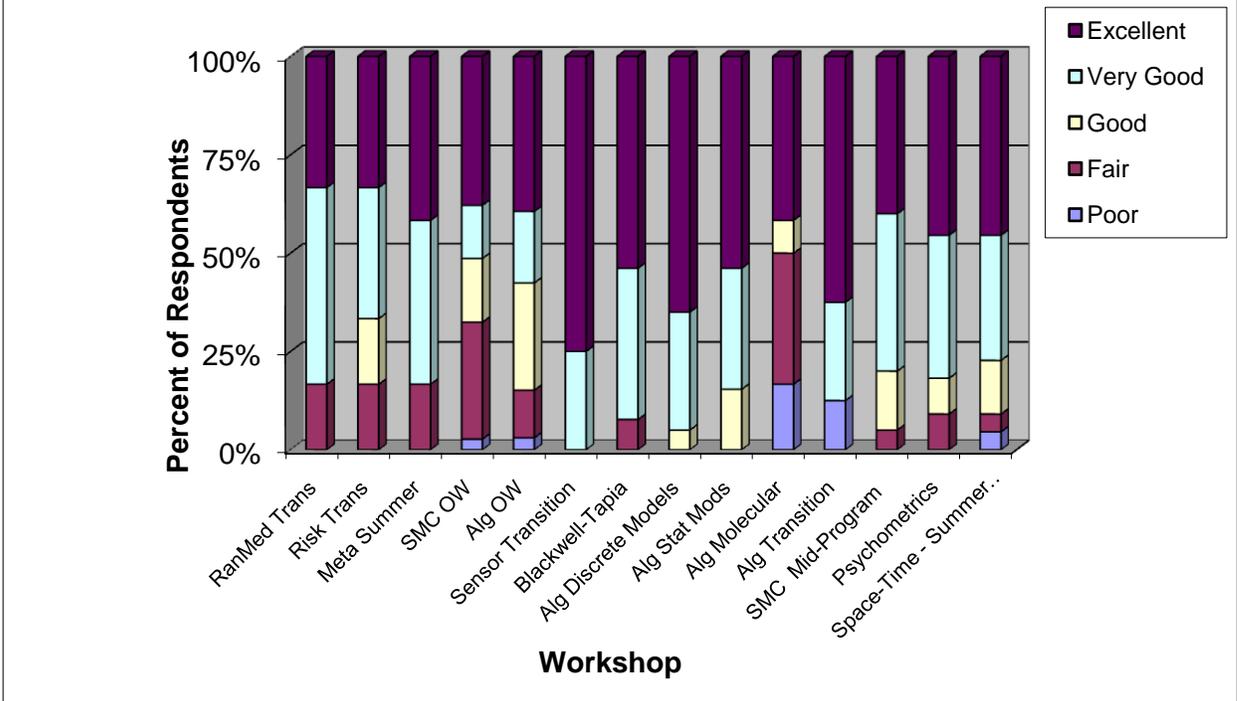


Evaluation of Lodging at E&O Workshops



F.6 Evaluation of Transportation

Evaluation of Transportation at Program Workshops



Evaluation of Transportation at E&O Workshops

