NSF Annual Report 2010-2011

Submitted to the National Science Foundation

Table of Contents

NSF Annual Progress Report for 2010-2011

As outlined in the terms of grant DMS-0635449, the following is the Annual Progress Report for the Statistical and Applied Mathematical Sciences Institute (SAMSI), for the period August 1, 2010 – July 31, 2011. Past activities that concluded during this period and future activities of SAMSI are also discussed.

0. Executive Summary

A. Outline of Activities and Initiatives

1. Programs for 2010-2011

- Analysis of Object Data
  - Opening Workshop & Tutorials (9/12/10 – 9/15/10)
  - Interface Functional & Longitudinal Data Analysis (11/8/10 – 11/10/10)
  - AOOD Meets Evolutionary Biology (4/30/11 – 5/2/11)
- Complex Networks
  - Opening Workshops & Tutorials (10/20/10 – 10/22/10)
  - Dynamics of Networks (1/10/11 – 1/12/11)
  - Pedestrian Traffic Flow (2/14/11 – 2/16/11)
  - Dynamics on Networks (3/21/11 – 3/23/11)
- Summer Program: SAMSI/Sandia Summer School of Uncertainty Quantification

Education and Outreach

- 2-Day Workshop for Undergraduates (10/29/10-10/30/10)
- 2-Day Workshop for Undergraduates (2/25/11-2/26/11)
- Graduate Student Probability Workshop (4/29/11 – 5/1/11)
- Interdisciplinary Workshop for Undergraduates (5/16/11-5/20/11)
- The Industrial Mathematical and Statistical Modeling Workshop for Graduate Students (7/7/11-7/15/11)
- Graduate Courses at SAMSI
  - Complex Networks, Fall 2010
  - Analysis of Object Data, Spring 2010
  - Analysis of Object Data, Spring 2011

2. Programs for 2011-2012

- Uncertainty Quantification: Climate Change Workshop
  - Opening Workshop & Tutorials (8/29/11 - 8/31/11)
  - UQ: Observations Workshop (1/17/12 – 1/19/12)
- Uncertainty Quantification: Methodology
  - Opening Workshop & Tutorials (9/7/11 – 9/10/11)

- High-Dimensional Approximation for Uncertainty Quantification (11/10/11)
- UQ: Simulation of Rare Events (2/13/12 – 2/14/12)
- UQ: Models with Complex & Uncertain Domains (3/22/12 – 3/23/12)
- UQ: Uncertain Quantification for High-Performance Computing (5/2/12 – 5/4/12)

- Uncertainty Quantification: Engineering & Renewable Energy
  - Opening Workshop & Tutorials (9/19/11 – 9/21/11)
  - Scientific Problems for the Smart Grid (10/3/11 – 10/5/11)
- Uncertainty Quantification: Geosciences Applications
  - Opening Workshop & Tutorials (9/21/11 – 9/23/11)
- Summer Program: Nonlocal Continuum Models for Diffusion, Mechanics, and other Applications (6/24/12 – 6/29/12)

Education and Outreach

- 2-Day Workshop for Undergraduates (10/28/11-10/29/11)
- 2-Day Workshop for Undergraduates (2/24/12-2/25/12)
- Interdisciplinary Workshop for Undergraduates (5/14/12-5/18/12)
- The Industrial Mathematical and Statistical Modeling Workshop for Graduate Students (7/16/12-7/24/12)
- Graduate Courses at SAMSI
  - Numerical Methods for Uncertain Quantification, Fall 2011
  - Numerical Methods for Uncertain Quantification – Part 2, Spring 2012

3. Programs for 2012-2013

- Program on Data-Driven Decisions in Healthcare
- Program on Statistical and Computational Methodology for Massive Datasets
- Summer Program: Nonlocal Continuum Models for Diffusion, mechanics, and other Applications, (6/25/12 – 6/29/12)
- Summer Program: Computational Advertising (8/6/12 – 8/17/12)

Education and Outreach

- 2-Day Workshop for Undergraduates 10/12
- 2-Day Workshop for Undergraduates 2/13
- Interdisciplinary Workshop for Undergraduates 5/13
- The Industrial Mathematical and Statistical Modeling Workshop for Graduate Students 7/13
- Graduate Courses at SAMSI
  - Operations Research Methods in Healthcare, Fall 2012
  - Computational and Inferential Methods for High Dimensions and Massive Datasets, Fall 2012

C. Directorate's Summary of Challenges and Responses

SAMSI's activities in 2010-11 included two year-long research programs that led to significant new research developments in statistics, probability and applied mathematics; as well as interdisciplinary research in such diverse areas as evolutionary biology, brain imaging and traffic flow. SAMSI's impact on the research community is felt in many ways: on the careers of researchers at all levels who visit SAMSI or participate in working groups; on young researchers who attend SAMSI as postdocs and graduate fellows, on graduate students through the IMSM workshop, and on undergraduates through attendance at one the three undergraduate workshops held each year. This year, for the first time, participation in the week-long May undergraduate workshop was extended also to teachers from undergraduate colleges. No summer program was held in 2011, but a summer school (aimed primarily at graduate students) was held at Sandia National Laboratory as a precursor to the 2011-12 Uncertainty Quantification program.

Changes in the Directorate. At the end of 2010, Nell Sedransk, who had served as Associate Director of SAMSI as well as Associate Director of NISS, moved to Washington as part of a planned relocation of her NISS position. As a result, she was replaced as Associate Director of SAMSI by Alan Karr, Director of NISS.

In May, 2011, Rick Durrett, who had taken up the position of Associate Director of SAMSI after moving from Cornell to Duke in the summer of 2010, resigned from that position and was replaced by Ezra Miller, also of the Mathematics Department at Duke. Ezra will serve out the remainder of Rick's term, through the summer of 2013.

Meanwhile Pierre Gremaud, whose position as Deputy Director had originally been scheduled to end in the summer of 2011, agreed to stay on for an additional year, while the Governing Board approved the search for an external Deputy Director to take over the position from July 2012. In anticipation of this change, the Governing Board also approved the appointment of a new Associate Director from NCSU. In July 2011, Ilse Ipsen, Professor of Mathematics at NCSU, took up this position. The new Deputy Director who will take over the position in July 2012 is Snehalata Huzurbazar, who will take a two-year leave from the University of Wyoming.

Other Personnel Changes. On the Governing Board, Robert Calderbank, Dean of Natural Sciences at Duke, replaced John Simon as the Duke representative following John's new appointment as Provost at the University of Virginia. Jim Landwehr came to the end of his term as chair of the NISS Board of Trustees and was replaced by Susan Ellenberg of the University of Pennsylvania; as a result, Susan also becomes the NISS representative on the Governing Board. Don Estep, representing SIAM, has agreed to serve a second three-year term (up to the end of 2014).

On the National Advisory Committee, Carlos Castillo-Chavez and Bin Yu each came to the end of their terms as co-chair and retired from the NAC after making many strong contributions over the years. They were replaced as co-chair by Mac Hyman of Tulane

University and Susan Murphy of the University of Michigan. Also joining the NAC were Adrian Raftery of the University of Washington and Habib Najm of Sandia National Lab.

As reported last year, SAMSI also has a Local Department Liaisons committee which includes one representative from each department in one of the partner universities that is not otherwise represented on the Directorate. The current members of this committee are Howard Bondell (NCSU, Statistics), Greg Forest (UNC, Mathematics), Stephen George (Duke, Biostatistics and Bioinformatics), Robert Wolpert (Duke, Statistics) and Donglin Zeng (UNC, Biostatistics).

Long-term Participants. The following table shows the numbers of postdocs, graduate fellows, faculty fellows and visitors to SAMSI in 2010-11:

	Postdocs	Postdoc Assoc	Grad Local	Grad Visiting	Visitors	Faculty Fellows	New Researchers	Total
Year	8	5	12		3	12	1	41
Fall				1	39		1	41
Spring				1	27			28

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.

Communications and Marketing

A tri-fold marketing brochure was produced that gives a brief explanation of what SAMSI is and the programs that are offered. It is used as part of SAMSI's overall marketing efforts and given to people at who visit SAMSI's exhibit at the math and statistics conferences.

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. The posters are distributed to colleges and universities across the United States.

SAMSI's newsletter was mailed out quarterly to a list of about 460 colleges and universities. Another 2,643 email contacts receive the email edition of the newsletter as well using the Bronto e-mail marketing system. Roughly 22 messages were sent using the Bronto system to communicate with our target market. In addition to the newsletter, these messages included announcements of upcoming workshops and programs, and any special announcements the organization wanted to convey.

Two press releases were written and sent to local and national media. The first was about the hiring of additional postdoctoral fellows and was used as part of a larger effort by all 13 mathematical institutes to publicize the new fellows that were able to be hired thanks to an increase in funds by the NSF. This release went out on May 11, 2009 and was picked up by 30 media outlets including dBusinessNews, Carolina Newswire, Business Leader and others. This release was picked up mainly by local media, such as the Durham Herald Sun and the Duke Chronicle, but also was run in the ISM Bulletin.

SAMSI's efforts in the social media realm were recognized by the North Carolina chapter of the Public Relations Society of America as the best social media campaign. SAMSI is using three main areas of social media, which include LinkedIn, Facebook and Twitter to communicate with its target market. See below for statistics on the reach that SAMSI has had during the Fiscal 09-10 year.

A "Welcome to SAMSI" brochure was produced to give to new visitors to the area. The brochure includes information such as area restaurants, places to live, places to shop, recreation and places to see and a guide for parents. It was given to each new visitor to SAMSI during the 2009-10 year.

SAMSI Twitter Followers

Date	Number of Followers
9/13/10	695
11/4/10	743
1/14/11	828
5/31/11	1000
6/12/11	1020

Facebook Friends

Date	Number of Friends	Number of Visits That Week
9/13/10	139	50
10/25/10	154	75
11/22/10	168	74
1/17/11	172	65
3/14/11	187	40
4/26/11	193	87
5/24/11	210	115
7/26/11	233	96

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

D.1 Research

During 2010-2011, SAMSI organized two year-long research programs, Analysis of Object Oriented Data (AOOD) and Complex Networks (CN). The AOOD program was focused on the development of statistical methods for data objects that go beyond numbers and vectors, such as curves and images for which there is by now a well developed theory, and also objects such as trees and graphs for which a systematic statistical theory is still only just being developed. A particular achievement of the program was to recognize that many of the objects of interest can be characterized as structured manifolds and this insight led to new mathematics as well as new statistical methodology. With eight very active working groups, this was also one of the largest programs SAMSI has organized to date. The CN program was focused more on developments in probability and applied mathematics, though there were statistical contributions here to problems as diverse as traffic flow and power grids. Particularly noteworthy were the two working groups on dynamically evolving networks, one focused on cases where the network structure is static but a stochastic process is evolving on its nodes (dynamics ON networks), the other where the network itself is evolving in time (dynamics OF networks).

In the following descriptions, we describe each of these programs in some detail. The full reports from both programs are included elsewhere in this report.

D.1.1 Analysis of Object Oriented Data (AOOD)

In recent years the subject of statistics has expanded to encompass many types of "objects" beyond the numbers and vectors of elementary statistical methods. One topic that is already well established is *functional data analysis* (FDA), in which random curves are considered as individual data points in a broad framework of statistical analysis techniques. In recent years, however, this topic has expanded to include data on various kinds of manifolds (e.g. directional data), image analysis, and even data structures such as trees and graphs which cannot be described at all by standard Euclidean geometry. The SAMSI program on AOOD evolved around five broad themes: FDA; time dynamics data objects (for example, curves evolving in time according to differential equations); data on surfaces and manifolds; image analysis (including applications to fMRI and brain imaging); and the statistical analysis of trees and graphs.

The program leaders were Hans-Georg Müller and Jane-Ling Wang, both of the University of California at Davis; Ian Dryden of the University of South Carolina; James O. Ramsay of McGill University; and Steve Marron of the University of North Carolina. Marron also served as local coordinator for the program; the directorate liaison was Jim Berger (Richard Smith after July 2010) and the National Advisory Committee liaison was Jianqing Fan of Princeton.

Workshops

The Opening Workshop (September 12-15) was one of SAMSI's most highly subscribed ever, and featured opening-day tutorial lectures on each of the five main themes, followed by five half-day sessions, one on each of the five main themes of the program. The Opening Workshop also featured a couple of "two-minute madness" sessions in which speakers were invited to speak for just two minutes with at most two or three slides. This proved a popular way of allowing the workshop to feature a larger number of speakers without detracting from the main invited sessions. The workshop also featured an evening poster session and, on Wednesday afternoon, the formation of working groups.

Three other workshops also took place within the program:

1. "The Interface of Functional and Longitudinal Data Analysis" took place on Nov 8-10, at SAMSI, organized by Marie Davidian of NC State, Fang Yao of the University of Toronto, and Hans-Georg Müller of UC Davis. Longitudinal Data Analysis is very well established in biostatistics, referring to studies in which a group of individuals is followed through time with data collected from each individual at a number of time points. This workshop was organized to explore the concept that by embedding longitudinal data analysis within a FDA perspective, one could significantly expand the scope of the analysis, for example though richer classes of dependence structures, more detailed analysis of dynamics, and allowing the data to determine the most suitable model. The workshop featured 14 invited talks, as well as a poster session and five discussions, where the interactions with working group activities could be explored at greater length.

2. "AOOD Meets Evolutionary Biology" was held at SAMSI, April 30-May 2, with the objective of connecting the AOOD themes with function value traits in evolutionary biology. There is already a well-established link with FDA methods in this field of application, but the workshop extended this to many other forms of data objects, examples being fly wing shapes, caterpillar growth trajectories and viral phylogenetic trees. The workshop began with a "two-minute madness" session to allow all the participants to introduce their work; there followed a total of eight talks from biologists interspersed with five talks by AOOD participants.

3. The Transition Workshop from June 9-11 was the last formally organized activity of the program, and once again featured a separate session on each of the five main themes of the program, including invited speakers and discussants in each. One of the things to emerge from this workshop was the extent of interconnectedness among the five themes. For example, curve, surface and volume registration problems in FDA and brain imaging are essentially shape analysis under constraints. Another example is viewing FDA and image data as dynamical systems evolving under ordinary or partial differential equations.

Courses

Two graduate courses were taught in connection with this program, entitled "Analysis of Object Data I" (fall) and "Analysis of Object Data II" (spring). The first course featured all five program leaders as instructors and covered FDA in various guises, as well as a section on shape analysis. The second course was split between Ramsay and Marron and concentrated on three more detailed topics: dynamical systems, manifold data and tree-structured data.

Undergraduate Workshop

There was also a two-day undergraduate workshop in February on AOOD, featuring lectures and demonstrations from program leaders, other local faculty members and postdocs.

Working Groups

This was an exceptionally productive program: the final report produced by the program leaders listed 39 papers published or accepted for publication; 43 submitted; and 84 either in preprint form or in preparation. These numbers clearly show the high level of research commitment shown by the large number of individuals who were involved in one or more of the working groups.

The working groups themselves were initially organized around the five main themes of the program, but several split into subgroups that in effect became independent working groups. The program report (Section I.E of this report) details the interrelations among the working groups as well as giving very detailed accounts of their activities; for the present account, we list them as separate working groups, and briefly outline the main achievements of each.

Working Group on Statistical Inference for Functional Data

This was the first of two working groups on FDA, led by Feng Yao (Toronto) and Hao Helen Zhang (NC State). Functional data provide challenges for statistical inference because of the high dimension and often complex structure of the datasets involved. This working group focused on regression and classification approaches with particular attention to variable selection, model choice, sparse estimation, experimental design and robustness. The group involved many new or junior researchers and produced many papers. Specific research achievements included:

- a new class of semiparametric functional regression models for jointly modeling functional and non-functional predictors;
- a regularization framework that can automatically select important predictors for nonparametric functional regression;

- residence at SAMSI allowed Graciela Boente and Jane-Ling Wang to complete a long-standing project combining robust projection pursuit with various smoothing methods in the setting functional data analysis.

Working Group on Hierarchical Modeling

This working group, led by Jeff Morris of the M.D. Anderson Cancer Center, provided a Bayesian counterpoint to the work of the previous group.

Hierarchical modeling allows flexible unified models that account for multiple sources of variability and structure. Past work on hierarchical models for FDA has used both Bayesian and frequentist approaches, but the present group eventually focused on three topics with a Bayesian focus: Bayesian nonparametric regression applied to functional data; object regression; and hierarchical models for images. Results of the working group led to numerous publications in statistical journals, presentations at several workshops, and a topic contributed paper session at the Joint Statistical Meetings.

An example of the sort of applied problem to which these techniques have been applied is the paper by Lim and Dass (2011), about fingerprint matching. In this paper, Bayesian analysis was applied to fit a marked point process model for minutiae – small imperfections that are used as a diagnostic in comparing fingerprints. The result is a probabilistic measure for deciding when two fingerprints are from the same person, an important issue in forensic science. Other papers written by members of the group, such as Morris (2011) or Zhu et al. (2011), emphasized applications in proteomics such as mass spectrometry and 2D gel electrophoresis.

Working Group on Dynamics and Inference

This group was concerned with statistical methods for nonlinear dynamical systems. Such system pose special challenges for statistical inference when expressed in terms of systems of equations that do not admit an analytic solution. Classical approaches using iterative numerical solutions are gradually being replaced by methods using the equations themselves to define regularization penalties, relying on relaxation strategies similar to those being developed in other areas of numerical analysis. Wider interest in these problems has been stimulated by genomic and proteomic data, dynamical models for diseases, and the explosive growth of real-time data monitoring. The working group brought together a diverse group of statisticians who have worked on these problems from different perspectives, as well as researchers new to the area. As a result, much of the time was spent on reviews and comparisons of different approaches.

During the program, Jim Ramsay translated the CollocInfer package from R to Matlab; he and Giles Hooker continue to develop the CoolcInfer package. Jim also presented eight 2-hour classes as part of the spring graduate course. As with several other working groups, numerous papers were either written during the program or are in preparation.

Working Group on Data Analysis on Sample Spaces with Manifold Stratification

This working group, one of three within the general area of Shapes and Manifolds, was led by Ezra Miller (Mathematics, Duke) and Vic Patrangenaru (Statistics, Florida State). It was one of the most active of all the working groups and clearly illustrates the advantages of having researchers of different expertise working together. The group arose from discussions at the opening workshop among Miller, Patrangenaru and Stephan Huckeman (Institute for Mathematical Stochastics, Göttingen, Germany), that led to the identification of a common class of mathematical objects, known as *stratified sample spaces*, that could be used as the basis for mathematical, probabilistic and statistical research on object data. The mathematical part of the research led to a multi-authored paper with the novel title *Sticky Central Limit Theorems on Open Books* (about probability limit theorems on a certain class of metric spaces), but the main focus of the research was statistical, including nonparametric statistics for various kinds of stratified spaces, contributions to efficient computation, and an emerging theory of principal components analysis on manifolds. The group contributed to the Evolutionary Biology workshop that was part of the SAMSI program and at the transition workshop, as well as at least seven papers at the 2011 Joint Statistical Meetings. Group members were also instrumental in proposing and organizing a workshop on *Statistics, Geometry, and Combinatorics on Stratified Spaces Arising from Biological Problems*, held at the Mathematical Biosciences Institute (May 21-25, 2012).

Working Group on Metrics on Shape Spaces

The second of three groups on Shapes and Manifolds was led by John Kent (Leeds, UK) and Anuj Srivastava (University of Florida). This was another very diverse group including experts in engineering, image analysis, pure and applied mathematics, and statistics. A "shape" is mathematically defined as a geometric object that is invariant under certain classes of transformations; shape spaces have a manifold structure that may be endowed with various metrics. A key issue for this group was the influence of the choice of metric for studying different features of shape spaces. The discussions covered mathematical themes such as differential geometry and curve fitting, through to applications to protein shape, computer vision and neuroscience spike trains, among several others. Outcomes includes contributions to functional data analysis, including a model for a video sequence of dancing figure; projective geometry; directional data analysis (including two papers stimulated by Professor Kanti Mardia's visit to a biochemistry lab at Duke); and feature selection in images.

Working Group on Geometric Correspondence

The third working group on Shapes and Manifolds was led by Stephan Huckemann (Göttingen) and Ross Whitaker (University of Utah). The focus here was on statistical analysis of geometrical objects. The group eventully split into six subgroups studying subjects such as measuring landmark importance, modeling locus and shape variation, and shape classification. Outcomes included several new papers, and contributions to the transition workshop and a Banff International Research Station workshop on *Geometry for Anatomy*.

Working Group on Brain Imaging

This workshop was chaired by John Aston of the University of Warwick (UK). The initial intention was to focus on the link between functional data analysis and brain imaging, but that quickly got expanded into a much broader set of statistical issues in brain imaging datasets. The three main themes settled down to be deconvolution and design-free analysis, spatial and temporal modeling, and hierarchical analysis.

Among the specific outcomes of this working group were a paper on functional correlation measures for resting state FMRI data from UC Davis; a collaboration with a Psychology and Neuroscience professor at Duke that led to a grant submission on functional imaging genetic data anlaysis; and a paper on voxel by voxel nonparametric inference for brain images.

Working Group on Tree Structured Data Objects

The final working group of this program was led by Steve Marron (UNC), and focused on three main approaches: combinatoric, phylogenetic trees and Dyck Path methods. The Dyck Path work was particularly important for the dissertation of Dan Shen, one of Marron's PhD students at UNC.

D.1.2 Complex Networks (CN)

Network Science is an emerging area of the mathematical sciences that includes contributions in applied mathematics, statistics, computer science and statistical physics, as well as applications in biology, engineering, computer science and the social sciences. This program featured several interconnected research foci identifying key mathematical and statistical issues underlying the different approaches to the field.

The program leaders were Eric Kolaczyk (Boston University) and Alex Vespignani (Indiana University); local scientific coordinators were Alun Lloyd (NCSU) and Peter Mucha (UNC); directorate liaisons were Rick Durrett (Duke) and Pierre Gremaud (NCSU); the National Advisory Committee liaison was Bin Yu of Berkeley.

In contrast to AOOD, the Complex Networks program was more focused on applied mathematical and probabilistic analysis including classical discrete probability models such as percolation, the voter model and epidemic processes. Another subtheme was *agent-based models,* which are models for social and economic behavior in which the small-scale dynamics of individuals are linked to the large-scale dynamics of systems, usually through simulation. There were also statistical contributions, especially in the Sampling/Modeling/Inference working group, and applications to traffic flow and power grids as well as epidemiology and disease dynamics.

Workshops

The Opening Workshop was held at the Radisson Hotel, August 29-September 1, with around 150 participants. Tutorial lectures were given by Kolaczyk, Vespignani, Durrent and Michael Mahoney (Stanford). The workshop focused on five complementary themes which served as defining themes for the whole program:

- Network Sampling and Inference
- Dynamic Networks
- Percolation and Diffusion on Networks
- Spectral Analysis and Geometric Algorithms
- Biological Applications of Networks

Each of these was the subject of a session featuring at least three invited speakers, supplemented by a panel discussion in several cases, and there were also presentations by beginning researchers.

The remaining workshops of the program were as follows:

- A workshop on Complex Networks Modeling was held at SAMSI, October 20-22, 2010, organized by David Banks of Duke and Eric Kolaczyk of Boston University. The focus of this workshop was statistical analysis of network data, and especially the sampling, modeling and inference of networks. The workshop included twelve invited talks and a poster reception.

- A workshop on Dynamics OF Networks took place at SAMSI, January 10-12, 2011. There is a growing mathematical and statistical theory about networks that evolve in time. One example is the network of contacts for the spread of an infectious disease. Other examples arise in communications networks and in political networks. Compared with static networks, statistical and computation tools for dynamically evolving networks are relatively few in number, and the workshop served to highlight the considerable potential of this topic. The workshop featured a program of twelve invited talks, several discussions of working group issues, and a poster reception.

- A workshop on Pedestrian Traffic Flow was held at SAMSI, February 14-16, 2011, organized by Alina Chertock (NCSU), Pierre Degond (CNRS, France) and Alexander Kurganov (Tulane). Studies of human crowds have demonstrated complex behavior (such as the spontaneous formation of lines) but have so far lacked both adequate data and mathematical models to understand their dynamics. However, through individual (agent) based models, a framework exists to study of medium- and large-scale behavior from assumptions at the individual level. The main goal of this workshop was to study the relationship among the three scales of modeling, in particular showing how PDEs to describe macroscale behavior are influenced by assumptions at the microscopic level. The workshop featured a program of seven speakers, plus discussions.

- A workshop on Dynamics ON Networks was held at SAMSI, March 21-23, 2011, organized by Rick Durrett, Alun Lloyd, Peter Mucha and Alex Vespignani. Unlike the workshop on Dynamics OF Networks, this workshop studied cases where the network itself is static, but there are various processes taking place on the nodes of the network – examples are the spread of opinions, and epidemics. The key focus for this workshop was how the structure of the network affects the processes taking place on the network. The main objective was to bring long distance collaborators to SAMSI and others whose work was important to that of the working group. The workshop featured a program of 16 invited speakers, discussion sessions, and a poster reception.

The final workshop of the program was the Transition Workshop, held at SAMSI on June 6-7, organized by Eric Kolaczyk, Michael Mahoney, Mason Porter (Oxford University) and Alex Vespignani. This workshop featured half-day sessions by each of the five working groups. It included twelve invited talks, discussion sessions and a poster reception.

Courses

A one-semester course for graduate students on *Complex Networks: Theory and Applications* was held in Fall 2010. Ten students took the course for credit and another ten individuals attended on a regular basis. The course focused on mathematical and statistical analysis and modeling of both static and dynamic networks, with biological, technological and social science applications. The lecturers were Eric Kolaczyk, Peter Mucha, James Moody (Department of Sociology, Duke University), Alun Lloyd and Rick Durrett.

Undergraduate Workshop

A two-day undergraduate workshop on Complex Networks was held at SAMSI on October 29-30, 2010. Ten lectures were given, including talks on Matlab and R and a talk about career options.

Working Groups

Since the program naturally organized itself around five main themes, each of those themes became the focus of a working group. In the following, we present a brief discussion of the aims and outcomes of each of those working groups.

Working Group on Sampling/Modeling/Inference

This group, let by Eric Kolaczyk and with 15 members, focused on statistical inference problems connected with networks, primarily static network but also with some consideration of dynamically evolving networks. The group identified six projects with a diverse focus, of which we briefly describe two. David Banks, Bruce Rogers and Cosma Shalizi considered parameter estimation for agent-based models, which study the behavior of large systems by simulating the dynamics of individuals (or agents) in the system. Such models are expensive to rerun at every possible combination of parameter values; but the calculations may be simplified by developing an approximation to the true system known as an emulator, thus borrowing ideas from the Bayesian analysis of computer models. A second project led to a paper by Eric Kolaczyk and Pavel Krivitsky on effective sample size – a quantity that characterizes how the performance of estimators scales with the size of the network. Yet another project resulted in Edo Airoldi and David Banks submitting an NSF proposal on minimum description length inference for latent space models.

Working Group on Dynamics ON Networks

This group was led by David Banks, James Moody and Peter Mucha. The topics overlapped substantially with those of other working groups, and this was the primary motivation for the decision to merge this group with the OF group in the second semester. Meanwhile, three projects were developed as distinctive foci for this group. The first was the phenomenon of *explosive percolation*, characterized by the emergence of a giant component in a seemingly sharp transition as the network dynamics change. Much of the focus was on developing the mathematical machinery to produce rigorous results. A second project concerned *voter models* that incorporate two different processes, a rewiring dynamic that defines the evolution of the network, and a voter interaction process that would tend to push the network towards complete agreement within components. A key concept is that of a critical rewiring frequency parameter that controls the transition between regions in which one of the two processes is dominant over the other. A third project focused on the role of concurrent partners in the spread of a sexually transmitted disease, and focused on the utility of different network representations.

Working Group on Dynamics OF Networks

This group was led by Rick Durrett of Duke and Alun Lloyd of NCSU with 17 other members. The group worked on at least six different projects studying the large-scale behavior of networks on which different types of random process are taking place. Examples include: an epidemic problem involving two communities with different infection rates within and between the communities; voter models with nonlinear response functions; the Axelrods voter model, in which both the number of opinions per voter and the set of possible values of those opinions are large; and understanding cascades in interacting systems. Products of this group include two published papers in *PNAS*, and other papers forthcoming.

Working Group on Geometrical/Spectral Analysis

This group was led by Mauro Maggioni of Duke and Michael Mahoney of Stanford, with eight other participants. The focus was on spectral and multiscale techniques for graphs, and their geometric and algorithmic implications. Specific topics were: multiscale conductance and time series networks; soft clustering; tree decompositions; community detection; scan statistics and sparse regression; and epidemic/reactive networks.

Working Group on Modeling Flows

This group was led by Taufiquar Khan of Clemson, with nine other members, and was concerned with two areas of application of flows in networks: traffic flows and smart grids. The main conception behind the traffic flows building up large-scale dynamics (modeled with PDEs) from the small-scale dynamics of the system. The pedestrian flow workshop already mentioned (February 14-16, 2011) was closely related to this topic. The “smart grids” topic was concerned with mathematical and statistical techniques for developing an efficient and sustainable power network. Specific topics included reduced graph models for power networks, optimal load management in the network, and identification of critical paths. A related Smart Grids workshop was help in the Fall of 2011, as part of the Uncertainty Quantification program, and several papers are in preparation.

D.2 Human Resource Development

SAMSI’s impact on human resource development 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.

SAMSI’s postdoctoral fellows and associates again in 2010-11 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.

Of the eleven postdocs who were in residence during 2010-11, two are currently second-year SAMSI postdocs; four went on to other postdoc-level positions; two went on to visiting faculty positions; and three to tenure-track faculty positions.

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

The impact of new technology for remote participation in SAMSI working groups continues to be steady; 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.
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 63 research fellows (non-local visitors who played a major role in program activities) plus 5 new researcher fellows, 9 local faculty fellows, 13 postdoctoral fellows and associates, and 15 graduate fellows (including visiting graduate fellows). The research workshops associated with the major programs attracted a total of 790 participants; the education and outreach workshops a total of 139; the IMSM graduate research workshop had 44 participants; the planning workshop for Massive Datasets had 25, and the 2011 summer school on Uncertainty Quantification has 80 participants. Altogether, this makes nearly 1,200 individuals who participated in SAMSI in some form during the year.

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.

Pierre Gremaud serves 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 David Auckly (MSRI). The Institutes Diversity Coordination Committee has been working together to promote diversity in the Mathematical Sciences at national conferences and through other special events.

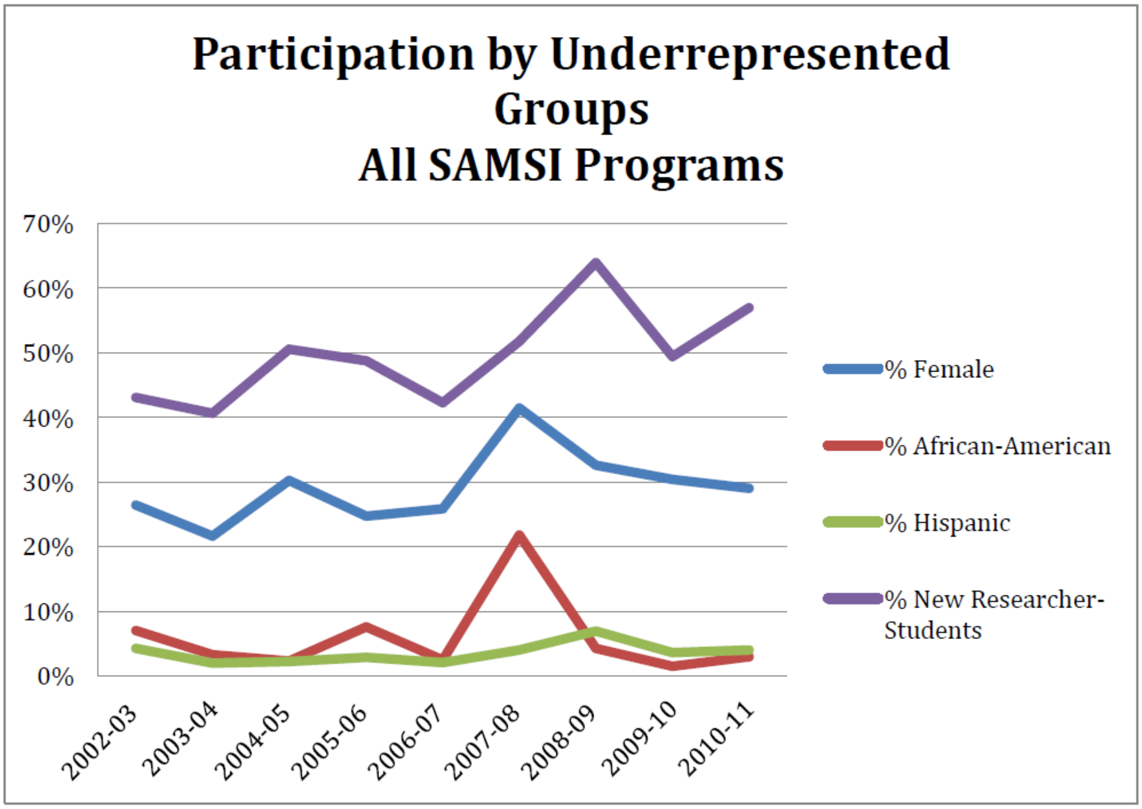
Together with the other NSF mathematical sciences institutes, SAMSI has recently secured supplemental NSF funding that will allow for the long term planning of an entire portfolio of activities including the Modern Math workshop series, the Workshop Celebrating Diversity series, the Minority Professional Development workshop, the Careers for Minorities workshops, the Careers for Women workshops, the Blackwell Tapia Conferences and some activities in conjunction with the Association for Women in Mathematics. SAMSI took part in the Modern Math program at the 2009 SACNAS National Convention in Dallas, TX, which included a presentation by one of the SAMSI postdoctoral fellows.

Minority Participation in SAMSI Programs

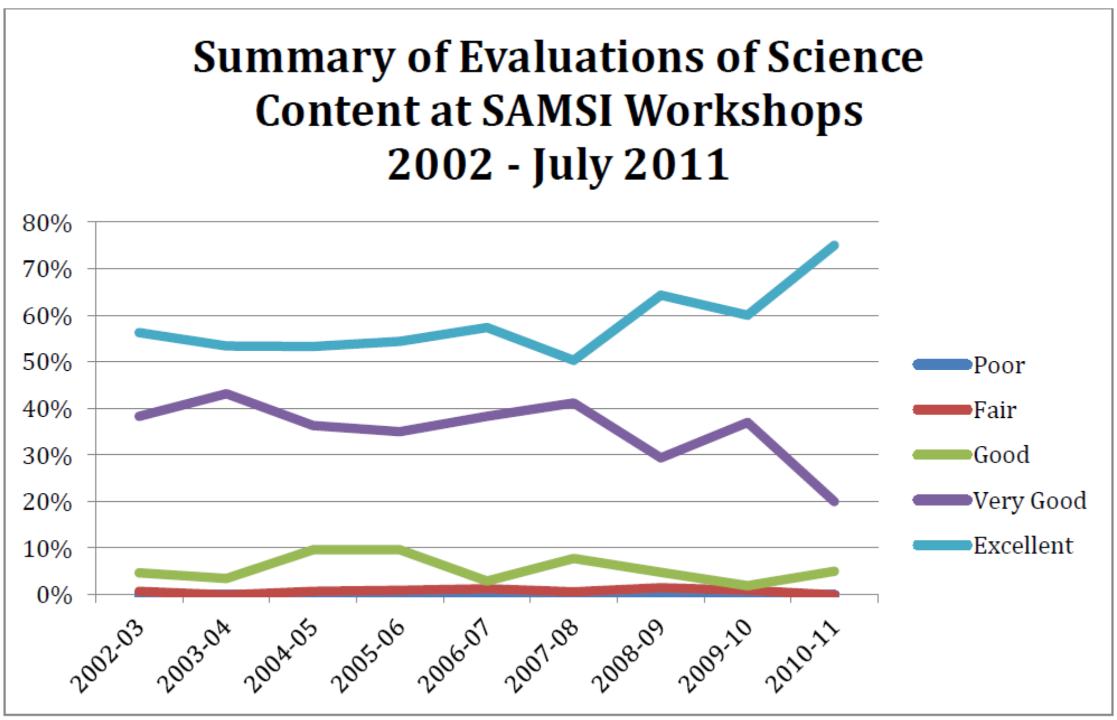
For the 2010-11 Research programs, of six postdocs hired, two are women and two are minorities. SAMSI creates additional possibilities for under-represented groups through its education and outreach activities.

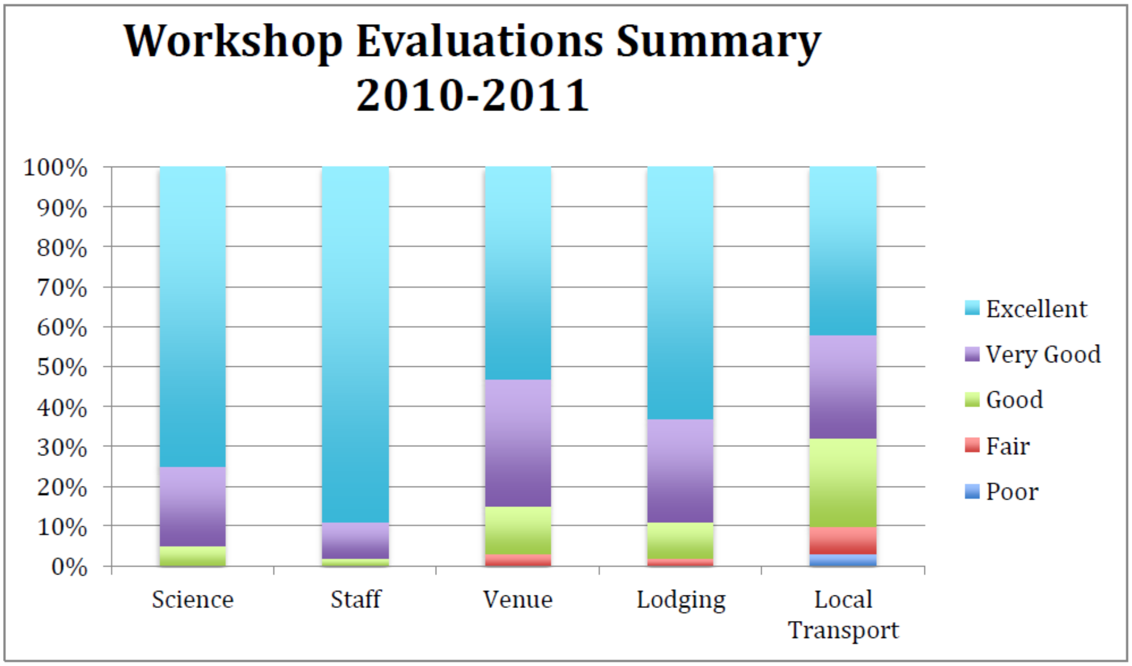
Overall Participations in Workshops by Underrepresented groups

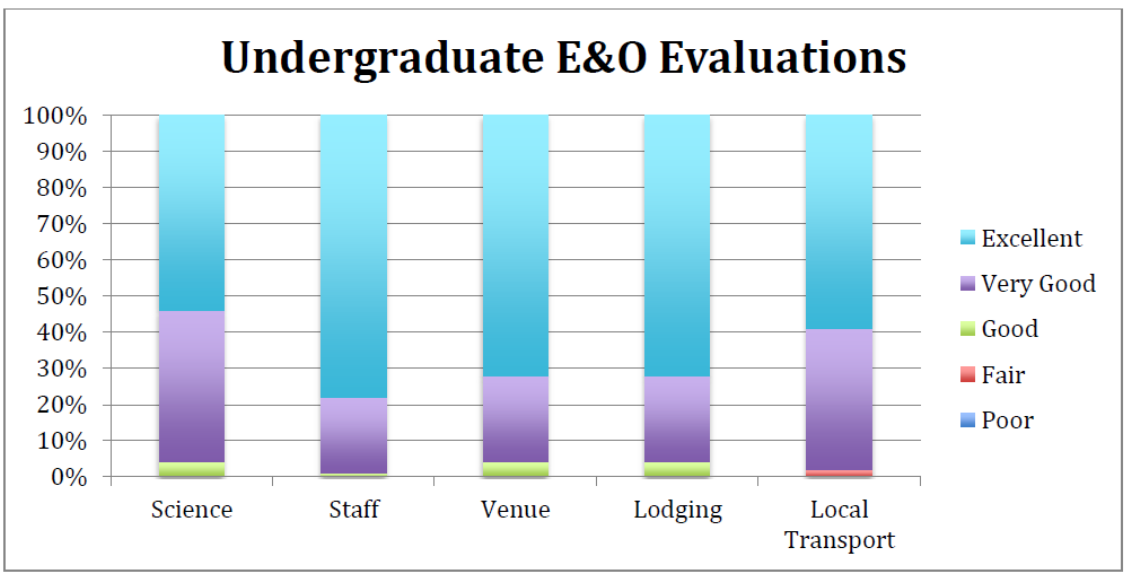
The following chart shows an overall summary of the participation by underrepresented groups at SAMSI events. The large spike in female and African-American participants in 2007-08 was partly due to the Infinite Possibilities conference, which focused specifically on African-American women.



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







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
2009-2010	190	107	45	342
2010-2011 (thru 7/31/11)	154	130	31	315
TOTAL	2,000	968	370	3,338
Percentage of all funded participants	60.0%	29.0%	11.1%	

- Includes all funded visitors and workshop participants. Total funded = 431 (316 + 115 people with no citizenship information.)

2010-2011 SAMSI Participation			
Funded Participants	**Home Institution by DMS Funding Level**		
	Top 50 DMS Funded	51-200 DMS Funded	Other
# of Institutions	45	50	63
# of People	133	98	85
% People	42.1%	31.0%	26.9%
All Participants			
	Top 50 DMS Funded	51-200 DMS Funded	Other
# of Institutions	46	59	93
# of People	194	131	133
% People	42.4%	28.6%	29.0%

D.3 Education and Outreach

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 Analysis of Object Data and Complex Networks. 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.
   a. 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.
   b. The workshop provided students with an intensive introduction to the synergy between applied mathematics and statistics in the context of physical applications.
   c. 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:
   a. Expose mathematics and statistics students to current research problems from government laboratories and industry which have deterministic and stochastic components;
   b. Expose students to a team approach to problem solving.
   c. For the 2010 workshop, research problems were presented by scientists from the National Center for Atmospheric Research, the Environmental Protection Agency, Sandia National Laboratory, MIT Lincoln Laboratory, Progress Energy and the Batelle Memorial Institute.. Each team gave a 30 minute oral presentation summarizing their results on the final day of the workshop and written reports were compiled as a SAMSI Technical Report (http://www.samsi.info/sites/default/files/IMSM_2010.pdf).

(iv) The Kenan Fellows Progam 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. During 2011, Lori Craven, a science teacher at the Cedar Creek Middle School in Youngsville, NC, was the Kenan Fellow working with Richard Smith on a project concerned with the health consequences of climate change. Over the course of the summer, she developed statistical analysis and course tools covering the relationship between temperature and mortality and projecting future trends under various scenarios of climate change. During the 2011-12 school year, she continued her work under this project developing these tools in the classroom. A new Kenan Fellow has been appointed for 2012; Michele Josey will be working with Alan Karr on “Why are surveys so hard?”

E. Evaluation by the SAMSI Governing Board

(Robert Calderbank, George Casella, Michael Crimmins, Susan Ellenberg, Don Estep, Dan Solomon (Chair))

Tragically, the Governing Board lost a valued member with the passing of Dr. George Casella on June 17, 2012. He provided valued service to SAMSI from 2005 until his death and was a champion for quality. We will miss his steady hand.

The Governing Board provides broad oversight for the Institute's administration, finances, and evaluation, and for relationships among the partnering institutions. This year's evaluation follows as responses to three broad questions:

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

Consistent with its founding vision, 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 those from other areas of the mathematical sciences. The results of these efforts lie not only in the production of 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, statistics and other disciplines for which primary activity ended or started in the past year, and that are noted in the annual report, are listed below.

- The Analysis of Object Data (AOOD) program was primarily about statistical analysis on objects more complicated than scalars and vectors: for example, curves, images and shapes. This, however, raised many mathematical questions about the objects being studied, and in some cases led to fruitful interactions between statisticians and experts in various areas of mathematics.
    - One particularly productive Working Group was the one on Data Analysis on Sample Spaces with Manifold Stratification. The three leaders of this group were a statistician (Vic Patrangenaru of Florida State), an expert on stochastics (Stephan Huckeman of Göttingen) and a mathematician with particular interest in combinatorial aspects of geometry and algebra (Ezra Miller of Duke, who subsequently became an Associate Director of SAMSI). Their particular achievement was to define a common class of mathematical spaces, known as stratified

sample spaces, which could be used to define a structure for a number of object data problems. Products of the group included a probability paper on “sticky central limit theorems,” several ongoing papers on nonparametric statistics and a proposal for a workshop focused on biological applications, which subsequently took place at the Mathematical Biosciences Institute.
    - The Dynamics and Inference Working Group of AOOD was another that combined mathematical and statistical elements, focusing on statistical inference problems arising in dynamical systems defined by families of differential or difference equations.

- In contrast, the program on Complex Networks was led more by the applied mathematics and probability communities, but also included substantial forays into statistics, both from a theoretical viewpoint (e.g. the effective sample size in a process defined on a network) and in several areas of applications such as epidemics, traffic flow and the analysis of power grids.
    - The Sampling/Modeling/Inference group included, apart from the above-mentioned work on effective sample size, a group looking at parameter estimation in agent-based models.
    - Traffic flow was the theme both of a Working Group and of a workshop focused on pedestrian traffic. The underlying theme is that of agent-based models, a topic popular in the social sciences but receiving only limited attention from mathematical scientists so far. A feature of these models is to understand how microscale dynamics, which define interactions among individuals, lead to macroscale behavior defined by PDEs. There is therefore a natural link between the statistics needed to study individual parameters and the applied mathematics needed for the large-scale dynamics.
    - Smart Grids was another theme involving both statistics for microscale behavior and probability and optimization for large-scale properties, in this case applied to power grids. There was a Working Group on this theme that led to a workshop held at SAMSI in the fall of 2011.

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. As above, consistent with its founding vision, 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 AOOD program included a Working Group studying fMRI and brain imaging, led by a visiting researcher from Warwick University in the UK (John Aston). This started out looking at the applications of one particular area, functional data analysis (FDA), but soon broadened to include other areas of AOOD such as image analysis and shape analysis.
- The AOOD Working Group on Hierarchical Models was motivated by a number of problems in scientific domains. Group leader Jeffrey Morris from the MD Anderson Cancer Center in Houston is an expert on proteomics and worked on a number of applications such as mass spectrometry and 2D gel electrophoresis. Another member of the working group (Sarat Dass from Michigan State) wrote a paper about an application in forensic science focused on an improved diagnostic for determining when two fingerprint patterns belong to the same person.
- Visiting professor Kanti Mardia from Leeds (UK) wrote a paper on directional data analysis motivated by a visit to a Biochemistry research group at Duke.
- The Working Group in Geometric Correspondence led to a presentation at a Banff International Research Station workshop on "Geometry for Anatomy."
- The Complex Networks program also generated research focused on broad science themes. One particularly striking example was the paper by Brummitt et al. that appeared in *PNAS* (2012). This paper, which analyzes how the interdependence between two grids affects cascading behaviors, was motivated by problems arising in power grids but also has applications to earthquakes, landslides and forest fires, among others.
- Another paper that arose out of Complex Networks and was published in *PNAS* (Durrett et al.) concerned social networks whose individuals hold one of two opinions that may change as a result of interaction with neighbors. The evolution of the network continues until there are no longer any edges of the network joining voters with discordant opinions; the central result of the paper was the relationship between the final proportion of voters holding each opinion and the initial proportion and interaction dynamics of the network.
- The program also developed results in epidemiology and disease dynamics; for example, understanding the role of concurrent partners in the spread of a sexually transmitted disease.

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. As reported by researchers in the AOOD program, there were 39 accepted publications over the year, with an additional 43 papers submitted and 84 papers in various stages of preparation. At the time of writing, we have only incomplete numbers for the Complex Networks program, but they include two papers published in *PNAS* (as reported above), and several papers written during the program that are going through the journal review process.

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 remain at stable levels. For 2010-2011, the participants were 29% female, 3% African American and 4% Hispanic. Participation by new researchers and students this year is at 57%. The SAMSI website now offers information about its diversity programs at http://www.samsi.info/about/diversity-statement .

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. Less than half of funded participants in SAMSI programs are from top 50 DMS funded institutions.

The detailed participant lists for concluded programs provide ample evidence of the national and international draw of SAMSI activities. SAMSI programs attracted 63 research fellows (non-local visitors who played a major role in program activities) plus 5 new researcher fellows, 9 local faculty fellows, 13 postdoctoral fellows, 15 graduate fellows and an overall total of almost 1,200 in all types of SAMSI activities.

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

The directorate model continues to serve SAMSI well, and transitions in the directorate have gone smoothly. Rick Durrett served as Associate Director for only nine months but following his resignation, Ezra Miller (Duke, Mathematics) has ably fulfilled that role. Ilse Ipsen (NCSU, Mathematics) was appointed Associate Director in 2011 in order to ensure that there would be a Directorate representative from NC State University well before Pierre Gremaud stepped down as Deputy Director in June 2012. Meanwhile a national search for his replacement was conducted and resulted in the appointment of Snehalata Huzurbazar from the University of Wyoming, for a two-year term beginning July 1, 2012. On the NISS side, Nell Sedransk stepped down as Associate Director of SAMSI upon her move to Washington and was replaced by Alan Karr who thus resumed a role he had held for several years in the early days of SAMSI.

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 (Estep). A process is being initiated to identify a successor for the ASA slot. The Duke representative on the Governing Board became Robert Calderbank, Dean of Natural Sciences at

Duke, who took over from John Simon following the latter's appointment as Provost of the University of Virginia.

The Governing Board Chair and the SAMSI Director continue to have regular telephone conferences at which administrative and personnel matters are discussed and issues addressed where they have arisen. There is also good cooperation among the partner universities and NISS to ensure that obligations are met and that SAMSI continues to flourish. The pending change of PI institution (with the new grant period starting in August 2012) has also been handled smoothly with all three institutions readily agreeing to continue their cost-sharing support.

I. Annual Progress Report

These Year 9 programs were: Complex Networks, and Analysis of Object Data. Their final reports are in Appendices A and B, respectively.

A. Program Personnel

1. Program and Activity Organizers

Program Organizers

Program	Name	Affiliation	Field
Complex Networks	Rick Durrett	Cornell University	Statistics
	Eric Kolaczyk	Boston University	Statistics
	Alex Vespignani	Indiana University	Physics
	Pierre Degond	Institut de Mathématiques	Mathematics
2010-11 SAMSI Program	Stephan Fienberg	Carnegie Mellon	Statistics
	Martina Morris	University of Washington	Sociology
	Alun Lloyd	NCSU	Mathematics
	Peter Mucha	UNC	Mathematics
	Pierre Gremaud	SAMSI / NCSU	Mathematics
	Bin Yu	UC Berkeley	Statistics
Analysis of Object Data	Hans-Georg Müller	UC Davis	Statistics
	Jane-Ling Wang	UC Davis	Statistics
	Ian Dryden	U South Carolina	Statistics
	Jim Ramsay	McGill University	Psych / Statistics
	Steve Marron	UNC	Statistics
2010-11 SAMSI Program	Nell Sedransk	NISS / SAMSI	Statistics
	Jianqing Fan	Princeton	Statistics
Education & Outreach Program	Negash Begashaw	Benedict College	Mathematics
	Carlos Castillo-Chavez (ex officio)	Arizona State U	Mathematics
	Karen Chiswell	NCSU	Statistics
	Cammey Cole	Meredith College	Mathematics & CS
	Anne Fernando	Norfolk State U	Mathematics
	Leona Harris	College of NJ	Mathematics
	Gabriel Huerta	University of NM	Statistics
2010-11 SAMSI Program	Marian Hukle	U of Kansas	Biological Sciences
	Masilamani Sambandham	Morehouse College	Mathematics

Activity Organizers

2009-10 Programs		
Program Year	**Activity**	**Name(s)**
Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change		
2009-10	Transition Workshop – *October 11-13, 2010*	Alan Gelfand (Duke), Richard Smith (SAMSI)
Stochastic Dynamics		
2009-10	Transition Workshop – *November 17-19, 2010*	Peter Kramer (RPI), Jonathan Mattingly (Duke), Sorin Mitran (UNC)
2010-2011 Programs		
Complex Networks		
2010-11	Tutorials and Opening Workshop – *August 29-September 1, 1010*	Eric Kolaczyk (Boston U), Michael Mahoney (Stanford), Peter Mucha (UNC), Rick Durrett (Cornell)
2010-11	Modeling Workshop – *October 20-22, 2010*	Eric Kolaczyk (Boston U), David Banks (Duke)
2010-11	Dynamics of Networks – *January 10-12, 2011*	Raissa D'Souza (UC Davis), Stephan Fienberg (CMU), Eric Kolaczyk (Boston U), Jim Moody (Duke), Peter Mucha (UNC), Mason Porter (Oxford)
2010-11	Pedestrian Traffic Flow – *February 14-16, 2011*	Pierre Degond (Inst. Of Mathematics), Alina Chertock (NCSU), Alexander Kurganov (Tulane)
2010-11	Dynamics on Networks – *March 21-23, 2011*	Rick Durrett (Duke), Alun Lloyd (NCSU), Peter Mucha (UNC), Alex Vespignani (Indiana U)
2010-11	Transition – *June 6-7, 2011*	Eric Kolaczyk (Boston U), Michael Mahoney (Stanford), Mason Porter (Oxford), Alex Vespignani (Indiana U)
Analysis of Object Data		
2010-11	Tutorials and Opening Workshop – *September 12-15, 2010*	Hans Mueller (U of California, Davis), Jane-Ling Wang (U of California, Davis), Ian Dryden, (U of South Carolina), James Ramsey (U of Toronto), J. Stephen Marron (UNC), Nell Sedransk (SAMSI)
2010-11	Interface, Functional and Longitudinal Data Analysis – *November 8-10, 2010*	Hans Mueller (U of California, Davis), Fang Yao (U Toronto), Marie Davidian (NCSU)

2010-11	AOOD Meets Evolutionary Biology – *April 30-May 2, 2011*	Steve Marron (UNC)
2010-11	Transition – *June 9-11, 2011*	Hans Mueller (U of California, Davis), Jane-Ling Wang (U of California, Davis), Ian Dryden, (U of South Carolina), James Ramsey (U of Toronto), J. Stephen Marron (UNC)
Uncertainty Quantification Summer School		
2010-11	UQ Summer School – June 20-24	James Stewart (Sandia National Laboratory), Pierre Gremaud (North Carolina State Univ)
Education and Outreach		
2010-11	Two-Day Undergraduate Workshop -- *October 29-30, 2010*	Pierre Gremaud (SAMSI)
2010-11	Two-Day Undergraduate Workshop -- *February 25-26, 2011*	Pierre Gremaud (SAMSI)
2010-11	SAMSI/CRSC Interdisciplinary Workshop for Undergraduates -- *May 16-20, 2011*	Pierre Gremaud (SAMSI), Cammey Cole Manning (Meredith)
2010-11	CRSC/SAMSI Workshop for Graduate Students – *July 7-15, 2011*	Pierre Gremaud (NCSU), Ilse Ipsen (NCSU), Ralph Smith (NCSU)

2. Program Core Participants and Targeted Experts

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.

Complex Networks Program

Program Core Participants and Targeted Experts

Last Name	First Name	Gender	Affiliation	Department	Status
Airoldi	Edo	M	Harvard University	Statistics	RF
Balachandrian	Kash	M	Duke University	Mathematics	GF
Bayarri	Maria	F	University of Valencia	Statistics	RF
Blitzstein	Joe	M	Harvard University	Statistics	RF
Bortz	David	M	University of Colorado	Applied Mathematics	RF
Brummitt	Charles	M	University of California-Davis	Mathematics	GRF
Chipman	Hugh	M	Acadia University	Mathematics & Statistics	RF
Cintron-Arias	Ariel	M	East Tennessee University	Mathematics & Statistics	RF
Diaz	Oliver	M	Duke University	Mathematics	RF
Dukic	Vanya	F	University of Colorado	Applied Mathematics	RF
Estrada	Ernesto	M	University of Strathclyde	Mathematics & Statistics	RF

Jain	Ankita	F	University of Houston	Mathematics	GRF
Jung	Sungkyu	M	UNC-CH	Statistics	GF
Khan	Taufiquar	M	Clemson University	Mathematics	RF
Kolaczyk	Eric	M	Boston University	Mathematics & Statistics	RF
Li	Yingbo	F	Duke University	Statistics	GF
Lloyd	Alun	M	NC State University	Mathematics	FF
Lynch	James	M	University of South Carolina	Statistics	RF
Maggioni	Mauro	M	Duke University	Mathematics & Computer Science	FF
McSweeney	John	M	Concordia University	Mathematics & Statistics	NRV
Michailidis	George	M	University of Michigan	Statistics	RF
Michael	Robert	M	NC State University	Mathematics	GF
Mucha	Peter	M	UNC-CH	Mathematics	FF
Porter	Mason	M	University of Oxford	Mathematics	RF
Shafiei	Mahdi	M	Acadia University	Computer Science	NRV
Rogers	Bruce	M	Duke University	Mathematics	PF
Shi	Bill	M	UNC-CH	Mathematics	GF
Shojaie	Ali	M	University of Michigan	Statistics	RF
Sivakoff	David	M	Duke University	Mathematics	PF
Spencer	Bruce	M	Northwestern University	Statistics	RF
Spickenheuer	Anne	F	IPA	Epidemiology	GF
Sun	Yi	M	SAMSI	Mathematics	PF

Timofeyev	Ilya	M	University of Houston	Mathematics	RF
Traud	Amanda	F	NC State University	Biostatistics	GF
Van Haren	Ken	M	Duke University	Statistics	GF

Analysis of Object Data

Program Core Participants and Targeted Experts

Last Name	**First Name**	**Gender**	**Affiliation**	**Department**	**Status**
Aston	John	M	University of Warwick	Statistics	RF
Boente	Graciela	F	Ciudad University	Mathematics	RF
Brunel	Nicolas	M	University d'Evry	Statistics	RF
Campbell	David	M	Simon Fraser University	Statistics and Actuarial Science	RF
Cao	Jiguo	M	Simon Fraser University	Statistics and Actuarial Science	RF
Chiou	Jeng-Min	M	Academia Sinica	Statistics	RF
Degras	David	M	NCSU	Statistics	PF
Ding	Jimin	F	Washington University of St. Louis	Mathematics	RF
Dryden	Ian	M	University of Nottingham	Statistics	RF
Fearnhead	Paul	M	Lancaster University	Mathematics & Statistics	RF
Girolami	Mark	M	University of Glasgow	Computing Science	RF
Hall	Peter	M	University of Melbourne	Mathematics & Statistics	RF
Hooker	Giles	M	Cornell University	Statistics	RF
Hotz	Thomas	M	Georgia Augusta University of Goettingen	Mathematical Stochastics	RF

Huckeman	Stephan	M	Georgia Augusta University of Goettingen	Mathematical Stochastics	RF
Huzurbazar	Snehelata	M	University of Wyoming	Statistics	RF
Jank	Wolfgang	M	University of Maryland	Department of Decisions, Operations & Information Technologies	RF
Jiang	Ci-Ren	M	NCSU	Statistics	PF
Joshi	Sarang	M	University of Utah	Scientific Computing and Imaging Institute	RF
Kang	Emily	F	NC State University	Mathematics	PA
Kent	John	M	University of Leeds	Statistics	RF
Kim	Yongdai	M	Seoul National University	Statistics	RF
Kneip	Alois	M	University of Bonn	Statistics	RF
Kurtek	Sebastian	M	Florida State University	Statistics	VGF
Le	Huiling	F	University of Nottingham	Mathematical Sciences	RF
Lee	Jaeyong	M	Seoul National University	Statistics	FF
Liu	Wei	M	Florida State University	Statistics	VGF
Ma	Junheng	M	NCSU	Statistics	PF
Maldonado	Yolanda	F	Michigan Technological University	Mathematical Sciences	NRF
Mardia	Kanti	M	University of Leeds	Statistics	RF
Marron	Steve	M	UNC-CH	Statistics	FF
Miller	Ezra	M	Duke University	Mathematics	FF
Morris	Jeffrey	M	University of Texas	Statistics	RF

Mueller	Hans-Georg	M	University of California-Davis	Statistics	RF
Ogden	Todd	M	Columbia University	Biostatistics	RF
Olhede	Sofia	F	University College London	Statistical Science	RF
Panaretos	Victor	M	Swiss Federal Institute of Technology	Mathematics	RF
Park	Byeong	M	Seoul National University	Statistics	RF
Park	Juhyun	F	Lancaster University	Mathematics & Statistics	RF
Patrangenaru	Vic	M	Florida State University	Statistics	RF
Ramsay	James	M	University of Ottawa	Mathematics & Statistics	RF
Rodriguez	Daniela	F	SAMSI	Mathematics	RF
Ruymgaart	Frits	M	Texas Tech University	Statistics	RF
Samworth	Richard	M	Cambridge University	Mathematics	RF
Sangalli	Laura	F	Politecnico di Milano	Mathematics	RF
Seiler	Christof	M	University of Bern	Institute for Surgical Technology & Biomechanics	GF
Senturk	Damla	F	Pennsylvania State University	Statistics	NRF
Shaby	Ben	M	Duke University	Statistics	PA
Shen	Haipeng	M	UNC-CH	Statistics	FF
Srivastava	Anuj	M	University of Florida	Statistics	RF
Stadtmueller	Uli	M	Ulm University	Dept. of Number and Probability Theory	RF
Tchumtchoua	Sylvia	F	Duke University	Statistics	PF
Wang	Haonan	M	Colorado State University	Statistics	RF

Wang	Jane-Ling	F	University of California-Davis	Statistics	RF
Wang	Xiaojing	F	Duke University	Statistics	GF
Whitaker	Ross	M	University of Utah	Statistics	RF
Wilkinson	Darren	M	Newcastle University	Mathematics & Statistics	RF
Wood	Andy	M	University of Nottingham	Mathematical Science	RF
Wu	Hulin	M	University of Rochester	Biostatistics	RF
Wu	Ping-Shi	M	Lehigh University	Mathematics	RF
Wu	Rongling	M	Pennsylvania State University	Statistics	NRF
Wu	Yuefeng	M	Cornell University	Biostatistics & Computational Biology	PA
Yao	Fang	M	University of Toronto	Statistics	RF
Zeng	Donglin	M	UNC-CH	Biostatistics	FF
Zhang	Bo	M	UNC-CH	Mathematics	GF
Zhang	Jun	M	UNC-CH	Statistics	PA
Zhang	Tingting	F	University of Virginia	Statistics	RF
Zhu	Hongtu	M	UNC-CH	Biostatistics	FF
Zhu	Hongxiao	F	Duke University	Statistics	PF

B. Postdocs

1. Overview

SAMSI had a total of six new postdocs entering this year, to join a group of three second-year postdocs. In this section, we present our goals for the program, followed by the revised mentoring plan in Section B.2, and then the detailed postdoc reports.

Our goals for postdocs are to:

(i) Facilitate a transition from being a graduate student who works on problems suggested by their adviser to an independent researcher who can formulate their own problems.

(ii) Achieve a broader perspective on the use of mathematics in applications.

(iii) Improve their presentations on research through participation in the postdoc seminar and the undergraduate workshops, giving talks at local universities, and at regional and national meetings.

(iv) Enhance their teaching skills by teaching in their second year.

(v) Experience the pleasures of service to the community and enhance their communication skills through outreach activities.

Program participation is the primary activity, and the one that attracts outstanding postdocs to SAMSI. Postdocs typically participate intensively in two Working Groups in their "home program," working with national and international leaders. One of these groups is their main research focus, while the other may be a secondary interest or a topic about which they wish to learn. Each postdoc is responsible, for one Working Group, for maintaining the web page and for managing the use of WebEx. Postdocs attend all courses associated with their program, which provides the background needed to carry out their research and broadens their scientific horizons. They are required to attend all workshops conducted by their home program, and encouraged to attend the opening workshops of other concurrent programs, as well as events at NISS and the university partners.

Postdocs also actively engage in SAMSI's Education & Outreach activities. They are presenters and mentors in the week-long Undergraduate Modeling Workshop and the Two-Day Undergraduate Workshop associated with their home program. The benefits to the postdocs are improved communication skills, and experience serving to the mathematics and statistics community. Some postdocs function as team mentors for the IMSM workshops for graduate students, and on occasion, as team leaders.

Community engagement is the third key postdoc activity. From the day of their arrival at SAMSI, postdocs join a community of SAMSI postdocs, NISS postdocs and other early career researchers in the NISS-SAMSI complex. In some ways, the informal relationships are the most potent component of the community. But, there are also more formal activities, especially a weekly postdoc seminar. A typical seminar has two parts, one focusing on general postdoc issues and upcoming events, and the second comprising a 30-minute presentation by one of the postdocs on his or her research.

2. Postdoc Mentoring Plan

Two mentors. Each SAMSI postdoc has two mentors, a Scientific Mentor and an Directorate Mentor. The latter is a member of the SAMSI Directorate and is the same for both years. The Scientific Mentor in the first year is one of the program organizers or a long-term visitor. In many cases the first year Scientific Mentor is a visitor from outside the Research Triangle, while in the second year the Scientific Mentor is a professor at the postdoc's associated university with closely related interests. As the names indicate, the Scientific Mentor advises the postdoc on his or her research and in the second year on teaching, while the Directorate Mentor is concerned with other aspects of professional development and supervises the postdoc's performance of duties at SAMSI. For example, while the Directorate Mentor will explain what happens in the job search process and when the steps occur, the Scientific Mentor is in a better position to give advice about specific positions that the postdoc may be considering.

Postdoc Orientation. Before the scientific program begins the postdocs (and graduate students associated with the current programs) come to SAMSI for an afternoon of orientation activities. Part of this is technical – many people participate in Working Groups remotely using WebEx, and the Working Group relies on its web page to facilitate the collaboration, so the postdocs and graduate students need to be trained in the use of these tools. The orientation also covers other technical aspects such as computing and the use of SAMSI's webpages. Another important aspect of orientation is to spell out in detail what the postdocs will do during their first year.

The postdoc seminar plays an important role in the mentoring of postdocs. The seminar meets once a week and is run by one or two members of the Directorate. All of the first year postdocs, and some in their second year, participate in this seminar. In a typical meeting, the first fifteen minutes or so are devoted to informal discussion of general postdoc issues and upcoming events. After this, there is a 30-minute presentation by one of the postdocs on his or her research. Owing to the wide variety of research interests of the postdocs, this must be a talk for a general audience, which introduces background and carefully explains new concepts. Participants are encouraged to ask questions when they do not understand, and to give the speaker constructive criticism at the end of the presentation. An important goal of the seminar is to teach the postdocs how to give clear talks accessible to a wide audience at conferences and seminars.

On some occasions, there is a more structured discussion of some topic such as refereeing papers, applying for jobs, or writing grant applications. The seminar is not an appropriate place to discuss personal issues, so the Directorate Mentor will meet with each of his or her postdocs roughly once a month in an informal setting over lunch or coffee to make sure things are going smoothly.

Postdoc reports are important for documenting activities and are also useful to ensure that postdocs articulate their plans. Each first-year postdoc writes three reports: (i) a research plan for the year due by October 15; (ii) a first progress report and updated research plan due by January 21; and (iii) a final report on the year's activities due by July 1. A second-year postdoc writes two reports: (i) a plan describing the postdoc's research objectives and job search due by October 15, and (ii) a final report, to be turned in before leaving SAMSI, which indicates the postdoc's job placement and describes the work done while at SAMSI.

3. List of 2010-11 Postdocs

Continuing:

Oliver Diaz-Espinosa (Ph.D., Applied Mathematics, 2006, University of Texas – Austin)

SAMSI Program: Stochastic Dynamics
Research Mentor: Jonathan Mattingly
Directorate Mentor: Michael Minion
Current Position: Lecturer, University of Houston, Victoria Campus

Emily Kang (Ph.D., Statistics, 2009, Ohio State University)

SAMSI Program: Space-Time Analysis
Research Mentor: John Harlim
Directorate Mentor: Pierre Gremaud
Current Position: Assistant Professor, Department of Mathematical Sciences, University of Cincinnati

Bruce Rogers (Ph.D., Mathematics, 2009, Arizona State University)

SAMSI Program: Stochastic Dynamics
Research Mentor: Peter Mucha
Directorate Mentor: Michael Minion
Current Position: Fellowship Instructor, Augustana College.

Ben Shaby (Ph.D., Statistics, 2009, Cornell University)
SAMSI Program: Space-Time Analysis
Research Mentor: Alan Gelfand
Directorate Mentor: Richard Smith

Yi Sun (Ph.D., Applied and Computational Mathematics, 2006, Princeton University)
SAMSI Program: Stochastic Dynamics
Research Mentor: Pierre Gremaud
Directorate Mentor: Pierre Gremaud
Current Position: Assistant Professor, Department of Mathematics, University of South Carolina

New:

David Degras (Ph.D., Mathematics, University Paris 6, 2007)
SAMSI Program: Analysis of Object Data
Research Mentor: Giles Hooker
Directorate Mentor: Richard Smith
Current Position: Assistant Professor, Department of Mathematical Sciences, De Paul University

Ci-Ren Jiang (Ph.D., Statistics, University of California – Davis, 2009)
SAMSI Program: Analysis of Object Data
Research Mentor: Jeffrey Morris
Directorate Mentor: Nell Sedransk
Current Position: Assistant Research Fellow, Academia Sinica, Taipei, Taiwan.

Junheng Ma (Ph.D., Statistics, Case Western Reserve University, 2010)
SAMSI Program: Analysis of Object Data
Research Mentor: Hans Mueller
Directorate Mentor: Richard Smith
Current Position: Postdoc at NISS.

David Sivakoff (Ph.D., Mathematics, University of California – Davis, 2010)

SAMSI Program: Complex Networks
Research Mentor: Rick Durrett
Directorate Mentor: Rick Durrett
Current Position: Second-year Postdoc at SAMSI

Sylvie Tchumtchoua (Ph.D., Statistics, University of Connecticut, 2009)

SAMSI Program: Analysis of Object Data
Research Mentor: Jeffrey Morris
Directorate Mentor: Richard Smith
Current Position: Second-year Postdoc at SAMSI

Hongxiao Zhu (Ph.D., Statistics, Rice Univeristy, 2008)

SAMSI Program: Analysis of Object Data
Research Mentor: Jeff Morris
Directorate Mentor: Nell Sedransk
Current Position: Postdoctoral Associate, Department of Statistical Science, Duke University.

SAMSI Postdoctoral Fellow Research Plan

Date:

1. **Name:** David Degras

2. **Ph.D. Program**
University & Department: University Paris 6, France
Dissertation Advisor: Daniel Pierre-Loti-Viaud
Year of Ph.D.: 2007

3. **SAMSI Research**
SAMSI Program Title: Analysis of Object Data
SAMSI Research Mentor: Giles Hooker
SAMSI Administrative Mentor: Richard Smith

4a. **SAMSI Activities**
Course(s) (fall & spring): AOD I & AOD II
Workshops Attended (and Workshop Support Tasks): Opening workshop, Interface Functional and Longitudinal Data Analysis
Postdoc-Grad Student Seminar – Presentation(s): 2010-10-27
Undergraduate Workshop(s) – Participation (specifics to be added later):

4b. **Other Activities** (e.g., teaching)

5a. **Working Group I: Statistical Inference of Functional Data**
Special Tasks for Working Group:
Presentations to Working Group: 09-23 / 10-13
Research Area – Plans: Survey Sampling with Functional Data
We study the use of survey sampling to estimate a population mean function in the context of distributed sensors. Specifically, we deal with situations in which large numbers of data streams are observable in real time but where observation and storage of the entire population data is infeasible for technical and cost issues. This type of problem is very relevant e.g. for utility companies having digital meters installed at the customers' sites. See the state-of-the art paper of Cardot and Josserand (2009) for an example with the french electricity company EDF. Our goal here is to take advantage of the functional nature of the data to devise a time-varying sampling scheme that allows to adaptively explore the population structure. Specifically, we let the estimator adapt to three sources of variability: between predefined population strata, between units within the strata, and over time. In building such an adaptive procedure our hope is to achieve a nearly optimal estimation.
My collaborators for this project are Herve Cardot, Etienne Josserand (Universite de Bourgogne, France) and Yolanda Munoz-Maldonado (Michigan Technological University, visiting faculty at SAMSI this year). One paper on (not time-varying) stratified sampling of functional data is already under way with H. Cardot and E.

Josserand and we have started writing a new one on time-varying sampling. We hope to get the company Electricite De France to provide us high quality data as they have done in the past with H. Cardot. Alternatively we may explore other options with US based utility companies as the problem under study may be of important economic interest for them.

5b. Working Group II: Dynamics and Inference

Special Tasks for Working Group: Webmaster

Presentations to Working Group: not yet

Research Area – Plans: I will collaborate with Giles Hooker in the study of dynamic systems of the form dX/dt = f(X(t),theta) + e(t), where X(t) is the stochastic process under study, f is a nonlinear dynamic function, theta is an unknown vector of parameters, and e(t) is an error process with suitable regularity properties (smooth sample paths, stationarity). We are interested in this kind of model because in many applications, it appears more realistic than ODE while generalizing them thanks to the stochastic process e(t). Note that our model leaves the distribution of e(t) unspecified in contrast to SDE where the usual assumption is that e(t) is a Wiener process. The gradient matching technique (e.g. Ellner, 2002) for estimating theta is both efficient and easy to implement in the case of ODE with noisy observations. We believe that it may still yield consistent estimators in our more general model. Our first aim is to determine under what conditions on e and the design this consistency effectively holds. This requires to simultaneously study infill and expanding domain asymptotics in order to recover the trajectory X by smoothing its noisy observations and then estimate theta by nonlinear least squares. Secondly, when consistent estimation is possible, we would like to elaborate limit results such as convergence rates and asymptotic normality for statistical inference.

5c. Working Group III (if appropriate) NA

Special Tasks for Working Group:

Presentations to Working Group:

Research Area – Plans:

6. Other Research

Work on Papers from Ph.D. Research: NA

Other Research started or continued at SAMSI:

* Working paper "Local polynomial estimation of the average growth curve with functional data", with Karim Benhenni, Universite de Grenoble.

* Research on nonparametric smoothing techniques and simultaneous confidence bands for fMRI data. I am currently in discussion with Martin Lindquist (Columbia University), met at the AOOD opening workshop, to decide which one of two potential projects to pursue. One involves testing goodness-of-fit of parametric models for the Hemodynamic Response Function. This would come as a nice application of a recent paper of mine on SCB and as an extension of Martin's previous work, thus requiring little work. The other project relates to a parametric in time, nonparametric in space model for brain activity. The goal here would be to estimate relevant parameters simultaneously over the whole brain to overcome some

limitations of the prevalent two-stage approaches (voxel-level/brain-level). This would also be based on recent work from Martin. In either project, Martin's role will essentially be to provide data and consulting help.

Continuing Collaborations while at SAMSI:
* Working paper "Horvitz-Thompson estimators for sampled noisy functional data", with Herve Cardot and Etienne Josserand, Universite de Bourgogne, France.

Presentations of Other Research:
Here are 4 interesting conferences where I consider to go and present papers.
* 2011 ENAR/IMS meeting, March 20-23, Miami, FL
* 2011 IISA Conference on Probability, Statistics, and Data Analysis, April 21-24, Raleigh, NC.
* Second International Workshop on Functional and Operatorial Statistics (IWFOS 2011), Santander, Spain, June 16-18.
* JSM 2011, July 30 – August 4. Miami, FL
More discussions with my SAMSI mentors would be useful to determine the interest and feasibility of all this. In particular we should look at the financial resources available at SAMSI and elsewhere, the time investment required from my part, the submission deadlines, etc.

Research Progress Report & SAMSI Program Final Report

Date: June 18, 2011

7a-z. Research Contributions – Current Projects (grouped by Working Group)

Working Group I: Statistical Inference of Functional Data
Research Project Title: Longitudinal Survey Methods for Functional Data
Collaborator(s) & Mentor(s): none
Specific Goals & Accomplishments (results): Develop new sampling designs to improve the estimation of population functional parameters (e.g. trend) in survey sampling. Results: 1) Development of two sampling designs (termed partial and full resampling) based on time-varying samples, 2) Theoretical derivation of the bias and the covariance functions of the related Horvitz-Thompson, 3) Stabilization of the integrated squared error (ISE): comparison between the two novel sampling designs and usual time-invariant sampling designs. Tremendous improvement: theory shows that the proposed methods reduce the variance of the ISE by orders of magnitude. These findings are confirmed by artificial and real data analyses.
Research Contributions (publication submissions, articles in preparation, etc.):
Article in preparation (final stages)
Presentations outside SAMSI (including invitations for future talks):
ENAR 2011, Miami FL.

Working Group II: Dynamics and Inference
Research Project Title: Nonparametric Smoothing for Robust Inference in Ordinary Differential Equations
Collaborator(s) & Mentor(s): Giles Hooker.
Specific Goals & Accomplishments (results): See above. Mostly, work in progress. We have already established the consistency of the first step in gradient matching (recovery of the trajectories and their derivatives). More will be done this summer.

Research Project Title: Parameter Estimation from Locally Enforced Differential Equation Models
Collaborator(s) & Mentor(s): Nicolas Brunel, Dave Campbell, and Jiguo Cao.
Specific Goals & Accomplishments (results): Development of the estimation methodology based on ideas of profiling and multiple shooting. Development of related code in Matlab. Theoretical study in progress.

8. Future Research Plans (after completion of SAMSI Program)

Research Area – Plans:
1. Survey sampling methods for functional data
2. Neuroimaging
3. Bandwith selection in the nonparametric estimation of the mean function of a random process
4. Dynamical systems

Continuing Collaborations (if appropriate):
Herve Cardot, Universite de Bourgogne (1)
Martin Lindquist, Columbia University (2)
Karim Benhenni, Universite de Grenoble (3)
Giles Hooker, Nicolas Brunel, Dave Campbell, Jiguo Cao (4)

Presentations outside SAMSI:

SAMSI Postdoctoral Fellow Research Plan

Date: June 5, 2011

1. **Name** Oliver R. Diaz Espinosa

2. **Ph.D. Program**
   University & Department: University of Texas at Austin, Mathematics
   Dissertation Advisor: Rafael de la Llave
   Year of Ph.D.: 2006

3. **SAMSI Research**
   SAMSI Program Title: Stochastic Dynamics/Complex Networks
   SAMSI Research Mentor: Jonathan Mattingly
   SAMSI Administrative Mentor: Rick Durret

4a. SAMSI Activities
Course(s) (fall & spring): Complex Networks: Theory and Applications
Workshops Attended (and Workshop Support Tasks): Complex Networks Opening Workshop technical support

Undergraduate Workshop(s) – Participation:
• Education and Outreach Program: SAMSI Interdisciplinary Workshop for Undergraduate Students and Faculty (May, 16-20 2011). *Small Random Perturbations of Chaotic Dynamical Systems and the Central Limit Theorem. R-Lab session.*

4b. Other Activities
Teaching: M103, Duke University.
Talk: Central Limit Theorem and Renormalization of chaotic maps,
October 11, 2010 Probability seminar at NCSU.

Workshops Attended:

• October 24-30, 2010. Oberwolfach Seminar: The Ergodic Theory of Markov Processes.
• June 20-30, 2011. New Directions Short Course: Invariant Objects in Dynamical Systems and its Applications.

6. **Other Research**
   • Work on Papers from Ph.D. Research: Central Limit Theorem and renormalization of maps with Siegel disks.

6 b. Other Research started or continued at SAMSI:
• J. Mattingly (Duke University) Discrete Time Gibbsian Dynamics

Continuing Collaborations while at SAMSI:

• S. Schmidler (Duke University) Bounds for replica-exchange Langevin dynamics in metastabale systems

• A. Budhijara (UNC) and X. Wang (Texas A&M). Differentiation of T cells into T1 and T2 subpopulations.

Research Progress Report & SAMSI Program Final Report

Date:

7a-z. Research Contributions – Current Projects (grouped by Working Group)

Stochastic Dynamics:

Research Project Title: Discrete Time Gibbsian Dynamics

Collaborator(s) & Mentor(s): Jonathan Mattingly

Specific Goals & Accomplishments (results):

The goal is to develop a technique, based in simple techniques from statistical mechanics, to study the convergence of Markov chains on Product spaces that have a spacial structure called Gibbsian dynamics. This structure considers the whole "past" of a given non Markovian process construct the new position of a process to give a Markov structure in "past" space. The invariant measures of this infinite dimensional Markov process have identical marginals which are the target of our study. Some applications are simple derivations of algorithms to sample distributions, derivation of rate of convergence of such algorithms.

We expect to conclude the first part of this work by the end of the summer of 2011.

Biological Dynamics:

Research Project Title: Differentiation of T cells into T1 and T2 subpopulations.

Collaborator(s) & Mentor(s): Amarjit Budhijara and Xueying Wang

Specific Goals & Accomplishments (results):

The development of a multicellular organism from a single fertilized egg cell to specialized cells depends on programs of gene expression. Following the initial stage of cell determination is a maturation process called differentiation by which cells acquire specific recognizable phenotype and functions. In particular, the T lymphocytes of the immune system, upon maturation, differentiate into either Th1 or Th2 cells that have different functions.

We are investigating the stochastic effects on T cell differentiation in the models described above. Numerical simulations suggests that the stochastic model with multiplicative noise have one of the following equilibria: either extinction or normal differentiations, that is to say both types of Th cells die or only one type of Th cell matures. The resulting stochastic system of equations that we consider is of the McKean–Vlasov and we are using techniques from particle systems to analize the long term behaviour of the system: existence and uniqueness of invariant measures and limiting distribution as the size of noise goes to zero.

We expect to complete this work by the Fall of 2011.

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

• (with J. Mattingly) Discrete Time Gibbsian Dynamics (in preparation)

8. Future Research Plans (after completion of SAMSI Program)

Research Area – Plans:

Complete the following projects:

• *Differentiation of T cells into T1 and T2 subpopulations.*

• *Bounds for replica-exchange Langevin dynamics in metastable systems.*

Work on the following problems:

• *Directed Polymers in Product Form Random Environment.*

• *Random Pertubations of Holomorphic Maps with Siegel Disks*

Continuing Collaborations (if appropriate):

 Currently the in the project "Discrete Time Gibbsian Dynamics" we are analyzing the convergence of different MCMC methods through Gibbsian Dynamics. The next step is to study the rate of convergence of this methods using different metrics such as Total variation and Wasserstein metrics.

• *Konstantin Khanin: Directed Polymer Systems with product form.* A directed polymer is a classical model in statistical mechanics in which paths of stochastic processes interact with impurities, or quenched disorder. we are studying the behavior of the polymer for random environments of the form. We suspect that in this situation, the random walk will be more localized and likely non-diffusive.

• *Walter Craig and Catherine Sulem: Water waves over a rough bottom, 3-d problem.* The problem of water waves under a variable bottom is a classical problem in fluid dynamics and it is relevant to coastal engineering and ocean wave dynamics. For reasons of applications, there is interest in wave motion in basins with non constant bathymetry.The presence of bottom topography in the fluid domain introduces additional space and time scales to the classical perturbation theory. Our goal in this problem is to derive the effective Kadomtsev–Petviashvili (KP) equations for the components of the solution. The KP equation is the three–dimensional analog of the KdV equation, which is derived under the assumption that transverse variations of the wave motions in the x2 direction are weaker than those in the x1 direction

SAMSI Postdoctoral Fellow Research Plan

Date: 06/15/2011

1. Name: Ci-Ren Jiang

2. Ph.D. Program

University & Department: University of California at Davis
Dissertation Advisor: Jane-Ling Wang
Year of Ph.D.: 2009

3. SAMSI Research

SAMSI Program Title: Analysis of Object Data
SAMSI Research Mentor: Jeffrey Morris
SAMSI Administrative Mentor: Nell Sedransk

4a. SAMSI Activities

Course(s) (fall & spring): AOOD

Workshops Attended (and Workshop Support Tasks):

1. 2010-2011 Analysis of Object Data Opening Workshop and Tutorials (help speakers to upload files to the computers)
2. AOD: Interface Functional and Longitudinal Data Analysis(help speakers to upload files to the computers) - November 8-10, 2010
3. AOOD Meets Evolutionary Biology – April 30 – May 2, 2011
4. AOOD Transition Workshop – June 9-11, 2011

Postdoc-Grad Student Seminar – Presentation(s):

1. "Functional Single Index Models for Longitudinal Data" on 02/09/2011
2. "Nonparametric response function estimation via FPCA with an application to Dynamic PET Data" on 06/11/2011 (AOOD Transition workshop)

Undergraduate Workshop(s) – Participation (specifics to be added later):
02/25/2011: Matlab Demo
05/16/2011~05/20/2011: LSE, Regression and SIR.

4b. Other Activities (e.g., teaching)

1. Review service for Computational Statistics and Data Analysis.
2. Review service for the book volume entitled "Festschrift in Honor of Professor P. K. Bhattacharya on the Occasion of his 80th Birthday"
3. Review service for Bernoulli.

5a. Working Group I: Brain Imaging
Special Tasks for Working Group: Webmaster

Presentations to Working Group:
09/29/2010: Smoothing dynamic positron emission tomography time courses using functional principal components

Research Area – Plans:
Positron emission tomography (PET) is a nuclear medicine imaging technology producing 3D image of functional processes in the body. The isotope concentration in a tissue over time is measured and the local biochemical and physiologic activities can be evaluated via it. In general, the concentration function is modeled as a convolution between a known arterial input function and an unknown impulse response function. The parameter of interest here is the integral of the impulse response function that indirectly provides the tissue metabolic activity information. Recently, some de-convolution methods have been applied to estimate the response function at voxel level. However, doing so did not take advantage of the information at hand that voxels are spatially correlated. In this project, we would like to propose a de-convolution approach based on FPCA framework and also take the spatial dependency among voxels into account. This method is not only restricted to dynamic PET data; with appropriate input functions, it can also be extended for functional magnetic Resonance Imaging (fMRI) data.

5b. Working Group II: Hierarchical Methods for Object Data
Special Tasks for Working Group: N.A.
Presentations to Working Group:

Research Area – Plans:
High dimension and low sample size data are becoming very common in many fields, such as medical imaging and genetics data. Principal component analysis (PCA) is a popular method for dimension reduction, and it has been extended to functional data analysis recently. In practice, PCA was found unable to capture local features of functions. Similar to principal components, wavelets are also families of orthonormal basis functions that are used to present other functions parsimoniously. These two approaches target at different function features. Wavelets are well-known with the property of capturing local features of functions; on the other hand, principal components can be used to model the global changes of functions. In this project, we hope to develop a hierarchical model to analyze functional data and the hierarchy is in the sense of global and local features.

5c. Working Group III (if appropriate)
Special Tasks for Working Group:
Presentations to Working Group:

Research Area – Plans:

6. Other Research

Other Research started or continued at SAMSI:

Continuing Collaborations while at SAMSI:

1. Hierarchical multi-label classification (joint work with Prof. Haiyan Huang at UC Berkeley)

In multi-label classification problems, some statistics, such as posterior probabilities, are calculated for each class and used to classify the label of the respective class. In general, an identical criterion (e.g. rejection region) is applied to all the classes. However, one important issue has been ignored by doing so. Take the posterior probability for example. The posterior distributions of different classes are obtained given different amount of information (e.g. different sample sizes of positive or negative class). Therefore, using the same criterion may not lead to the overall optimal results for the classifiers and in many cases, the performances of these classifiers are seriously underestimated. Also, using grid search for each class is not feasible when the number of class is not too small. To resolve this issue, we propose an efficient approach based on local precision to evaluate the overall performance of a multi-label classifier. The method is demonstrated with numerical studies (both simulation studies and real data analysis).

2. Sparse Linear Modeling of RNA-seq Data for Isoform Discovery and Abundance Estimation (joint work with Jingyi Jessica Li, James B. Brown, Haiyan Huang, Peter J. Bickel)

Since the inception of RNA-Seq technology, various attempts have been made to utilize RNA-Seq data to assemble full-length mRNA isoforms de novo and estimate abundance of isoforms in annotations. For genes with more than a few exons, the problem tends to be very challenging, and often involves identifiability issues of statistical modeling. We have developed a statistical method called Sparse Linear modeling of RNA-Seq data for Isoform Discovery and abundance Estimation (SLIDE) that takes exon boundaries and RNA-Seq data as input to discern the set of mRNA isoforms that are most likely to present in an RNA-seq sample. SLIDE is based on a linear model with a design matrix that models the generating probability of RNA-Seq data from different possible mRNA isoforms. To tackle the unidentifiability issue in the model, SLIDE uses a modified Lasso procedure for parameter estimation. Compared with de novo mRNA isoform assembly algorithms mostly using deterministic approaches, SLIDE is a new statistical method which takes known aspects of gene structure into account, and hence is more robust to noise and has better precision and recall rates for curated annotations. Moreover, SLIDE can use the same linear model to estimate the abundance of discovered mRNA isoforms simultaneously with high accuracy. We note that SLIDE also has the flexibility of incorporating multiple types of genomic data and can be used downstream of de novo assembly algorithms in order to integrate newly discovered exons into extant gene models.

Presentations of Other Research:

1. "Covariate Adjusted Functional Principal Component Analysis", National Chung Cheng University at Taiwan (01/05/2011)
2. "Covariate Adjusted Functional Principal Component Analysis", National Cheng Kung University at Taiwan (01/06/2011)
3. "Covariate Adjusted Functional Principal Component Analysis", Academia Sinica at Taiwan (job talk 01/13/2011)
4. "Covariate Adjusted Functional Principal Component Analysis", Michigan State University (job talk 02/15/2011)
5. "Covariate Adjusted Functional Principal Component Analysis", University of California at Riverside (job talk 03/1/2011)
6. "Smoothing dynamic positron emission tomography time courses using functional principal components", Interface 2011 at Cary, NC (Invited Talk)

Research Progress Report & SAMSI Program Final Report

Date: 06/15/2011

7a-z. Research Contributions – Current Projects (grouped by Working Group)

Research Project Title: Nonparametric response function estimation via FPCA with an application to Dynamic PET Data

Collaborator(s) & Mentor(s): John Aston, Jane-Ling Wang

Specific Goals & Accomplishments (results):

In this project, we would like to propose a de-convolution approach to estimate the parameter of interest (the integral of the impulse response function that indirectly provides the tissue metabolic activity information) based on FPCA framework and also take the spatial dependency among voxels into account. The idea is first to extend the current covariate adjusted FPCA to handle 3D coordinate information and then is to apply de-convolution technique in the FPCA framework to estimate the impulse response function. Once the impulse response function is estimated, the parameter of interest can be easily obtained. The difficulty is to handle the issues of memory usage and computational time since the data size is huge and we plan to perform 4D smoothing and 5D smoothing.

So far, the MatLab codes of a 4D smoother for mean estimator and a 5D smoother for covariance function have been written and tested. The out-of-memory issue has been resolved. The algorithms have been further improved to reduce the *computation time. Some simulation studies have been done. The results look all right, but not good enough. Some results are presented in the AOOD transition workshop. Therefore, we are investigating new approaches to have more accurate de-convolution results though it is not an easy work.*

Once it is done, the paper will be in preparation.

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

The project is still ongoing.

Research Project Title: Not decided yet
Collaborator(s) & Mentor(s): Jeff Morris

Specific Goals & Accomplishments (results):

In this work, we would like to make the point that frequently it is not possible to estimate all modes of variability in the data reliably, especially when the data are complex (meaning there are lots of "true features" or "dimensions" p^* or modes of variability) when the sample size N is not large. The limitations are obvious when the true number of features $p^* > N$, but even when $N > p^*$, many of the estimated principal component directions may be spurious. We would like to propose a method to select the number of principal components "that were estimated reasonably".

We accomplish this by performing a permutation test. The key is that under this null hypothesis of no structure, all eigenvalues are equal, but due to randomness, in reality there will be some apparent structure that is in fact spurious. We also explored whether under the null spherical hypothesis, we could actually derive the distribution of the eigenvalues analytically and we found it would not be too tractable. Thus, we will stick with the permutation approach.

The simulation studies show that one cannot reliably estimate too many directions when $N \ll p^*$ or even when $N \sim p^*$. We will try to prove a basic theorem that our approach will choose the right number of principal components as N goes to infinity.

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

The project is still ongoing.

Presentations outside SAMSI (including invitations for future talks):

8. Future Research Plans (after completion of SAMSI Program)
Research Area – Plans:

Continuing Collaborations (if appropriate):
Presentations outside SAMSI:

SAMSI Postdoctoral Fellow Research Plan

Date: June 15, 2011

1. Name Junheng Ma

2. Ph.D. Program
University & Department: Case Western Reserve University / Statistics
Dissertation Advisor: Prof. Joe Sedransk and Prof. Jiayang Sun
Year of Ph.D.: January, 2011

3. SAMSI Research
SAMSI Program Title: AOOD
SAMSI Research Mentor: Prof. Hans Mueller
SAMSI Administrative Mentor: Prof. Richard Smith

4a. SAMSI Activities
Course(s) (fall & spring): AOOD fall course 2010 and AOOD II spring course 2011
Workshops Attended (and Workshop Support Tasks): AOOD opening workshop, AOOD: Interface Functional and Longitudinal Data Analysis, Two-Day Undergraduate Workshop (introduction to R), AOOD meets evolutionary Biology, Interdisciplinary Workshop for Undergraduate Students and Faculty (lab support)
Postdoc-Grad Student Seminar – Presentation(s): Feb. 23, 2011
Undergraduate Workshop(s) – Participation (specifics to be added later):
Two-Day Undergraduate Workshop (introduction to R) and Interdisciplinary Workshop for Undergraduate Students and Faculty (lab support)

4b. Other Activities (e.g., teaching)

5a. Working Group I (AOOD: Statistical Inference)
Special Tasks for Working Group: Webmaster
Presentations to Working Group:

Research Area – (1) Functional/Longitudinal Data clustering
(2) Functional data Classification

5b. **Working Group II** (AOOD: Hierarchical Methods)
Special Tasks for Working Group:
Presentations to Working Group:
Research Area – Plans:

5c. **Working Group III** (if appropriate) (AOOD: Brain Imaging)
Special Tasks for Working Group:
Presentations to Working Group:
Research Area – Plans:

6. **Other Research**
Work on Papers from Ph.D. Research:
(1) Sample Size Determination for Unordered Categorical Data
Finished and handed to my advisor Prof. Jiayang Sun for submission.
(2) nFCA: Numerical Formal Concept Analysis
Working on the final draft.
(3) Bayesian Predictive Inference for Finite Population Quantities under Informative Sampling
Working on the draft.
Other Research started or continued at SAMSI:
(1) Bayesian predictive inference for finite population quantities
Continuing Collaborations while at SAMSI:
Prof. Joe Sedransk and Prof. Jiayang Sun
(2) Hierarchical Bayesian Spatio-temporal Analysis of Salmon Population
Started at SAMSI with Prof. Jiguo Cao
Presentations of Other Research:

Research Progress Report & SAMSI Program Final Report

Date: June 15, 2011

7a-z. **Research Contributions – Current Projects** (grouped by Working Group)
Research Project Title: Clustering methods for Functional/Longitudinal Data
Collaborator(s) & Mentor(s): Prof. Fang Yao and Prof. Helen Zhang
Specific Goals & Accomplishments (results):

The goal of this project is to develop a set of methodology to cluster both dense/regular and sparse/irregular observations from functional/longitudinal data. For dense/regular observations from functional data, a variety of metrics such as L2 distance can be readily chosen to measure the similarity or dissimilarity between subjects. However, finding a suitable distance measure for sparsely and irregularly observed data can be a challenge. Peng and Muller (2008) developed a distance for such data. The basic idea is take the expectation of L2 distance conditioning on the observed data. Inspired by this conditioning distance, we propose a weighted L2 distance conditioning on the data. The key of defining a weighted distance is to choose the weights. The weights can be either some function of the eigenvalues or can be chosen by minimizing the ratio of within-clustering distance and between-clustering distance. There also exist other alternative choices of weights. In our research project, we choose the weights as the power transformation of the eigenvalues, which can be estimated from functional principal component analysis (FPCA). Starting with conventional K-mean for the estimated scores obtained from FPCA, the basic idea is to classify each curve iteratively based on the weighted conditional L2 distances between the predicted curves from each cluster and the corresponding cluster means curves. The optimal power transformation can be found by minimizing the ratio of within-clustering distance and between-clustering distance. We are still investigate the performance of this method for different situations.

Research Contributions (publication submissions, articles in preparation, etc.):
Presentations outside SAMSI (including invitations for future talks):

8. **Future Research Plans** (after completion of SAMSI Program)
Research Area – Plans: I will continue working on functional data analysis and hopefully to focus on adapting conventional multivariate data analysis methods to functional data as well as developing methodology for longitudinal data based on methods for functional data.

Continuing Collaborations (if appropriate): I would like to continue my collaboration with Prof. Fang Yao, Prof. Helen Zhang, Prof. Hans Mueller and Prof. Yichao Wu on clustering/classification for functional and longitudinal data.

Presentations outside SAMSI:

June 15, 2011

Current Projects:

Dynamics OF working group.

Project: Applying community detection to a changing network

Collaborators: James Moody and Peter Mucha

Results: We are applying Peter Mucha's multi-slice modularity algorithm to a data set of friendship nomination networks among children at various grade schools. This is an exercise in mathematical sociology; we are interested in understanding how friendship groups change over time and the groups' relationship to other social outcomes such as graduation from high school, smoking etc.

Mathematically, the work has already been done by Mucha's group, but we are developing a flexible software version of the algorithm that will allow domain specialists to apply it easily.

Article in preparation, probably for the journal of Social Networks.

Sampling, Modeling, Inference working group

Project: Parameter estimation and model selection for agent based models

Collaborators: David Banks and Ken Van Haren

Results: Agent-based models describe collective behavior through modeling the actions of individuals. They have been widely used in biology, economics and sociology, but are not a class of models often given much thought by mathematicians and statisticians because we lack a inferential framework for free parameters.

We have developed such an inferential framework using Approximate Bayesian Computation (ABC). We are still in the "proof of concept" phase, but we are fairly certain that the ABC framework provides the ability to give point estimates and confidence intervals for complicated agent-base models.

Presentations: I have submitted an abstract to talk on this subject at the Joint Mathematical Meeting in January 2012.

Future Research Plans:

Once we more thoroughly understand how to use Approximate Bayesian Computation to estimate parameters for agent-based models, we will build an agent-based model to understand the social dynamics and politics involved in troop fission among baboon in Kenya. There is ample field data to fit such a model.

SAMSI Postdoctoral Fellow Research Update and Final Report

Date: 6-15-2011

1. David Sivakoff

2. Ph.D. Program

University & Department: University of California, Davis, Department of Mathematics
Dissertation Advisor: Prof. Janko Gravner
Year of Ph.D.: 2010

3. SAMSI Research

SAMSI Program Title: Complex Networks
SAMSI Research Mentor: Prof. Rick Durrett
SAMSI Administrative Mentor: Prof. Rick Durrett

4a. SAMSI Activities

Course(s) (fall & spring): Complex Networks course

Workshops Attended (and Workshop Support Tasks): Complex Networks Opening Workshop (I was in charge of assisting the presenters during one of the sessions), Dynamics OF Networks Workshop (I was in charge of assisting the presenters on the second day), Dynamics ON Networks Workshop (I delivered a talk about my original research related to this group), Complex Networks Transition Workshop.

Postdoc-Grad Student Seminar – Presentation(s): Random Site Subgraphs of the Hamming Torus (on 9-22-10), Bootstrap Percolation on the Hamming Graph (on 4-27-11).

Undergraduate Workshop(s) – Participation (specifics to be added later): I presented an introduction to networks and random graphs at the first undergraduate workshop. I helped to plan and present a lecture/lab about the multi-population SIR model for the week-long undergraduate workshop in May. I also helped the students with their projects on the Thursday afternoon of the workshop.

4b. Other Activities (e.g., teaching) None yet.

5a. Working Group I: Dynamics ON Networks

Special Tasks for Working Group: I am the webmaster for this group.

Presentations to Working Group: I presented an overview of some modified voter models that could be amenable to rigorous analyses. I also presented some of my work on bootstrap percolation, as it relates to the problem of cascades on networks.

Research Area – Plans: As part of a project from one of last year's programs, Bruce Rogers and John McSweeney ran simulations of an interesting contact process model on modular networks. They presented their problem to this working group, and I have been working with Rick Durrett to prove rigorous results about their model that will explain the behavior observed in their simulations. I have made a lot of progress on this problem, and I presented my results at the March workshop. My analysis incorporates applications of branching process theory, mixing times for random walks on random graphs, and an isoperimetric inequality for random graphs. I will be giving a talk about my work on this problem at the Cornell Probability Summer School in July.

5b. Working Group II: Dynamics OF Networks

Special Tasks for Working Group:

Presentations to Working Group: I presented a simulation of the dynamic voter model that I ran on my computer.

Research Area – Plans: I am studying the "evolving voter model". This is a hybrid model that incorporates both a dynamic network and a voter process on that network, which affect one another to varying degrees depending on a parameter. Our primary goal is to demonstrate the existence (or nonexistence) of a critical value for this parameter, at which the model changes behavior between the two extremes of a purely voter model dynamic and a purely rewiring dynamic. I have written a computer simulation of this model, and Bill Shi and Alun Lloyd have extended these simulations, which demonstrate a very interesting threshold behavior in the model. Along with Rick Durrett and Peter Mucha, we are attacking this problem as a group both computationally and analytically to further understand its behavior. We have a plan to publish our computational results, and are working to derive analytical results later this year.

While Charlie Brummitt was visiting SAMSI this semester, he and I started work on a mathematical model of puzzle-solving that we have termed "Jigsaw Percolation". In this model, each person in a social (or academic) network of people holds one piece of a puzzle. The puzzle pieces, in a separate, structured graph, are related to one another in some way. Two sets of connected puzzle pieces can be joined to one another if at least one pair of people (one from each group) are acquaintances in the social network. We have already proved some rigorous results for this model in a special case, and we are seeking to generalize our results.

6. Other Research

Work on Papers from Ph.D. Research: Submitted in April

Other Research started or continued at SAMSI: Prof. Jonathan Mattingly and I are in the formative stages of a collaboration on a project investigating how Stein's method might be applied to two sequences of discrete random variables that approach one another more quickly than they approach a fixed distribution.

Along with Prof. Rick Durrett, I have started a collaboration with Dr. Kurtis Sobush, a Fellow in Pediatric Pulmonary Medicine, and Prof. Barry Stripp, both of whom are at the Duke University Medical Center. I am running simulations to help guide their research on epithelial progenitor cells in the lungs.

Continuing Collaborations while at SAMSI: My graduate advisor, Prof. Janko Gravner, and I are continuing our work on the problem of bootstrap percolation on the Hamming torus. Also, we have started thinking about a dynamic network model of coalition formation using positive feedback. I visited him for a week this past semester, during which we wrote an outline for our bootstrap percolation paper, and made a plan for the coalition project.

Presentations of Other Research: I presented my thesis work at the Duke Probability seminar in September, and my bootstrap percolation work at the postdoc seminar in April.

Research Progress Report & SAMSI Program Final Report

Date: 6-15-2011

7a-z. Research Contributions – Current Projects (grouped by Working Group)

Research Project Title: Contact Process on Modular Networks

Collaborator(s) & Mentor(s): Rick Durrett

Specific Goals & Accomplishments (results): To prove convergence in distribution of the time to infect the second cluster of vertices for the contact process started in the first cluster of vertices. Each cluster is a densely connected Erdos-Renyi random graph, and the two clusters are connected by a finite number of edges. Also, en-route to this result, I proved exponential (in number of vertices) survival time for the super-critical contact process on an Erdos-Renyi random graph.

Research Contributions (publication submissions, articles in preparation, etc.): A paper to be submitted to a math journal is in preparation.

Presentations outside SAMSI (including invitations for future talks): Graduate Student Probability Conference (May 2011), Cornell Summer School in Probability (July 2011).

Research Project Title: Evolving Voter Model

Collaborator(s) & Mentor(s): Rick Durrett, Alun Lloyd, Peter Mucha, Bill Shi

Specific Goals & Accomplishments (results): We analyzed four slightly different versions of a model of opinion dynamics on a network that also changes in response to the opinions held by the actors. Through computer simulations, we demonstrated that the model undergoes a phase transition in the parameter controlling the relative rate at which opinions change verses the rate at which the network changes. We have some analytical evidence to explain our simulation results, though our goal is to have a complete analytical explanation.

Research Contributions (publication submissions, articles in preparation, etc.): We have two publications in preparation - one presenting our model and simulations for a general science audience, and another presenting our analytical results to a mathematics audience.

Presentations outside SAMSI (including invitations for future talks): So far we have only presented this work at SAMSI.

Research Project Title: Jigsaw Percolaion

Collaborator(s) & Mentor(s): Charlie Brummitt, Rick Durrett

Specific Goals & Accomplishments (results): This is a simple model of how a network of people might facilitate solving a puzzle or problem. We have analyzed this model on the ring graph, with promising results. We hope to generalize our results to high dimensional finite lattices, and later to more general graph structures.

Research Contributions (publication submissions, articles in preparation, etc.): We plan to publish our results in a math journal.

Presentations outside SAMSI (including invitations for future talks): This problem is quite new, and we have not talked about it anywhere outside of the Dynamics OF Networks working group yet.

8. Future Research Plans (after completion of SAMSI Program)

Research Area – Plans: I plan to continue working on the problems mentioned above (in sections 5 and 7) in addition to my new collaboration with Dr. Kurtis Sobush and Prof. Barry Stripp at the Duke University Medical Center (see section 6), as well as several other problems involving random processes on graphs.

SAMSI Postdoctoral Fellow Research Plan

Date: June. 15, 2011

1. Name: Yi Sun

2. Ph.D. Program
University & Department: Princeton University, Applied & Computational Mathematics
Dissertation Advisor: Bjorn Engquist Year of Ph.D.: 2006

3. SAMSI Research
SAMSI Program Title: Complex Networks (2010-2011)
SAMSI Research Mentor: Pierre Gremaud
SAMSI Administrative Mentor: Pierre Gremaud

4a. SAMSI Activities
Course(s) (fall & spring): Complex Networks Fall Course
Workshops Attended (and Workshop Support Tasks):

1. Complex Networks Opening workshop (support task on the second day)
2. Complex Networks Modeling workshop (support task on the first day & present a poster)
3. Stochastic Dynamics Transition Workshop (support task and present a poster)
4. Complex Networks Dynamics of Networks Workshop (support task and present a poster)
5. Complex Networks Dynamics on Networks Workshop (support task and present a poster)
6. Complex Networks Transition Workshop (present a poster)

Postdoc-Grad Student Seminar – Presentation(s): Talk on December 1, 2010.

Undergraduate Workshop(s) – Participation:

1. Presentation at Undergraduate Workshop on Complex Networks, October 29-30, 2010.
2. Interdisciplinary Workshop for Undergraduates, May 16-20, 2011

4b. Other Activities (e.g., teaching) No

5a. Working Group I: Modeling Flows
Special Tasks for Working Group: Webmaster
Presentations to Working Group: Kinetic Monte Carlo simulation for traffic flow model.
Research Area – Plans:
Collaborate with Ilya Timofeyev and Cory Hauck on traffic flow model.

5b. Working Group II: Dynamics ON Networks
Special Tasks for Working Group: No
Presentations to Working Group: No
Research Area – Plans: Study the problems about Dynamics ON Networks

6. Other Research
Work on Papers from Ph.D. Research:

1. “A multiscale method for epitaxial growth”, SIAM Multiscale Model. Simul, **9**, pp.335–354, 2011.

Other Research started or continued at SAMSI:

1. “Numerical study of singularity formation in relativistic Euler flows”. submitted to SIAM J. Sci. Comput., 2011
2. Coupling boundary conditions for numerical PDEs.

Continuing Collaborations while at SAMSI:

1. “Coarse-grained event tree analysis for quantifying Hodgkin-Huxley neuronal network dynamics”, accepted, in press, J. Comput. Neurosci.

Presentations of Other Research:

1. Present a poster at IMA workshop, Nov.1-4, 2010.
2. Two minisymposia presentation in The 7th International Congress on Industrial and Applied Mathematics (ICIAM 2011), Vancouver, Canada, July 2011.

Research Progress Report & SAMSI Program Final Report

Date:

7a-z. Research Contributions – Current Projects (grouped by Working Group)
Research Project Title, Collaborator(s) & Mentor(s), Specific Goals & Accomplishments (results), Research Contributions (publication submissions, articles in preparation, etc.):

1. (with Pierre Gremaud), Numerical study of singularity formation in relativistic Euler flows. submitted to SIAM J. Sci. Comput., 2011
2. (with Pierre Gremaud), Coupling boundary conditions for numerical PDEs. (in progress).
3. (with Ilya Timofeyev and Cory Hauck) Traffic flow model. (in progress).

Presentations outside SAMSI (including invitations for future talks):

1. Present a poster at IMA workshop, Nov.1-4, 2010.
2. Two minisymposia presentation in The 7th International Congress on Industrial and Applied Mathematics (ICIAM 2011), Vancouver, Canada, July 2011.

8. Future Research Plans (after completion of SAMSI Program)
Research Area – Plans:

1. Coupling boundary conditions for numerical PDEs with applications in blood flow simulations.
2. Traffic flow model, traffic network model.

Continuing Collaborations (if appropriate):

1. with Pierre Gremaud
2. with Ilya Timofeyev and Cory Hauck.

Presentations outside SAMSI:

SAMSI Program Final Report

Name: Sylvie Tchumtchoua

Date: June 16, 2011

7a-z. Research Contributions – Current Projects (grouped by Working Group)

Working Group I: Hierarchical Methods for Object Data

Research Project Title: Online Variational Bayesian Inference in Hierarchical Models for Correlated High-dimensional Data

Collaborator(s) & Mentor(s): Jeffrey Morris/ David Dunson

Specific Goals & Accomplishments (results)

High-dimensional data with hundreds of thousands of observations are becoming commonplace in many disciplines. The analysis of such data poses many computational challenges, especially when the observations are correlated over time and/or across space. In this paper we propose flexible hierarchical regression models for analyzing such data that accommodate serial and/or spatial correlation. We address the computational challenges involved in fitting these models by adopting an approximate inference framework. We develop an online variational Bayes algorithm that works by incrementally reading the data into memory one slice at a time. The performance of the method is assessed through simulation studies. We applied the methodology to analyze signal intensity in MRI images of subjects with knee osteoarthritis, using data from the Osteoarthritis Initiative.

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

Presentations outside SAMSI (including invitations for future talks):
The paper will be presented at the Joint Statistical Meetings in Miami, Florida.

Working Group I: Brain Imaging

Research Project Title: A Heterogeneous Dynamic Structural Equation Model with Application to Brain Connectivity

Collaborator(s) & Mentor(s): Jeffrey Morris/ David Dunson

Specific Goals & Accomplishments (results):

Structural equation models (SEM) are frequently used to investigate relationships between one or more independent variables and one or more dependent variables. In this paper we are concerned with the problem of investigating relationships between data that can be of mixed type. Such data arise in various fields. For instance, in psychology MRI scans are often obtained at the same time as socio-demographic data and behavioral measurements with interests lying in identifying the effects of factors such as age, experimental conditions, or other covariates on various regions of the brain. Alternatively, one could be interested in investigating causal interactions between two or more brain regions. SEM has been used to achieve these goals but existing literature is at the region of interest level and focuses on one subject at a time. Within region of interest spatial correlation is typically ignored and regions of interest have to be predefined. In addition, existing literature estimates a single set of relationships between regions of interest, which ignores heterogeneity among subjects in multiple-subject settings.

The purpose of this research is to propose a heterogeneous SEM where the measurement model is specified as a mixture model which helps identify the regions of interest by classifying the voxels. The structural model is formulated as a heterogeneous multivariate autoregressive model with the relationship between regions of interest allowed to vary with covariates. Inference is conducted in a Bayesian framework using an online variational Bayes procedure for approximating the posterior distributions of the model parameters.

We are currently assessing the performance of the method via simulation studies. The methodology will be exemplified through a study of age and experimental condition related changes in brain activity.

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

Presentations outside SAMSI (including invitations for future talks):

8. Future Research Plans (after completion of SAMSI Program)

Research Area – Plans:

Our plan is to finish the paper "A Heterogeneous Dynamic Structural Equation Model with Application to Brain Connectivity" by the end the summer.

Continuing Collaborations (if appropriate):

Presentations outside SAMSI:

SAMSI Postdoctoral Fellow Research Plan

Date: Jun 14, 2011

1. Name

Hongxiao Zhu

2. Ph.D. Program
University & Department: Rice University, Department of Statistics
Dissertation Advisor: Dennis D. Cox
Year of Ph.D.: 2009

3. SAMSI Research
SAMSI Program Title: Analysis of Object Data (AOD) 2010-2011
SAMSI Research Mentor: David B.Dunson
SAMSI Administrative Mentor: Richard Smith

4a. SAMSI Activities
Course(s) (fall & spring):
Analysis of Object Data, Fall course

Workshops Attended (and Workshop Support Tasks):
AOD Workshop, Sept. 12-15,2010
AOD Transition Workshop, Jun. 9-11, 2011

Postdoc-Grad Student Seminar – Presentation(s):
Mar. 2, 2011, Title: Robust, Adaptive Models for Complex Functional Data

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

The two day undergraduate workshop, Oct. 29-30, 2010, provided R tutoring session.
The two day undergraduate workshop, Feb. 25-26, 2011, provided a tutoring session on functional object data.
The Interdisciplinary Workshop for Undergraduate Students and Faculty, May 16-20, 2011, provided tutoring session on Introduction to Basic Statistics.

4b. Other Activities (e.g., teaching)

Not scheduled.

5a. Working Group I

Special Tasks for Working Group: Statistical Inference for Functional Data. I primarily focused on the topic of model selection in functional data analysis. The work is under the supervision of Dr. Fang Yao and Dr. Helen Hao Zhang.

Presentations to Working Group:

Variable Selection in Functional Data Context, presented in the group meeting, Sept. 22, 2010.
Component Selection in Functional Additive Models, presented in the group meeting, May 24, 2011.

Research Area – Plans:

In this working group, I primary focused on developing novel statistical methods for model selections/variable selections in functional data analysis. Model selection has been well developed in recent years for regression problems when the number of covariates is excessively large. Brilliant proposals such as LASSO, SCAD have been intensively studied and their properties have been well explored. In functional data analysis, especially when the functions contain massive, redundant information, it is also desirable to obtain sparse estimation or select regions or part of the function for effective inference.

In this research, we consider the model selection in functional data regression. We focus on the case when the predictor is a function and the response is a scalar. We propose a framework that couples functional additive model (FAM) with component selection, in which the additive functional components are penalized through the Component Selection and Smoothing Operator (COSSO), so that the ``unimportant" additive components are shrunk towards zero. Different from the initially proposed FAM, in our proposed method, the FAM was constructed in the context of Reproducing Kernel Hilbert Space (RKHS), which can be decomposed to finite number of orthogonal subspaces. The orthogonal subspaces are assumed to be the second order Sobolev Hilbert Spaces, although more general assumptions is applicable. The component selection is achieved through penalized least squares, with the penalty chosen as the sum of the component norms, while conditioning on the functional principal component (FPC) estimates of the functional predictors. The proposed methods naturally facilitate a two-step algorithm which combines a FPC step and a component selection step. Asymptotic properties of the new estimator are studied, and extensive simulation studies are performed to assess the performance of the new method. We finally apply the method to the tecator data set, where we used the near infrared absorbance spectra to predict the contents of protein in meat samples.

Recent update of the research progress: we will have the draft ready by Jul. 2011. We plan to get the paper submitted by Aug. 2011.

5b. Working Group II

Special Tasks for Working Group: Hierarchical methods for functional objects.

Presentations to Working Group:

Working group presentation: Robust Classification of Functional Data using Functional Mixed Models, Apr. 1, 2011.

Transition workshop presentation: Robust Classification of Functional and Quantitative Image Data using Functional Mixed Models, Jun. 9, 2011.

Research Area – Plans:

In this working group, I focus on two projects. One is on Bayesian graphical models for multivariate functional data. The other is the modeling of serially correlated FMRI data using functional mixed models. The first project is under the supervision of Professor David Dunson. The second is under the supervision of Professor Jeff Morris.

Project A: In a broad variety of application areas there is interest in inferring the dependence structure in multivariate functional data. For vector data, conditional independence relationships can be inferred through allowing zeros in the precision matrix through a Gaussian graphical model. Bayesian methods can be used to allow unknown locations of zeros, while placing a Blocked Hyper Inverse-Wishart prior on the non-zero elements of the precision matrix. We generalize these methods to define a new class of Gaussian process graphical models. A simple case corresponds to a separable covariance structure with a single precision matrix encoding the conditional independence structure, but we also define a general class of non-separable models. Properties of the proposed process are considered, and efficient methods are developed for posterior computation relying on Markov chain Monte Carlo algorithms. The methods are evaluated through simulation studies and applied to an epidemiology application. I have finished formulizing the model and prove some properties. I will work on the design of MCMC algorithm, coding and simulations over the summer. The draft is expected to be finished by the end of Aug., 2011.

Project B: Modeling of a FMRI data set from nicotine withdrawal clinical trial. This project is data-oriented. The data comes from a complex design of nicotine withdrawal clinical trial. In the experiment, 72 subjects randomly receive 3 types of treatment. For each subject, the FMRI and EEG signal are collected at 2 time points-- a baseline measurement is taken before the treatment and a post measurement is taken after two months of the treatment. In each measurement, 4 image types are collected, they are: nicotine, positive, negative, and neutral. For each image type, 16 images are collected. Therefore there are totally 16*4=64 images per measurement, which are time correlated. We'd like to model the effect of nicotine withdraws treatments based on the FMRI images.

We will adopt the functional mixed model (FMM) method and extend it to incorporate serial correlations. The FMM is a flexible framework which can accommodate various random and fixed effect. The introduction of serial correlation through the residual will help us model the effect of time. I have finished the derivations of the extension. I expect to have some coding done by Aug. 2011.

5c. Working Group III (if appropriate)
Special Tasks for Working Group: NA
Presentations to Working Group: NA
Research Area – Plans: NA

6. Other Research

Other Research started or continued at SAMSI: As I am half sponsored by Professor Jeff S. Morris for the academic year of 2010-2011, I continued working with Dr. Morris on the research of functional mixed models. We have finished the revision of the submitted papers on robust functional mixed models. It has now been accepted by the Journal of American Statistical Association. Another paper on functional mixed model classifications is in submission to Biometrics.

Continued Collaborations while at SAMSI: Some of my other collaborations included the work with Dr. Fengrong Wei(University of West Georgia) on the Group Coordinate Descent Algorithms for Nonconvex Penalized Regression, currently under revision; the work with Dr. Xiaowei Wu (University of Chicago) on a Bayesian Analysis of Copy Number Variations in Array CGH Data, currently in submission. I will also continue working with Dr. Dennis D. Cox (Rice University) on the research of covariance operators. I expect to get the paper submitted by Aug. 2011.

Presentations of Other Research: None.

SAMSI Postdoc Fellowship Evaluation Questionnaire

NAME:___David Degras____________________________________

1. Program Involvement:
Which SAMSI program(s) have you been involved with and at what level(s)? (e.g., "been doing active research on …" or "just went to a tutorial/workshop to learn about that area which is new to me")
I was a postdoc in the AOOD program at SAMSI (2011). I have been doing active research in two working groups and attended the meetings of a third. I have given talks and organized student labs in SAMSI workshops (both Education and Outreach & Research).

2. Interactions with Other Institutions:
Describe your interactions with other institutions while at SAMSI. (e.g., a Triangle university, NISS or CRSC)
I was affiliated with the Department of Mathematics at NCSU. I attended the seminars of the Department of Statistics (NCSU) on a regular basis.

3. High Points at SAMSI:
What have been the high points of your SAMSI experience?
- the opening workshop: all the talks in different fields and the wonderful opportunities to initiate new research projects and collaborations.
- the variety of working groups: unique occasion to be exposed to new fields with leading specialists.
- the availability of the researchers during their visit at SAMSI and after

4. Suggestions for Improvement:
How could your SAMSI experience have been improved?

In my case (initial 1 year postdoc contract), have the directorate tell me clearly and early enough whether my postdoc can be extended for another year. That would have avoided a painful misunderstanding that resulted in my very late jumping in the job market.

5. Mentoring:
SAMSI aims to provide solid mentoring of Postdocs. How successful has this been in your case?

I have had an amazing scientific mentor who helped me not only with scientific questions but also at all stages of my job search and regarding academic life. Heartfelt thanks to Giles Hooker!

6. SAMSI Benefits for the Future:
Has your SAMSI experience met your expectations in preparing you for your next position? Are there (new) benefits that you now perceive coming from your SAMSI experience that are relevant to your career in the future?

The answer to both questions is a big YES.

7. SAMSI in contrast with University Setting:
How do you think your SAMSI experience compares to what you might have encountered in a typical university setting?

SAMSI has the advantage of concentrating top experts from various fields in its programs. The opportunities for networking and scientific collaborations are much larger than in the typical university environment. Also, the postdocs have no teaching duties and very few obligations in their first year, which boosts the scientific productivity. Postdoc salaries are very competitive as well.

8. SAMSI in comparison to Other Experiences:
If you have had experience in an industrial, national or other lab setting, how would you compare that experience with your SAMSI experience?

An industrial/national lab experience would probably have been more focused in terms of possibilities for research projects and number of collaborators. In my view, SAMSI allows to open up and learn in new fields whereas an industrial/lab experience is more suitable for researchers who already have experience in a given field and seek specialization.

9. Other Research while at SAMSI:
If you have spent considerable time during your appointment with some other research group, please comment on the benefits and drawbacks of this experience. (e.g., if your appointment was extended via an appointment at NISS, CRSC or elsewhere)

NA

10. Other Issues:
Are there other issues or concerns you would like to bring up?

No.

SAMSI Postdoc Fellowship Evaluation Questionnaire

NAME:____Oliver R Diaz Espinosa_____________________________________

1. Program Involvement:
Which SAMSI program(s) have you been involved with and at what level(s)? (e.g., "been doing active research on …" or "just went to a tutorial/workshop to learn about that area which is new to me")
Stochastic Dynamics and Complex Networks.
I was involved in the SD dynamics by doing active research in the groups Qualitative Behavior group and in the Biological Dynamics group/ Particle systems subgroup.

I did not have time to participate more actively in the second program due to my teaching schedule and also because I focused all my energy to make progress in the projects that I started during my first year.

2. Interactions with Other Institutions:
Describe your interactions with other institutions while at SAMSI. (e.g., a Triangle university, NISS or CRSC) My interaction was very positive at with the friendly and helpful staff at NISS-SAMSI. With regards the triangle universities, I found several collaborators at Duke and UNC and got invited a couple of times to give talks at NCSU. I liked the relaxed atmosphere in each of those institutions and the high quality of their facilities.

3. High Points at SAMSI:
What have been the high points of your SAMSI experience?
Collaborating with other researches and finding a couple of good problems that I believe are doable and which I expect to complete soon.

4. Suggestions for Improvement:
How could your SAMSI experience have been improved?
I think the initial meetings to discuss projects and form the subgroups for the research projects should take less time. That way one can start to work in something more concrete much faster.

5. Mentoring:

SAMSI aims to provide solid mentoring of Postdocs. How successful has this been in your case?
Academically my mentoring was very good, I learned a lot from my academic mentor and other collaborators.
My only concern is with the mentoring for other matters. such as writing grants, job search and all that was null the first year and only in my second year, towards the end of the spring semester some mentoring of that type occurs.

I think that SAMSI should reach out also to some of the several Industries around the triangle area. That would open other opportunities for people who may want to transition from academia to industry.

6. SAMSI Benefits for the Future:
Has your SAMSI experience met your expectations in preparing you for your next position? Are there (new) benefits that you now perceive coming from your SAMSI experience that are relevant to your career in the future?

I am prepared to achieve my next goal and the preparation that I got at SAMSI will be useful to that goal.

7. SAMSI in contrast with University Setting:
How do you think your SAMSI experience compares to what you might have encountered in a typical university setting?

The main advantage is that it is a center where people from different places meet and in that way it is great to develop scientific connections.
The only disadvantage is that is a little isolated in comparison to other institutes such as the Fields or IMA.

8. SAMSI in comparison to Other Experiences:
If you have had experience in an industrial, national or other lab setting, how would you compare that experience with your SAMSI experience?
NA

9. Other Research while at SAMSI:
If you have spent considerable time during your appointment with some other research group, please comment on the benefits and drawbacks of this experience. (e.g., if your appointment was extended via an appointment at NISS, CRSC or elsewhere)
NA

10. Other Issues:
Are there other issues or concerns you would like to bring up?

It was a positive experience for me and I will have fond memories of my days at SAMSI.
I hope to come back one day for a visit.

SAMSI Postdoc Fellowship Evaluation Questionnaire

NAME:____Ci-Ren Jiang___________________________________

1. Program Involvement:
Which SAMSI program(s) have you been involved with and at what level(s)? (e.g., "been doing active research on …" or "just went to a tutorial/workshop to learn about that area which is new to me")

AOOD
I have been doing research in Brain Imaging (collaborating with John Aston and Jane-Ling Wang) and Hierarchical Method (collaborating with Jeff Morris) working groups.

2. Interactions with Other Institutions:
Describe your interactions with other institutions while at SAMSI. (e.g., a Triangle university, NISS or CRSC)

I have continued the two projects initiated at UC Berkeley (with Haiyan Huang, Peter Bickel and Jessica Li). Now, both two projects are about to be submitted.

3. High Points at SAMSI:
What have been the high points of your SAMSI experience?

The active workshops brought a lot of visitors with interesting topics.

4. Suggestions for Improvement:
How could your SAMSI experience have been improved?

The officemate system does not work well. For example, if you have to discuss with your collaborators, you need to meet in a seminar room.

5. Mentoring:
SAMSI aims to provide solid mentoring of Postdocs. How successful has this been in your case?

It works ok.

6. SAMSI Benefits for the Future:

Has your SAMSI experience met your expectations in preparing you for your next position? Are there (new) benefits that you now perceive coming from your SAMSI experience that are relevant to your career in the future?

SAMSI experience definitely benefits much. By participating SAMSI programs, I learned more about the fields that I am interested in and I am more prepared for doing independent research or collaborating with others. However, one year is a bit short as one just starts to get familiar with the research topics.

7. SAMSI in contrast with University Setting:
How do you think your SAMSI experience compares to what you might have encountered in a typical university setting?

I also had postdoc experience with UC Berkeley. I think it works differently. At university, you will focus on the topic that your supervisors are interested and you might have little time to do your own research. It might be a productive year. But, at SAMSI you can concentrate on your own research and you might initiate new research projects with any visitors. SAMSI provides great opportunities for collaboration. Both are great and it depends on which one you prefer.

8. SAMSI in comparison to Other Experiences:
If you have had experience in an industrial, national or other lab setting, how would you compare that experience with your SAMSI experience?

Answered in question 7.

9. Other Research while at SAMSI:
If you have spent considerable time during your appointment with some other research group, please comment on the benefits and drawbacks of this experience. (e.g., if your appointment was extended via an appointment at NISS, CRSC or elsewhere)

Continuing the projects from UC Berkeley definitely took some time and might distract me from current SAMSI projects. Therefore, it slowed things down. But, I get to have further collaborations (new project) with previous collaborators.

10. Other Issues:
Are there other issues or concerns you would like to bring up?

It is a bit inconvenient to borrow books as SAMSI postdocs are affiliated with different universities. To borrow a book, I will need to drive to NCSU. Other than that, I think it’s a great place.

SAMSI Postdoc Fellowship Evaluation Questionnaire

NAME:_Junheng Ma____

1. Program Involvement:
Which SAMSI program(s) have you been involved with and at what level(s)? (e.g., "been doing active research on …" or "just went to a tutorial/workshop to learn about that area which is new to me")

I have been involved with SAMSI AOOD (Analysis of Object Data) program. As a webmaster of Statistical Inference subgroup, I have been focused on clustering for functional and longitudinal data.

2. Interactions with Other Institutions:
Describe your interactions with other institutions while at SAMSI. (e.g., a Triangle university, NISS or CRSC)

I have been working with several faculties in other universities on different projects, such Prof. Joe Sedransk and Prof. Jiayang Sun at Case Western Reserve University, Prof. Jiguo Cao at Simon Fraser University.

3. High Points at SAMSI:
What have been the high points of your SAMSI experience?

The most enjoyable part of my SAMSI experience is the opportunities to meet and work with some leading researchers in my area. I really appreciate those great opportunities.

4. Suggestions for Improvement:
How could your SAMSI experience have been improved?

My experience at SAMSI is invaluable and beyonds my expectations. Everything here at SAMSI is well organized. I couldn't think of anything to be improved.

5. Mentoring:
SAMSI aims to provide solid mentoring of Postdocs. How successful has this been in your case?

In my case, the mentoring provided by SAMSI is superb. I have been given enough freedom to do my research and valuable advice on my career plan.

6. SAMSI Benefits for the Future:
Has your SAMSI experience met your expectations in preparing you for your next position? Are there (new) benefits that you now perceive coming from your SAMSI experience that are relevant to your career in the future?

Definitely, I have a wonderful experience at SAMSI. My SAMSI experience will be very helpful for my next position. Except conducting high-level research with leading researchers, SAMSI also provides great opportunities for me to get into some new areas.

7. SAMSI in contrast with University Setting:
How do you think your SAMSI experience compares to what you might have encountered in a typical university setting?

The biggest difference between SAMSI experience and the experience in a typical university setting is at SAMSI one has the opportunity to choose the projects and collaborators according to his/her own interests.

8. SAMSI in comparison to Other Experiences:
If you have had experience in an industrial, national or other lab setting, how would you compare that experience with your SAMSI experience?

Although I don't have experience other than with SAMSI, my experience at SAMSI beyonds my expectation and I benefit a lot from this wonderful experience.

9. Other Research while at SAMSI:
If you have spent considerable time during your appointment with some other research group, please comment on the benefits and drawbacks of this experience. (e.g., if your appointment was extended via an appointment at NISS, CRSC or elsewhere)

I have been focused on my research at SAMSI during my appointment and don't have any appointments with other research groups.

10. Other Issues:
Are there other issues or concerns you would like to bring up?

SAMSI Postdoc Fellowship Evaluation Questionnaire

NAME:_Bruce Rogers_____________

1. Program Involvement:
Which SAMSI program(s) have you been involved with and at what level(s)? (e.g., "been doing active research on …" or "just went to a tutorial/workshop to learn about that area which is new to me")

I've been doing active research in Complex Networks program in the Sampling, Modeling, Inference and Dynamics OF working groups.

2. Interactions with Other Institutions:
Describe your interactions with other institutions while at SAMSI. (e.g., a Triangle university, NISS or CRSC)

I'm deeply involved in research at the RAMA robotics lab in the Duke Mechanical Engineering Department and Duke Network Analysis Center (mostly sociology). Also, I've been working with the Political Blog group at UNC.

3. High Points at SAMSI:
What have been the high points of your SAMSI experience?

The workshops are incredible. The smartest people in the world come to me.

4. Suggestions for Improvement:
How could your SAMSI experience have been improved?

At the end of the year, Richard Smith gave a very nice talk on "How to give a research talk." It would have been helpful to hear that at the START of the year, before all the post-doc seminars.

5. Mentoring:
SAMSI aims to provide solid mentoring of Postdocs. How successful has this been in your case?

Peter Mucha has been an incredible mentor.

6. SAMSI Benefits for the Future:
Has your SAMSI experience met your expectations in preparing you for your next position? Are there (new) benefits that you now perceive coming from your SAMSI experience that are relevant to your career in the future?

I feel that I'm ready to be a decent researcher. However, I want to be an educator first and a researcher second.

7. SAMSI in contrast with University Setting:
How do you think your SAMSI experience compares to what you might have encountered in a typical university setting?

I'm not sure; I've spent most of my training at institutions that encourage Centers and seminars (with little classroom work). So I don't know what a "typical University Setting" is.

8. SAMSI in comparison to Other Experiences:
If you have had experience in an industrial, national or other lab setting, how would you compare that experience with your SAMSI experience?

The working groups are strange compared to other labs where the divisions are not so formal. The groups help bring like minded people together, but it's always nice to encounter a non-like minded person to bring in fresh ideas. That is to say, there's not often alot of cross-fertilization between working groups (some of that is the fault of the post docs)

9. Other Research while at SAMSI:
If you have spent considerable time during your appointment with some other research group, please comment on the benefits and drawbacks of this experience. (e.g., if your appointment was extended via an appointment at NISS, CRSC or elsewhere)

I've spent lots of time with other groups (see #2), and essentially, SAMSI made those interactions possible through chance encounters.

10. Other Issues:
Are there other issues or concerns you would like to bring up?

No, thank you.

SAMSI Postdoc Fellowship Evaluation Questionnaire

NAME: David Sivakoff

1. Program Involvement:
Which SAMSI program(s) have you been involved with and at what level(s)? (e.g., "been doing active research on …" or "just went to a tutorial/workshop to learn about that area which is new to me")

I have been doing active research in the Complex Networks program. Specifically, in the Dynamics OF Networks and Dynamics ON Networks working groups.

2. Interactions with Other Institutions:
Describe your interactions with other institutions while at SAMSI. (e.g., a Triangle university, NISS or CRSC)

I attended seminars frequently and worked closely with people in the Mathematics Department at Duke University

3. High Points at SAMSI:
What have been the high points of your SAMSI experience?

I very much enjoyed working with the undergraduates at both of the undergraduate workshops with which I was involved (October and May). Also, the working groups that I was a member of, as well as the corresponding workshops, were fantastic sources for project ideas, and great places to interact with researchers with diverse backgrounds who were all interested in the same topic.

4. Suggestions for Improvement:
How could your SAMSI experience have been improved?

5. Mentoring:
SAMSI aims to provide solid mentoring of Postdocs. How successful has this been in your case?

I think that the mentoring aspect of the program was quite good. I have worked (and will continue to work) closely with my mentor, who has advised me not only about my research, but also paper writing, grant writing, conferences and presentations.

One comment that I would like to make, though: In the postdoc seminar, Richard Smith gave a great presentation about how to give a good scientific talk -- I think this should have been one of the first presentations of the year, rather than one of the last. I think the grant-writing seminar could also have appeared earlier.

6. SAMSI Benefits for the Future:
Has your SAMSI experience met your expectations in preparing you for your next position? Are there (new) benefits that you now perceive coming from your SAMSI experience that are relevant to your career in the future?

Yes, though my next position is to remain at Duke for another 1-2 years. SAMSI was a great place for me in terms of networking with people outside of my immediate field (Probability) that are working in similar research areas.

7. SAMSI in contrast with University Setting:
How do you think your SAMSI experience compares to what you might have encountered in a typical university setting?

I enjoyed that we had more long-term visitors at SAMSI, though I occasionally felt a bit isolated at SAMSI relative to how I feel in a university setting.

8. SAMSI in comparison to Other Experiences:
If you have had experience in an industrial, national or other lab setting, how would you compare that experience with your SAMSI experience?

N/A

9. Other Research while at SAMSI:
If you have spent considerable time during your appointment with some other research group, please comment on the benefits and drawbacks of this experience. (e.g., if your appointment was extended via an appointment at NISS, CRSC or elsewhere)

I liked that while at SAMSI I could attend seminars and work with people at Duke and UNC. The biggest drawback is that SAMSI is located far from any of these universities.

10. Other Issues:
Are there other issues or concerns you would like to bring up?

Thank you to the staff for making this a great year at SAMSI!

SAMSI Postdoc Fellowship Evaluation Questionnaire

NAME: Sylvie Tchumtchoua

1. Program Involvement:
Which SAMSI program(s) have you been involved with and at what level(s)? (e.g., "been doing active research on …" or "just went to a tutorial/workshop to learn about that area which is new to me")

I have been involved with the AOOD program and have done active research related to the Hierarchical methods for object data and Brain imaging working groups.

2. Interactions with Other Institutions:
Describe your interactions with other institutions while at SAMSI. (e.g., a Triangle university, NISS or CRSC)

I did not interact with other institutions.

3. High Points at SAMSI:
What have been the high points of your SAMSI experience?

Experience with new research areas; interactions with some of the most prominent people in the statistical field; a quiet working environment.

4. Suggestions for Improvement:
How could your SAMSI experience have been improved?

5. Mentoring:
SAMSI aims to provide solid mentoring of Postdocs. How successful has this been in your case?

My mentors were very helpful; they were always available when I had a question and always gave useful comments on my work.

6. SAMSI Benefits for the Future:
Has your SAMSI experience met your expectations in preparing you for your next position? Are there (new) benefits that you now perceive coming from your SAMSI experience that are relevant to your career in the future?

My SAMSI experience has strengthened my independent research skills.

7. SAMSI in contrast with University Setting:

How do you think your SAMSI experience compares to what you might have encountered in a typical university setting?

The possibility to meet people from various backgrounds makes the SAMSI experience unique.

8. SAMSI in comparison to Other Experiences:
If you have had experience in an industrial, national or other lab setting, how would you compare that experience with your SAMSI experience?

I do not have experience in an industrial, national or other lab setting.

9. Other Research while at SAMSI:
If you have spent considerable time during your appointment with some other research group, please comment on the benefits and drawbacks of this experience. (e.g., if your appointment was extended via an appointment at NISS, CRSC or elsewhere)

I did not have an appointment with some other research groups.

10. Other Issues:
Are there other issues or concerns you would like to bring up?

No.

SAMSI Postdoc Fellowship Evaluation Questionnaire

NAME: Hongxiao Zhu

1. Program Involvement:
Which SAMSI program(s) have you been involved with and at what level(s)? (e.g., “been doing active research on …” or “just went to a tutorial/workshop to learn about that area which is new to me”)

I have been doing active research in the 2010-2011 program on Analysis of Object Data (AOD).

2. Interactions with Other Institutions:
Describe your interactions with other institutions while at SAMSI. (e.g., a Triangle university, NISS or CRSC)

I have been collaborating from local faculties from NCSU (Dr. Helen Zhang), Duke (Dr. David Dunson) and University of Toronto (Dr. Fang Yao). I have also attended department seminars at Duke Universities.

3. High Points at SAMSI:
What have been the high points of your SAMSI experience?

(1) I have the opportunities to know many prestigious researchers from the world.
(2) I have the opportunities and flexibilities to work on topics of my best interests(which I either do not have the chance to do, or don't have good enough idea to do before I came here) through the working groups.
(3) The atmosphere of Triangle area is very encouraging. I get the chance to know many young researchers in the similar situation as me. Interaction with them helped me learned a lot on how to initiate my career.

4. Suggestions for Improvement:
How could your SAMSI experience have been improved?

The following are some suggestions on improving the interactions between SAMSI and the nearby universities:
(1) An improved orientation would be helpful. Not only an orientation on how the program would run, but also orientations held by the nearby universities (especially the department related), providing information on where to park, what resources they can provide, and how to use them.
(2) A email system that sending out seminar/conference information from all the nearby universities would be helpful.

5. Mentoring:
SAMSI aims to provide solid mentoring of Postdocs. How successful has this been in your case?

So far the mentoring went pretty smooth. My academic advisor, Dr. David Dunson has been meeting me regularly to talk about my research progress. We have been collaborating together too. My administrative advisor, Dr. Richard Smith has always been very helpful on providing suggestions on many aspects of my life.

6. SAMSI Benefits for the Future:
Has your SAMSI experience met your expectations in preparing you for your next position? Are there (new) benefits that you now perceive coming from your SAMSI experience that are relevant to your career in the future?

Yes. The major benefits that SAMSI provide is to enlarge the scope of my insight. I learned a lot from working on SAMSI projects and earned plenty of research experiences. I believe all these are priceless for my future career.

7. SAMSI in contrast with University Setting:
How do you think your SAMSI experience compares to what you might have encountered in a typical university setting?

In SAMSI, there are more flexibilities on what to work on, and who to work with. In a typical university setting, once a post-doc is hired, he/she either have to work with one boss, or have to work on a particular project.

8. SAMSI in comparison to Other Experiences:
If you have had experience in an industrial, national or other lab setting, how would you compare that experience with your SAMSI experience?

I had the experience of working in a cancer center before. Compared with experiences in SAMSI, I believe the SAMSI is much more open and flexible. I would expect to have chance coming back as visiting scholars in the future.

9. Other Research while at SAMSI:
If you have spent considerable time during your appointment with some other research group, please comment on the benefits and drawbacks of this experience. (e.g., if your appointment was extended via an appointment at NISS, CRSC or elsewhere)

N/A

10. Other Issues:
Are there other issues or concerns you would like to bring up?

Overall, I very much like my experience of staying in SAMSI. The SAMSI staff are all very friendly and helpful. They work very hard to make SAMSI a good experience for visitors. I am spreading SAMSI's new program info to people I know. Hopefully more people will know SAMSI and will join the research activities here.

I.C. Graduate Fellows

1 Analysis of Object Oriented Data

Matthew Avery (NCSU-Statistics, advisor: Helen Zhang)

Matt was a member of the functional data analysis working group. His research focuses on new methods for model selection with functional data. His approach is based on the fused LASSO method and works over a variety of different types of coefficient function. His approach shows better results than traditional functional regression methods for both prediction and model selection for simulated data. Perhaps most usefully, his work can be applied with any existing method resulting in an improved fit. Publications are in preparation.

Matt took two SAMSI AOD courses, participated in several workshops and assisted in the undergraduate graduate workshop in February 2011.

Sungkyu Jung (UNC-CH, Statistics and Operations Research, advisor: J.S. Marron)

Sungkyu was a member of the metrics on shape space working group His work is centered around the development of Principal Component Analysis (PCA) techniques for manifold-valued data, where the proposed methods intuitively capture major non-linear variations in lower dimension. Sungkyu has one paper currently submitted in the high dimension, low sample size context.

He has also been involved with Statistics on curved manifolds where particular applications include analysis of directional data, landmark-based shape data and medial representation of shapes for image analysis. Several papers are either submitted or in preparation.

Sungkyu took two AOD SAMSI courses. He is currently Assistant Professor at the Department of Statistics at the University of Pittsburgh.

Seonjoo Lee (UNC-CH., Statistics and Operations Research: advisors: Haipeng Shen, Young Truong)

Seonjoo was a member of the brain imaging working group. Her research focuses on developing independent component analysis (ICA), which can take into account the autocorrelation structures within the sources. In many applicatoins of ICA, the sources of interest have autocorrelation structures. Additionally, the future research includes combining spatial smoothing step into her ICA method with application to fMRI data. She has one related accepted paper and three papers in preparation.

She took part in three SAMSI workshops, took the Fall 2010 SAMSI courses on both AOD and Complex Networks. She is completing her PhD thesis on ICA on spectral domain.

Xiaojing Wang (Duke-Statistics, advisor: Jim Berger).

Xiaojing was a member of the hierarchical modeling group led by Jeff Morris. She works on the estimation of shape constrained functions using Gaussian Processes. The flexibility of these processes can be detrimental when the function is known to be in a shape-constrained class of functions. Gaussian processes with the squared exponential correlation function have the property that their derivatives are also Gaussian processes. This allows incorporation of shape constraints through placing constraints on the derivative process. This leads to a class of conditional Gaussian processes, which can be computationally handled via a Gibbs sampling scheme involving sampling from positive Gaussian random variables.

Xiaojing is completing her PhD work.

Bo Zhang (NCSU-Statistics, advisor: Lexin Li).

Bo was a member of the statistical inference for functional data working group. He took the Fall semester course on AOD and attended four SAMSI workshops.

2 Complex Networks

Prakash Balachandran (Duke-Mathematics, advisor: Mauro Maggioni)

Kash was an active member of both the Geometrical/Spectral Analysis working and group and the Modeling, Inference and Sampling working group. His research is centered around the theme of dimensionality reduction of networks data. He worked on two specific issues.

First, he was involved in multiscale analysis of graphs. Here, he's developed computer code to analyze the geometry of a network at multiple scales. He has developed a method to detect regions of low conductance in the network which are bottlenecks that trap the random walker. Upon detection, these regions are compressed out by constructing a non-Markov process that transitions between clusters preserving transitions of the original process between/within clusters. Current work involves iterating the procedure in order to obtain a multiscale decomposition of the transition density of the original Markov process in order to efficiently compute observables of the process at a fixed time.

The second issue involves research into local spectral clustering algorithms motivated by community detection of overlapping communities and the core/periphery problem. Current local spectral clustering algorithms deliver a cluster around a seed set of vertices of low conductance both when the network is directed and undirected. He has begun to compare various algorithms in order to understand the quality of their clusters in several different settings. It is hoped such empirical validation will aid in future research in using such algorithms to detect core/periphery structure in networks as well as overlapping communities in a controlled fashion.

Kash participated in four SAMSI workshops, gave a talk at the transition workshop, actively participated in two undergraduate events and is currently a postdoc at Boston University.

Yingbo Li (Duke-Statistics, advisor: David Banks)

Yingbo was a member of the sampling/modeling/inference working group. Her research is centered on a baboon network project (Dynamic Social Network Analysis of Baboon Troop Fission). Specific issues include reciprocity and transitivity. Her model can resolve the problem caused by unequal monthly observation intensity, explain the grooming interactions among baboons and capture the evolution of the grooming network. She has conducted a Markov chain Monte Carlo procedure to draw samples from the posterior distribution of latent positions, and aims at building the evolution across time.

She took part in several SAMSI workshops, was an active participant in the E&O activities and completing her PhD.

Michael Robert (NCSU-Mathematics, advisor: Alun Lloyd)

Michael was a member of the Dynamics ON/OF Networks. As part of his SAMSI involvement, Michael developed a canine social interaction network. Based on data he collected himself, he plans to study the structure of the static network and develop a dynamic network for studying disease spread. He is currently working toward the completion of his Ph.D. Dissertation which involves modeling of a vector-borne disease and strategies to control the disease, particularly by controlling the disease vectors.

He took part in five SAMSI workshops and was very actively involved in the 2011 week long undergraduate workshop on epidemiology.

Feng Bill Shi (UNC-CH, Mathematics, advisors: Peter Mucha and Greg Forest)

Bill was a member of the dynamics ON/OF working group. He works on robust scaling behavior in dynamics voter models. While there is a long history of study on dynamics of opinion formation on networks, little is known about the co-evolution of networks and opinions. He investigates a simple version of the dynamic voter model in which the dynamics of opinions are interwoven with the topology of the network and this interaction is controlled by a single parameter called rewiring probability. The results point to two types of robust scaling behaviors between the opinion density and the rewiring probability. In future study, he is currently working on a theoretical analysis of the dynamics of the system.

Bill took part in four SAMSI workshops, was an active participant in our SAMSI E&O activities and is completing his PhD. His work resulted in a *Proc. Nat'l. Acad. Sci.* paper. Other publications are under way.

Amanda Traud (NCSU-Mathematics, advisors: Alun Lloyd and Rob Dunn).

Mandi was a member of the Dynamics ON/OF Networks working groups. She is interested in communications over networks. She collaborated with Michael Robert, see above, on a canine social interactions model. She also performed a study of ant networks. More precisely, she looked into different patterns of communications between species for one genus, Formica. Using a network science approach, these studies compare observed networks of different species of Formica ants to networks simulated through Brownian motion. Mandi was also involved in the study of social structure of Facebook friendships, from which one published paper resulted.

She took the Fall Complex Networks course and was an invaluable help in the preparation and running two SAMSI undergraduate workshops. Mandi is currently completing her PhD work.

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.

I.E PROGRAM ACTIVITIES

1. **Analysis of Object Oriented Data**
2. **Complex Networks**
3. **Education and Outreach**

I.E.1: Final Report for the SAMSI Program: Analysis of Object Oriented Data (2010 - 2011)

Background

Modern science is generating a need to understand, and statistically analyze, populations of increasingly complex types. The term "Analysis of Object Oriented Data" (AOOD) is aimed at encompassing a broad array of such methods. This SAMSI program brought 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 to motivate and synthesize new approaches.

This was 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.

Organizers: Hans-Georg Müller (University of California, Davis), Jane-Ling Wang (University of California, Davis) (Co-Leaders); Ian Dryden (University of South Carolina), Steve Marron (University of North Carolina), Jim Ramsay (McGill University).

Local Scientific Coordinator: Steve Marron (UNC).

Directorate Liaison: James Berger (SAMSI); Richard Smith (after 7/1/10)

National Advisory Committee Liaison: Jiangqing Fan (Princeton).

1 Workshops

1.1 Opening Workshop September 12-15, 2010

1.1.1 Summary

The Opening Workshop for the SAMSI program on Analysis of Object Data (AOOD) was held on Sunday-Wednesday, September 12-15, 2010, at the Radisson RTP in Research Triangle Park, NC.

Tutorial sessions for each of five threads took place on Sunday, September 12. Invited talks were presented Monday to Wednesday. There was a poster session and reception on Monday, September 13. Immediately following the workshop, on Thursday and Friday, research working groups convened for initial meetings at SAMSI.

1.1.2 Activities

The workshop focused on five threads that exemplify the AOOD idea of generalizing the functional data analysis concept of random curves as data points to more general objects as data points. These include objects that are Euclidean, i.e., (constant length) vectors of real numbers, mildly non-Euclidean, i.e. points on a manifold and shapes, or strongly non-Euclidean, i.e. tree or graph structured objects. The five focus areas of the workshop were

1. Functional Data Analysis, theory and applications for samples of curve and surface data in the life, social, environmental and physical sciences and the interface with longitudinal data analysis;

2. Time Dynamics Data, with an emphasis on methodology for the analysis of observations that are governed by differential equations and the modeling of the dynamics of growth, gene expression, infections or online auctions;

3. Shape Analysis and Manifold Data, including the analysis of landmarks, curves, surfaces, and volumes and the analysis of data that naturally lie on a manifold such as directional data;

4. Image Analysis, with a focus on applications where data consist of a sample of images, including fMRI data, diffusion tensor imaging and diffeomorphisms for brain mapping;

5. Tree and Graph Structured Data, which are "strongly non-Euclidean" and require the development of statistical analysis from the ground up, to analyze samples of trees and graphs.

The workshop aimed at emphasizing synergies and interactions between these threads, and culminated in the formation of research working groups in the afternoon of Wednesday, September 15. These working groups then met individually at SAMSI on Thursday and Friday to further address specific research objectives to be addressed by the working group over the ensuing year. These meetings also established modes of cooperation for the working groups, via web or teleconference, to facilitate full participation of all members, regardless of residence status at SAMSI.

The workshop was full to capacity with about 150 participants. On the first day five tutorial lectures were held to provide extensive background of the five themes of the program. In particular the tutorials were:

- FDA Tutorial: Functional Data Analysis and Related Topics, Fang Yao, University of Toronto
- Dynamics Tutorial: Models for Output-Buffered Systems: An Introduction to Dynamics, Jim Ramsay, McGill University
- Images Tutorial: Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychology, Sarang Joshi, University of Utah and Martin Lindquist, Columbia University
- Shapes and Manifolds Tutorial: Shape Analysis and Manifold Data, John Kent, University of Leeds
- Trees Tutorial: Trees as Data, J.S. Marron, University of North Carolina

Videos of the expository lectures are available on the SAMSI website.

There then followed five half-day sessions on each of the themes. Each session consisted of lectures on future challenges from distinguished experts and also excellent contributions from new researchers.

- **Functional Data Analysis**

  Chair: Hans-Georg Müller, University of Calfornia, Davis.

  - Alois Kneip, University of Bonn, "Challenges for FDA Theory and Methodology."
  - Naisyin Wang, University of Michigan, "Challenges for FDA in Longitudinal Studies."
  - Shuang Wu, University of Rochester, " FDA for Non-traditional Data."
  - Ciprian Crainiceanu, John Hopkins University, " My first 100 Terabytes of Data: Challenges for FDA Modeling."

- **Dynamics**

  Chair: Jim Ramsay, McGill University.

  – Aaron King, University of Michigan, " Plug-and-play Inference for Stochastic Dynamical Systems."
  – Sy-Miin Chow, University of North Carolina, " Applications of Differential Equation Modeling in the Social Sciences."

- **Images**

  Chair: Jane-Ling Wang, University of California-Davis.

  – Dubois Bowman, Emory University, "Statistical Modeling of Brain Imaging Data: An Overview, Challenges, and Future Directions."
  – Laurent Younes, Johns Hopkins University, "Shape Analysis of Diastolic/Systolic Paired Cardiac Images."
  – Armin Schwartzman, Harvard School of Public Health, " Data Objects in Diffusion Tensor Imaging."
  – Jonathan Taylor, Stanford University, " Overview of Statistical Inference, Multiple Comparisons and Gaussian Random Fields for Neuroimaging Data."

- **Trees**

  Chair: J.S. Marron, University of North Carolina - Chapel Hill.

  – Burcu Aydin, Hewlett Packard Research, "Principal Component Analyses for Trees."
  – Sean Skwerer, University of North Carolina, "Analysis of Object Data: Averaging Metric Trees."
  – Shanker Bhamidi, University of North Carolina, " Dyck Path Correspondence and the Statistical Analysis of Brain Vascular Networks."
  – Haonan Wang, Colorado State University, "Smoothing and Branching Process Inference on Trees."

- **Shapes and Manifolds**

  Chair: Ian Dryden, University of South Carolina.

  – Huiling Le, University of Nottingham, " Some Aspects of Statistics on Riemannian Manifolds from the Perspective of Shape Analysis."
  – Anuj Srivastava, Florida State University, " Towards Statistical Modeling of Shapes of Curves and Surfaces."
  – Victor Panaretos, EPFL Lausanne, " Statistical Shape and Random Tomography in Structural Biology."
  – Alain Trouvé, Ecole Normale Supérieure de Cachan, " From Shape Comparison to Shape Evolution with a Geometrical and Statistical Perspective."

During the workshop there were a couple of Minute Madness sessions, where a large number of the participants had two minutes to present a topic of interest with a very strict time limit and a couple of slides. These lively sessions gave a showcase to the broad range of research interests of the participants. In addition, there was a poster session and reception, with a large number of excellent posters.

A key aspect of the opening workshop was the formation of the working groups, which was critical to the subsequent developments of the program.

The AOOD program was privileged to have such an outstanding workshop to set the scene for the year long research activities. Each theme's activities included presentations from some of the leading researchers in each area, from outstanding new researchers and involvement and interaction with a very knowledgeable and expert audience. The chairs of each session ensured that there was a lively amount of discussion throughout, and the workshop gave the perfect start to the program.

1.2 Workshop on the Interface Functional and Longitudinal Data Analysis - November 8-10, 2010 at SAMSI

1.2.1 General Description

Longitudinal and time-dynamic data are collected throughout the life sciences, social sciences and environmental sciences to explore time-dynamic aspects of phenomena that evolve over time. Longitudinal data analysis is a core technique of biostatistics, and has seen fast development over the last decades, primarily in the area of random effects modeling. Several researchers have begun more recently to explore the potential of adopting a functional perspective for the analysis of longitudinal data. The promise is that Functional Data Analysis provides flexible approaches dependence structures, includes derivatives and the analysis of dynamics, and permits the data to determine the most suitable model.

The workshop integrated these various perspectives and participants identified promising technology and prospective areas of interest for future research along this interface, emphasizing both methodology and applications. The workshop brought together researchers working in separate communities and explored and harnessed synergies between the various approaches. The workshop succeeded in generating lively discussions and engagement by all participants. Throughout these discussions, several key issues such as the role of functional principal components and the signficance of functional data analysis in biomedical longitudinal studies were thoroughly debated.

Organizers: Marie Davidian (N.C. State University), Fang Yao (University of Toronto), Hans-Georg Müller (Univ. of California-Davis)

Topics:

- Multivariate longitudinal/functional data
- Multilevel longitudinal/functional data
- Functional components in joint modeling of longitudinal and survival outcomes
- Longitudinal/functional data as components in regression models
- Longitudinal/functional models for dynamics
- Variable selection in longitudinal/functional models

1.2.2 Talks and Events

1. Ray Carroll, Texas A&M University: Generalized Functional Linear Models with Semiparametric Single-Index Interactions
2. Damla Şentürk, Penn State University: Functional Varying Coefficient Models for Longitudinal Data
3. Wensheng Guo, University of Pennsylvania: Functional Mixed Effects Spectral Models
4. Discussion and Connections to Working Groups
5. Tailen Hsing, University of Michigan: Uniform Convergence Rates for Principal Component Analysis in Functional/ Longitudinal Data
6. Jeff Morris, University of Texas: Adaptive, Robust Functional and Image Regression in Functional Mixed Models
7. Fang Yao, University of Toronto: Additive Modeling of Functional Regression and its Gradients
8. Discussion and Connections to Working Groups
9. Poster Session and Reception
10. Xihong Lin, Harvard University: Likelihood Based and Estimating Equation Based Methods for Variable Selection
11. Helen Zhang, North Carolina State University: Variance Component Selection in Linear Mixed Models
12. Graciela Boente, Universidad de Buenos Aires and CONICET: Robust Functional Principal Components: a projection-pursuit approach
13. Discussion and Connections to Working Groups
14. Jim Ramsay, McGill University: Economical Models for Functional Covariation
15. Jie Peng, University of California-Davis: Fitting Ordinary Differential Equation Models with Longitudinal Data
16. Huilin Wu, University of Rochester: Comparing Functional Data Analysis Approach and Nonparametric Mixed-Effects Modeling Approach for Longitudinal Data Analysis
17. Discussion and Connections to Working Groups
18. Ana-Maria Staicu, N.C. State University: Skewed Functional Processes and their Applications
19. Jimin Ding, Washington University: Time-varying Coefficient Cox model with Nonparametric Longitudinal Covariates
20. Discussion and Connections to Working Groups

1.3 Workshop on the AOOD Meets Evolutionary Biology – April 30-May 2, 2011 at SAMSI

1.3.1 General Description

The goal of this workshop was to build a bridge between activities in the SAMSI Analysis of Object Data Program, and the area of Function Valued Traits in Evolutionary Biology. The latter field already has strong ties with Functional Data Analysis, and this workshop was intended to define the new area of Object Valued Traits in Evolutionary Biology. Specific data contexts where the object valued viewpoint was specifically aimed included fly wing shapes (intersecting with the Shape and Manifold Research in AOOD), caterpillar growth trajectories (intersecting with Time Dynamics in AOOD) and rat brain blood vessel trees, plus families of viral phylogenetic trees (intersecting with the Tree Research in AOOD).

1.3.2 Highlights

Highlights of the meeting were:

- An Opening 2-Minute Madness Session, where each participant gave a 2 minute introductory talk. This enabled this diverse group of people to understand the breadth of attendees, and also to find people they had not previously known, to engage in individual conversations.
- A series of talks by biological leaders in Function Valued Traits, aimed at generating discussion among OODA researchers:
  - Joel Kingsolver, University of North Carolina, Genetic Variation and Evolution of Function-Valued Traits
  - Jay Beder, Univ. of Wisconsin-Milwaukee, Estimating the selection gradient of a function-valued trait
  - David Houle, Florida State University, Connecting Phenomic Objects to Genomic Predictors
  - Patrick Carter, Washington State University, Evolution of the Integrated Phenotype: A Function Valued Approach
  - Nancy Heckman, University of British Columbia, Dependence in Functional Data Analysis
  - Heather Jamniczky, University of Calgary, Quantification of Unusual Biological Shapes in Three Dimensions
  - Washington Mio, Florida State University, Spectral Methods in Shape Analysis
  - Saunak Sen, Univ. of California-San Francisco, Genetic mapping of function-valued traits
- A series of talks by OODA members on recently developed ideas related to evolutionary biology:
  - Daniel Gervini, University of Wisconsin, Semiparametric Curve Registration
  - Sarang Joshi, University of Utah, Towards Imaging Based Biomarkers
  - John Aston, Warwick University, A step towards a function-valued typology for language
  - J.S. Marron, University of North Carolina, OODA of Tree Structured Objects
  - Ezra Miller, Duke University, Stratified statistics for evolutionary biology

- At various points there were discussions at a number of levels, with excellent interactions between the two main groups.

1.4 Transition Workshop, June 9-11, 2011

1.4.1 Summary

The goals for this workshop were to review and discuss the progress made during the year on the various program projects and their synergies. Current status and future directions of research in the program were assessed. The workshop featured sessions on the five thematic areas of this program: functional data analysis, dynamics, data on manifolds, brain imaging and trees; and it brought together participants from the various working groups in these areas.

The organized discussions during the Workshop and the informal discussions after paper presentations revealed many important connections between these five themes. It was recognized that curve, surface and volume registration problems in functional data analysis and brain imaging were essentially shape analysis problems under some specific constraints, and a follow-up workshop at the next Joint Statistical Meetings was discussed as an outcome. The possibility of representing many functional and spatial data analysis problems as dynamic systems through ordinary and partial differential equations was frequently commented upon, and considerable progress along these lines during the year was outlined.

Although the phrase "strongly non-Euclidean" was often used to describe trees and other graphical models, in fact close links were seen with registration, shape analysis, and other themes represented elsewhere in the program. The overall objective of the SAMSI AOOD program of bringing together hitherto disparate research areas was well realized both, within the Transition Workshop and in new research programs launched during the year.

1.4.2 Activities

The program for the workshop was as follows:

- **Functional Data Analysis** *Hans-Georg Müller, organizer; Jane-Ling Wang, chair*
  - Jeff Morris, Hierarchical methods for the analysis of object data
  - Hongxiao Zhu, Robust classification of functional and quantitative image data using functional mixed models
  - Sylvie Tchumtchou, Online variational Bayesian inference in hierarchical models for correlated high-dimensional data
  - Discussion: Jane-Ling Wang
- **Dynamics** *J.O. Ramsay, organizer and chair*
  - Jiguo Cao, Quantitative trait loci mapping with differential equation models
  - David Campbell, Parameter estimation from locally enforced differential equation models
  - Hulin Wu, High-dimensional ODEs for dynamic gene regulatory networks
  - Jim Ramsay, Reflections on impacts and issues for statistical methodology for dynamic models generated by the AOOD project
- **Trees** *J.S. Marron, organizer and chair*

- Steve Marron, Driving example, background and research overview
- Sean Skwerer, Phylogenetic trees and stickiness
- Dan Shen, Dyck path and branch length analysis
- Lingsong Zhang, Non-negative matrix factorization approach to tree analysis
- Yongdai Kim, Thread bridging example: Pseudo-Bayesian factor analysis
- John Aston, Thread bridging example: Phylogenetic trees with dialects as leaves D
- Discussion: J. S. Marron, John Aston, Lingsong Zhang and Hernando Ombao

- **Shapes and Manifolds** *I. Dryden, organizer and chair*
  - Victor Patrangenaru, Statistical analysis of object data
  - Sungkyu Jung, Principal nested shape spaces and an application to reduction of the number of landmarks
  - Sebastian Kurtek, Registration of functional data using the Fisher-Rao metric
  - Jingyong Su, Fitting optimal curves to time-indexed, noisy observations of stochastic processes on nonlinear manifolds
  - Ian Dryden, Metrics, manifolds and geometric correspondence
  - Discussion: Ian Dryden, J. S. Marron, Victor Patrangenaru

- **Brain Imaging** *J-L. Wang organizer, John Aston and Hernando Ombao, chairs*
  - Heipeng Shen, Hemodynamic response function modeling
  - Tingting Zhang, Nonparametric inference of hemodynamic response for fMRI data with inhomogeneous variances through kernel smoothing
  - Ci-Ren Jiang, Nonparametric response function estimation via FPCA with an application to dynamic PET data
  - John Aston, Spatial functional data, temporal sequences and populations of change points for fMRI analysis
  - Seonjoo Lee, Independent component analysis for autocorrelated sources with an application to fMRI
  - Discussion: John Aston, Ian Dryden, Jeff Morris, Hernando Ombao, Heipeng Shen, Jane-Ling Wang

2 Courses and Workshop for Students

Two one semester courses for graduate students and a workshop for undergraduate students were offered under the auspices of this program.

2.1 Graduate Courses

- *Analysis of Object Data I*

  Fall Semester 2010

  Principal Instructors: I. Dryden, J.S. Marron, H.G. Müller, J. O. Ramsay, J.L. Wang

Lectures were given at SAMSI on Wednesdays, 4:30-7:00 p.m.

This was the first of two courses associated with the SAMSI program on Analysis of Object Data, offered for graduate students at University of North Carolina, Duke University and North Carolina State University. These courses provided an introduction into selected areas of object oriented data analysis, aiming at the following topics:

1. Introduction to Object Oriented Data (Marron)
2. Shape Analysis and Related Topics (Dryden) Introduction to Statistical Shape Analysis, Non-Euclidean Shape Spaces, Distances and Shape Co-ordinates, Procrustes Analysis, Principal Components Analysis and Geodesics.
3. Functional Data (Mller) Introduction to Functional Data, Functional Regression Models, Time Warping, Empirical Dynamics.
4. Functional and Longitudinal Data (Wang) Functional Principal Component Analysis, Modeling with Covariates, Interface Functional and Longitudinal Data Analysis.
5. Functional and Dynamic Data (Ramsay)

- *Analysis of Object Data II*

  Spring Semester 2011

  Principal Instructors: J. S. Marron, J. O. Ramsay

  Lectures were given at SAMSI on Wednesdays, 4:30-7:00 p.m.

  This was the second graduate course associated with the SAMSI program on Analysis of Object Data. It covered topics different from those covered in the course taught in the Fall and consisted of three main segments:

  1. Dynamic systems: (Ramsay) Models for multivariate functional data that explicitly model change through the involvement of one or more derivatives in the model specification. Topics include the anatomy of a dynamic system, dynamic systems as extensions of functional models, and parameter estimation and inference for systems where analytical solutions are impossible.
  2. Manifold data: (Ramsay and Marron) This topic extends functional data analysis in a number of ways, where data are typically distributed over time and/or space, to situations where the data are distributed over manifolds embedded within higher dimensional spaces. Along with a quick review of classic subjects such as principal components analysis and test theory, more advanced topics include medial shape representations, diffeomorphisms in image analysis, and diffusion tensor imaging.
  3. Tree-structured data: (Marron) This section contrasts the very diverse combinatorial, folded Euclidean and Harris correspondence approaches for this new area of data analysis.

2.2 Undergraduate Workshop

A two-Day Undergraduate Workshop was held at SAMSI on February 25-26, 2011, with a teaching program organized by Hans-Georg Müller. This workshop was part of SAMSI's Education and Outreach Program for 2010-2011.

The focus was the topic *Analysis of Object Data*, and the following lectures and events took place:

- Pierre Gremaud, N.C. State University and SAMSI, Welcome and Introduction
- Jim Ramsay, McGill University, Introduction to Analysis of Object Data
- Ci-Ren Jiang, SAMSI, MATLAB demo
- Snehalata Huzurbazar, University of Wyoming, Exmples of Collaborative Research Projects which use Analysis of Object-Oriented Data
- Hulin Wu, University of Rochester Dynamics, Modeling as a Weapon to Defend Ourselves Against Threats from Infectious Diseases and Bioterrorist Attacks
- Junheng Ma, SAMSI, R introduction
- Hongxiao Zhu, SAMSI , Introduction to Statistical Analysis of Functional Data
- Snehalata Huzurbazar, University of Wyoming, and Sylvie Tchumtchoua, SAMSI, MATLAB lab
- Pierre Gremaud, N.C. State University and SAMSI, Career Options
- Yolanda Muñoz-Maldonado, Michigan Tech, Sample Size Calculation for Functional Data
- David Degras, SAMSI, Topics in Functional Data Analysis
- Yolanda Muñoz-Maldonado, Michigan Tech and David Degras, SAMSI R lab

3 Report for the Working Groups on Functional Data Analysis

Leaders: Jeffrey Morris (University of Texas, Houston), Hans-Georg Müller (University of California, Davis), Jane-Ling Wang (University of California, Davis), Fang Yao (University of Toronto) and Hao Helen Zhang (North Carolina State University)

Functional data objects have been studied for quite some time now. In recent years, interest in this area has substantially increased, with a constant influx of new researchers and also established researchers starting to work in this area. Among the various types of object data, functional data are the most Euclidean and the main complication in their analysis stems from their infinite-dimensional nature, so that these objects lie in an infinite-dimensional Hilbert space. Moreover, in practice, the sample of random functions is usually not fully observed. Rather, observations are often sparse and noisy, especially when the functional data are observed in longitudinal studies. There are often complex dependence structures that one needs to deal with. Problems such as functional clustering and classification are not fully understood at this time. In addition, theoretical problems arise in function spaces, such as the non-existence of a Lebesgue density and these require creative solutions.

The mix of applied and theoretical challenges one faces in Functional Data Analysis at the cross-roads of stochastic processes, functional analysis, multivariate analysis and longitudinal and hierarchical modeling requires teams of researchers that bring mathematical, computational and various areas of statistical expertise to the table, and the SAMSI environment proved excellent for addressing some of the key issues. Two major working groups formed in this area, one focusing on more general topics and especially regression and classification problems for functional data (Leaders: Fang Yao and Helen Hao Zhang), and the other one on hierarchical and Bayesian functional modeling (Leader: Jeff Morris). Both groups shared a sizeable number of members and gained from mutual interactions and discussions between members. In addition, a workshop was held that emphasized the interface between longitudinal and functional data analysis. Details about this workshop can be found in the Workshops section of the report

The research in Functional Data Analysis profited from the Shape and Manifolds thread in individual discussions and joint membership in working groups. There were also close ties with the Dynamics thread and especially with the Brain Imaging thread, as functional data analysis is immediately applicable for many problems in Brain Imaging. In turn, many challenging problems for research on functional data analysis are motivated by the need to analyze various types of brain images, and both of these threads profited immensely from exchanges and cross-interactions between their respective members.

3.1 Report for the Working Group: Statistical Inference for Functional Data

Leaders: Fang Yao (University of Toronto) and Hao Helen Zhang (North Carolina State University)

3.1.1 Topics

Functional data pose unique challenges to statistical inferences because of their high-dimensional and complex structure. A further complication are sparse and noisy observations of the underlying random functions. The working group focused on two main goals: (1) to study inferential properties of existing approaches for functional data analysis (FDA) and to understand their fundamental behaviors; (2) to develop new, flexible, and powerful tools for FDA, including regression, classification, clustering, principal component analysis, and dimension reduction for such data. Situations of both densely and sparsely sampled functional data were considered. A variety of particularly relevant research topics and important future issues were identified and became topics for research conducted by group members. These topics include variable selection, model selection, sparse estimation, robust estimation, and experiment design for functional data. The new methods are motivated from real-world problems and have important applications in various fields such as image data, speech data, and shape analysis, for which connections with other workgroups proved useful. Concepts that were studied by group members also included time warping and manifolds with connections to other workgroups. Furthermore, extensions to user-friendly software and packages were discussed and partly implemented.

3.1.2 Participants

Participants in the work group were mostly new and junior researchers. The participants included SAMSI visitors, postdoctoral fellows, graduate students, local faculty and scientists: John Aston, Matt Avery, Graciela Boente, Herve Cardot, Jeng-Min Chiou, David Degras, Jimin Ding, Pang Du, Xingdong Feng, Kaushik Ghosh, Jinjiang He, Giles Hooker, Sungkyu Jung, Seonjoo Lee, Lexin Li, Yufeng Liu, Wenbin Lu, Junheng Ma, Hans-Georg Müller, Yolanda Munoz, Todd Ogden, Juhyun Park, Philip Reiss, Richard Samworth, Damla Şentürk, Haipeng Shen, Joon Jin Song, Ulrich Stadtmüller, Wenwen Tao, Sylvie Tchumtchoua, Haonan Wang, Jane-Ling Wang, Judy Wang, Liwei Wang, Yishi Wang, Hulin Wu, Shuang Wu, Yichao Wu, Fang Yao, Nuen Tsang Yang, Bo Zhang, Hao Helen Zhang, Lingsong Zhang, Jun Zhang, Hongxiao Zhu, Frank Zou, Jian Zou.

3.1.3 Activities and Research

The working group met throughout the entire AOOD program during 2010-2011 with weekly group discussions and presentations. Several subgroups were formed, which focused on specific topics. Close collaborations were established among the group participants. Group members were actively engaged in a variety of activities, including teaching special topic courses, group meetings, talks and presentations, and general as well as research discussions. Major activities and achievements in chronological order were as follows:

In September 2010, the working group was formed at the AOOD Opening workshop. Several important topics were identified in the area of statistical inference:

- Variable selection for functional linear regression models (FLM)
- Model selection for functional additive models (FAM)
- Classification and clustering for functional data
- Extensions of functional principal component analysis
- Experimental design for functional data
- Domain selection for functional data

From September to November of 2011, the entire working group met weekly to discuss important issues, study areas of general interest and to address theoretical and computational challenges. Each meeting focused on one topic and started with a couple of talks about the background for the problem, aiming to put all the participants on the same page. Existing works and open problems were then discussed, which stimulated the participant's interest to work on these problems and to establish collaborations with each other.

During November 8-10, 2010, a workshop on the Interface between Longitudinal and Functional Data was held at SAMSI under auspices of this working group, organized by Marie Davidian, Hans-Georg Müller and Fang Yao (for more details, see Section 1.2).

In Spring 2011, several small research groups were formed through interactions between participants, each focused on a narrower topic with the goal to do research on the topic. These small focused groups met regularly to collaborate on the selected problem, share research findings and progress, and discuss relevant issues. Manuscripts, software, and research papers were prepared or submitted. Some more detailed descriptions of the activities of a few such focus groups follow. These groups are a selection from a much larger number of such groups that were formed in Fall 2010; see the more comprehensive list of papers below, where the results of nearly all such groups are collected.

1. Variable Selection in Semiparametric Functional Linear Models

   Participants: Dehan Kong, Fang Yao and Hao Helen Zhang

   The goal of this group was to study the variable selection problem for semiparametric functional linear models. The group developed a new class of semiparametric functional regression models for jointly modeling the functional and non-functional predictors, identifying important scalar covariates while taking into account functional covariates. The new method takes advantage of modern shrinkage technique to achieve sparse estimation for parametric terms. The algorithm for this method is convenient for implementation and very efficient in identifying important scalar covariates. A preprint of the paper has been prepared.

2. Model Selection for Functional Additive Models

   Participants: Hongxiao Zhu, Fang Yao, Hao Helen Zhang

   Nonparametric functional data regression provides a flexible alternative to linear models in functional data analysis. This group focused on model selection in this setting, which is much more challenging than functional linear models, due to the unspecified form of effects for the functional predictors. The group proposed a new type of regularization framework which can automatically select important predictors and hence build sparse estimation models. Large sample properties of the new estimator are studied.

3. Experimental Design & Survey Sampling for Functional Data

   Participants: David Degras, Yolanda Munoz Maldonado, Herve Cardot

   The project deals with survey sampling techniques for functional data. The goal is to develop time-varying sampling schemes that improve statistical inference in finite populations. This type of work has important applications in sensor networks where huge numbers of signals can be observed at fine time scales but full observation and/or analysis of the data are operationally impossible. A good example is to estimate the electricity consumption in a large population based on digital meter readings. The group proposed two time-varying stratified sampling schemes: partial or full resampling. Population mean function was estimated by the Horvitz-Thompson (HT) survey estimator. The research activities included: (1) to determine the bias and covariance of the HT estimator, and comparisons with time-invariant samples; (2) to establish asymptotic theory and determine conditions on sampling rate, resampling frequency, and renewal rate needed for convergence results; (3) numerical studies with real and simulated data sets.

4. Functional Data Classification and Clustering

   Participants: Yichao Wu, Yufeng Liu, Junheng Ma, Fang Yao, Hao Helen Zhang

   The group has been working on two projects, respectively focusing on functional data classification and clustering. In the first project, for data with a functional predictor and a categorical response, the group proposes functional robust support vector machines. With the aid of functional principal component analysis provided by the PACE package, functional robust support vector machines can accommodate sparse and irregular functional data. One paper has been submitted on this project. In the second project, a new weighted distance is proposed for functional data clustering based on principal component analysis.

5. Domain Selection for Functional Data

   Participants: Giles Hooker and Damla Senturk

   The problem of domain selection belongs to an important class of problems and motivated from real-world problems. Typical examples where domain selection is crucial include: predicting the lifespan of medflies from their fecundity, estimating the length of hemodynamic response function, and vehicle exhaust measurement dependence of an engine function. The group proposed a profiled estimate and studied its theoretical properties including consistency and CLT. For practical implementation, various tuning methods were studied and compared.

6. Robust Functional Principal Components

   Participants: Graciela Boente and Jane-Ling Wang

   In many situations, data are recorded over a period of time and may be regarded as realizations of a stochastic process. In this paper, robust estimators for the principal components are considered by adapting the projection pursuit approach to the functional data setting. Our approach combines robust projection–pursuit with different smoothing methods. Consistency of the estimators are shown under mild assumptions. The performance of the classical and robust procedures are compared in a simulation study under different contamination schemes.

   This is a long term project that began many years ago but due to the different continents the collaborators reside, the progress has been slow. Both Boente and Wang visited SAMSI in

Fall 2010 and completed the project. Boente gave a presentation of the results in a SAMSI workshop "Functional approaches for longitudinal data" in November, 2010 and a paper was submitted shortly afterwards to the Annals of Statistics. They have received an invitation for a revision. The revision is still under review at this time.

7. Varying coefficient Cox model with functional and longitudinal covariates

   Participants: Jiming Ding and Jane-Ling Wang

   Ding and Wang worked on this project while they were both visiting SAMSI in Fall 2010. The project involves joint modeling of survival data and functional/longitudinal covariates. A varying coefficient Cox model is assumed for the survival data and nonparametric mixed-effects model is used to model the functional/longitudinal covariates in the Varying coefficient Cox model. All unknown components/quantities are modeled nonparametrically through a joint nonparametric likelihood and their asymptotic properties investigated. We have completed all theory and numerical analysis and have started to write a draft paper to be submitted to the Annals of Statistics. The plan is to have a preprint in a few months.

3.1.4 Additional Activities

- Participation in the Opening Workshop, Sep 12-14, 2010
  - Hans Müller and Jane-Ling Wang were co-organizers
  - Fang Yao gave Tutorial Lecture 1: Functional Data Analysis and Related Topics. (9:00-10:30am, Sep 12, 2010)
- Participation in the SAMSI-AOOD course:
  - Hans Müller gave several lectures for "Analysis of Object Data I", Fall 2010
  - Jane-Ling Wang gave several lectures for "Analysis of Object Data I", Fall 2010
- Participation in the AOD Workshop: Interface Functional and Longitudinal Data Analysis - November 8-10, 2010 (for more details see separate report below)
  - Hans Müller and Fang Yao served as co-organizers
  - Damla Senturk gave an invited talk "Functional Varying Coefficient Models for Longitudinal Data" (10:00-10:20am, Nov 8, 2010)
  - Fang Yao gave an invited talk "Additive Modeling of Functional Regression and its Gradients" (3:30-4:00pm, Nov 8, 2010)
  - Hao Helen Zhang gave an invited talk "Variance Component Selection in Linear Mixed Models" (10:00-10:30am, Nov 9, 2010)
  - Graciela Boente gave an invited talk "Robust Functional Principal Components: a projection-pursuit approach" (11:15-11:45am, Nov 9, 2010)
  - Jimin Ding gave an invited talk "Time-varying Coefficient Cox model with Nonparametric Longitudinal Covariates" (9:05-9:25am, Nov 10, 2010)
- Participation in the AOOD transition workshop, June 9-11, 2011
  - Hans Müller and Jane-Ling Wang served as co-organizers
  - Hongxiao Zhu gave an invited talk "Robust Classification of Functional and Quantitative Image Data using Functional Mixed Models" (9:30-9:50am, June 9, 2011)

– Jane-Ling Wang gave an invited talk “Discussion and Connections to FDA Working Groups” (10:10-10:40am, June 9, 2011)
– Yichao Wu gave an invited talk “Continuously Additive Models for Functional Regression Analysis” (11:10-11:40am, June 9, 2011)
– David Degras gave an invited talk “Longitudinal Survey Methods for Functional Data” (11:40-12:00pm, June 9, 2011)
– Jane-Ling Wang served on the panel for “Discussion and Future Developments” (12:00-12:30pm, June 9, 2011)

3.1.5 Working Papers and Publications

The works listed in the following benefitted from our work group at SAMSI in one or several of the following ways: (1) research inspired by work group discussions or workshops; (2) research resulting from collaborations among work group participants that were begun during participant's stay at SAMSI; (3) research continued or completed (submitting or revising papers) during participant's stay at SAMSI.

Published/Accepted

1. Acar, E., Craiu, R.V., Yao, F. (2011). Dependence calibration in conditional copula: a nonparametric approach. *Biometrics*, accepted.

2. Chen, K., Chen, K., Müller, H.G., Wang, J.L. (2011). Stringing high-dimensional data for functional analysis. *Journal of American Statistical Association* **106**, 275–284.

3. Chen, D., Hall, P., Müller, H.G. (2011). Single and multiple index functional regression models with nonparametric link. *Annals of Statistics* **39**, 1720–1747.

4. Wu, S., Müller, H.G. (2011). Response-adaptive regression for longitudinal data. *Biometrics*, accepted

5. Chen, K., Müller, H.G. (2011). Conditional quantile analysis when covariates are functions, with application to growth data. *J. Royal Statistical Society B*, accepted

6. Ding, J.-M., Symanzik, J., Sharif, Wang, J. -L., Duntley, Shannon, W. (2011). Powerful Actigraphy Data Through Functional Representation. *Chance*, accepted.

7. Park, S. Y. and Liu, Y. (2011). Robust penalized logistic regression with truncated loss. *The Canadian Journal of Statistics*, accepted.

8. Li, P.L. and Chiou, J.M. (2011) Identifying cluster numbers for subspace projected functional data clustering. *Computational Statistics and Data Analysis* **55**, 2090-2103.

9. Liu, Y. and Yuan, M. (2011). Reinforced multicategory support vector machines. *Journal of Computational and Graphical Statistics*, accepted.

10. Liu, Y., Zhang, H. H., and Wu, Y. (2011). Soft or hard classification? Large margin unified machines. *Journal of the American Statistical Association* **106**, 166-177.

11. Lu, W., Zhang, H. H., and Zeng, D. (2011). Variable selection for optimal treatment decision. *Statistical Methods in Medical Research*, accepted.

12. Müller, H.G. (2011). Functional data analysis. *International Encyclopedia of Statistical Science*, Ed. Lovric, M. Springer Science Business Media, Heidelberg. (Extended version available in StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies, id 242).

13. Müller, H.G., Sen, R., Stadtmüller, U. (2011). Functional data analysis for volatility. *Journal of Econometrics*, acceped

14. Wu, Y. (2011). An ordinary differential equation-based solution path algorithm . *Journal of Nonparametric Statistics*, **23**, 185-199.

15. Wu, Y. and Li, Lexin (2011). Asymptotic Properties of Sufficient Dimension Reduction with A Diverging Number of Predictors. *Statistica Sinica* **21**, 707-730.

16. Wu, Y. and Liu, Y. (2011). Non-crossing large-margin probability estimation and its application to robust SVM via preconditioning. *Statistical Methodology* **8**, 56-67.

17. Yang, W., Müller, H.G., Stadtmüller, U. (2011). Functional singular component analysis. *J. Royal Statistical Society B* **73**, 303–324.

18. Yao, F., Fu, Y., Lee, T.C.M. (2011). Functional mixture regression. *Biostatistics*, **12**, 341-353.

19. Zhang, H. H., Cheng, G., and Liu, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *Journal of the American Statistical Association*, accepted.

20. Zhang, Z., Müller, H.G. (2011). Functional density synchronization. *Computational Statistics and Data Analysis* **55**, 2234–2249.

21. Zhu, H., Brown, P. J., Morris, J. S. (2011). Robust, adaptive regression in functional mixed models framework. *Journal of the American Statistical Association*, accepted.

Submitted/In Revision

1. Ahn, M., Zhang, H. H., and Lu, W. (2011) Moment-based method for random effect for selection in linear mixed models. *Statistica Sinica*, revised.

2. Bosca, L., Boente, G., Tyler, D. and Wang, J.-L. (2011) . Robust functional principal components. *Annals of Statistics*, revised.

3. Cardot, H., Degras, D., and Josserand, E. (2011). Confidence bands for Horvitz-Thompson estimators using sampled functional data. Submitted.

4. Chen, D., Müller, H.G. (2011). Nonlinear manifold representations for functional data. Submitted.

5. Ding, A.A. and Wu, H. (2011). Estimation of ODE parameters using constrained local polynomial regression. Submitted.

6. Fang, Y., Wu, H., Zhu, L. (2011). A Note on Data Augmentation-Based Pseudo-Least Squares Estimation for ODE Models. Submitted.

7. Müller, H.G., Wu, Y. Yao, F. (2011). Continuously additive models for functional regression. Submitted.

8. Şentürk, D., Ghosh, S. and Nguyen, D. V. (2011). Exploratory time varying lagged regression for longitudinal data. *Biometrics*. Submitted.

9. Wei, F. and Zhu, H. (2011). Group Coordinate Descent Algorithms for Nonconvex Penalized Regression. *Computational Statistics & Data Analysis*. Submitted.

10. Wu, Y. and Liu, Y. (2011). Functional robust support vector machines for sparse and irregular longitudinal data. Revised.

11. Wu, H., Miao, H., Xue, H., Topham, D.J., and Zand, M. (2011). Quantifying immune response to influenza virus infection via multivariate nonlinear ODE models with partially observed state variables and time-varying parameters. *Journal of American Statistical Association*, submitted.

12. Wu, H., Xue, H., Kumar A. (2011), Numerical algorithm-based estimation methods for ODE models via penalized spline smoothing, *Biometrics*, revised.

13. Zhu, H., Brown, P. J., Morris, J. S. (2011). Robust Classification of Functional and Quantitative Image Data using Functional Mixed Models. *Biometrics*, revised.

In Preparation/Preprint

1. Asencio, M., Hooker, G. and Gao, H. (2011). Functional Convolution Models. *Preprint*.

2. Avery, M, Zhang, H. H. , Wu, Y. (2011) . Sparse estimation in functional data analysis. *In preparation*.

3. Chen, Y. and Samworth, R. J. (2011). Smoothed log-concave maximum likelihood estimation with applications. *Preprint*.

4. Chiou, J.M., Ma, Y. and Tsai, C.L. (2011). Functional time-varying random effects models for longitudinal data. *Preprint*.

5. Chiou, J.M. and Müller, H.G. (2011). Linear manifold modeling of multivariate functional data, with application to traffic flow analysis. *Preprint*.

6. Degras, D. (2011). Longitudinal survey sampling for functional data. *In preparation*.

7. Ding, J.-M. and Wang, J.-L. (2011). Varying coefficient Cox model with functional and longitudinal covariates. *In preparation*.

8. Hooker, G.. (2011) Domain selection for functional linear models. *In preparation*.

9. Kong, D. , Yao, F., Zhang, H. H. (2011). Semiparametric functional linear model. *In preparation*.

10. Lee, W. and Liu, Y. (2011). Simultaneous multiple response regression and inverse covariance matrix estimation via penalized Gaussian maximum likelihood. *Preprint* .

11. Li, L., Yao, F., Craiu, R. V. (2011). Minimum description length principle for correlated data. *Preprint*.

12. Li, L., Zhou, H., and Zhu, H. (2011). Tensor regression with applications in neuro-imaging data analysis. *In preparation*.

13. Li, N., and Zhang, H. H. (2011) Sparse learning in multi-class classification. *Preprint.*

14. Lin, C. Y., Zhang, H. H., Bondell, H., and Zou, H. (2011) Nonparametric variable selection in quantile regression. *Preprint.*

15. Liu, C., Ray, S., Hooker, G. and Friedl, M. (2011) . Functional factor analysis for periodic remote sensing data. *Preprint.*

16. McLean, M., Hooker, G., Saicu, A. and Ruppert, D. (2011) Functional generalized additive models. *Preprint.*

17. Du, P., Cheng, G., and Zhang. H. H. (2011) . Structure selection for nonparametric survival analysis. *In preparation.*

18. Samworth, R. J. (2011). Optimal weighted nearest neighbor classifiers. *Preprint.*

19. Şentürk, D. (2011). Efficient estimation for generalized varying coefficient models with longitudinal data. *In preparation.*

20. Shah, R. and Samworth, R. J. (2011), Variable selection with error control: Another look at stability selection. *Preprint.*

21. Staicu, A. M., Şentürk, D., Carroll, R. J. (2011). Generalized time-varying spatial regression of multilevel functional data. *In preparation.*

22. Wang, L., Wu, Y., and Li, R. (2011). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Preprint.*

23. Wong, R.K.W., Yao, F., Lee, T.C.M. (2011). Robust estimation for generalized additive models. *Preprint.*

24. Wu, Y. (2011). Elastic net for Cox's proportional hazards model with a solution path algorithm. *Preprint* .

25. Zhou, H. and Wu, Y. (2011) A general path algorithm for regularized statistical estimation. *Preprint.*

26. Zhu, H., Dunson, D. B. (2011). Bayesian graphical models for multivariate functional data. *Preprint.*

27. Zhu, H., Yao, F., Zhang, H. H. (2011). Component Selection in Functional Additive Models. *Preprint.*

3.2 Report for the Working Group: Hierarchical Modeling

Leader: Jeffrey S. Morris (University of Texas, Houston)

3.2.1 Goals

This working group is interested in developing hierarchical modeling approaches for object data, including functions, images, and more general structures like shapes and trees. The goal is to develop inferential methodology motivated by specific applications yielding complex, structured data. The idea of hierarchical modeling implies flexible, unified models that can simultaneously take into account variability and structure from multiple sources in the data set, within and between objects, and induced by the design or other measured covariates. Both Bayesian and frequentist approaches have been considered, and discussion of the connections and distinctions among existing Bayesian and frequentist approaches in the literature done. As longer term goals, our objective is to build new unified methodology for modeling and performing inference for functional, image, shape, and other object data using as building blocks various modeling tools that have been used in modeling in simpler contexts, but not yet used together. We will also make data and software available for these methods so the work can have impact both in terms of stimulating more methodological research and producing tools investigators can use to analyze their object data.

3.2.2 Participants

Jeffrey S. Morris, Syvie Tchumtchaoua, Jaeyong Lee, Jim Berger, Susan Bayarri, Ci-Ren Jiang, Xia Wang, Brian Reich, Howard Bondell, Ana-Maria Staicu, Veera Baladandayuthapani, Michele Guindani, Todd Ogden, Philip Reiss, Scott Holan, Josue Martinez, Hongtu Zhu, Raymond Carroll, Giles Hooker, Ian Dryden, Darren Wilkenson, Dubois Bowman, Hongxiao Zhu, Xiaojing Wang, Sarat Dass, Genevera Allen, Xiaojing Wang, David Dunson, Jiguo Cao.

3.2.3 Activities

Throughout the year, we had a number of presentations and discussions involving various types of hierarchical modeling approaches that have been applied and/or can potentially be applied to the analysis of object data, including the following:

- Bayesian Variable Selection Methods
- Sparsity Priors and Penalties
- Multilevel Functional Principal Components
- Functional Mixed Models
- Bayesian Nonparametrics
- Density and Quantile Regression
- Spatial Models for Dependency
- Independent Components Analysis
- Generalized Least Squares Matrix Decomposition
- Product Kernels

These talks were done weekly throughout the fall, and then there were a few more in the spring. These talks provided general knowledge to all attendees, and sparked interesting conversations as well as new collaborations, which spun off the group and worked on research manuscripts. Some of these new collaborations have already led to manuscripts that have been submitted, and others that are in preparation (see list below).

In the spring, we split the working group into three subgroup areas and continued to meet every other week as subgroups. The groups and their defined aims were as follows:

1. FDA and Bayesian Nonparametrics: Develop methods applying various Bayesian nonparametric modeling approaches to the analysis of functional and other object data.

2. Object Regression: Develop flexible methods and models for regression analysis for object data such as functional, image and shape data.

3. Hierarchical Methods for Image Analysis: Develop hierarchical methods for the analysis of large-scale image data sets, including biomedical imaging data and spatial data.

During our subgroup meetings, we had some pre-arranged presentations, but also brainstormed research ideas and reviewed literature in various areas with the goal of stimulating new collaborations and research. Following are some of the projects within these areas, which include some projects that have been completed, some with substantial progress, and some that have just been started.

- FDA and Bayesian Nonparametrics
  - Development of Bayesian nonparametric approaches to analyze functional and image data.
  - Development of a general theoretical framework for Bayesian nonparametric functional data analysis, describing important support properties in this framework, assessing which existing methods possess these properties, and use these as motivations for developing new methods.
  - Develop methods to perform Functional density regression, which model covariate effects on the entire distribution of the function, including mean, covariance, and extreme quantiles.
  - Detect change points using functional Dirichlet Processes.
- Object Regression
  - Develop methods incorporating local variable selection in nonparametric regression and functional data settings, using priors with positive probability given to flat regions of the curves.
  - Develop method to analyze longitudinal shape data using functional mixed models.
  - Develop methods for fitting functional mixed models using adaptively regularized sparse functional principal components, and using hybrid transforms that include both global and local components.
  - Develop methods to model functional data with spatially heterogeneous shape characteristics.
  - Develop sufficient dimension reduction methods, which aim to identify the smallest set of linear combinations of the predictors that retain all information in the predictors about the response distribution, and apply the new methods to develop climate indices for forecasting yearly hurricane intensity.
  - Investigate the performance of functional principal components in settings with complex, high dimensional functional and image data, to provide guidance on when these models should be used and how many dimensions to model.

- Develop hierarchical methods to model spatially correlated functional and image data.
- Develop new method for regression of scalar responses on high dimensional functional binary predictors, and apply to the analysis of GWAS data.
- Write review chapter on multi-level functional data analysis methods.
- Develop method for analysis of nonstationary time series using functional mixed models involving spectrograms and wavelet transforms.
- Develop method to perform robust functional inference for spatially heterogeneous functional and image data.
- Develop methods to classify individuals based on functional data observed in nested hierarchical structure.
- Develop methods to analyze functional data observed on a spatial grid.
- Develop methods for analyzing spatially correlated binary longitudinal data within a multilevel model.

- Hierarchical Methods for Image Analysis
  - Develop hierarchical model-based methods to simultaneously perform subject- and group-level analyses of complex, extremely high dimensional quantitative image data (e.g. fMRI) using functional mixed models.
  - Develop new methods for multivariate longitudinal data using flexible semiparametric Bayes models.
  - Motivated by massive dimensional dynamic functional data, develop flexible online Bayesian methods, which work by reading in a slice of data at one time and approximating the posterior based on these data, and then updating the approximation as additional data are read in.
  - Develop methods to identify genetic pathways that affect a subject's functional connectivity in response to a series of external stimuli, using a hierarchical Bayesian model combining information across fMRI and genetic data.
  - Develop Bayesian Independent Components Analysis (ICA) methods that exploit prior information about spatial association and known connectivity patterns between anatomic features.

3.2.4 Additional Activities

- Participated in longitudinal/functional data workshop, with the following invited talk:
  - 2:30-3:00 Jeffrey S. Morris, University of Texas MD Anderson Cancer Center Adaptive, Robust Functional and Image Regression in Functional Mixed Models

- Participated in transitional workshop, with the following talks presented:
  - 9:00-9:30 Jeffrey S. Morris, MD Anderson Hierarchical Methods for Analysis of Object Data
  - 9:30-9:50 Hongxiao Zhu, SAMSI Robust Classification of Functional and Quantitative Image Data using Functional Mixed Models
  - 9:50-10:10 Sylvie Tchumtchou, SAMSI Online Variational Bayesian Inference in Hierarchical Models for Correlated High-dimensional Data

  - 12:00-12:30 Discussion and Future Developments: Steve Marron, UNC; Jeff Morris, MD Anderson; Jim Ramsay, McGill University; Jane-Ling Wang, University of California, Davis

- Organized a Topic Contributed Session for 2011 Joint Statistical Meetings in Miami Beach, Florida, as follows: Hierarchical Methods for Object Data: SAMSI Object Data Program Working Group Organizer(s): Jeffrey S. Morris, The University of Texas MD Anderson Cancer Center Chair(s): Denis Larocque, HEC Montréal
  - 10:35 AM Bayesian Hierarchical Functional Models for High-Dimensional Genomics Data Veera Baladandayuthapani, The University of Texas MD Anderson Cancer Center
  - 10:55 AM Classification of Unknown Powders Using a Support Vector Machine Classification Model - Jessi Cisewski, The University of North Carolina at Chapel Hill ; Jan Hannig, The University of North Carolina at Chapel Hill ; Emily Snyder, Environmental Protection Agency
  - 11:15 AM Joint and Individual Variation Explained (JIVE) for Integrated Analysis of Multiple Datatypes - Eric Frazer Lock, The University of North Carolina at Chapel Hill
  - 11:35 AM Estimating Shape-Constrained Functions Using Gaussian Processes - Xiaojing Wang, Duke University ; James Berger, Duke University
  - 11:55 AM An Application of Nonparametric Function Estimation for Planning Reconstructive Surgical Procedures for the Skull - Daniel Osborne, Florida State University ; Victor Patrangenaru, Florida State University ; Xiuwen Liu, Florida State University ; Hillary W. Thompson, Louisiana State University
  - 12:15 PM Floor Discussion

3.2.5 Publications

* Also associated with FDA Inference working group ** Also associated with Brain Imaging working group

Published/Accepted:

1. Cao, J., Cai, J. and Wang, L. (2011). Estimating Curves and Derivatives with Parametric Penalized Spline Smoothing, Accepted by *Statistics and Computing.*

2. Dass, S. C., Lim, C. Y. and Maiti, T. (2011). Default Bayesian Analysis for Multivariate Generalized CAR Models, To appear in *Statistica Sinica.*

3. Lim, C. Y. and Dass, S. C. (2011). Assessing Fingerprint Individuality Using EPIC: A Case Study In The Analysis Of Spatially Dependent Marked Processes,*Technometrics*, vol. 53, no. 2, pp. 112-124.

4. **Morris JS, Baladandauthapani V, Herrick RC, Sanna PP, and Gutstein HG (2011). Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomic data, *Annals of Applied Statistics*, to appear.

5. *Staicu, A. M., Crainiceanu, C., Reich, D. and Ruppert, D. (2011). Modeling functional data with spatially heterogeneous shape characteristics, *Biometrics*, to appear.

6. Wu, R., Cao, J., Huang, Z., Wang, Z., Gai, J. and Vallejos, E. (2011). Systems mapping: how to improve the genetic mapping of complex traits through design principles of biological systems, *BMC Systems Biology* 5:84, 1-24.

7. Zhu, H., Brown, P. J., and Morris, J. S. (2011). Robust, Adaptive Functional Regression in Functional Mixed Model Framework. *JASA*, to appear.

Submitted/In Revision:

1. Crainiceanu, C. M., Caffo, B. S., and Morris, J. S. (2011). Multilevel Functional Data Analysis. Submitted to *SAGE Handbook for Multilevel Modeling.*
2. Dass, S. C., Lim, C. Y and Maiti, T. (2011). Detecting Change Points of Cancer Incidence Rates using Functional Dirichlet Processes. *Submitted.*
3. Li, J.J., Jiang, C.R., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Sparse linear modeling of RNA-seq data for isoform discovery and abundance estimation. In revision for *PNAS.*
4. Martinez, J. G., Bohn, K. M., Carroll, R. J. and Morris, J. S. (2011). A study of Mexican free-tailed bat syllables: Multi-domain modeling of nonstationary time series with high frequency content using Bayesian functional mixed models. *Under revision.*
5. *McLean, M. W., Hooker, G., Staicu, A. M., and Ruppert, D. (2011). Functional Generalized Additive Models. *Submitted.*
6. Montagna, S., Tokdar, S.T., Neelon, B. and Dunson, D.B. (2011). Bayesian latent factor regression for functional and longitudinal data. *Submitted.*
7. Morris, J. S. (2011). Statistical Methods for Proteomic Biomarker Discovery using Feature Extraction or Functional Data Analysis Approaches. Under Revision for *Statistics and its Interface.*
8. Wang, L. and Cao, J. (2011). Estimating Delay Differential Equations. Under revision for *Journal of Agricultural, Biological, and Environmental Statistics.*
9. *Staicu, A. M., Lahiri, S., and Carroll, R. (2011). Tests of significance for spatially correlated multilevel functional data. Submitted.
10. Tchumtchoua, S. (2011). "Bayesian Semiparametric Functional Generalized Linear Models for Longitudinal Data". *Submitted.*
11. Zhang, L., Baladandayuthapani,V., Mallick, B. K., Thompson, P., Bondy, M. and Do, K.A. (2011). Bayesian hierarchical structured variable selection methods with application to MIP studies in breast cancer. *Under revision.*
12. Zhu, H., Brown, P. J., and Morris, J. S. (2011). Robust classification of functional and quantitative image data using functional mixed models. *Submitted.*

In Preparation:

1. Baladandayuthapani, V., Morris, J.S., Coombes, K. R. and Abruzzo, L. (2011). Bayesian Adaptive Functional Linear Models for Copy Number Data. *In preparation.*
2. **Baladandayuthapani, V., Bharath, K., Baggerly, K., Czerniak, B., and Morris, J. S. (2011). Bayesian Spatial-functional Models for High-throughput genomic data. *In preparation.*

3. Cai, J. and Cao, J. (2011). Estimating Damage Accumulation Dynamic Models Hierarchical Bayesian Methods for Parameter Estimation in Dynamic Duration of Load Models. *In preparation.*

4. Chkrebtii, O. and Cao, J. (2011). Generalized Additive Modeling of Spatio-Temporal Trends in Salmon Productivity. *In preparation.*

5. Jiang, C.R., Aston, J.A. and Wang, J.L. (2011). Nonparametric response function estimation via FPCA with an application to dynamic PET data. *In preparation.*

6. Jiang, C.R. and Morris, J.S. (2011). Choosing number of principal components in high dimension, low sample size situations for complex functional data. *In preparation.*

7. Morris, J. S. and Allen, G. (2011). Analysis of fMRI data using functional mixed models and sparse empirically determined basis function representations. *In preparation.*

8. Pati, D. and Dunson, D.B. (2011). Bayesian closed surface fitting through tensor products. *In preparation.*

9. Serban, N., Staicu, A. M., and Carroll, R. J. (2011). Multilevel Spatially Correlated Binary Longitudinal Data. *In preparation.*

10. Staicu, A. M., Senturk, D., and Carroll, R. J. (2011). Generalized Time-Varying Spatial Regression of Multilevel Functional Data. *In preparation.*

11. **Tchumtchoua, S., Dunson, D. B., and Morris, J. (2011). Online Variational Bayes Inference for High-Dimensional Correlated Data.*In preparation.*

12. Tchumtchoua, S., Dunson, D. B., and Morris, J. (2011). A Heterogeneous Dynamic Structural Equation Model With Application to Brain Connectivity. *In preparation.*

13. Wang, Xia, Chen, M.-H., Dey, D. K., and Kuo C.-Y. (2011). Space-Time Modeling of Atlantic Cod Abundance in the Gulf of Maine. *In preparation.*

14. Wang, Xia and Sedransk, N. (2011). Bayesian Models on Biomarker Discovery Using Spectral Count Data in the Label-Free Shotgun Proteomics. *In preparation.*

15. Wang, Xia, Sedransk, N. and Tabb, D. (2011). Variability of Base Peak Intensities in Shotgun Proteomics Experiment: Perspectives from Functional Data Analysis. In preparation.

16. Wang, Xiaojing and Berger J. (2011). Estimated shape constrained functions using Gaussian Processes. *In preparation.*

17. Wei, F. and Zhu, H. (2011). Group coordinate descent algorithms for nonconvex penalized regression. *In preparation.*

18. Zhu, H. and Dunson, D. (2011). Bayesian graphical models for multivariate functional data. *In preparation.*

19. Zhu, H. and Morris, J. S. (2011). Functional Mixed Models for Serially Correlated Functional Data. *In preparation.*

4 Report for the Working Group: Dynamics and Inference

Leaders: David Degras (SAMSI), Giles Hooker (Cornell University) and James Ramsay (McGill University)

4.1 Participants

Karthik Bharath, Nicolas Brunel, David Campbell, Jiguo Cao, Oksana Chkrebtii, Sy-Miin Chow, Sarat Dass, Jimin Ding, Paul Fearnhead, Xingdong Feng, Kaushik Ghosh, Mark Girolami, Andrea Gottlieb, Serge Guillas, Snehalata Hurzurbazar, Hachem Kadri, Hamid Krim, Jaeyong Lee, Seonjoo Lee, Lexin Li, Tao Lu, Peter Marcy, Yonada Munoz, Juhyun Park, Debashis Paul, Jie Peng, Simon Preston, Sofia Olhede, Laura Sangalli, Damla Senturk, Jim Sethna, Valentina Staneva, Wenwen Tao, Michael Wierzbicki, Darren Wilkinson, Andy Wood, Donglin Zeng, Jay Wang, Naisyin Wang, Xiaohui Wang, Hulin Wu, Yuefeng Wu, Nuen Tsang Yang, Tingting Zhang, Xiaoke Zhang, Ji Zhu, Jian Zou

4.2 Topics

The field of statistical inference for nonlinear dynamic systems associated with observational data is a relatively new area of statistical research, although the fundamental ideas can be found in time-domain time series analysis, econometrics, pharmacological modeling and a few other topics with considerable histories. There are no general monographs on the modeling of data over continuous time that target specifically the statistical community.

Dynamic systems pose the wider challenge of statistical methodology for mathematical models expressed as systems of equations that do not admit an analytical solution. Classical approaches have involved iterative numerical approximation of solutions, but new methods are being proposed that use the equations themselves in regularization penalties in inverse problems, and consequently do not require repeated attempts to approximate exact solutions to these equations. Rather, these methods rely on relaxation strategies that have proven invaluable in other areas of numerical analysis.

Wider interest in the topic has been stimulated by new methods for parameter estimation and inference, the modeling of genomic and proteomic data, dynamic systems modeling of the spread and time course of disease, the use of stochastic dynamics beyond financial analysis, developments in functional data analysis, and the explosive growth in real-time data monitoring over space, time, and other continua and manifolds. This working group brought together several groups of statisticians who have separately developed methods from quite different perspectives, and several researchers interested in learning more about the problems in the area and how to get involved.

Because of these new developments, the working group was an invaluable opportunity for the consolidation of efforts, discussion of methodological and computational issues, and the exchange of perspectives. The working group therefore spent some considerable time reviewing literature and approaches and identifying new areas of potential interest. In particular, sources of interest were:

1. The distinction between deterministic and stochastic models of system dynamics and differences between them with respect to methodological approaches and fundamental conceptualization of modeling and inference problems.

2. Dealing with lack of fit due to simplifications or partial understanding inherent in any dynamic model for real world data.

3. The question of what is desirable to fit – data as observed, or (what is more frequently of interest to applications) qualitative features (cycles, periods, chaos) of the system dynamics.

4. The identifiability of parameters in dynamic systems models, and the problem of designing experiments to improve parameter estimation.

5. The design of software. This is particularly challenging in nonlinear dynamics: the models employed are non-linear and generally unique to each problem, this is coupled to models of

observation processes that are similarly problem-specific and these models need to be easily incorporated into computationally complex methods as well as being easily modified. Additionally, all current computational methods can suffer from failures, either of optimization or of stochastic sampling and useful diagnostic processes for these failures need to be developed.

6. The use of partial differential equations in spatial and spatial-temporal data analysis to model variation over space and time.

7. Ways of facilitating the entry of new researchers into this area, including books, collections of applications, workshops, on-line seminars and so on. The inclusion of nine lectures of 2 hours each on dynamical systems modeling in the second semester graduate seminar at SAMSI during the project was especially valuable as a stimulant to the consideration of methods for dissemination.

8. General methods for parameter estimation and inference for both continuous and discrete time dynamic systems.

9. A wide range of specific applications, including systems biology, spread of disease, viral dynamics, clinical trials, population dynamics in mathematical biology, glaciology, dynamic systems with covariates, chemical reactions, and transport models.

10. Ways of avoiding the heavy computational overhead and long development times typical of application of Markov Chain Monte Carlo methods for Bayesian analysis.

11. Parallel computing strategies for dynamic systems

12. A library of real-world datasets and models for testing software and methodology

The working group focussed on providing introductions to these problems and discussions of varying approaches to dealing with them within different modeling frameworks. In addition, several areas of statistical inference have been identified as future areas of research:

- Diagnostics for model lack of fit.
- Adaptive experimental design for nonlinear dynamics.
- Incorporating random effects into repeated dynamic systems.
- Optimal choice of qualitative features to assess both in terms of information obtained and robustness to model choices.
- The extension of methods for ordinary differential equations to partial differential equations. In this context, an example data set involving ice-melt on glaciers has been identified as a relevant and approachable problem.

4.3 Achievements, Working Papers and Publications

Achievements

Jim Ramsay translated the CollocInfer package in R into Matlab in January, and did considerable additional work on the CollocInfer manual and on various test problems. Giles Hooker and Jim Ramsay are continuing to collaborate on the development of this software for parameter estimation and inference for dynamic systems.

Published/Accepted

1. Cao, J., J. Cai and L. Wang (2011) Estimating Curves and Derivatives with Parametric Penalized Spline Smoothing, Accepted by Statistics and Computing.
2. Hooker, G., S. P. Ellner, L. Roditi and D. J. D. Earn (2011) Parameterizing State-space Models for Infectious Disease Dynamics by Generalized Profiling: Measles in Ontario, Journal of the Royal Society Interface, 8:961-975.
3. Wilkinson, D. J. (2011) Stochastic Modeling for Systems Biology, second edition, Boca Raton, Florida: Chapman and Hall/CRC Press, in press.
4. Wilkinson, D. J. (2011) Stochastic dynamical systems, in Handbook of Statistical Systems Biology, M.P.H. Stumpf, M. Girolami, D.J. Balding (eds), Wiley, in press.
5. Wu, R., J. Cao, Z. Huang, Z. Wang, J. Gai and E. Vallejos (2011) Systems mapping: how to improve the genetic mapping of complex traits through design principles of biological systems, BMC Systems Biology 5:84, 1-24.

Preprints/Technical Reports

1. Golightly, A., Wilkinson, D. J. (2011) Bayesian Parameter Inference for Stochastic Biochemical Network Models using Particle MCMC, in submission to Interface Focus.
2. Luo, W., J. Cao, M. Gallagher and J. Wiles (2011) Estimating the Intensity of Ward Admission and its Effect on Emergency Department Access Block, Submitted
3. Ratmann, O., P. Pudlo, S. Richardson and C. Robert, Monte Carlo Algorithms for Model Assessment via Conflicting Summaries, submitted.
4. Hooker, G. and S. P. Ellner, (2011) On Forwards Prediction Error, Technical Report BU-1679-M, Department of Biological Statistics and Computational Biology, Cornell University.
5. Hooker, G., J. O. Ramsay and L. Xiao, (2011) CollocInfer: An R Library for Collocation Inference for Continuous and DiscreteTime Dynamic Systems, R library and manual.

Working Papers/In Progress

1. Ramsay, J. O. (2011) A Functional Estimate of a Functional Variance-Covariance Matrix and its Inverse. *Paper in preparation.*
2. Sangalli, L., Ramsay, J.O. and Ramsay, T.O. (2011) Spatial Spline Regression Models. *Paper in preparation.*
3. Armagan, A., D. Dunson and J. Lee, Posterior consistency of Bayesian regression model for high-dimensional data.
4. Chkrebtii, O. and J. Cao, Generalized Additive Modeling of Spatio-Temporal Trends in Salmon Productivity
5. Cai, J., and J. Cao, Estimating Damage Accumulation Dynamic Models Hierarchical Bayesian Methods for Parameter Estimation in Dynamic Duration of Load Models
6. Dass, S., J. Lee, and K. Lee, Fast Computation for Regression Models with Ordinary Differential Equations.

7. Hooker, G., Rogers, B., Lin, K. and Ng, T., Control Theory and Optimal Adaptive Experiments in Nonlinear Stochastic Models

8. Ramsay, J. O. A Functional Estimate of a Functional Variance-Covariance Matrix and its Inverse

9. Sangalli, L., Ramsay, J.O. and Ramsay, T.O. Spatial Spline Regression Models Thorbergson, L. and Hooker, G., Optimal Adaptive Experimental Design in Hidden Markov Models

10. Wu, Y. and S. Ghosal, Convergence Rates of Multivariate Density Estimation by Dirichlet Mixture Priors

11. Wu, Y. and G. Hooker, Generalized Profiling, Stochastic Differential Equations and Higher-Order Stochastic Runge Kutta Schemes

5 Report for the Working Groups on Shapes and Manifolds

Leaders: Ian Dryden (University of South Carolina), Ezra Millar (Duke University), Victor Patrangenaru (Florida State University), John Kent (university of Leeds), Anuj Srivastava (Florida State University), Stephan Huckermann (University of Göttingen), Ross Whitaker (University of Utah).

5.1 Introduction

The Shapes and Manifolds theme involved three Working Groups:

1. Data analysis on sample spaces with a manifold stratification. Leaders: Ezra Miller, Victor Patrangenaru.

2. Metrics on shape spaces. Leaders: John Kent, Anuj Srivastava.

3. Geometric correspondence. Leaders: Stephan Huckemann, Ross Whitaker.

Individual reports from the working group leaders follow, but collectively many important issues were addressed during the program. There was a strong emphasis on sample spaces that have lower dimensional strata or sub-spaces, that themselves are manifolds. Particular examples include: metric phylogenetic trees on a fixed set of n taxa, where the space can be viewed as having structure like an open book, with pages attached to a book spine; covariance matrices with some equal eigenvalues; and the shapes of three dimensional landmarks, where lower dimensional shapes such as collinear points are also on a manifold. Some unexpected central limit theorem results were obtained, including various types of 'sticky results' where the Fréchet mean remains in a stratum with probability 1, as in some tree data examples. This contrasts with the more usual situation where Fréchet means are almost surely on the highest dimensional manifold, e.g. in the 3D shape case for certain Procrustes distances. Careful characterization of these issues was made by the Manifold Stratification Working Group, and this was an example of a project crossing both the Trees and Manifolds themes.

Excellent progress was made on the registration of functional data using curve shape analysis methods. This work initially arose out of a series of lectures and discussions in the Metrics working group. By adapting the Fisher-Rao metric based procedures for curves, some promising function registration methods were developed. This topic straddled both the Functional Data Analysis and Manifolds themes.

Careful examination of the topology and geometry of projective shape space was carried out by various members of the three groups. A key aspect of the work was to distinguish between an orientated versus an unorientated camera, and an axial versus directional camera. Using a particular representation called Tyler standardization the group was able to characterize all special cases. An alternative method of representing projective shapes based on projective frames was also investigated. These two approaches pave the way for practical methodologies for the statistical analysis of projective shapes from digital camera images.

A sub-group of the Metrics Group developed methodology for a type of spline on manifolds. By making use of the Palais metric it is possible to obtain an expression for the gradient of the associated objective function, which leads to a practical fitting algorithm. Details were worked out for various manifolds, including the space of rotations in 3D, Kendall's shape space and symmetric positive definite matrices with determinant 1. Also a new metric for the space of symmetric positive definite matrices was explored.

Dimension reduction was explored through influential landmarks. A sub-group of the Geometric Correspondence group developed a method for dimension reduction of shapes, by considering nested shape spaces where the most important landmarks are selected at each level. A large number of additional projects were also carried out on the general theme of Shapes and Manifolds, undertaken jointly by group members visiting SAMSI or resulting from discussions between groups members at SAMSI. A complete list is given in the individual working group reports.

All three Working Groups held regular seminar series, and extensive discussions. Coherent groups of researchers continue to work on topics of interest in this theme, and there were two sessions at the JSM 2011 in Miami Beach, Florida, which contained summaries of SAMSI material from the Shapes and Manifolds theme. A future workshop will be held in May 2012 at the Mathematical Biosciences Institute in Ohio, and the organizers of this workshop were all key participants in the Shapes and Manifolds theme.

The following subsections are reports from the working group leaders on the activities of the three manifolds working groups.

5.2 Report of Working Group: Data Analysis on Sample Spaces With Manifold Stratification

Leaders: Ezra Miller and Vic Patrangenaru

5.2.1 Overview and general research questions

At the opening workshop in Fall 2010, out of conversations between Vic Patrangenaru, Ezra Miller, and Stephan Huckemann following their presentations in 2-minute madness, it became clear that the objects the speakers in the meeting were studying all shared the desirable property of being represented as points on certain metric spaces that admit a *topological stratification*: a decomposition as a disjoint union of manifolds satisfying substantive tameness requirements (roughly: the singularities are "locally constant" on each stratum). Arguably, all object data may be represented as points on such spaces; therefore they should play a key role in twenty-first century Statistics.

Stratified sample spaces include real algebraic varieties (shape spaces or configuration spaces), polyhedral complexes (such as phylogenetic tree space), spaces of positive semidefinite matrices (from diffusion tensor imaging or any application producing distributions of covariance matrices). Our study involved a mix of geometry, probability, combinatorics, and computation, along with more usual statistical methods. The emerging field of *(geometrically) stratified statistics* elucidates phenomena in nonparametric multivariate statistical data analysis that arise when the sample space is singular. For example, tangential approximation does not yield standard statistics on Euclidean vector spaces, because singular spaces are not locally approximated by linear spaces. Therefore more geometric substitutes or analogues are required for notions such as mean, variance, principal component analysis, multi-dimensional scaling, and nonparametric bootstrap.

Historically, early results on nonparametric data analysis on Euclidean spaces, spheres, projective spaces, and orthogonal groups were pioneered by Daniel Bernoulli, deMoivre, Gauss, Cramer, Fréchet, G. Watson, David Kendall, Rudy Beran, Nick Fisher, Peter Hall, Andy Wood, John Kent, Huiling Le, Harrie Hendriks, Frits Ruymgaart, Peter Kim, Ted Chang, Herbert Ziezold, and other researchers. Data analysis on submanifolds of numerical spaces was first considered by Harrie Hendriks and Zinoviy Landsman. The first approach to nonparametric data analysis on abstract manifolds was due to Harrie Hendriks, Vic Patrangenaru, and Rabi Bhattacharya, and, later, to Peter Kim, Bruno Pelletier, Abhishek Bhattacharya, Ian Dryden, and others. On the other hand, nonparametric methods for data analysis on certain sample spaces with singularities (tree spaces or shape spaces) were developed by Susan Holmes, Steve Marron, Scott Provan, Megan Owen, Ezra Miller, Axel Munk, Stephan Huckemann, and others. These developments led us to the natural idea of developing a general approach to nonparametric data analysis on topologically stratified spaces, to extend the above methodologies for object data analysis, and to potentially address functional data analysis as well as analysis of shapes of curves or surfaces.

Collaboration between mathematicians, applied mathematicians, biostatisticians, statisticians, and computer scientists became a major theme (and goal) of the WG.

5.2.2 Key ideas and developments

1. **Statistics on negatively curved stratified spaces**

   The most fundamental phenomenon distinguishing singular sample spaces from smooth ones is that of *sticky means*: when a space is "negatively curved" near a singularity, large-sample empirical Fréchet sample means of arbitrary distributions can rest exactly at the Fréchet population mean, rather than asymptotically approaching the population mean as it does classically. Many of the WG discussions centered around quantifying this phenomenon and defining concepts so as to make it precise and gauge its generality. In particular, the WG investigated the shapes of central limit theorems in this context, particularly on spaces related to tree space (see the next subsection).

2. **Statistics on tree space**

   In the context of the previous paragraph, the motivating example for the WG of a negatively curved ("CAT(0)", or "globally nonpositively curved") singular space was the space T_n of metric phylogenetic trees on a fixed set of n taxa (Billera, Holmes, and Vogtmann 2001). It is a polyhedral complex whose number of facets is roughly factorial in the number of leaves. An algorithm to compute shortest paths in T_n efficiently was published shortly before the WG started (Owen and Provan 2011), and algorithms for centroid computation were in progress (Miller, Owen, and Provan 2011; Holmes 2011). Evolutionary biologists study distributions on T_n arising from gene tree or species tree reconstruction using genetic, morphological, or proteomic data. Notions of statistics on T_n are not new (see the survey Holmes (2003): *Statistics for phylogenetic trees*), but many issues remain open. The WG aimed to develop stratified methods for statistics and visualization specific to tree space and applications in evolutionary biology and medical imaging. Members of the WG carried out a number of computational experiments on statistical behavior of Fréchet means and other phenomena on polyhedral spaces. The tree space aspect of this WG overlapped substantially with the WG on Trees.

3. **Probability and statistics on positively curved stratified spaces**

Configurations of points can be viewed as discrete approximations to shapes in two or more dimensions. The set of equivalence classes of point configurations under group operations such as isometries, direct similarities, affine transformations, projective transformations, or other non-linear transformations is a *shape space*. Such spaces are, by definition, quotients of vector spaces modulo Lie group actions. Kendall shape spaces in dimension 3 or higher are important examples of stratified sample spaces with positive curvature. The role of singularities and stratifications in the positively curved context are markedly different from negatively curved situations, because Fréchet means tend to run away from the singularities here. In addition, the possibility of a nontrivial cut locus complicates the probability and statistics, since in this case there are no known conditions for the existence of a Fréchet mean.

4. **Probability and statistics on other stratified spaces**

   The WG spent substantial time discussing affine shape and projective shape, with applications in (for example) pattern recognition and machine vision in mind. Projective shape is widely accepted as the most appropriate notion of shape by computer scientists and others dealing with machine vision (Hartley and Zisserman). For this reason, two approaches were discussed by the WG for dealing with a statistical analysis of projective shapes in general position: one extends Patrangenaru's projective frame approach (J. Multivariate Anal. 2010), in which a specific base frame is fixed for the computation, and the other is based on a representation of projective shape by "Tyler standardization" (Tyler, Annals of Statistics, 1987; Biometrika 1987), in which every shape is brought to a projectively equivalent position that is optimal in a particular sense.

   The WG also discussed spaces of positive semidefinite matrices (covariance matrices), which play an important role in Diffusion Tensor Imaging. Again, two points of view were discussed: as a convex set in the space of matrices, and stratified according to which subsets of the eigenvalues are equal.

5. **Computational Issues**

   What type of data analysis is computationally faster on manifolds—and by extension on stratified spaces—was another issue discussed in the WG. Conversations on this topic compared and contrasted intrinsic and extrinsic analysis on shape and tree spaces as well as base-frame and Tyler-standardized approaches to image analysis. In the context of tree space, the WG had many discussions about the efficiency of various iterative algorithms to compute or approximate means.

6. **PCA on manifolds**

   The WG studied variants of PCA on manifolds, including the "forward approach" due to Fletcher et al. and the "backward approach" due to Huckeman et al. These two approaches both attempt to construct optimal nested families of submanifolds, generalizing the way PCA constructs a chain of vector subspaces by using increasing numbers of eigenvalues. The difference is that the forward approach constructs the nested submanifolds from low dimension to high dimension (essentially by PCA on the tangent space to the barycenter), while the backward approach starts with high dimension and proceeds downward by non-PCA optimization methods. Generalizing to singular stratified situations requires substantial ingenuity; it is the subject of some research projects generated by the WG.

5.2.3 Working Group participants

The following list includes everyone who registered as a participant in the working group and attended at least one session, either physically or remotely. (The SAMSI website listed 40 WG members, but a few members did not actively participate after signing up to inspect the details of the WG.) Participation ranged from nearly always absent to nearly always present. The WG administrators were Vic Patrangenaru (Florida State U) and Ezra Miller (Duke). The other participants were Karthik Bharath, Marius Buibas, Michael Crane, Arturo Donate, Ian Dryden, Leif Ellingson, David Groisser, Harrie Hendriks, Stephan Huckemann, Sungkyu Jung, John Kent, Peter Kim, Yongdai Kim, Huiling Le, Xiuwen Liu, Steve Marron, John Moriarty, Daniel Osborne, Megan Owen, Frits Ruymgaart, Armin Schwartzman, Gabe Silva, Ross Whitaker, Andy Wood, Nuen Tsang Yang, Hongtu Zhu.

5.2.4 Activities

We had a two-hour meeting every week from mid-September 2010 until the beginning of June 2011, with a one-month hiatus for winter break. Usually each two-hour period was split into two pieces, devoted to separate topics. Typically, one piece consisted of a presentation by a WG member, or a guest, on relevant past achievements or work in progress, while the other was active discussion on research in development by the WG, although sometimes both pieces were presentations, or the WG only discussed research. Occasionally the meetings were joint with the Tree WG. During the last part of the program, the Geometric Correspondence WG joined our WG, and a few broadcast meetings in Spring 2011 were focused on related Geometric Correspondence topics.

The research in the WG was highlighted in the presentations by Patrangenaru (including contributions by many WG members), by Dryden (including contributions by Kent), and by Jung and Marron (joint work with Huckemann and Hotz) at the June 2011 SAMSI Transition Workshop.

Two WG members (Marron and Miller) presented relevant WG results at the AOOD Meets Evolutionary Biology conference at SAMSI, April 30–May 2, 2011. At least seven papers were presented by WG members at the 2011 Joint Statistical Meetings in Miami Beach, FL. Four WG members (Hendriks, Huckemann, Miller, Patrangenaru) attended a conference on "Nonparametrics and Geometry", in Prague, Czech Republic. Three of them presented research connected to the WG at the conference.

A workshop proposal on the topics of the WG, written by Ezra Miller and submitted to the Mathematical Biosciences Institute (MBI), was funded to take place on 21–25 May 2011, with the title "Workshop on statistics, geometry, and combinatorics on stratified spaces arising from biological problems". The organizers are Miller along with Huckemann, Le, Owen, and Patrangenaru. At least 17 of the WG participants, including all of the key players, will attend that workshop, but it is important to note that two dozen additional prominent researchers in related areas have accepted invitations to participate or speak.

5.2.5 Working Group research output

The group has been extremely productive and most of the resulting papers in preparation or published are listed in the following annotated references.

1. Thomas Hotz, Stephan Huckemann, Huiling Le, Steve Marron, Jonathan Mattingly, Ezra Miller, James Nolen, Megan Owen, Victor Patrangenaru, Sean Skwerer (2011), Sticky central limit theorems on open books, *preprint*, (with some subset of the above as authors).

2. Stephan Huckemann, Jonathan Mattingly, Ezra Miller, James Nolen (2011), Central limit theorems in codimension 1 on nonpositively curved stratified spaces, *work in progress.*

3. Steve Marron, Ipek Oguz, Sean Skwerer (2011), Smoothing in phylogenetic tree space using simple iterated pairwise geodesics, *work in progress.*

4. Stephan Huckemann, Steve Marron, Ezra Miller, Yolanda Munoz, Megan Owen, Victor Patrangenaru, Sean Skwerer (2011), Multidimensional scaling (MDS) with curved targets, *work in progress.*

5. Megan Owen, Sean Skwerer (2011), Fréchet means in tree space, *work in progress.*

6. Rabi Bhattacharya, Leif Ellingson, Xiuwen Liu, Vic Patrangenaru, Michael Crane (2011), Extrinsic analysis on manifolds is computationally faster than intrinsic analysis, with applications to quality control by machine vision, *to appear in Appl. Stochastic Models in Business and Industry.*

7. Steve Marron, Ezra Miller, Megan Owen, Scott Provan, Sean Skwerer (2011), Towards PCA on tree spaces, *work in progress.*

8. Thomas Hotz, Stephan Huckemann (2011), Intrinsic means on the circle: uniqueness, locus, and asymptotics. *http://arxiv.org/abs/1108.2141 [stat.ME] [math.PR]*

9. Leif Ellingson, Frits Ruymgaart, Vic Patrangenaru (2011), Nonparametric estimation for extrinsic mean shapes of planar contours,*revision submitted to Ann. Statist.*

10. John Kent, Kanti Mardia (2011), The geometric approach to projective shape and the cross ratio,*paper in preparation.*

11. John Kent, Thomas Hotz, Stephan Huckemann, and Ezra Miller (2011), The geometry and topology of projective shape spaces, *preprint.*

12. Thomas Hotz, Stephan Huckemann, Huiling Le, Vic Patrangenaru (2011), Hyperbolic data analysis, *work in progress.*

13. Leif Ellingson, Vic Patrangenaru, Sean Skwerer (2011), Computational methods for extrinsic mean on phylogenetic tree spaces, *work in progress.*

14. Leif Ellingson, Vic Patrangenaru (2011), Extrinsic PCA on manifolds, *work in progress.*

15. Leif Ellingson, David Groisser, Daniel Osborne, Vic Patrangenaru, Armin Schwartzman (2011), Data analysis on spaces of positive definite matrices with an application to dyslexia detection, *preprint.*

16. Harrie Hendriks (2011), Two sample problem for mean location, *http://nonparam11.karlin.mff.cuni. cz/bookabs20110719.pdf*

17. Ezra Miller, David Houle, Paul Bendich (2011), Quantifying shape differences in fruit fly wing morphology using persistent homology, *in progress.*

18. Marius Buibas, Michael Crane, Leif Ellingson, Vic Patrangenaru (2011), A projective frame based shape analysis of a rigid scene from noncalibrated digital camera imaging outputs, *to appear in Proc. of JSM, 2011, Miami, FL.*

19. Sungkyu Jung, Stephan Huckemann, J. S. Marron, Thomas Hotz (2011), Principal nested shape spaces and an application to reduction of number of landmarks, *work in progress, http://www.samsi.info/workshop/aood-transition-workshop-june-9-11-2011.*

20. Daniel Osborne, Victor Patrangenaru, Xiuwen Liu, Hillary Thompson (2011), 3D size-and-reflection shape analysis for planning reconstructive surgery of the skull, *to appear in Proc. of JSM, 2011, Miami, FL.*

5.3 Report on Working Group: Metrics on Shape Spaces

Leaders: John Kent and Anuj Srivastava

"Shape" consists of the information in a geometric object that is invariant under a group of transformations. For objects consisting of a set of labeled points in $\mathbb{R}^d$ (typically $d = 1, 2$ or 3), these groups can include some or all of the following operations: translation, scaling, rotation, reflection and, more generally, projective transformations. For continuous objects consisting of unlabeled points such as curves and surfaces, changes to the parameterization of the object can also be included (e.g. time-warping of curves).

It is also possible to think about shape more generally, where in addition to the location of each point of an object, there is also extra information recorded at each point. Examples include a direction or a positive definite matrix. The space of shapes forms a manifold-like structure and it is important to be able to quantify differences in shape. Procrustes methods generalize Euclidean distance. For continuous objects, nonlinear deformations can also be included.

The overall goal of the working group was to explore the implications of the choice of metric in various shape spaces. A metric often induces a Riemannian structure, which in turn leads to geodesics and parallel transport. The reason for our interest in metrics is that different metrics can highlight different features of shape differences.

5.3.1 Key ideas and developments

The participants in the group had a wide variety of backgrounds ranging from engineering and image analysis to pure and applied mathematics and statistics. We discovered that many of us were tackling similar problems from different points of view, and the working group provided an opportunity for us to deepen our understanding and appreciate connections to other areas, and to build new collaborations. The unifying theme of the working group is perhaps best described as new methods to deal with "shape", defined very generally.

Several themes emerged through the discussions: the importance of key concepts in differential geometry (Riemannian metrics exponential function, parallel transport), time-warping (Fisher-Rao metric), fitting curves (especially growth curves with more general links to functional data analysis), methods for positive definite matrices, a deeper understanding of projective shape and advances in the related field of directional data analysis.

Application areas include human activity modeling using image analysis, object identification in image analysis, growth curves, handwritten signatures, protein shape, computer vision, diffusion weighted magnetic resonance imaging, neuroscience spike trains, and gene expression signals. Here is a summary of the main achievements, where the numbering of the references refers to Section 5.3.4.

1. Functional data

   Ideas originally developed for the shapes of curves (dealing with the deformation of time using the Fisher-Rao metric) have been adapted to align curves in functional data analysis. A good example is growth curves, where growth spurts may occur at different times in different individuals. [4]

   Splines are a tool for fitting nonparametrically smooth functions of time. Classical spline theory provides exact methods for fitting functions with values in Euclidean spaces; simple approximate methods have also been developed for functions with values in simple manifolds such as spheres. New work gives more sophisticated fitting methods on more general manifolds. Applications include modeling a video sequence of a dancing figure. [1,2]

2. Projective geometry

   A projective shape consists of the features of a configuration of points that are invariant under different camera views. Classically projective shape has been studied using projective invariants such as the cross ratio. However, projective invariants do not possess a natural metric structure to enable the comparisons of different projective shapes. A standardized representation of the configuration has recently been proposed to facilitate such comparisons, with links to similarity shape analysis. As a result of the Samsi program, a much deeper understanding of the topology and geometry of this standardization is now available, especially at singularities. [5,6] [This work also fits into the theme of the working group on Data Analysis on Sample Spaces with a Manifold Stratification.]

3. Directional data analysis

   Classically this subject is concerned with the directions of points on a circle or sphere, but there are natural extensions to other objects involving an orientation, such as the eigenvalues and eigenvectors of a symmetric matrix. New nonparametric kernel density methods using mixtures of Bingham distributions have been developed, with applications to the use of diffusion weighted magnetic resonance imaging in the brain to model white matter fiber orientation. [3]

   A protein consists of a sequence of amino acids in three-dimensional space and its shape can be described in terms of a collection of angles on the circle. New methods based on kernel density estimation and on mixture modeling have been developed in this setting to give a deeper understanding of protein structure. [10,11]. This work is somewhat tangential to the main interests of the working group, but was stimulated during Mardia's visit to Samsi by his contacts with the Richardson Lab (Professors David and Jane Richardson) at Duke University.

4. Feature selection in images

   Differential invariants are a version of projective invariants which have been developed to pick out features of curves in images which are invariant under different camera views. [7]

   Applications such as human activity analysis can be viewed as curves in a similarity shape space (a manifold). Using ideas such as parallel transport, such paths can be transformed into curves in Euclidean space for the purposes of statistical identification, comparison and dimension reduction. [8,9]

5.3.2 Participants

John Kent was in charge of organizing the group's activities, with Anuj Srivastava also being a joint leader. The following list includes everyone who showed an interest in the working group, with participation ranging from slight to regular: James Damon, Ian Dryden, Jinjiang He , Stephan Huckemann, Sungkyu Jung, Hachem Kadri, John Kent, Irina Kogan, Hamid Krim, Sebastian Kurtek, Huiling Le, Kanti Mardia, Jeffrey Morris, Hans Mueller, Megan Owen, Victor Panaretos, Hailin Sang, Christof Seiler, Anuj Srivastava, Valentina Staneva, Jonathan Taylor, Alain Trouvé, Jane-Ling Wang, Andy Wood, Ross Whittaker, Christine Xu, Nuen Tsang Yang, Laurent Younes, Hongtu Zhu.

5.3.3 Subgroups and their topics, activities

We had a two-hour meeting every week, from mid-September until the end of November 2010, with occasional, largely remote, meetings in the period Jan - March 2011. Typically one person would make a presentation about current achievements or work in progress. This would sometimes inspire smaller groups to work on problems in more detail, with summary results often reported back to the working group. Some talks and discussions:

- Anuj Srivastava: Shapes of elastic curves.
- Hans Mueller: Function registration.
- Sebastian Kurtek: Riemannian framework for function registration.
- Hamid Krim: Squigraph modeling of shapes
- Irina Kogan: Geometric transformation, invariance, curve matching.
- John Kent: Basis expansion of shape. Projective shapes.
- Valentina Staneva: Diffeomorphisms and RKHS.
- Ian Dryden: data-based distance metrics.

A summary of the work from the Metrics groups was presented by Ian Dryden at the Transition Workshop, as well as talks by Vic Patrangenaru, Sungkyu Jung, Sebastian Kurtek and Jingyong Su. Also, work was presented at two sessions at JSM, 2011.

5.3.4 List of papers (preprints) of work done at SAMSI or inspired by SAMSI

Several research papers have been substantially influenced by discussions and interactions at Samsi.

1. J. Su, I. L. Dryden, E. Klassen, H. Le and A. Srivastava (2011), Fitting Smoothing Splines to Time-Indexed, Noisy Points on Nonlinear Manifolds. *Submitted to Journal of Image and Vision Computing.*
2. J. Su, I. L. Dryden, E. Klassen, H. Le and A. Srivastava, A new metric for Symmetric Positive Definite Matrices. *Technical report in preparation.*
3. Dryden, I.L. and Olhede, S. (2011), A kernel density estimator for the orientation distribution function in diffusion weighted imaging, using Bingham Distributions. *Technical report in preparation.*
4. A. Srivastava, W. Wu, S. Kurtek, E. Klassen, and J. S. Marron, Registration of Functional Data Using Fisher-Rao Metric. *Technical report.*
5. J T Kent and K V Mardia, The geometric approach to projective shape and the cross ratio. *Technical report in preparation.*
6. John Kent, Thomas Hotz, Stephan Huckemann and Ezra Miller, The geometry and topology of projective shape spaces. *Technical report in preparation.*
7. Burdis J. M., Kogan, I.A., Object-image correspondence for curves under finite and affine cameras.
   Technical Report. http://arxiv.org/abs/1004.0393
8. Sheng Yi, Hamid Krim, and Larry K. Norris, Human Activity as a Manifold Valued Random Process. *IEEE Transactions on Image Processing*

9. Sheng Yi, Hamid Krim, Larry K. Norris, A Invertible Dimension Reduction of Curves on a Manifold.
Technical report. arXiv:submit/0291586 [cs.CV] 29 Jul 2011

10. Charles C. Taylor, Kanti V. Mardia, Marco Di Marzio, Agnese Panzera, Validating protein structure using kernel density estimates. *Submitted to Statistical Applications in Genetics and Molecular Biology.*

11. K. V. Mardia, J. T. Kent, Z Zheng, C. C. Taylor and T. Hamelryck, Mixtures of concentrated sine distributions with applications to bioinformatics. *Technical report in preparation.*

5.4 Report for Workgroup: Geometric Correspondence

Leaders: Stephan Huckemann and Ross Whitaker

5.4.1 Participants and Meetings

Stephan Huckemann (administrator), Ross Whitaker (administrator)

Andy Wood, Anuj Srivastava, Christine Xu, Christof Seiler, Clarisa Williams, Cong Xu, Daniel Gervini, Gosh Debashis, Ezra Miller, Giseon Heo, Hachem Kadri, Haonan Wang, Ian Dryden, James Damon, Jeffrey Morris, John Kent, John Moriarty, Kanti Mardia, Kaushik Ghosh, Leif Ellingson, Nuen Tsang Yang, Sean Skwerer, Serge Guillas, Steve Marron, Sue McDonald, Sungkyu Jung, Valentina Staneva, Victor Patrangenaru, Yongdai Kim

After spinoff the group met weekly (Thursdays 1–3 pm) with the bulk of participants having physically left SAMSI, joining over the webex interface. A period of lively discussions and formation of working subgroups followed in the months of fall. When finally also the administrators physically left SAMSI it was decided to continue on working subgroup level while the main group meetings were joined with group weekly meetings of the working group “Data Analysis on Sample Spaces with a Manifold Stratification”.

5.4.2 Thematic Content

After singling out key issues of interest, a series of talks followed which ultimately led to the formation of five working subgroups. The working subgroups began to have regular meetings independent of the main group meetings which were eventually joined with another workgroup. The working subgroups are currently still very active and meet on a regular basis, often weekly.

1. **Identification of Key Issues**

   Our first meetings were dominated by identifying key issues occurring in the analysis of geometrical objects in view of geometric correspondence. Four main topics emerged. The first two reflecting different means of representations, the third an overall underlying statistical approach and the fourth emphasizing the connection of research with driving problems in applications.

2. **Alignment of Functions**

   In this context it became clear that future work would confront the difficult issue of “alignment” without “knots”. Discussions focussed around more strongly exploiting the joint information given by ensembles of functions of interest. Also, the group got excited about an idea proposed by the group member Daniel Gervini going from the traditional “horizontal alignment” toward a “vertical alignment”.

3. **Alignment of Shapes**

   It was agreed that all infinite dimensional shape representations should be dealt with in the above context. Here, we would concentrate on shape representations given by a finite number of landmarks. Key issues not having found a satisfactory solution to date consist in the alignment of "mismatching configurations". This comprises the problem of treating occlusions, topologcial differences, unknown labeling or even unknown positioning. It became very clear that ideas borrowing from statistics could be powerful and robust enough to tackle the enormous computational challenges involved. In this context the group grew very interested in previous work of the group leader Ross Whitaker, optimizing an information theoretic content as well as in the non-deterministic "pseudo landmarking" introduced by the group member Vic Patrangenaru. Also, based on the method of "principal nested spheres" recently developed by the group members Steve Marron and Sunkyu Jung, a potential for the indentification of the statistical importance of a single landmark was noted by the group leader Stephan Huckemann.

4. **Statistical Issues**

   The above topics let to intense discussions of statistical issues and their intertwining with geometrical ideas. Within a non-Euclidean geometry which is a natural surrounding in the context of the problems at hand, elementary statistical issues e.g. regression, avoiding overfitting, the role of bias and the use of fair "p"-values turn into non-trivial challenges. In the specific geometric context, identification of information can be obtained by marginalizing over the deformation versus residual variability. Here, it was suggested to investigate an iterative "minimax" algorithmic approach.

5. **Driven by Applications**

   Several group members, including Kanti Mardia, stressed the urgent neccessity to develop sophisticated tools to address current problem in the areas of protein matching and folding. Fairly recent approaches based on angular information lead to tori model spaces which turn out to be unsatisfactory e.g. for PCA like methods. It was expected that the discussion of the above collected issues provides for a methodology to link in a genuine 3D fashion deformaton with chemistry, possibly exploiting a matching of "blobs" between structures. Such methodology may also turn out to be successful for any other comparison between organisms or parts thereof, be it ensembles of comparatively smooth surfaces of biological shapes (bones, brains, faces, hippocampi, etc.) or branching structures occuring among others in medial skeletal modeling and phylogenetic trees.

   An important aspect was added by the group member Vic Patrangenaru, pointing out that all human visual information is not based on similarity shape but on projective shape. This led to a feed back to the "Alignment of Shape" subgroup to develop a "frame-free" formulation of projective shape. Motivated by applications, we also touched the issue of treatment of shape discontinuities, and the issue of alignment "under constraints", e.g. growth.

5.4.3 Series of Talks

After a series of rather informal discussions identifying key issues of interest, group meetings provided for ample space for extended talks and intense discussions. Until the end of November 2010 the following talks and follow up extended discussions were held by core group members from their individual expertise, touching the above key goals.

- Automated correspondences (Ross Whitaker),

- functional alignment (Daniel Gervini),
- curve matching landmarks (Leif Ellingson and Victor Patrangenaru),
- alignment within Gaussian processes (John Moriarty) and
- correspondence using random fields (Ian Dryden).

5.4.4 Working Subgroup Formation

Following these talks and discussions, certain collaborative work materialized in specific research projects. The corresponding subgroups began to meet regularly.

- **Measuring Landmark Importance for Planar Shapes** Planar landmark based Kendall shape space is a complex projective space. A space with less landmarks is a lower dimensional complex projective space that can be embedded in various ways. From a given data sample with k landmarks, in previously unpublished research, a method has been developed to compute the best fitting abstract shape space with $k-1$ landmarks and in this the best abstract shape space with $k-2$ landmarks, etc. The entire procedure until the shape space of triangles is reached, or even further, up to a single geodesic and a single point can be called a *nested backward shape space principal component analysis* (cf. Jung et al., 2011, submitted). One way of assessing landmark importance is to compare the total intrinsic variances in these abstract spaces with the total intrinsic variances of concrete spaces obtained from leaving out one and more specific landmarks. Algorithmic work has begun as well as some applications to Ian Dryden's digits 3 dataset (Dryden and Mardia, 1998, Wiley) have been performed. Further application are planned to leaf data (cf. Huckemann et al., 2010, IEEE PAMI) in order to automatically identify a minimal number of landmark placment for specific discrimination problems arising in forestry. First results have been presented at the AOOD transition workshop and will be submitted for publication shortly:

  Huckemann, S., Jung S., Marron, S.(2011) Principal Nested Shape Spaces and an Application to the Reduction of the Number of Landmarks.

- **Modeling Locus and Shape Variation by Partial Rotations**

  A goal in computational anatomy is to keep track of rigid body motions as well as of deformations of internal organs, one application being radiation therapy of the male prostate. In a novel approach, we are estimating few axes of rotations of parts of organs in order to nearly exhaustively model rigid body rotation, and rotations of parts of the organ, e.g. bending and pinching. Preliminary research based on modifications of *principial arc analysis* for spoke data on medial skeletons (cf. Jung et al., 2010, submitted) has shown to have a very promising potential. Estimating dominating axes of (internal) rotations may also have applications to protein folding.

  This currently very active research has recently begun to attract numerous new collaborators and resulted in a first publication in a conference proceeding

  Pizer, S., Jung, S., Goswami, D., Zhao, X., Chaudhuri, R., Damon, J., Huckemann, S. and Marron, S.J. (2011). Nested sphere statistics of skeletal models, *in Proc. Dagstuhl Workshop on Innovations for Shape Analysis.*

  This research will also be presented at the Birs workshop, "Geometry for Anatomy" at Banff in August 2011.

- **A Frame-Free Projective Shape Space**

  All shape spaces come typically as equivalence classes on a "top" space from which – in order to obtain a meaningful topology – a certain "singularity set" has to be removed. For similarity shapes this singularity set is rather obvious. For projective landmark based shapes, a typical choice has been to select a frame in general position and to prohibit any other landmark to coincide with any of the frame points (e.g. Mardia and Patrangenaru, 2005). Obviously a statistical analysis is dependent on the frame chosen. Furthermore, it turns out that the resulting shape space may be disconnected. The components may be connected when certain landmark correspondences are tolerated. Then, among others, the resulting topology has to be carefully analyzed. This research has resulted in a collaboration with two other working groups "Metrics on Shape Spaces" and "Data Analysis on Sample Spaces with a Manifold Stratification". In fact the problems touched are very deep and fundamental for any shape analysis building on any groups other than the similarity group conveying "shape equivalence". A first preprint is in its finalizing stage:

  Kent, J., Hotz, T., Huckemann, S. and Miller, E. (2011) The topology and geometry of projective shape spaces, *Preprint.*

- **Automatic Landmark Extraction for Planar Contours**

  For the infinite dimensional shape spaces of closed contours, to date only an intrinsic statistical analysis is available coming along with the typical and quite considerable numerical challenges to compute intrinsic quantities. Moreover, for the minimization process involved, there is no guarantee that a local minimum found actually corresponds to the desired global minimimum. In order to make an extrinsic statistical methodology available also for infinite dimensional spaces, with simple numerical procedures yielding guaranteed global minimizers, for the space of planar curves one can consider the limit $k \to \infty$ for the extrinsic Veronese-Whitney embedding of Kendall's planar shape space Σ_2^k with k landmarks. Additionally, a desired boundary correspondence can be obtained by placing the k landmarks according to a suitable probabilistic model. Ongoing research into this direction has resulted at this point to

  Ellingson, L., Patrangenaru, V. and Ruymgaart F. (2011) Automatic Landmark Extraction for Planar Contours, *Preprint submitted for publication.*

- **Generative Models for Registering Ensembles of Functions and Images**

  Suppose we have a set of functions f_i of which have sampled values f_{ij}. Underlying are unwarped signals $\widetilde{f}_i = f_i \circ T_i^{-1}$ with a warping T_i. We want to estimate the T_i from the data by setting up a *generative process*, that in the first instance we consider Gaussian

  $$P(T, f) = G\big(\Sigma_{\widetilde{f}}, \mu_{\widetilde{f}}\big)G\big(\Sigma_T, \mu_T\big)$$

  which suggests a EM-approach to identify hidden parameters. E.g., such an approach for images warps images not to each other but rather to a mean. Currently active researchers are Daniel Gervini and Ross Whitaker.

- **Minimax Correspondence for Shape Classification**

  Many applications deal with a specific classification problem. Here we present a generic method of addressing this issue, which is of interest for applications in forestry, assessing leaf variation of leaf shape over genotype, temperature gradient and leaf location in the tree (e.g. crown or breast hight). Previous research using landmark based shape spaces has shown that for powers of tests the "correct" number and placement of landmarks is of great importance. In this project for a specific classification problem given, an automated landmarking scheme is to be developed that places landmarks such that the intra-group variation is minimal while simultaneously the variation over the groups is maximal. This method is to be based on purely statistical features (Cates et al., 2006, Int Conf Med Image Comput Comput Assist Interv). Another potential application of this method lies in the early diagnosis of degenerative brain processes such as Alzheimer's desease. Research in this project is currently at a very conceptional level, both working out a stringent mathematical formulation and making feasible a numerical approach. Currently active researchers are Stephan Huckemann, Ross Whitaker, Martin Styner.

6 Report for the Working Group: Brain Imaging

Leaders: John Aston (*Chair*), Jeffrey Morris, Hans Müller, Haipeng Shen and Jane-Ling Wang. SAMSI Postdoc and Group Webmaster: Ci-Ren Jiang.

6.1 Summary

This working group started out specifically focused on the link between functional data analysis and brain imaging. However, during the SAMSI AOOD year, the group extended its focus to all statistical issues associated with brain imaging data sets. In particular three subgroups were formed to examine the aspects of Deconvolution and Design-Free analysis, Spatial and Temporal Modeling, and Hierarchical analysis in Brain Imaging (this last subgroup being also associated with the Hierarchical Modeling SAMSI working group). The group had regular weekly meetings and talks during the Fall semester and more focused working group meetings (including talks and increasingly presentations on current work) during the Spring semester. Throughout the year, there were very close connections with the Functional Data Analysis groups, in particular the WG on hierarchical functional data also developed an emphasis on brain imaging and the resulting papers could also be co-opted by this WG.

6.2 Participants

The working group consisted of 103 members who regularly received the email notifications and updates on the group activities, making it one of the largest SAMSI working groups. Not all of these were active members, but there was still considerable activity throughout the year.

Active Participants: Adrian Bowman, Alois Kneip, Armin Schwartzmann, Bill Schucany, DuBois Bowman, Giles Hooker, Fan Li, Hernando Ombao, Hongtu Zhu, Ian Dryden, Juhyun Park, Lexin Li, Nuen Tsang Yang, Phil Reiss, Richard Samworth, Seonjoo Lee, Sylvie Tchumtchoua, Tingting Zhang, Todd Ogden, Xia Wang. Many additional participants joined for particular talks or group meetings, as well as the six or so outside speakers, who gave presentations on their work to the group.

6.3 Activities

As mentioned, there were 3 subgroups

1. Deconvolution and Model Free Analysis in Brain Imaging - leader Haipeng Shen. This working group mainly examined the problem of estimating parameters in time series analysis of functional Magnetic Resonance Imaging (fMRI) data where part of the experimental design

was unknown. One particular focus was Hemodynamic Response Function (HRF) estimation. In usual fMRI analysis, the HRF is assumed to be known and convolved with the experimental design before linear modeling takes place. The working group examined under different modeling assumptions how best to account for the HRF when it was assumed to be unknown. This lead to several innovative techniques being developed during the year, as was the focus of 2 talks during the overall SAMSI AOOD transition workshop, with an additional talk being focused on deconvolution models in Positron Emission Tomography. A second area of interest was the converse problem of when the HRF was assumed known but where the experimental design was now unknown. This can be framed in a change point context and several talks and projects throughout the year investigated this area.

2. Spatial Temporal Subgroup - leader John Aston. This subgroup explored the issues of spatial and temporal correlations in brain imaging and particular as to how this might affect techniques such as functional data analysis, which are typically assumed to have independent errors. This subgroup especially benefitted from many speakers from outside areas talking about their experiences for different data and the subgroup then explored how these ideas might be more focused on imaging data. Particularly areas of research included the problems of massive data structures on covariance estimation, accounting for temporal or spatial correlation in principal component analysis and functional basis approaches for representing imaging data.

3. Hierarchical Modeling - leader Jeffrey Morris. This working group emphasised looking at how multi-trial or multi-subject data could be analysed efficiently through various approaches. One aspect of this working group that was of particular interest in relation to the other subgroups was the inclusion of Bayesian modeling principles as well as frequentist approaches into the framework. fMRI was again the primary focus of this group, but other modalities were also considered. As mentioned earlier, this working group was somewhat of a cross-over subgroup between two working groups, the brain imaging working group and the hierarchical modeling working group.

The working group also took an active role in some of the larger workshops that occurred during the year. Several of the working group leaders (Hans Mller and Jane-Ling Wang) were also very much involved in the longitudinal data analysis workshop (with many brain imaging group members being active participants in this workshop). In the final AOOD transition workshop, the Brain Imaging working group (particular Jane-Ling Wang) organised the final half-day session with five talks in all areas covered by the working group. In addition, there was a final panel discussion at the end of the AOOD workshop primarily focussing on how to make the statistical methodologies of groups such as the Brain Imaging group more readily available and accessible to those working in the field of Brain Imaging as practitioners.

A few examples of projects inspired and/or carried out by the Brain Imaging Working Group

- UCD team (in alphabetical order: O. Carmichael, J. He, H.-G. Müller and J.-L. Wang) - We study resting state FMRI data that were collected at UC Davis and develop functional correlation measures. The goal is to quantify spatial dependencies of the FMRI signals at various voxels, aiming at improved connectivity measures and subject classification. The work is still ongoing and benefitted from discussions of the SAMSI Brain Working group. Part of the work is performed while Mueller and Wang visited SAMSI in Fall 2010, so we would like to acknowledge SAMSI and the brain working group.

- Tingting Zhang - I participated in SAMSI 2010-11 Program on Analysis of Object Data, because my research interest of brain imaging data analysis falls within the applications areas emphasized by the program. Under its support, I was able to visit SAMSI in 2011 spring, and developed collaborative projects with Dr. Ahmad Hariri, Professor of Psychology and Neuroscience and Investigator in the Institute for Genome Sciences and Policy at Duke University. My collaborator Dr. Fan Li from Duke University and I were motivated by SAMSI working group discussions. We submitted an NSF grant on functional imaging genetic data analysis. In addition, we are going to finish a paper on nonparametric inference of functional magnetic resonance imaging (fMRI) data. In a short summary, we have greatly benefited from the opportunities presented by the SAMSI program

- Phil Reiss - My work on voxel-by-voxel nonparametric inference for samples of brain images (in collaboration with Lei Huang of NYU) was inspired by discussions at the September and November 2010 workshops on Analysis of Object Data. In particular, a conversation with Ciprian Crainiceanu at the opening workshop led to a visit to his working group at Johns Hopkins, which gave me an opportunity to present an early version of the work and to have further helpful discussions. The final meeting of the Brain Imaging working group provided me with a valuable opportunity to present two functional data analytic aspects of this project: (1) a procedure for clustering of voxelwise function estimates, and (2) a modified smoothing procedure that borrows strength across neighboring voxels. The feedback received during my presentation helped a great deal to fine-tune this work, which is currently being prepared for publication.

6.4 Papers/Preprints/Work in Preparation

1. Aston JAD and Kirch C. Detecting and estimating epidemic changes in dependent functional data, *submitted.*

2. Aston JAD and Kirch C. Estimation of the distribution of change-points with application to fMRI data, *in preparation.*

3. Bunea F, She Y Ombao H, Gongvatana W, Devlin K and Cohen R. (2011). Penalized Least Squares Regression Methods and Applications to Neuroimaging, *NeuroImage*, (55), 1519-1527.

4. Fiecas, M. and Ombao, H. (2011). The Generalized Shrinkage Estimator for the Analysis of Functional Connectivity of Brain Signals, *Annals of Appl Stat, in press.*

5. Gorrostieta C and Ombao H.. General Spectral Measures of Cross-Dependence in Multivariate Time Series, *JASA, under revision.*

6. Gorrostieta C, Ombao H, Bedard P and Sanes J.N. Investigating Stimulus-Induced Changes in Connectivity Using Mixed Effects Vector Autoregressive Models, *NeuroImage, under revision.*

7. Gorrostieta C, Ombao H, Rrado R, Patel S and Eskandar E. Coherence Analysis of Local Field Potentials, *Journal of Time Series Analysis, under review.*

8. Jiang C, Wang JL and Aston JAD. Functional PCA based Deconvolution for Position Emission Tomography, *in preparation.*

9. Kang H, Ombao H, Linkletter C, Long N and Badre D. Spatio-Spectral Mixed Effects Model for Functional Magnetic Resonance Imaging Data, *JASA, under revision.*

10. Lee S, Shen H, Truong Y, Lewis M, and Huang X (2011) Independent Component Analysis Involving Auto-correlated Sources with an Application to Functional Magnetic Resonance Imaging, *Journal of the American Statistical Association, accepted.*

11. Morris JS, Baladandauthapani V, Herrick RC, Sanna PP, and Gutstein HG (2011). Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomic data, *Annals of Applied Statistics, 5(2A), 894-923.*

12. Nam CFH, Aston JAD and Johansen AM, *Quantifying uncertainty in change points, in revision.*

13. Ogden, R. T., Zhao, Y., and Reiss, P. T. Wavelet-based functional principal component regression, *In preparation.*

14. Shen H, Tian S, and Huang J. A Two-Way Regularization Method for MEG Source Reconstruction, *under review at Annals of Applied Statistics.*

15. Shen H, Trong Y, and Lee S. Asymptotics of Colored Independent Component Analysis, *under review at Annals of Statistics.*

16. Shen H, Zhang L and Huang J. Two-Way Robust Functional Data Analysis, *to be submitted shortly.*

17. Shen H, Truong Y and Chen W. Hemodynamic Response Function Estimation, *in preparation.*

18. Shi, XY, Zhu, HT, Ibrahim JG, Styner M, Intrinsic regression models for median representation of subcortical structures.

19. Tchumtchoua, S., Dunson, D. B., and Morris, J. Online Variational Bayes Inference for High-Dimensional Correlated Data, *In preparation (to be submitted soon).*

20. Tchumtchoua S, A Heterogeneous Dynamic Structural Equation Model with Application to Brain Connectivity, *in preparation.*

21. Van Lunen D, Ombao H and Aston JAD. Online Detection Methods: A Model Selection Framework. *In preparation.*

22. Yuan, Y., Zhu, H.T., Lin, W. L., and Marron, J. S. Local polynomial regression for symmetric positive definitive matrices.

23. Yuan, Y., Zhu, H.T., Styner, M., J. H. Gilmore., and Marron, J. S. Varying coefficient model for modeling diffusion tensors along white matter bundles.

24. Zhao, Y., Ogden, R. T., and Reiss, P.T. Wavelet-based LASSO in functional linear regression, *Submitted manuscript.*

25. Zhao, Y., Bagiella, E., and Ogden, R. T. A functional approach to analysis of RR interval variability, *In preparation.*

26. Zhu, H., Brown, P. J., and Morris, J. S. Robust classification of functional and quantitative image data using functional mixed models, *Submitted.*

7 Report for the Working Group: Tree Structured Data Objects

Leader: J.S. Marron

7.1 Overview

Following the Opening Workshop, and follow-on discussion, three main approaches emerged:

1. Combinatorial: There were some discussions of this approach, but the main players seemed to move outside the SAMSI framework, perhaps mostly due to SAMSI restrictions on meeting times.

2. Phylogenetic Trees: This group continued very actively through the entire year, and there were very interesting interactions with the Stratified Manifolds Working Group. Half of the meeting time each week was devoted to this effort.

3. Dyck Path: This group was also very active throughout the year, and occupied the other half of the meeting time.

In both of the latter two groups, activities centered around weekly discussions. In each case, there were both presentations of related work by mostly irregular participants, and discussion of on-going work led by regular participants. Much of the on-going work was done by a single graduate student, who would present results, and then get feedback as to what to do for the following week.

Major Topics Discussed / Explored by the Phylogenetic Tree Sub-Group:

- John Aston's phylogenies with covariance matrix representations of dialects as leaves.
- Metrics for covariance matrices as data.
- Susan Holmes gave excellent overview of phylogenetic methods and ideas.
- Phylogenetic Bootstrapping and Multi-Dimensional Scaling
- Discussed Tom Nye's approach to PCA in phylogenetic tree space
- Explored MDS of Brain Vessel trees, and used to probe curvature of tree space. Motivated idea of MDS with
- "curved target space".
- Stickiness fed into heuristic improvement over Sturm's algorithm.
- Visualization of geodesic paths.
- Smoothing in Phylogenetic Tree Space.
- Gene Trees vs. Species Trees.
- Scott Schmidler's phylogenies with shape objects as leaves.
- Variations on Sturm's Algorithm.
- Challenges in defining "convex set".

Major Topics Discussed / Explored by the Dyck Path Sub-Group:

- Correspondence issues, motivated by standard shape considerations.

- Branching processes as probability models.
- Graphical representations of Dyck Paths
- Dyck path mean, and unsuitability of it (some potential fixes)
- Node length representation and analysis
- Tree Pruning Analysis, for deeper insights into population structure.
- Explored transformed Node Length Representations
- Problem with Dyck Path PCA leaving Tree Space. Solution: Non-Negative Matrix Factorization
- Neuromorph Group: Neurons as tree structured data objects.
- Kim's Bayesian Gaussian Latent Factor Model Approach

7.2 Participants

Participants, Phylogenetic Tree Sub-Group: The pivotal graduate student, who did most of the work in this group was Sean Skewerer, UNC Statistics and Operations Research. Active Regular Participants were: J. S. Marron (UNC), Ezra Miller (Duke), Megan Owen (UC Berkeley), John Aston (Warwick), Huiling Le (Nottingham) , Snehalata Huzurbazar (Wyoming). Irregular participants included: Ian Dryden (South Carolina), Susan Holmes (Stanford), Scott Schmidler (Duke), Chris Challis (Duke), Sridevi Polavaram (NeuroMorph).
Participants, Dyck Path Sub-Group: The central graduate student, who did most of the work in this group was Dan Shen, UNC Statistics and Operations Research. Active Regular Participants were: J. S. Marron (UNC), Lingsong Zhang (Purdue), Haipeng Shen (UNC), Yolanda Munoz (Michigan Tech.), Yongdai Kim (Seoul National). Irregular participants included: Haonan Wang (Colorado State) Sarabdeep Singh (Wyoming), Elizabeth Bullitt (UNC), Andrew Wood (Nottingham) , Ruchi Parekh (NeuroMorph).

I.E.2 Final report: SAMSI Program on Complex Networks (2010-11)

Background

This year-long program focussed 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.

The program considered several 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.

Overall Program Leaders: Eric Kolaczyk (Boston U.), Alessandro Vespignani (Indiana U.)

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: Rick Durrett (Cornell/Duke) and Pierre Gremaud (NCSU)

National Advisory Committee Liaison: Bin Yu (UC Berkeley)

1 Workshops

1.1 Opening workshop

1.1.1 Summary

The Opening Workshop for the SAMSI program on Complex Networks was held on Sunday-Wednesday, August 29-September 1, 2010, at the Radisson RTP in Research Triangle Park, NC. There were around 150 participants. Tutorial sessions were held on Sunday, August 29 on

- Eric Kolaczyk (Boston University), *statistical analysis of network data*,
- Alessandro Vespignani (Indiana University), *diffusion and epidemic processes in complex techno-social networks*,
- Rick Durrett (Duke University), *some features of the speed of epidemics and opinions on a random graph*,
- Michael Mahoney (Stanford University), *geometric tools for identifying structure in large social and information networks*.

Videos of these presentations are available at the SAMSI website. Invited talks were presented Monday to Wednesday. There was a poster session and reception on Monday, August 30. Immediately following the workshop, on Thursday and Friday, research working groups convened for initial meetings at SAMSI.

1.1.2 Activities

The workshop focused on five complementary themes at the forefront of current research in complex networks, incorporating perspectives ranging from theory to applications, in disciplines spanning applied mathematics, computer science, probability, statistics, and statistical mechanics. The five focus areas were

Network Sampling and Inference: Theory and methods relating to the sampling of network data and the corresponding inference of network characteristics, including applications to tasks like community detection and estimating the size of 'hard-to-count' populations. This session was led by Eric Kolaczyk and featured the following presentations

- Peter Bickel (UC-Berkeley) *statistical inference for unlabelled graphs*,
- Mark Newman (U. of Michigan) *community structure and link prediction in networks*,
- Matt Salganik (Princeton U.) *improvements to the network scale-up method for estimating the sizes of hard-to-count populations*,
- Panel discussion: Stan Wasserman (Indian University), Liza Levina (U. of Michigan) and Bruce Spencer (Northwestern U.)

Dynamic Networks: Modeling and inference of networks in the context of dynamical systems evolving in time, such as time-varying gene regulatory interactions or email social networks. Steve Fienberg (Carnegie Mellon U.) led this session. The following talks were given

- Raissa D'Souza (UC Davis) *what are dynamic networks?*
- Tom Snijders (Oxford U.) *actor-oriented models for network dynamics*,
- Sidney Redner (Boston U.) *dynamics of voter models on heterogeneous networks*,
- Panel discussion: Peter Mucha (U. of North Carolina), Hugh Chipman (Acadia U.), Josh Socolar (Duke U.).

Percolation and Diffusion on Networks: Theory regarding the behavior of phenomena like social contact processes and epidemics on networks. This session was directed by Rick Durrett; the following talks were offered

- Zoltan Toroczkai (U. of Notre Dame) *modeling functional networks for the primate cortex*,
- Erik Volz (U. of Michigan) *simple models for infection disease epidemics in complex networks*,
- Pierre Degond (CNRS) *continuum models for complex systems*,
- Panel discussion: James Moody (Duke U.), Ginestra Bianconi (Northeastern U.), Mason Porter (Oxford U.).

Spectral Analysis and Geometric Algorithms: Topics in network analysis integrating aspects of spectral graph theory and algorithms based on geometric embeddings. Michael Mahoney chaired this session which included the following presentations

- Fan Chung Graham (UC San Diego) *PageRank algorithms with applications to graph sparsification and partitioning*,
- Aaron Clauset (Sante Fe Institute) *the trouble with community detection*,
- Mauro Maggioni (Duke U.) *multi scale methods for analysis of graphs.*

Biological Applications of Networks: Various applications of network-based approaches to biological problems, such as in the context of cell biology or the epidemic spread of disease. This session was chaired by Alex Vespignani and featured the following presentations

- Eric Xing (Carnegie Mellon U.) *time varying networks; reverse engineering and analyzing rewiring social and genetic interactions*,
- Hongzhe Li (U. of Pennsylvania) *statistical methods for network-based analysis of genomic data*,
- Michelle Girvan (U. of Maryland) *effects of network topology in simple models of gene regulation*,
- Desmond Higham (U. of Strathclyde) *algorithms for evolving networks*.

In addition, several shorter talks were given by beginning researchers

- Crystal Linkletter (Brown U.) *explaining network structure: the importance of modeling pair-wise preferences*,
- Joe Bliztstein (Harvard U.) *respondent-driven sampling: degrees of uncertainty with with uncertain degrees*,
- Alexander Gutfraind (Los Alamos NL) *dark networks and vital infrastructure*,
- Natalia Katenka (Boston U.) *the impact of partial information on network inference and characterization*,
- Edoardo Airoldi (Harvard U.) *integer polytope samplers with applications to network analysis*.

1.1.3 Working groups

The opening workshop resulted in the formation of five working groups which we now briefly describe. Report from each individual group are included in Section 3.

Sampling/modeling/inference This group aimed to work towards moving the current state of knowledge on these inter-related tasks – in the specific context of networks – to rest on a more principled and integrated mathematical and statistical foundation. We are pursuing this goal by focusing on a handful of specific prototype problems in the context of certain application areas, ranging from information networks to animal communities to neuroscience.

The group leader was Eric Kolaczyk (Boston U.).

Dynamics OF networks The dynamics of networks working group explored a variety of mathematical and statistical approaches for describing and understanding the changing connection topology of networks over time, the interplay of these network dynamics with other dynamic processes on the network, and the connections between these different mathematical and statistical methodologies.

The group leaders were David Banks (Duke U.), James Moody (Duke U.) and Peter Mucha (U. of North Carolina).

Dynamics ON networks Random graphs are useful models of social and technological networks. To date most of the research in this area has concerned geometric properties of the graphs. This working group will focus on processes taking place ON the network. In particular we are interested in how their behavior on networks differs from that in homogeneously mixing populations or on regular lattices of the type commonly used in ecology and physics.

The group leaders were Rick Durrett (Duke U.) and Alun Lloyd (NCSU).

Geometrical/spectral analysis This working group was concerned with the following topics: detection of communities in networks, multiscale spectral methods for the analysis of the geometry of networks, algorithms that simplify graphs into simpler graphs in order to speed up certain optimization problems, metrics for comparing graphs, and multiscale homogenization of random walks. These topics have applications biology and to spread of "epidemics" in financial networks.

The group leaders were Mauro Maggioni (Duke U.) and Michale Mahoney (Stanford U.).

Modeling flows This working group consisted of two subgroups: (i) modeling traffic flows and (ii) modeling smart grid networks. In the traffic flows subgroup, people were interested in both stochastic and deterministic models to represent microscopic and macroscopic behavior of traffic flows. In modeling smart grid networks, the subgroup worked on mathematically formulating a graph/topology reduction problem of representing the larger network to a reduced number of local area networks, then modeling the corresponding dynamics and parameter estimation for the reduced system.

The group leader was Taufiquar Khan (Clemson U.).

1.2 Complex networks modeling workshop, October 20-22, 2010, SAMSI

1.2.1 General description

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 sampling, modeling and inference in practice. In the context of network data, however, such an infrastructure is only beginning to be developed.

The goal of this workshop was to bring together researchers working on the sampling, modeling, and inference of networks, for the purpose of helping move the current state of knowledge on these inter-related tasks to rest on a more principled and integrated mathematical and statistical foundation. Topics of focus include recent advances in network sampling (e.g., respondant driven sampling), inference from partially sampled (e.g., ego-centric) network data, and the confluence of traditional models (e.g., stochastic block-models, Gaussian graphical models) with modern tools for high-dimensional data analysis (e.g., l1-penalized optimization, spectral partitioning).

Organizers:

David Banks (Duke U.) and Eric Kolaczyck (Boston U.).

1.2.2 Talks

- Stephen Fienberg (Carnegie Mellon U.) *statistical challenges in network modeling*,
- Edo Airoldi (Harvard U.) *network representation*,
- Tian Zheng (Columbia U.) *statistical methods for studying social networks using aggregated relational data*,
- Purnamitra Sarkar (Carnegie Mellon U.) *theoretical justification of popular link prediction heuristics*,
- Andrew C. Thomas (Carnegi Mellon U.) *exploring the limits of conditionally independent dyadic network models*,

- Lucy Robinson (Johns Hopkins U.) *change point detection in networks*,
- Bruce Spencer (Northwestern U.) *sampling research questions*,
- Krista Gile (U. of Washington) *self-consistent network model-assisted prevalence estimation from respondent-driven sampling data*,
- Bin Yu (UC Berkeley) *spectral clustering and the high-dimensional stochastic Block model*,
- Aarti Singh (Carnegie Mellon U.) *identifying graph-structured network activitations*,
- Stephane Robin (Agro Paris Tech) *uncovering structure in interactions networks*,
- Denise Scholtens (Northwestern U. Medical School) *sequential sampling designs for estimating local connectivity in bait-prey graphs.*

In addition, a poster reception took place on Wednesday, Oct. 20., 2010.

1.3 Workshop on the dynamics of networks, January 10-12, 2011, SAMSI

1.3.1 General description

The changing structure of networks over time impact and are indeed inherent in the study of a broad array of network phenomena. The network of contacts for the spread of an infectious disease varies in time, with that variation playing a potentially important role in the course of the disease. Ad hoc communications networks between roaming elements must continously readjust and renavigate between nodes according to the changing landscape of connections. Political networks of association connections or voting similarities vary from one legislative session to the next.

The detailed local social and/or technological processes underlying each of these example applications obviously differ, but many of the basic mathematical and statistical questions regarding such networks and the generalized information they carry are similar. Though the importance of dynamics in networks has of course been long recognized, renewed interest has emerged in part due to the increasing accessibility of dynamic network data, ranging from longitudinal data waves to complete time histories of network evolution. Additionally, most of the theoretical modeling work that has been 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, statistical and computational tools for analyzing 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 thus remains an ongoing need and opportunity for more thorough mathematical and statistical analysis and modeling of dynamic networks. This workshop aimed to bring together researchers interested in pushing forward this extremely fertile area of research.

Organizers:

Raissa D'Souza (UC Davis), Stephen Fienberg (Carnegie Mellon U.), Eric Kolaczyk (Boston U.), Jim Moody (Duke U.), Peter Mucha (U. of North Carolina), Mason Porter (Oxford U.)

1.3.2 Talks

- Peter Mucha (U.of North Carolina) *the "OF" working group*,
- Carter Butts (UC Irvine) *modeling complex social interaction within and across settings via relational events*,

- Reka Albert (Penn State) *modeling the dynamics of biological signaling networks*,
- Carsten Wiuf (Aarhus U., Denmark) *likelihood and likelihood-free inference for certain growing network models*,
- J.P. Onnela (Harvard U.)
- Eric Kolaczyk (Boston U.) *analysis of time-indexed networks in epilepsy*,
- Skye Bender-deMoll (Self-employed) *escaping from the matrix: storing, exploring and explaining dynamic networks*,
- Peter Hoff (U. of Washington) *multiway array models for dynamic networks and relational data*,
- Pavel Krivitsky (Carnegie Melon U.) *a separable model for dynamic networks*,
- Ginestra Bianconi (Northeastern U.) *dynamics of social interactions at short timescales*,
- Dani Bassett (UC Santa Barbara) *dynamic community structure in adaptive systems*,
- Katy Borner (Indiana U.) *visualizing the structure and evolution of science and technology*.

In addition, time was reserved for the discussion of working group specific issues and for a poster reception.

1.4 Workshop on pedestrian traffic flow, February 14-16, 2011, SAMSI

1.4.1 General description

Human crowds and pedestrian groups exhibit complex and coordinated spatio-temporal patterns such as the spontaneous organization of pedestrian ows into lines and the oscillations of fluxes at gates or intersections. Understanding these phenomena requires a deeper knowledge of the laws governing the interactions of individuals with both themselves and their environment. Although a number of experiments can be found in the literature, the available data does not lend itself well to systematic study. The lack of completeness and varying quality of available experimental data make it nearly impossible to identify the fundamental principles underlying pedestrian behavior. Motion capturing techniques have become available in the recent years and permit more complex experiments, yielding comprehensive recordings of individual positions. Such experiments generate a formidable amount of data. Mathematical modeling and numerical simulation can facilitate the extraction of information from such data, helping to identify relevant observables by which a set of biological hypotheses can be validated. Such a dialectic use of experiment and simulation results in the ability to extract from traffic data an ensemble of behavioral rules, which can then be inserted into individual (agent)-based simulation models. For crowds involving several thousand (or more) individuals, continuum fluid-mechanical-like models are attractive, from both the computational and theoretical viewpoints. Deriving these models from the individual behavior by means of statistical physics guarantees their biological relevance and opens the way to qualitative analysis and understanding of many observed emerging phenomena in crowds.

The workshop aimed at bringing together experts from different elds with the main objective to systematically develop and analyze a hierarchy of pedestrian traffic models. The studied hierarchy of models consists of three levels: (i) Microscopic description, (ii) Mesoscopic description and (iii) Macroscopic PDE description. The microscopic description is similar to spin-flip models, an interacting system of cellular automata. All modeling assumptions are only made at the microscopic level. Mesoscopic and macroscopic models are rigorously derived through appropriate averaging and limiting procedures; no ad-hoc assumptions on the density and velocity of the flow are used in the derivation. One of the main goals of the workshop was to naturally link three levels of description, allowing for a detailed analysis of how various terms in the macroscopic PDE model are affected by the microscopic assumptions.

Organizers:

Alina Chertock (NCSU), Pierre Degond (CNRS, France), Alexander Kurganov (Tulane U.)

1.4.2 Talks

- Jian-Guo Liu (Duke U.) *dynamics of orientational alignment and phase transition*,
- Vladislav Panferov (California State U., Northridge) *continuum description for systems of self-propelled particles used in modeling fish migration*,
- Anthony Polizzi (Tulane U.) *modern freeway traffic flow models*,
- Ilya Timofeyev (U. of Houston) *group report on traffic flow modeling*,
- Pierre Degond (CNRS) *current challenges in pedestrian dynamics and crowd modeling*,
- Alexander Kurganov (Tulane U.) *numerical methods for traffic and pedestrian flow models*,
- Sébastien Motsch (U. of Maryland) *a traffic model for pedestrian and its comparison with experimental data.*

In addition, time was reserved for the discussion of working group specific issues.

1.5 Workshop on dynamics on networks, March 21-23, 2011, SAMSI

1.5.1 General description

It has now been clearly established that many social and technological systems are complex networks. However after the structures have been estimated and the geometric properties of the graphs such as their "small world" nature have been studied, there remains the question: How does the structure of the network affect the behavior of processes taking place on the network? The SAMSI working group Dynamics ON Networks studied this question for evolutionary games and various models of the spread of opinions and epidemics, both for tree like random networks and for clustered networks. The main purpose of this workshop was to facilitate progress on group projects by bringing long distance collaborators to SAMSI and by bringing researchers whose work had been important to the group's investigations.

Organizers:

Rick Durrett (Duke U.), Alun Lloyd (NCSU), Peter Mucha (U. of North Carolina) and Alex Vespignani (Indiana U.)

1.5.2 Talks

- Vittoria Colizza (ISI and INSERM) *human mobility in an emerging epidemic: a key aspect for response planning*,
- Elizabeth Leicht (Oxford U., UK) *interacting networks: formalism and significance*,
- Joel Miller (Harvard U.) *epidemic spread in networks with one equation*,
- Duygu Balcan (Indiana U.) *phase transition in contagion processes mediated by recurrent mobility patterns*,

- Charlie Brummitt (UC Davis) *sandplie cascades on interacting tree-lie networks*,
- James Gleeson (U. of Limerick, Ireland) *analytical results for cascades on networks*,
- Mason Porter (Oxford U., UK) *cascades on networks*,
- Sid Redner (Boston U.) *the role of reinforcement in social dynamics*,
- Nicolas Lanchier (Arizona State U.) *The Axelrod model for the dissemination of culture revisited*,
- Sara Solia (Northwestern U.) *fast and slow dynamics in neural networks with small-world connectivity*,
- Deena Schmidt (Ohio State U.) *network structure and dynamics of sleep-wake regulation*,
- John McSweeney (Concordia U. and SAMSI) *single seed cascades on networks with triangles*,
- Rachel Kranton (Duke U.) *strategic interactions and networks*,
- David Sivakoff (SAMSI) *contact process on modular networks*,
- Shirhsendu Chatterjee (Cornell U.) *latent voter model on random regular graphs*,
- Shankar Bhamidi (U. of North Carolina) *flows, first passage percolation and random disorder in networks*.

In addition, time was reserved for the discussion of working group specific issues and for a poster reception.

1.6 Transition workshop, June 6-7, 2011, SAMSI

1.6.1 General description

This workshop provided a time to look back over the achievements of the program and to highlight directions for future research. This two-day workshop featured half-day sessions on the topics of the five working groups: sampling, modeling and inference; geometric and spectral properties; dynamics of networks; dynamics on networks; and flows and networks.

Organizers:

Eric Kolaczyk (Boston U.), Michael Mahoney (Stanford U.), Mason Porter (Oxford U.) and Alex Vespignani (Indiana U.)

1.6.2 Talks

- Kash Balachandran (Duke U.) *comparison of local spectral clustering algorithms*,
- Alan Lenarcic (University of North Carolina) *multiple latent trait model for interaction in expander networks*,
- Blair Dowling Sullivan (Oak Ridge NL)
- David Banks (Duke U.) *grooming networks in a baboon troop*,
- Bruce Rogers (SAMSI) *fitting monkey models*,
- Amanda Traud (NCSU) *ant modeling*,

- Tyler McCormick (Columbia U.) *surveying hidden populations through sampled respondents in a social network: a comparison of two strategies*,
- Ali Shojaie (U. of Michigan) *reconstructing directed regulatory networks from multiple steady-state and perturbed gene expression profiles*,
- Eric Kolaczyk (Boston U.) *some progress on asymptotics for ERGMs*
- Tipan Verella (U. of Virginia) *the dual of the random intersection graph*,
- Bill Shi (U. of North Carolina) *robust scaling behavior in dynamics voter models*,
- David Sivakoff (SAMSI) *evolving voter results.*

In addition, time was reserved for the discussion of working group specific issues and for a poster reception.

2 Courses and workshop for students

A one semester course for graduate students and a workshop for undergraduate students were offered as part of this program.

2.1 Graduate course: Complex Networks: Theory and Applications

This course was offered during the Fall semester of 2010. Ten graduate students from the University of North Carolina at Chapel Hill , Duke University and North Carolina State University took the course for credit while another ten people attended the class on a regular basis. Lectures were given at SAMSI on Tuesdays, 4:30-7:00 p.m.

This course focused on the mathematical and statistical analysis and modeling of networked systems, such as arise in biological, social, and technological contexts. Both static and dynamic perspectives were studied. Specific topics included network graph construction and relevant sampling issues, characterization of networks, community detection, and network modeling and inference. Various applications were considered, including in social networking, biology, and epidemiology.

Texts included

- Kolaczyk, E.D. (2009). Statistical Analysis of Network Data: Methods and Models. Springer, New York.
- Durrett, R. (2006). Random Graph Dynamics. Cambridge University Press.
- Durrett, R., (2010), "Some features of the spread of epidemics and information on a random graph", Proc. Nat. Acad. USA, (107) 2010, pp. 4491-4498

In addition, various additional publications, handouts, etc, were provided. The course outline and the corresponding instructors are listed below

1. September 7: Introduction and motivation; network mapping. (Eric Kolaczyk)
2. September 14: Network characterization. (Eric Kolaczyk)
3. September 21: Network sampling and inference. (Eric Kolaczyk)
4. September 28: Community detection (Peter Mucha)

5. October 5: Community detection continued. (Peter Mucha)
6. October 19: Network role/positional analysis (James Moody)
7. October 26: Network change: How/why in social settings (James Moody)
8. November 2: Characterization of dynamic networks (James Moody)
9. November 9: Epidemic processes on networks (Alun Lloyd)
10. November 16: Epidmic processes on networks continued. (Alun Lloyd)
11. November 30: Dynamics on networks (Rick Durrett)
12. December 7: Dynamics on networks continued (Rick Durrett)

2.2 Undergraduate student workshop

A two-Day Undergraduate Workshop was held at SAMSI on October 29-30, 2010, on the general theme of Complex Networks. This workshop was part of SAMSIs Education and Outreach Program for 2010-2011. The following lectures and events took place

- Pierre Gremaud (SAMSI) *welcome and introduction*,
- Rick Durrett (Duke U.) *what are graphs and why are they important?*
- Yi Sun (SAMSI) *MATLAB demonstration*,
- David Sivakoff (SAMSI) *random graphs*,
- Mauro Maggioni (Duke U.) *graphs, high dimensional data and visualization*,
- Hongziao Zhu (SAMSI) *R demonstration*,
- Joshua Mendelsohn (Duke U.), *using R for scientific research; R lab*,
- Pierre Gremaud (SAMSI) *career options*,
- David Banks (Duke U.) *a history of network modeling*,
- Bruce Rogers (SAMSI) and Mandi Traud (NCSU) *graph clustering with random walks; MATLAB lab*.

3 Working group reports

3.1 Sampling/modeling/inference working group

Leader: Eric Kolaczyk (Boston U.)

3.1.1 Participants

Edo Airoldi (Harvard U.), Shankar Bhamidi (UNC-CH), David Banks (Duke U.), Shirshendu Chatterjee (Cornell U.), Eric Kolaczyk (Boston U.), Pavel Krivitsky (Carnegie Melon), Yingbo Li (Duke U.), Crystal Linkletter (Brown U.), Tyler McCormick (Columbia U.), Bruce Rogers (SAMSI), Cosma Shalizi (Carnegie Melon U.), Ali Shojie (U. of Michigan), Sean Simpson (NCSU), Bruce Spencer (Northwestern U.), Tian Zheng (Columbia U.).

3.1.2 Research activities and results

The working group focused on topics that arguably involve some or all of sampling, modeling, and inference of networks. Much of the focus was restricted to static networks, but some projects had nontrivial overlap with interests in the Dynamics OF working group as well. Members of the working group are active on six group projects.

Project 1: Inference on Agent-Based Models for Networks Agent based models (ABM) specify local interaction rules between individuals and implement these rules through computer simulation. Agents are placed together in a simulated environment, and then one can observe repeated realizations of the evolution of social networks among the agents. But such applications need methods for statistical inference that relate the parameters (agent rules) to the network behavior. Important open problems include how to estimate parameters given a time series of the (simulated) observables or how assess the goodness-of-fit of the model from the given data. These challenges are difficult because each agent can have its own set of parameters, so the dimension of the parameter space grows geometrically in the number of agents. To perform dimension reduction in parameter space and make inference possible, David Banks, Bruce Rogers and Cosma Shalizi built statistical emulators for agent based models using a Bayesian framework for the calibration of computer models. As a proof of concept, they used the "Sugarscape" agent-based model from Axtel and Epstein's book Growing Artificial Societies.

Project 2: Fit Assessment for Attachment Models Much of the recent literature on real-world networks concerns mathematical models which relate their characteristics, including degree distributions, small worldedness, and clustering coefficients to parameters in various preferential attachment models. At the simplest level new nodes enter the system and decide to link themselves to nodes in the pre-existing network with probability proportional to some functional of the vertices in the network, for example their present degree. Models based on such schemes arise in a number of areas, ranging from economic systems wherein nodes try to optimize two competing functionals to evolutionary biology (Yule processes). Often one looks at the degree distribution of the real network at hand and matches it to an appropriate attachment model and then uses various properties of the model to infer properties about the network (e.g., epidemic durations). However, there is no mathematical theory adequate to judge such fits. Shankar Bhamidi, Shirshendu Chatterjee and Eric Kolaczyk explored this area in order to provide the necessary inferential theory.

Project 3: Baboon Grooming Data A consortium of primatologists has collected data on social interactions in twelve baboon troops in Kenya. A subgroup consisting of David Banks, Yingbo Li, Bruce Rogers and Ali Shojie is attempting to fit a dynamic latent space model to grooming relationships over time, using covariate information on kinship, relatedness, age, health, gender, reproductive status, dominance hierarchy, and rainfall. The key aim is to product the observed fission in the baboon troop, which entails a block model for the the proximity matrix in the latent space.

Project 4: Effective Sample Size in Network Modeling Stochastic models for network structure and processes have applications spanning the social, informational, and biological sciences. However, complex networks have complex structure, and when sophisticated models reflect this in their own dependence structure, ordinary sampling and asymptotic results used to quantify uncertainty often do not apply. Indeed, the only unambiguously independent sample for many network processes is repeated independent observations of realizations of that network over the same set of actors — something that is, with a few exceptions, not possible. And yet, observing more actors in a network clearly provides more information about the network's structure than observing fewer. Pavel Krivitsky, Sean Simpson, Crystal Linkletter, and Eric Kolaczyk worked toward establishing notions of effective sample size for network models, taking into account the dependence structure and parametrization of these models and the process by which the network is observed.

Project 5: Estimation of the Degree Distributions The degree distribution of a sampled network can differ from the degree distribution of the underlying population network, as pointed out, for example, in Subnets of Scale- Free Networks are Not Scale-Free: Sampling Properties of Networks Stumpf, Wiuf, May

(PNAS, 2005). Work by Frank has shown how to obtain unbiased estimates of the number of nodes of any degree given a simple random sample of nodes, and normalizing that distribution yields an estimate of the degree distribution. Bruce Spencer and Eric Kolaczyk worked to extend those results in two directions. First, the mean and variance of the degree distribution can be calculated from the triad census, and hence they can be estimated from an induced graph sample. Work is underway to improve Frank's estimators by adjusting them to estimates of the mean and variance based on the sample triads. Second, work is underway to extend the results from samples where nodes are included with equal probabilities to samples based on unequal probability sampling.

Project 6: Comparison of Aggregated Relation Data and Ego-Centric Sampling Tyler McCormick, Tian Zheng, and Eric Kolaczyk studied a framework for inference in multi-relational data measured through standard surveys. Our goals are (i) formalizing the information contained in two sampling methods (aggregated relational data and ego-centric fixed-size nominations) (ii) exploring the relationship between these two methods (iii) learning about the relationship between two nested networks. Mini-Project: Minimum Description Length Inference Edo Airoldi and David Banks have written an NSF proposal that includes a study of minimum description length inference for a class of latent space models. If funded, this project will surely move forward; if not, its pursuit will depend upon available energy and time.

3.1.3 Working papers and publications

1. Krivitsky, P.N. and Kolaczyk, E.D. (2012). On the question of effective sample size in network modeling. Submitted.
2. McCormick, T.H., He, R., Kolaczyk, E.D., and Zheng, T. (2012). Surveying hard-to-reach groups through sampled respondents in a social network: a comparison of two survey strategies. *Statistics in Biosciences* Special Issue on Networks, 4:1, 177-195

3.2 Dynamics OF networks working group

Leaders: David Banks (Duke U.), James Moody (Duke U.) and Peter Mucha (UNC-CH)

3.2.1 Participants

See Dynamics ON networks list.

3.2.2 Research activities and results

This working group formed as a natural outgrowth of common interests at the Opening Workshop. The group included significant overlap of interests with the other CN working groups, including but not limited to analysis of longitudinal data (overlapping with the Geometrical/Spectral group), the modeling of networks via agent based models (overlapping with Sampling/Inference), and the exploration of abstract voter models (as just one specific project overlapping with the Dynamics ON Networks group). Indeed, because of the significant overlap with the ON group, the weekly meeting of the two groups were merged for the Spring semester.

Starting at the OW, there has been intense early interest in the working group on bridging the different network perspectives commonly attributed to communities in statistics, probability, and statistical physics. Much of this interest in the group has focused on the phenomenon of explosive percolation (described at the OW in the presentation of Raissa DSouza), namely the observation in some constructive models where the giant component emerges in a seemingly sharp transition, cf. in classic models such as the Erdos-Renyi random graph model. Quantifying this notion mathematically, understanding the scaling window and understanding how small components merge into larger components eventually leading to the formation of the giant component brings in very interesting connections to the famous Smoluchowski equations in colloidal chemistry, and Markov processes such as the multiplicative coalescent. One of the hopes of the working group is to develop the techniques that will enable us to prove the first mathematically rigorous results in this context. Our early efforts in this direction after the opening workshop included conversations in the group led by Steve Fienberg and Raissa DSouza, among others. Our ongoing efforts aimed towards rigorous results are currently led by Shirsendu Chatterjee, Charlie Brummitt, Shankar Bhamidi, Raissa DSouza, and Rick Durrett.

A second model problem of focus in the working group is the analysis of Dynamic Voter Models [Holme & Newman, 2006], hybrid models that incorporate both a rewiring dynamic, so the network changes in time, and a voter interaction process which modifies the node/actor opinion to drive agreement between connected nodes in the network. When the selected rewiring dynamic is to drop connections between disagreeing actors, the two processes cooperate in pushing the network towards a final state of complete consensus within each component, but the distribution of properties of 228 those components depends on the relative frequency of activity of the rewiring and interaction processes (codified by a parameter). Our primary goals include the demonstration of the existence of a critical value for this rewiring-frequency parameter, at which the model changes behavior between the two extremes of a purely voter model dynamic and a purely rewiring dynamic, the development of model moment-closure equations for approximating this process, and the rigorous analysis of the process (especially at and near criticality). Numerical simulations are being used to explore the behavior of this system, including the different behaviors observed at different rewiring frequencies, edge densities, and opinion initial conditions. We are using a pair- approximation approach to gain a heuristic understanding of the process and the location of the critical value. We are also using a rigorous large deviations approach to demonstrate the existence of the critical value. This work is led by David Sivakoff, Bill Shi, Rick Durrett, Alun Lloyd, Peter Mucha, and Mason Porter.

A third subgroup project is investigating the role of concurrent partners (overlapping in time) in the spread of sexually transmitted diseases. STD transmission through the dynamic contact network must obviously obey an arrow of time, requiring that the start time and duration of each edge must be respected, and the resulting questions about the influence of concurrency in enhancing rapid spread are important for modeling such systems and for properly targeting behavioral and biological interventions. Multiple researchers have developed competing measures for the level of concurrency in a networks, and a wide variety of approaches have been used to study disease dynamics as a function of concurrency, including but not limited to extensive use of simulations. One part of this working groups activity in this area is in addressing the utility of different network representations thereof. For instance, the temporal information can be used to create a directed network of possible transmissibility (either exactly or nearly obeying all temporal rules). An alternative approach is to explore the allowed permutations in a fully ordered edge list (those which do not change the available transmission paths). These representations will be used to develop formal models of dynamic sexual networks, targeting bounds on the size of the largest or expected disease outbreak as a function of an appropriate global measure of concurrency. This work is led primarily by Bruce Rogers, Alun Lloyd, Jim Moody, and Peter Mucha.

3.2.3 Working papers and publications

1. Rombach, M.P., Porter, M.A., Fowler, J.H., Mucha, P.J. Core-periphery structure in networks. Submitted.

3.3 Dynamics ON networks working group

Leader: Rick Durrett (Duke U.) and Alun Lloyd (NCSU)

3.3.1 Participants

Shankar Bhamidi (UNC), Charlie Brummitt (UC Davis), Shirshendu Chatterjee (Cornell visiting Duke), Ariel Cintron-Arias (East Tennessee State U), Luis Gordillo (U Puerto Rico), Christian Gromoll (Virginia), Peter Kramer (Rensselaer), Jim Lynch (South Carolina), John McSweeney (Concorida), Peter Mucha (UNC), Mason Porter (Oxford), Michael Roberts (NC State), Puck Rombach (Oxford), Bill Shi (UNC), David Sivakoff (SAMSI), Mandy Traud (NC State), Tipan Verella (Virginia)

3.3.2 Research activities and results

A number of projects have emerged for investigation.

SIS models on networks with communities. During the Stochastic Dynamics program, John McSweeney and Bruce Rogers considered the contact process (SIS epidemic) on a strongly clustered graph, described as a stochastic block model. In the simplest case there are two communities with the probability of an edge connecting two individuals within the same community p1 and the probability of an edge connecting two individuals between communities p2 much smaller than p1. In a range of parameters, a plot of the number of infected sites versus time shows a stair step: the epidemic reaches equilibrium in one community before spreading to the other. Work on this problem is continuing in the Dynamics ON Networks working group with David Sivakoff and Rick Durrett. A paper has been published.

Nonlinear voter models. The ordinary voter model with a linear response function is now fairly well understood on complex networks due to its duality with coalescing random walk. Redner et al have studied various nonlinear versions. Chatterjee and Durrett will try to use ideas of Cox, Durrett, and Perkins to show that the mean field predictions for the nonlinear voter model are accurate when the nonlinear response function is almost linear. They think that this method can be used to cover the latent voter model where voters who have just changed their opinions enter an inert state for an exponential amount of time with mean λ if the rate λ is large. Although this is a small perturbation of the voter model, the behavior changes discontinuously, the density converges to 1/2 starting from any positive level.

In **Axelrods model**, voters have a vector of F opinions, each of which can take one of q values. Each voter x at rate 1 picks a neighbor y, and one of his opinions i. If the two agree on issue i then x picks a j and imitates ys opinion on that issue. When $q = F = 2$ this reduces to the model with leftists (00), centrists (01, 10) and rightists (11). Lanchier has proved some results in one dimension about the clustering in the case $q = F = 2$ and freezing when q is larger than F . Durrett and J.C. Li will work on generalizing the second result to two dimensions, starting with the case when both q and F are large.

Pair approximation. Mandy Traud and Michael Roberts are coding and testing the full pairwise approximation model for SIS epidemics using the equations exhibited in Keeling and Eames - Modeling Dynamic and Network Heterogeneities in the Spread of Sexually Transmitted Diseases. They plan to test actual network data with these equations, code few different closure approximations, and test different selections of parameters.

Suppressing cascades of load in interdependent networks Charlie Brummitt and his collaborators studied cascades. Understanding how interdependence among systems affects cascading behaviors is increasingly important across many fields of science and engineering. Inspired by cascades of load shedding in coupled electric grids and other infrastructure, they studied the BakTangWiesenfeld sandpile model on modular random graphs and on graphs based on actual, interdependent power grids. Starting from two isolated networks, adding some connectivity between them is beneficial, for it suppresses the largest cascades in each system. Using a multitype branching process and simulations they showed these effects and estimated the optimal level of interconnectivity that balances their trade-offs.

The Dual of a Random Intersection Graph. Christian Gromoll and Tipan Verella are interested in studying the emergence of a giant component in the dual of a generalized random intersection graph.

Conceptually, given a set of nodes and a set of attributes, a random intersection graph is a graph on nodes induced by constructing a random bipartite graph between the nodes and the attributes. Interest in random intersection graphs, has increased in recent years. However, very little attention has been devoted to studying the dual graph naturally constructed by the same process that produces the random intersection graph.

3.3.3 Working papers and publications

1. Brummitt, C.D., D'Souza, R.M., Leicht, E.A. Suppressing cascades of load in interdependent networks. *Proc. Nat'l. Acad. Sci.* 109 (2012), E680-E689.

2. Durrett, R., Gleeson, J., Lloyd, A., Mucha, P., Shi, F., Sivakoff, D., Varghese, C. Graph fission in an evolving voter model. *Proc. Nat'l. Acad. Sci.* 109 (2012), 3682-3687.

3. Durrett, R. and Sivakoff, D. Contract process on modular networks. In progress.

4. Verella, J.T. A randomized approximation algorithm for community detection in complex networks. In preparation.

3.4 Geometrical/spectral analysis working group

Leaders: Mauro Maggioni (Duke U.) and Michael Mahoney (Stanford U.)

3.4.1 Participants

Prakash Balachandran (Duke), Pietro Poggi Corradini (Kansas State), Mihail Cucuringu (Princeton), Raymond Falk, Mauro Maggioni (Duke), Michael Mahoney (Stanford), Peter Mucha (UNC Chapel Hill), Mason Porter (Oxford), Ali Shojaie (U. Wisconsin Madison), Blair Sullivan (Oak Ridge Lab).

3.4.2 Research activities and results

The group has been studying a number of topics in spectral and multiscale techniques for graphs, and their geometric and algorithmic implications. It focused on several problems.

1. Multiscale conductance and time-series Networks. Inspired by recent work on multislice networks, and on multiscale conductance, the group studied the behavior of multiscale methods (e.g. conductance, community detection, diffusions) on multislice networks.

2. Soft clustering. Most existing community-detection algorithms find hard partitions in the network, while in some situations it seems more natural to construct soft- partitions. Few methods do that, but it is a new topic and little is well-understood. It would also be interesting to apply new ideas to the case of multislice networks.

3. Tree decompositions. These decompositions simplify graphs, leading to faster optimization/solution of problems which are computationally hard. They are related to important problems in statistics (especially in graphical models) but such ties are not well- understood. Moreover, these decomposition algorithms are right now very combinatorial in nature, and one may wonder if they would become easier/more stable by weaking some of the requirements, while still being useful.

4. Community Detection. Spectral techniques based on intermediate eigenvectors of the Laplacian and/or modularity may be used to detect communities.

5. Scan Statistics/sparse regression. Scan statistics on graphs is not well understood and only few technique exist, albeit many applications seem to be in need of such tools, such as epidemics (early detection), genomics (identification of pathways relevant in a disease), etc...
6. Epidemic/Reactive Networks. Corradini collected and is processing data from questionnaires that allow mapping social interactions and behavior in a community, both under normal conditions and in the case of the emergence of an epidemics. Many questions arise: would the network change enough to prevent the epidemics? How would the network change? What would be efficacious ways of modifying the network to prevent the spread?

3.4.3 Working papers and publications

3.5 Modeling flows

Leader: Taufiquar Khan (Clemson U.)

3.5.1 Participants

Aranya Chakrabortty (FREEDM Center, NCSU), Alina Chertock (NCSU), Alex Kurganov (Tulane U.), George Michailidis (U. of Michigan), Anthony Polizzi (Tulane U.), Sean Simpson (NCSU), Yi Sun (SAMSI), Ilya Timofeyev (U. of Houston), Zifang Wang (UC Davis).

3.5.2 Research activities and results

The group looked at two specific "flow applications" on networks: traffic flows and smart grids.

Traffic flows

Most of the activities of that subgroup were centered around the derivation of new continuum models based microscopic approaches. In particular, new stochastic optimal velocity models were proposed and analyzed. Several numerical issues linked to the resolution of nonlocal interactions were also studied. The group also considered a new multi-class traffic flow model with Arrhenius look-ahead dynamics. They first derived a cellular automatum model and study its numerical solution obtained by the Monte Carlo simulations. They then passed to the PDE limit, which is a hyperbolic system of conservation laws with global fluxes. The solution of the PDE system was numerically studied using the central-upwind scheme. Finally, the group also considered pedestrian traffic flows. One-dimensional models for the behavior of pedestrians in a narrow street or corridor were proposed and investigated. The corresponding model was formulated at the microscopic level as a cellular automata model with explicit rules for pedestrians moving in two opposite directions. Coarse-grained analogs was derived leading to the coupled system of partial differential equations for the density of the pedestrian traffic. The group compared and contrasted the behavior of the microscopic stochastic model and the resulting coarse-grained equations for various parameter settings and initial conditions. Diffusive higher-order corrections were also discussed. A related workshop was held on February 14-16, 2011, see Section 1.4.

Smart grid

There is tremendous growth in scientific research efforts in "smart grid related field. This new trend in growth is particularly clear from the number of organized groups, centers, consortia developed just in the past few years (in addition to the existing ones). The electrical network, a good example of a complex network, has many different aspects and fits well into several areas of applied mathematics and statistics: control and dynamical systems, network optimization and power quality, network communication/information theory.

These aspects of the power network and the new development in the area of power grid brings new challenges that have significant overlap with existing applied mathematical and statistical techniques as well as there is a need to develop new techniques to tackle the relevant pressing problems to develop an efficient and sustainable power network. In particular, the group investigated the following issues.

1. Graph and Model Reduction for Large Scale Power Systems Network. In this thrust, the groups aimed at reducing the complex system of generators, buses with loads etc to a reduced graph problem and then formulate the model reduction problem for the dynamics of the major components.

2. Optimal Load Management in the Network. An optimization problem for the power flow equation was formulated in order to extend to dynamic settings the usual steady state setting (which is more or less already done).

3. Identification of Critical Paths in Power Network. Here is the goal was to identify which vertices/paths of the graph have critical power connections. The group then investigated how the system configuration changes if one or more of these critical paths are removed from the network for example does it have significant impact to cause blackouts etc.

SAMSI held a related Smart grid workshop in the Fall of 2011 as part of the Uncertainty Quantification program. Several papers are in preparation.

3.5.3 Working papers and publications

1. Sun, Y. and Timofeyev, I. New stochastic models for traffic flows. In preparation.

2. Chertock, A., Kurganov, A., Polizzi, A. Multi-Class Traffic Flow Model with Look-Ahead Dynamics. Submitted.

3. Chertock, A., Kurganov, A., Polizzi, A., Timofeyev, I. Pedestrian Flow Models with Slowdown Interactions. Submitted.

I.E.3 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 in July.

Undergraduate Outreach Days

The two outreach workshops are held annually to expose undergraduates from institutions 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.

Complex Networks: October 29-30, 2010

The Fall outreach workshop focused on topics from the SAMSI Program on Complex Networks. 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, Rick Durrett (Duke) gave a general introduction to the importance and ubiquity of graph theory. Yi Sun (SAMSI postdoc) then led a tutorial on MATLAB. This was followed by a more focused and research oriented talk by David Sivakoff (SAMSI) on random graphs. In the afternoon, Mauro Maggioni (Duke) gave a very nice overview on graphs, high dimensional data and visualization. This was followed by an introduction to R given by Hongziao Zhu (SAMSI). Most of the reminder of the afternoon was devoted to a presentation and R lab organized by Josh Mendelsohn from the Dept. of Sociology at Duke. Josh did a fantastic job giving the students a flavor of research work in a interdisciplinary setting. 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. First, David Banks (Duke) gave the students an overview of the history of network modeling. This was followed by a presentation from Mandi Traud (NCSU) and Bruce Rogers (SAMSI) on graph clustering with random walks. Mandi and Bruce's presentation was paired with an interactive session in MATLAB. Details regarding the workshop can be obtained at http://www.samsi.info/workshop/two-day-undergraduate-workshop-october-29-30-2010

Of the 30 students participating, 20 were female, with no minorities. A total of 47 persons, including faculty and mentors, participated.

Analysis of object data: February 25, 26, 2011

The Spring outreach workshop was devoted to the analysis of object data (AOD). 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. The first talk in the workshop was given by Jim Ramsay (McGill) who gave an overview of functional data analysis. This was followed by an introduction to MATLAB given by Ci_Ren Jiang (SAMSI). Snehelata Huzurbazar (U. of Wyoming) then described several interdisciplinary projects pertaining to AOD, ranging from glaciology to bioinformatics. After lunch, Hulin Wu (U. of Rochester) discussed dynamics modeling of bioterrorism. This was followed by a short introduction to the marvels of R given by Junheng Ma (SAMSI). Hongxiao Zhu (SAMSI) gave the students an introduction to the analysis of functional data objects, a key part of the AOD program. The students then worked on a MATLAB lab related to Huzurbazar's presentation in the morning. The lab was overseen by both Huzurbazar and Sylvie Tchumtchoua (SAMSI). 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. On Saturday morning, Yolanda Munoz-Maldonado (Michigan Tech) discussed her personal research projects in functional data. The rest of the morning was devoted to an R lab related to the analysis of a specific (functional) data set. The lab was led by both Yolanda and David Degras (SAMSI). Details regarding the workshop can be obtained at http://www.samsi.info/workshop/two-day-undergraduate-workshop-february-25-26-2011

Of the 27 students participating, 11 were female, 3 were African-American, 2 were Hispanic. A total of 36 persons, including faculty and mentors participated.

Undergraduate modeling workshop: May 17-21, 2010

The undergraduate modeling workshop is a yearly week event. The focus is different from the outreach workshop. Here, we consider one application and guide the students through the entire modeling process, from the design of a model, its mathematical analysis, its statistical properties, its implementation and finally discussion of the results. 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 and real data.

As was the case in 2010, we decided to build on the strengths present at SAMSI and consider an application related to epidemiology. This field pertains to one of the 2010-11 programs (Complex Networks). The students were first guided through introductory presentations related to classical epidemiology models (SIR and the likes). They were also given background in basic statistics and probability, R and MATLAB. A first interactive session allowed to the students to test their skills on a "perfect" data set: a well known boarding school influenza example for which perfectly mixed models such as SIR work very well. After some background on complex network theory and parameter estimation methods, the students were made to work on actual CDC flu data. That data set corresponds to about ten regions in the US for which aggregated flu counts are given at regular time intervals. The students had the task of adapting simple models to this situation. For the part of the workshop, the students work in groups (of about three). Some groups successfully adapted the models to specific regions, some others applied new tools from network theory. On the final day of the workshops, each student team presented the results obtained during the week.

The workshop was a resounding success, so much so in fact that the 2011 version will be devoted to epidemiology again. The 2011 heavily involved all SAMSI postdocs and SAMSI graduate fellows who interacted with the undergraduate students through lectures, interactive session and socially. In the middle of the week on Wednesday morning, two panels are assembled to discuss professional development issues. One panel gathers faculty with various background and addresses graduate school opportunities. The other panel is devoted to career options after studies are completed; the panelists come from various industries and companies. More details about the workshop can be found at http://www.samsi.info/content/student-schedule-interdisciplinary-workshop-undergraduate-students-and-faculty-may-2011

For the first time, we ran a parallel workshop, from Monday to Wednesday, for faculty from primary teaching institutions. The benefits of this approach were twofold. First, it facilitates the participation of students from these primarily teaching institutions. Second, it allows the participating to put together a new teaching module. About 10 faculty participated. Their schedule was the same as the students on Monday and Wednesday morning, but we designed specific activities for them on Tuesday, say here more details http://www.samsi.info/content/faculty-schedule-interdisciplinary-workshop-undergraduate-students-and-faculty-may-2011

While the participating faculty commented in general very positively about the time at SAMSI (good organization, interesting plenary lectures, fantastic opportunity for their students, networking opportunities), several aspects of the workshop require some fine tuning. Among them, the recruitment of the involved faculty should be improved (the workshop is aimed at non expert); the participating faculty had very different background which made the proper design of a teaching module challenging.

Of the 32 students participating, 13 were female, 3 were African-American, 2 were Hispanic. A total of 44 persons participated, including faculty and mentors.

SAMSI Graduate Fellow Poster day, April 13, 2011

Through the year, the 11 SAMSI graduate fellows are involved in research projects directly related to working group activities. To recognize their achievements, the entire SAMSI community gathers to celebrate their achievements in a poster reception. Each student presents his/her results in the form of a poster. The students are encouraged to present their work at relevant conferences and make use of their SAMSI poster there.

Industrial Mathematical and Statistical Modeling (IMSM) Workshop: July 19-27, 2010

The ten-day Industrial Mathematical and Statistical Modeling Workshop for Graduate Students was the 16th year in the series and the 8th sponsored 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.

Both the undergraduate modeling workshop and the graduate IMSM workshop share (and 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.

Of the 32 students participating, 11 were female, 3 were African-American, and 1 was Hispanic. There were a total of 44 persons participating, including faculty and mentors.

The projects investigated during the workshop corresponded to current research problems presented by scientists from the Environmental Protection Agency, MIT Lincoln Laboratory, the National Center for Atmospheric Research, Progress Energy, Sandia National Research Laboratories and the Battelle Memorial Institute. Each team gave a 30 minute oral presentation summarizing their results on the final day of the workshop and written reports were compiled and are available at http://www.ncsu.edu/crsc/reports/ftp/pdf/crsc-tr10-16.pdf Details regarding the workshop can found at http://www.samsi.info/workshop/2010-industrial-mathstat-modeling-workshop-graduate-students

Diversity

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

Courses

See the program reviews in Section I.E for discussion of the SAMSI courses. Two courses were offered during the Fall semester in 2010 and one during Spring 2011; these are credited 3 credits/units at each of the participating Universities.

Graduate Course on Complex Networks. This course was offered dur during the Fall semester of 2010. Ten graduate students from the University of North Carolina at Chapel Hill , Duke University and North Carolina State University took the course for credit while another ten people attended the class on a regular basis. Lectures were given at SAMSI on Tuesdays, 4:30-7:00 p.m. This course focused on the mathematical and statistical analysis and modeling of networked systems, such as arise in biological, social, and technological contexts. Both static and dynamic perspectives were studied. Specific topics included network graph construction and relevant sampling issues, characterization of networks, community detection, and network modeling and inference. Various applications were considered, including in social networking, biology, and epidemiology.
Texts included

- Kolaczyk, E.D. (2009). Statistical Analysis of Network Data: Methods and Models. Springer, New York.
- Durrett, R. (2006). Random Graph Dynamics. Cambridge University Press.
- Durrett, R., (2010), ”Some features of the spread of epidemics and information on a random graph”, Proc. Nat. Acad. USA, (107) 2010, pp. 4491-4498

In addition, various additional publications, handouts, etc, were provided.

The course outline and the corresponding instructors are listed below

1. September 7: Introduction and motivation; network mapping. (Eric Kolaczyk)
2. September 14: Network characterization. (Eric Kolaczyk)
3. September 21: Network sampling and inference. (Eric Kolaczyk) .
4. September 28: Community detection (Peter Mucha)
5. October 5: Community detection continued. (Peter Mucha)
6. October 19: Network role/positional analysis (James Moody)
7. October 26: Network change: How/why in social settings (James Moody)
8. November 2: Characterization of dynamic networks (James Moody)
9. November 9: Epidemic processes on networks (Alun Lloyd)
10. November 16: Epidmic processes on networks continued. (Alun Lloyd)
11. November 30: Dynamics on networks (Rick Durrett)
12. December 7: Dynamics on networks continued (Rick Durrett)

Graduate course on the Analysis of Object Data I
Principal Instructors: I. Dryden, J.S. Marron, H.G. Mueller, J. O. Ramsay, J.L. Wang
Lectures were given at SAMSI on Wednesdays, 4:30-7:00 p.m.
Attendance: 46.

This was the first of two courses associated with the SAMSI program on Analysis of Object Data. These courses provided an introduction into selected areas of object oriented data analysis, aiming at the following topics:

1. Introduction to Object Oriented Data (Marron)
2. Shape Analysis and Related Topics (Dryden) Introduction to Statistical Shape Analysis, Non-Euclidean Shape Spaces, Distances and Shape Co-ordinates, Procrustes Analysis, Principal Components Analysis and Geodesics.
3. Functional Data (Mueller) Introduction to Functional Data, Functional Regression Models, Time Warping, Empirical Dynamics.
4. Functional and Longitudinal Data (Wang) Functional Principal Component Analysis, Modeling with Covariates, Interface Functional and Longitudinal Data Analysis.
5. Functional and Dynamic Data (Ramsay)

Graduate course on the Analysis of Object Data II
Principal Instructors: J.S. Marron, J. O. Ramsay
Attendance: 17.

This was the second graduate course associated with the SAMSI program on Analysis of Object Data. Lectures were given at SAMSI on Wednesdays, 4:30-7:00 p.m. It covered topics different from those covered in the course taught in the Fall and consisted of three main segments:

1. Dynamic systems: (Ramsay) Models for multivariate functional data that explicitly model change through the involvement of one or more derivatives in the model specifica- tion. Topics include the anatomy of a dynamic system, dynamic systems as extensions of functional models, and parameter estimation and inference for systems where analytical solutions are impossible.
2. Manifold data: (Ramsay and Marron) This topic extends functional data analysis in a number of ways, where data are typically distributed over time and/or space, to situations where the data are distributed over manifolds embedded within higher dimensional spaces. Along with a quick review of classic subjects such as principal components analysis and test theory, more advanced topics include medial shape representations, diffeomorphisms in image analysis, and diffusion tensor imaging.
3. Tree-structured data: (Marron) This section contrasts the very diverse combinatorial, folded Euclidean and Harris correspondence approaches for this new area of data analysis.

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.

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.

Publications and Technical Reports

Analysis of Object Data

Published/Accepted

1. Acar, E., Craiu, R.V., Yao, F. (2011). Dependence calibration in conditional copula: a nonparametric approach. *Biometrics*, accepted.

2. Rabi Bhattacharya, Leif Ellingson, Xiuwen Liu, Vic Patrangenaru, Michael Crane (2011), Extrinsic analysis on manifolds is computationally faster than intrinsic analysis, with applications to quality control by machine vision, *to appear in Appl. Stochastic Models in Business and Industry.*

3. Marius Buibas, Michael Crane, Leif Ellingson, Vic Patrangenaru (2011), A projective frame based shape analysis of a rigid scene from noncalibrated digital camera imaging outputs, *to appear in Proc. of JSM, 2011, Miami, FL.*

4. Bunea F, She Y Ombao H, Gongvatana W, Devlin K and Cohen R. (2011). Penalized Least Squares Regression Methods and Applications to Neuroimaging, *NeuroImage*, (55), 1519-1527.

5. Cao, J., Cai, J. and Wang, L. (2011). Estimating Curves and Derivatives with Parametric Penalized Spline Smoothing, Accepted by *Statistics and Computing.*

6. Chen, K., Chen, K., Müller, H.G., Wang, J.L. (2011). Stringing high-dimensional data for functional analysis. *Journal of American Statistical Association* **106**, 275–284.

7. Chen, D., Hall, P., Müller, H.G. (2011). Single and multiple index functional regression models with nonparametric link. *Annals of Statistics* **39**, 1720–1747.

8. Dass, S. C., Lim, C. Y. and Maiti, T. (2011). Default Bayesian Analysis for Multivariate Generalized CAR Models, To appear in *Statistica Sinica.*

9. Chen, K., Müller, H.G. (2011). Conditional quantile analysis when covariates are functions, with application to growth data. *J. Royal Statistical Society B*, accepted

10. Ding, J.-M., Symanzik, J., Sharif, Wang, J. -L., Duntley, Shannon, W. (2011). Powerful Actigraphy Data Through Functional Representation. *Chance*, accepted.

11. Fiecas, M. and Ombao, H. (2011). The Generalized Shrinkage Estimator for the Analysis of Functional Connectivity of Brain Signals, *Annals of Appl Stat, in press.*

12. Hooker, G., S. P. Ellner, L. Roditi and D. J. D. Earn (2011) Parameterizing State-space Models for Infectious Disease Dynamics by Generalized Profiling: Measles in Ontario, Journal of the Royal Society Interface, 8:961-975.

13. Lee S, Shen H, Truong Y, Lewis M, and Huang X (2011) Independent Component Analysis Involving Auto-correlated Sources with an Application to Functional Magnetic Resonance Imaging, *Journal of the American Statistical Association, accepted.*

14. Li, P.L. and Chiou, J.M. (2011) Identifying cluster numbers for subspace projected functional data clustering. *Computational Statistics and Data Analysis* **55**, 2090-2103.

15. Lim, C. Y. and Dass, S. C. (2011). Assessing Fingerprint Individuality Using EPIC: A Case Study In The Analysis Of Spatially Dependent Marked Processes, *Technometrics*, vol. 53, no. 2, pp. 112-124.

16. Liu, Y. and Yuan, M. (2011). Reinforced multicategory support vector machines. *Journal of Computational and Graphical Statistics*, accepted.

17. Liu, Y., Zhang, H. H., and Wu, Y. (2011). Soft or hard classification? Large margin unified machines. *Journal of the American Statistical Association* **106**, 166-177.

18. Lu, W., Zhang, H. H., and Zeng, D. (2011). Variable selection for optimal treatment decision. *Statistical Methods in Medical Research*, accepted.

19. Morris JS, Baladandauthapani V, Herrick RC, Sanna PP, and Gutstein HG (2011). Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomic data, *Annals of Applied Statistics, 5(2A), 894-923.*

20. Müller, H.G. (2011). Functional data analysis. *International Encyclopedia of Statistical Science*, Ed. Lovric, M. Springer Science Business Media, Heidelberg. (Extended version available in StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies, id 242).

21. Müller, H.G., Sen, R., Stadtmüller, U. (2011). Functional data analysis for volatility. *Journal of Econometrics*, accepted

22. Daniel Osborne, Victor Patrangenaru, Xiuwen Liu, Hillary Thompson (2011), 3D size-and-reflection shape analysis for planning reconstructive surgery of the skull, *to appear in Proc. of JSM, 2011, Miami, FL.*

23. Park, S. Y. and Liu, Y. (2011). Robust penalized logistic regression with truncated loss. *The Canadian Journal of Statistics*, accepted.

24. Pizer, S., Jung, S., Goswami, D., Zhao, X., Chaudhuri, R., Damon, J., Huckemann, S. and Marron, S.J. (2011). Nested sphere statistics of skeletal models, *in Proc. Dagstuhl Workshop on Innovations for Shape Analysis.*

25. *Staicu, A. M., Crainiceanu, C., Reich, D. and Ruppert, D. (2011). Modeling functional data with spatially heterogeneous shape characteristics, *Biometrics*, to appear.

26. Wilkinson, D. J. (2011) Stochastic Modeling for Systems Biology, second edition, Boca Raton, Florida: Chapman and Hall/CRC Press, in press.

27. Wilkinson, D. J. (2011) Stochastic dynamical systems, in Handbook of Statistical Systems Biology, M.P.H. Stumpf, M. Girolami, D.J. Balding (eds), Wiley, in press.

28. Wu, R., Cao, J., Huang, Z., Wang, Z., Gai, J. and Vallejos, E. (2011). Systems mapping: how to improve the genetic mapping of complex traits through design principles of biological systems, *BMC Systems Biology* 5:84, 1-24.

29. Wu, S., Müller, H.G. (2011). Response-adaptive regression for longitudinal data. *Biometrics*, accepted

30. Wu, Y. (2011). An ordinary differential equation-based solution path algorithm . *Journal of Nonparametric Statistics*, **23**, 185-199.

31. Wu, Y. and Li, Lexin (2011). Asymptotic Properties of Sufficient Dimension Reduction with A Diverging Number of Predictors. *Statistica Sinica* **21**, 707-730.

32. Wu, Y. and Liu, Y. (2011). Non-crossing large-margin probability estimation and its application to robust SVM via preconditioning. *Statistical Methodology* **8**, 56-67.

33. Yang, W., Müller, H.G., Stadtmüller, U. (2011). Functional singular component analysis. *J. Royal Statistical Society B* **73**, 303–324.

34. Yao, F., Fu, Y., Lee, T.C.M. (2011). Functional mixture regression. *Biostatistics*, **12**, 341-353.

35. Sheng Yi, Hamid Krim, Larry K. Norris, A Invertible Dimension Reduction of Curves on a Manifold. Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on Digital Object Identifier: 10.1109/ICCVW.2011.6130412; pp. 1378 – 1385.

36. Sheng Yi, Hamid Krim, and Larry K. Norris, Human Activity as a Manifold Valued Random Process. *IEEE Transactions on Image Processing*, to appear

37. Zhang, H. H., Cheng, G., and Liu, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *Journal of the American Statistical Association*, accepted.

38. Zhang, Z., Müller, H.G. (2011). Functional density synchronization. *Computational Statistics and Data Analysis* **55**, 2234–2249.

39. Zhu, H., Brown, P. J., Morris, J. S. (2011). Robust, adaptive regression in functional mixed models framework. *Journal of the American Statistical Association*, accepted.

Submitted/In Revision

1. Ahn, M., Zhang, H. H., and Lu, W. (2011) Moment-based method for random effect for selection in linear mixed models. *Statistica Sinica*, revised.

2. Aston JAD and Kirch C. Detecting and estimating epidemic changes in dependent functional data, *submitted.*

3. Bosca, L., Boente, G., Tyler, D. and Wang, J.-L. (2011) . Robust functional principal components. *Annals of Statistics*, revised.

4. Cardot, H., Degras, D., and Josserand, E. (2011). Confidence bands for Horvitz-Thompson estimators using sampled functional data. Submitted.

5. Chen, D., Müller, H.G. (2011). Nonlinear manifold representations for functional data. Submitted.

6. Crainiceanu, C. M., Caffo, B. S., and Morris, J. S. (2011). Multilevel Functional Data Analysis. Submitted to *SAGE Handbook for Multilevel Modeling.*

7. Dass, S. C., Lim, C. Y and Maiti, T. (2011). Detecting Change Points of Cancer Incidence Rates using Functional Dirichlet Processes. *Submitted.*

8. Ding, A.A. and Wu, H. (2011). Estimation of ODE parameters using constrained local polynomial regression. Submitted.

9. Leif Ellingson, Frits Ruymgaart, Vic Patrangenaru (2011), Nonparametric estimation for extrinsic mean shapes of planar contours,*revision submitted to Ann. Statist.*

10. Ellingson, L., Patrangenaru, V. and Ruymgaart F. (2011) Automatic Landmark Extraction for Planar Contours, *Preprint submitted for publication.*

11. Fang, Y., Wu, H., Zhu, L. (2011). A Note on Data Augmentation-Based Pseudo-Least Squares Estimation for ODE Models. Submitted.

12. Golightly, A., Wilkinson, D. J. (2011) Bayesian Parameter Inference for Stochastic Biochemical Network Models using Particle MCMC, in submission to Interface Focus.

13. Gorrostieta C and Ombao H.. General Spectral Measures of Cross-Dependence in Multivariate Time Series, *JASA, under revision.*

14. Gorrostieta C, Ombao H, Bedard P and Sanes J.N. Investigating Stimulus-Induced Changes in Connectivity Using Mixed Effects Vector Autoregressive Models, *NeuroImage, under revision.*

15. Gorrostieta C, Ombao H, Rrado R, Patel S and Eskandar E. Coherence Analysis of Local Field Potentials, *Journal of Time Series Analysis, under review.*

16. Hooker, G. and S. P. Ellner, (2011) On Forwards Prediction Error, Technical Report BU-1679-M, Department of Biological Statistics and Computational Biology, Cornell University.

17. Hooker, G., J. O. Ramsay and L. Xiao, (2011) CollocInfer: An R Library for Collocation Inference for Continuous and DiscreteTime Dynamic Systems, R library and manual.

18. Kang H, Ombao H, Linkletter C, Long N and Badre D. Spatio-Spectral Mixed Effects Model for Functional Magnetic Resonance Imaging Data, *JASA, under revision.*

19. Li, J.J., Jiang, C.R., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Sparse linear modeling of RNA-seq data for isoform discovery and abundance estimation. In revision for *PNAS.*

20. Luo, W., J. Cao, M. Gallagher and J. Wiles (2011) Estimating the Intensity of Ward Admission and its Effect on Emergency Department Access Block, Submitted

21. Martinez, J. G., Bohn, K. M., Carroll, R. J. and Morris, J. S. (2011). A study of Mexican free-tailed bat syllables: Multi-domain modeling of nonstationary time series with high frequency content using Bayesian functional mixed models. *Under revision.*

22. *McLean, M. W., Hooker, G., Staicu, A. M., and Ruppert, D. (2011). Functional Generalized Additive Models. *Submitted.*

23. Montagna, S., Tokdar, S.T., Neelon, B. and Dunson, D.B. (2011). Bayesian latent factor regression for functional and longitudinal data. *Submitted.*

24. Morris, J. S. (2011). Statistical Methods for Proteomic Biomarker Discovery using Feature Extraction or Functional Data Analysis Approaches. Under Revision for *Statistics and its Interface.*

25. Müller, H.G., Wu, Y. Yao, F. (2011). Continuously additive models for functional regression. Submitted.

26. Nam CFH, Aston JAD and Johansen AM, *Quantifying uncertainty in change points, in revision.*

27. Ogden, R. T., Zhao, Y., and Reiss, P. T. Wavelet-based functional principal component regression, *In preparation.*

28. Ratmann, O., P. Pudlo, S. Richardson and C. Robert, Monte Carlo Algorithms for Model Assessment via Conflicting Summaries, submitted.

29. Şentürk, D., Ghosh, S. and Nguyen, D. V. (2011). Exploratory time varying lagged regression for longitudinal data. *Biometrics*. Submitted.

30. Shen H, Tian S, and Huang J. A Two-Way Regularization Method for MEG Source Reconstruction, *under review at Annals of Applied Statistics.*

31. Shen H, Truong Y, and Lee S. Asymptotics of Colored Independent Component Analysis, *under review at Annals of Statistics.*

32. *Staicu, A. M., Lahiri, S., and Carroll, R. (2011). Tests of significance for spatially correlated multilevel functional data. Submitted.

33. J. Su, I. L. Dryden, E. Klassen, H. Le and A. Srivastava (2011), Fitting Smoothing Splines to Time-Indexed, Noisy Points on Nonlinear Manifolds. *Submitted to Journal of Image and Vision Computing.*

34. Charles C. Taylor, Kanti V. Mardia, Marco Di Marzio, Agnese Panzera, Validating protein structure using kernel density estimates. *Submitted to Statistical Applications in Genetics and Molecular Biology.*

35. Tchumtchoua, S. (2011). "Bayesian Semiparametric Functional Generalized Linear Models for Longitudinal Data". *Submitted.*

36. Wang, L. and Cao, J. (2011). Estimating Delay Differential Equations. Under revision for *Journal of Agricultural, Biological, and Environmental Statistics.*

37. Wei, F. and Zhu, H. (2011). Group Coordinate Descent Algorithms for Nonconvex Penalized Regression. *Computational Statistics & Data Analysis*. Submitted.

38. Wu, Y. and Liu, Y. (2011). Functional robust support vector machines for sparse and irregular longitudinal data. Revised.

39. Wu, H., Miao, H., Xue, H., Topham, D.J., and Zand, M. (2011). Quantifying immune response to influenza virus infection via multivariate nonlinear ODE models with partially observed state variables and time-varying parameters. *Journal of American Statistical Association*, submitted.

40. Wu, H., Xue, H., Kumar A. (2011), Numerical algorithm-based estimation methods for ODE models via penalized spline smoothing, *Biometrics*, revised.

41. Zhang, L., Baladandayuthapani,V., Mallick, B. K., Thompson, P., Bondy, M. and Do, K.A. (2011). Bayesian hierarchical structured variable selection methods with application to MIP studies in breast cancer. *Under revision.*

42. Zhao, Y., Ogden, R. T., and Reiss, P.T. Wavelet-based LASSO in functional linear regression, *Submitted manuscript.*

43. Zhu, H., Brown, P. J., Morris, J. S. (2011). Robust Classification of Functional and Quantitative Image Data using Functional Mixed Models. *Biometrics*, revised.

In Preparation/Preprint

1. Armagan, A., D. Dunson and J. Lee, Posterior consistency of Bayesian regression model for high-dimensional data.

2. Asencio, M., Hooker, G. and Gao, H. (2011). Functional Convolution Models. *Preprint.*

3. Aston JAD and Kirch C. Estimation of the distribution of change-points with application to fMRI data, *in preparation.*

4. Avery, M, Zhang, H. H. , Wu, Y. (2011) . Sparse estimation in functional data analysis. *In preparation.*

5. Baladandayuthapani, V., Morris, J.S., Coombes, K. R. and Abruzzo, L. (2011). Bayesian Adaptive Functional Linear Models for Copy Number Data. *In preparation.*

6. **Baladandayuthapani, V., Bharath, K., Baggerly, K., Czerniak, B., and Morris, J. S. (2011). Bayesian Spatial-functional Models for High-throughput genomic data. *In preparation.*

7. Burdis J. M., Kogan, I.A., Object-image correspondence for curves under finite and affine cameras.
   Technical Report. http://arxiv.org/abs/1004.0393

8. Cai, J. and Cao, J. (2011). Estimating Damage Accumulation Dynamic Models Hierarchical Bayesian Methods for Parameter Estimation in Dynamic Duration of Load Models. *In preparation.*

9. Chen, Y. and Samworth, R. J. (2011). Smoothed log-concave maximum likelihood estimation with applications. *Preprint.*

10. Chiou, J.M., Ma, Y. and Tsai, C.L. (2011). Functional time-varying random effects models for longitudinal data. *Preprint.*

11. Chiou, J.M. and Müller, H.G. (2011). Linear manifold modeling of multivariate functional data, with application to traffic flow analysis. *Preprint.*

12. Chkrebtii, O. and Cao, J. (2011). Generalized Additive Modeling of Spatio-Temporal Trends in Salmon Productivity. *In preparation.*

13. Dass, S., J. Lee, and K. Lee, Fast Computation for Regression Models with Ordinary Differential Equations.

14. Degras, D. (2011). Longitudinal survey sampling for functional data. *In preparation.*

15. Ding, J.-M. and Wang, J.-L. (2011). Varying coefficient Cox model with functional and longitudinal covariates. *In preparation.*

16. Dryden, I.L. and Olhede, S. (2011), A kernel density estimator for the orientation distribution function in diffusion weighted imaging, using Bingham Distributions. *Technical report in preparation.*

17. Du, P., Cheng, G., and Zhang. H. H. (2011) . Structure selection for nonparametric survival analysis. *In preparation.*

18. Leif Ellingson, David Groisser, Daniel Osborne, Vic Patrangenaru, Armin Schwartzman (2011), Data analysis on spaces of positive definite matrices with an application to dyslexia detection, *preprint.*

19. Leif Ellingson, Vic Patrangenaru, Sean Skwerer (2011), Computational methods for extrinsic mean on phylogenetic tree spaces, *work in progress.*

20. Leif Ellingson, Vic Patrangenaru (2011), Extrinsic PCA on manifolds, *work in progress.*

21. Harrie Hendriks (2011), Two sample problem for mean location, *http://nonparam11.karlin.mff.cuni. cz/bookabs20110719.pdf*

22. Hooker, G.. (2011) Domain selection for functional linear models. *In preparation.*

23. Hooker, G., Rogers, B., Lin, K. and Ng, T., Control Theory and Optimal Adaptive Experiments in Nonlinear Stochastic Models

24. Thomas Hotz, Stephan Huckemann (2011), Intrinsic means on the circle: uniqueness, locus, and asymptotics. *http://arxiv.org/abs/1108.2141 [stat.ME] [math.PR]*

25. Thomas Hotz, Stephan Huckemann, Huiling Le, Vic Patrangenaru (2011), Hyperbolic data analysis, *work in progress.*

26. Thomas Hotz, Stephan Huckemann, Huiling Le, Steve Marron, Jonathan Mattingly, Ezra Miller, James Nolen, Megan Owen, Victor Patrangenaru, Sean Skwerer (2011), Sticky central limit theorems on open books, *preprint.*

27. Stephan Huckemann, Steve Marron, Ezra Miller, Yolanda Munoz, Megan Owen, Victor Patrangenaru, Sean Skwerer (2011), Multidimensional scaling (MDS) with curved targets, *work in progress.*

28. Stephan Huckemann, Jonathan Mattingly, Ezra Miller, James Nolen (2011), Central limit theorems in codimension 1 on nonpositively curved stratified spaces, *work in progress.*

29. Huckemann, S., Jung S., Marron, S.(2011) Principal Nested Shape Spaces and an Application to the Reduction of the Number of Landmarks.

30. Jiang, C.R., Aston, J.A. and Wang, J.L. (2011). Nonparametric response function estimation via FPCA with an application to dynamic PET data. *In preparation.*

31. Jiang, C.R. and Morris, J.S. (2011). Choosing number of principal components in high dimension, low sample size situations for complex functional data. *In preparation.*

32. Jiang C, Wang JL and Aston JAD. Functional PCA based Deconvolution for Position Emission Tomography, *in preparation.*

33. Sungkyu Jung, Stephan Huckemann, J. S. Marron, Thomas Hotz (2011), Principal nested shape spaces and an application to reduction of number of landmarks, *work in progress, http://www.samsi.info/workshop/aood-transition-workshop-june-9-11-2011.*

34. John Kent, Thomas Hotz, Stephan Huckemann, and Ezra Miller (2011), The geometry and topology of projective shape spaces, *preprint.*

35. John Kent, Kanti Mardia (2011), The geometric approach to projective shape and the cross ratio,*paper in preparation.*

36. Kong, D. , Yao, F., Zhang, H. H. (2011). Semiparametric functional linear model. *In preparation.*

37. Lee, W. and Liu, Y. (2011). Simultaneous multiple response regression and inverse covariance matrix estimation via penalized Gaussian maximum likelihood. *Preprint* .

38. Li, L., Yao, F., Craiu, R. V. (2011). Minimum description length principle for correlated data. *Preprint.*

39. Li, L., Zhou, H., and Zhu, H. (2011). Tensor regression with applications in neuro-imaging data analysis. *In preparation.*

40. Li, N., and Zhang, H. H. (2011) Sparse learning in multi-class classification. *Preprint.*

41. Lin, C. Y., Zhang, H. H., Bondell, H., and Zou, H. (2011) Nonparametric variable selection in quantile regression. *Preprint.*

42. Liu, C., Ray, S., Hooker, G. and Friedl, M. (2011) . Functional factor analysis for periodic remote sensing data. *Preprint.*

43. K. V. Mardia, J. T. Kent, Z Zheng, C. C. Taylor and T. Hamelryck, Mixtures of concentrated sine distributions with applications to bioinformatics. *Technical report in preparation.*

44. Steve Marron, Ezra Miller, Megan Owen, Scott Provan, Sean Skwerer (2011), Towards PCA on tree spaces, *work in progress.*

45. Steve Marron, Ipek Oguz, Sean Skwerer (2011), Smoothing in phylogenetic tree space using simple iterated pairwise geodesics, *work in progress.*

46. McLean, M., Hooker, G., Saicu, A. and Ruppert, D. (2011) Functional generalized additive models. *Preprint.*

47. Ezra Miller, David Houle, Paul Bendich (2011), Quantifying shape differences in fruit fly wing morphology using persistent homology, *in progress.*

48. Morris, J. S. and Allen, G. (2011). Analysis of fMRI data using functional mixed models and sparse empirically determined basis function representations. *In preparation.*

49. Megan Owen, Sean Skwerer (2011), Fréchet means in tree space, *work in progress.*

50. Pati, D. and Dunson, D.B. (2011). Bayesian closed surface fitting through tensor products. *In preparation.*

51. Ramsay, J. O. (2011) A Functional Estimate of a Functional Variance-Covariance Matrix and its Inverse. *Paper in preparation.*

52. Samworth, R. J. (2011). Optimal weighted nearest neighbor classifiers. *Preprint.*

53. Sangalli, L., Ramsay, J.O. and Ramsay, T.O. (2011) Spatial Spline Regression Models. *Paper in preparation.*

54. Şentürk, D. (2011). Efficient estimation for generalized varying coefficient models with longitudinal data. *In preparation.*

55. Serban, N., Staicu, A. M., and Carroll, R. J. (2011). Multilevel Spatially Correlated Binary Longitudinal Data. *In preparation.*

56. Shah, R. and Samworth, R. J. (2011), Variable selection with error control: Another look at stability selection. *Preprint.*

57. Shen H, Zhang L and Huang J. Two-Way Robust Functional Data Analysis, *to be submitted shortly.*

58. Shen H, Truong Y and Chen W. Hemodynamic Response Function Estimation, *in preparation.*

59. Shi, XY, Zhu, HT, Ibrahim JG, Styner M, Intrinsic regression models for median representation of subcortical structures.

60. A. Srivastava, W. Wu, S. Kurtek, E. Klassen, and J. S. Marron, Registration of Functional Data Using Fisher-Rao Metric. *Technical report.*

61. Staicu, A. M., Senturk, D., and Carroll, R. J. (2011). Generalized Time-Varying Spatial Regression of Multilevel Functional Data. *In preparation.*

62. J. Su, I. L. Dryden, E. Klassen, H. Le and A. Srivastava, A new metric for Symmetric Positive Definite Matrices. *Technical report in preparation.*

63. **Tchumtchoua, S., Dunson, D. B., and Morris, J. (2011). Online Variational Bayes Inference for High-Dimensional Correlated Data.*In preparation.*

64. Tchumtchoua, S., Dunson, D. B., and Morris, J. (2011). A Heterogeneous Dynamic Structural Equation Model With Application to Brain Connectivity. *In preparation.*

65. Thorbergson, L. and Hooker, G., Optimal Adaptive Experimental Design in Hidden Markov Models

66. Van Lunen D, Ombao H and Aston JAD. Online Detection Methods: A Model Selection Framework. *In preparation.*

67. Wang, L., Wu, Y., and Li, R. (2011). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Preprint.*

68. Wang, Xia, Chen, M.-H., Dey, D. K., and Kuo C.-Y. (2011). Space-Time Modeling of Atlantic Cod Abundance in the Gulf of Maine. *In preparation.*

69. Wang, Xia and Sedransk, N. (2011). Bayesian Models on Biomarker Discovery Using Spectral Count Data in the Label-Free Shotgun Proteomics. *In preparation.*

70. Wang, Xia, Sedransk, N. and Tabb, D. (2011). Variability of Base Peak Intensities in Shotgun Proteomics Experiment: Perspectives from Functional Data Analysis. In preparation.

71. Wang, Xiaojing and Berger J. (2011). Estimated shape constrained functions using Gaussian Processes. *In preparation.*

72. Wei, F. and Zhu, H. (2011). Group coordinate descent algorithms for nonconvex penalized regression. *In preparation.*

73. Wong, R.K.W., Yao, F., Lee, T.C.M. (2011). Robust estimation for generalized additive models. *Preprint.*

74. Wu, Y. (2011). Elastic net for Cox's proportional hazards model with a solution path algorithm. *Preprint* .

75. Wu, Y. and S. Ghosal, Convergence Rates of Multivariate Density Estimation by Dirichlet Mixture Priors

76. Wu, Y. and G. Hooker, Generalized Profiling, Stochastic Differential Equations and Higher-Order Stochastic Runge Kutta Schemes

77. Yuan, Y., Zhu, H.T., Styner, M., J. H. Gilmore., and Marron, J. S. Varying coefficient model for modeling diffusion tensors along white matter bundles.

78. Yuan, Y., Zhu, H.T., Lin, W. L., and Marron, J. S. Local polynomial regression for symmetric positive definitive matrices.

79. Zhao, Y., Bagiella, E., and Ogden, R. T. A functional approach to analysis of RR interval variability, *In preparation.*

80. Zhou, H. and Wu, Y. (2011) A general path algorithm for regularized statistical estimation. *Preprint.*

81. Zhu, H., Dunson, D. B. (2011). Bayesian graphical models for multivariate functional data. *Preprint.*

82. Zhu, H., Yao, F., Zhang, H. H. (2011). Component Selection in Functional Additive Models. *Preprint.*

83. Zhu, H. and Dunson, D. (2011). Bayesian graphical models for multivariate functional data. *In preparation.*

84. Zhu, H. and Morris, J. S. (2011). Functional Mixed Models for Serially Correlated Functional Data. *In preparation.*

Complex Networks

At the time of preparing this report, we have only a very incomplete list of publications and reports from the Complex Networks program. A more complete list will be provided when it is available.

Published/Accepted

1. McCormick, T.H., He, R., Kolaczyk, E.D., and Zheng, T. (2012). Surveying hard-to-reach groups through sampled respondents in a social network: a comparison of two survey strategies. *Statistics in Biosciences* Special Issue on Networks, 4:1, 177-195

2. Brummitt, C.D., D'Souza, R.M., Leicht, E.A. Suppressing cascades of load in interdependent networks. *Proc. Nat'l. Acad. Sci.* 109 (2012), E680-E689.

3. Durrett, R., Gleeson, J., Lloyd, A., Mucha, P., Shi, F., Sivakoff, D., Varghese, C. Graph fission in an evolving voter model. *Proc. Nat'l. Acad. Sci.* 109 (2012), 3682-3687.

Submitted/In Revision

1. Chertock, A., Kurganov, A., Polizzi, A. Multi-Class Traffic Flow Model with Look-Ahead Dynamics. Submitted.

2. Chertock, A., Kurganov, A., Polizzi, A., Timofeyev, I. Pedestrian Flow Models with Slowdown Interactions. Submitted.

3. Krivitsky, P.N. and Kolaczyk, E.D. (2012). On the question of effective sample size in network modeling. Submitted.

4. Rombach, M.P., Porter, M.A., Fowler, J.H., Mucha, P.J. Core-periphery structure in networks. Submitted.

In Preparation/Preprint

1. Durrett, R. and Sivakoff, D. Contract process on modular networks. In progress.

2. Sun, Y. and Timofeyev, I. New stochastic models for traffic flows. In preparation.

3. Verella, J.T. A randomized approximation algorithm for community detection in complex networks. 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.

Participation in the NSF Institutes' Diversity Committee

Pierre Gremaud 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 David Auckly (MSRI). The Institutes Diversity Coordination Committee has been working together to promote diversity in the Mathematical Sciences at national conferences and through other special events.

Together with the other NSF mathematical sciences institutes, SAMSI has recently secured supplemental NSF funding that will allow for the long term planning of an entire portfolio of activities including the Modern Math workshop series, the Workshop Celebrating Diversity series, the Minority Professional Development workshop, the Careers for Minorities workshops, the Careers for Women workshops, the Blackwell Tapia Conferences and some activities in conjunction with the Association for Women in Mathematics.

SAMSI took part in the *Modern Math* program at the 2009 SACNAS National Convention in Dallas, TX. 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. Oliver Diaz-Espinosa, one of our SAMSI postdoctoral fellows and a member of the SAMSI program on Stochastic Dynamics presented an overview of his research. SAMSI was again a participant in the Modern Math program in Oct. of 2010, which will be reported in the next Annual Report of SAMSI.

Minority Participation in SAMSI Programs

For the 2010-11 Research program, of seven post-docs hired, two are women: Sylvie Tchumtchoua and Hongxiao Zhu and two are minorities David Degras and Sylvie Tchumtchoua. Of the six hired for 2011-12, three are women (Jenny Brynjarsdottir Chia Ying Lee and Ying Sun).

Education and Outreach Programs: SAMSI continues to create opportunities, through its E&O Program, to enhance its diversity efforts by recruitment of under-represented participants. We recruit from HBCU's for all programs. We collaborate with our National Advisory and Education and Outreach Committees to reach out to the Hispanic and Native American communities. Further, new efforts have been made to engage both faculty and students at primarily teaching colleges. The diversity breakdown in specific E&O workshops is as follows.

- Industrial Mathematical and Statistical Modeling (IMSM) Workshop (July 2010): From the 44 participants, 14 were female, 3 were African-American and 1 was Hispanic.
- 2-Day Undergraduate Workshop (Oct 2010): From the 47 participants, 23 were female.
- 2-Day Undergraduate Workshop (Feb. 2011): From the 36 participants, 15 were female, 3 were African American and 2 were Hispanic.
- Undergraduate Workshop (May 2011): From the 56 participants, 26 were female, 3 were African American and 2 were Hispanic.

		# Participants	# Female	# African-American	# Hispanic	# New Resrcher-Students
2010-11	Opening Workshop -- *August 29 - September 1,*	179	50	2	11	106
2010-11	Modeling Workshop -- *October 20-22, 2010*	54	16	2	1	33
2010-11	Dynamics of Networks -- *January 10-12, 2011*	37	7	0	1	20
2010-11	Pedestrian Traffic Flow -- *February 14-16, 2011*	15	1	0	0	7
2010-11	Dynamics on Networks -- *March 21-23, 2011*	45	11	0	2	23
2010-11	Transition Workshop -- *June 6-7, 2011*	36	6	1	1	28
2010-11	Opening Workshop -- *September 12-15, 2010*	177	50	5	3	90
2010-11	Interface Functional and Longitudinal Data --	75	25	3	3	44
2010-11	AOOD Meets Evolutionary Biology -- *April 30-May 2, 2011*	58	10	3	1	28
2010-11	Transition Workshop -- *June 9-11, 2011*	41	9	1	2	27

		# Participants	# Female	# African-American	# Hispanic	# New Resrcher-Students
2010-11	Two-Day Undergraduate Workshop -- *October 29-30,*	47	23	0	0	30
2010-11	Two-Day Undergraduate Workshop -- *February 25-26,*	36	15	3	2	27
2010-11	SAMSI/CRSC Interdisciplinary Workshop for Undergraduates -	56	26	3	2	32
2010-11	Industrial Stat/Math Modeling Workshop for	44	14	3	1	32
2010-11	Uncertainty Quantification Summer School - - *June 20-24,*	80	20	0	6	37
2010-11	Big Data Meeting -- *May 20, 2011*	25	7	0	2	4
		1005	**29%**	**3%**	**4%**	**57%**

Affiliates

1. Affiliate Involvement

1.1. 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. The entire directorate interacts directly with affiliates; NISS director Alan Karr and associate director Nell Sedransk have responsibility for operation of these programs.

Recent new affiliates include the Department of Statistical Science at Cornell University, the Department of Mathematics and Statistics at the University of Maryland Baltimore County, the Departments of Biostatistics and Statistics at the University of Pittsburgh and the Department of Statistics at Virginia Tech. A complete listing of affiliates as of May 2012 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 (as well as to industry and government), 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. All 2009-10 and 2010-11 contain components and activities suggested by affiliates. The 2012-13 program on Data-Driven Decisions in Healthcare and 2013-14 program on Computational Methods in the Social Sciences are strongly affiliate-influenced.

2.2 NISS Affiliates and NISS/SAMSI Affiliates

Affiliates marked with (*) joined the program directly as the result on participation in SAMSI events.

Corporations: AT&T Labs Research, Avaya Labs, GlaxoSmithKline, Eli Lilly, Merck Research Laboratories, MetaMetrics, Inc., PNYLAB (*), RTI International, SAS Institute, SPSS, Telcordia Technologies, and Yahoo! Labs

Government Agencies and National Laboratories: Bureau of Labor Statistics, Census Bureau, Energy Information Administration, National Agricultural Statistics Service, National Center for Education Statistics, National Center for Health Statistics/CDC, National

Geospatial-Intelligence Agency (*), National Security Agency, and Office of the Comptroller of the Currency (*)

Universities: University of California Berkeley, Department of Statistics; Carnegie Mellon University, Department of Statistics; Columbia University, Department of Biostatistics; University of Connecticut, Department of Statistics; Cornell University, Department of Statistical Science; Duke University, Department of Mathematics (*); Duke University, Department of Statistical Science; Emory University, Department of Biostatistics, 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; University of Maryland Baltimore County, Department of 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; University of Pittsburgh, Departments of Biostatistics and Statistics; Purdue University, Department of Statistics; Rice University, Department of Statistics; University of South Carolina, Department of Statistics; Stanford University, Department of Statistics; Texas A&M University, Department of Statistics; Virginia Commonwealth University, Department of Biostatistics; Virginia Tech, Department of Statistics (*)

2.3 Affiliate Participation

All SAMSI programs and events during 2010-11 had strong affiliate participation, nearing one-half of attendees at some workshops. Expenditures from Affiliates Reimbursement Accounts to attend SAMSI events exceeded $60,000.

2.4 Plans for the Future

SAMSI and NISS are conducting a full-scale review of the affiliates program in 2011-12, in order to ensure that it continues to deliver value to the affiliates. Multiple initiatives are expected to result.

Current Advisory Committee Members

Committee	Name	Affiliation	Field	Term
Directorate	Richard Smith	UNC (Director)	Statistics	Since 2010
	Snehalata Huzurbazar	Uof Wyoming (Deputy Director)	Statistics	2012-2014
	Ezra Miller	Duke (Associate Director)	Mathematics	2011-2014
	Ilse Ipsen	NCSU (Associate Director)	Mathematics	2011-2014
	Alan Karr	NISS (Associate Director)	Statistics	Since 2011
Governing Board	Robert Calderbank	Duke, Dean of Natural Sciences	Comp Sci/Math	Since 2011
	Michael Crimmins	UNC, Sen Assoc Dean Science	Chemistry	Since 2009
	Susan Ellenberg	NISS, Chair of Board	Biostatistics	Since 2011
	Don Estep	Colorado State (SIAM Rep)	Math and Stat	Since 2008
	Daniel Solomon (Chair)	NCSU, Dean	Statistics	Since 2002
	Vacancy	ASA Rep		
National Advisory Committee	Susanne Brenner	Lousiana State U	Mathematics	2010-2013
	Ricardo Cortez	Tulane U	Mathematics	2008-2014
	Jianqing Fan	Princeton U	Ops Research	2008-2014
	Mac Hyman (co-chair)	Tulane U, Los Alamos Nat Lab	Math	2010-2013
	Irene Gamba	U of Texas - Austin	Mathematics	2010-2013
	Max Gunzburger	Florida State U	Scient. Computing	2010-2013
	Diane Lambert	Google	Statistics	2010-2013
	Jun Liu	Harvard U	Statistics	2007-2013
	Susan Murphy (co-chair)	U Michigan	Stat/Psychiatry	2008-2013
	Habib Najm	Sandia Nat Lab	Systems Res.	2011-2014
	Adrian Raferty	U of Washington	Statistics/Sociology	2011-2014
	Kathryn Roeder	Carnegie Mellon U	Statistics	2010-2013
Local Development Committee	Howard Bondell	NCSU	Statistics	2011-2014
	Gregory Forest	UNC	Mathematics	2010-2013
	Stephen George	Duke	Biostat , Bioinfor.	2010-2013
	Robert Wolpert	Duke	Statistics	2010-2013
	Donglin Zeng	UNC	Biostatistics	2010-2013
Chairs Committee	Edward Carlstein	UNC	Statistics	Since 2009
	Elizabeth DeLong	Duke	Bioinfo & Stat	Since 2009
	Montse Fuentes	NCSU	Statistics	Since 2011
	Alan Gelfand	Duke	Statistics	Since 2008
	Loek Helmnick	NCSU	Mathematics	Since 2005
	Michael Kosorok	UNC	Biostatistics	Since 2006
	Harold Layton	Duke	Mathematics	Since 2009
	Peter Mucha	UNC	Mathematics	Since 2010
Education and Outreach Committee	Negash Begashaw	Benedict College	Mathematics	2004-2012
	Carlos Castillo-Chavez	Arizona State U	Mathematics	2004-2012
	Karen Chiswell	NCSU	Statistics	2004-2012
	Cammey Cole Manning	Meredith College	Math & CS	2004-2012
	Anne Fernando	Norfolk State U	Mathematics	2009-2012
	Pierre Gremaud (chair)	NCSU	Mathematics	2009-2012
	Leona Harris	College of NJ	Mathematics	2009-2012
	Gabriel Huerta	U of New Mexico	Statistics	2009-2012
	Marian Hukle	U of Kansas	Biological Sci	2004-2012
	Masilamani Sambandham	Morehouse College	Mathematics	2004-2010

Former Members of Advisory Committees

Committee	Name	Affiliation	Field	Term
Directorate	James Berger	Duke (Director)	Statistics	2002-2010
	J. Stephen Marron	UNC (Associate Director)	Statistics	2002-2004
	H. Thomas Banks	NCSU (Associate Director)	Mathematics	2002-2005
	Alan Karr	NISS (Associate Director)	Statistics	2002-2005
	Young Truong	UNC (Associate Director)	Biostatistics	2004-2005
	Christopher Jones	UNC (Associate Director)	Mathematics	2005-2007
	Ralph Smith	NCSU (Associate Director)	Mathematics	2005-2008
	Nell Sedransk	NISS (Associate Director)	Statistics	2006-2010
	Michael Minion	UNC (Associate Director)	Mathematics	2007-2010
	Pierre Gremaud	NCSU (Associate Director)	Mathematics	2008-2009
	Pierre Gremaud	NCSU (Deputy Director)	Mathematics	2010-2012
	Richard Durrett	Duke (Associate Director)	Mathematics	2010-2011
Governing Board	Bruce Carney	UNC, Assoc. Dean	Astronomy	2004-2006
	George Casella	U of Florida (ASA Rep)	Statistics	2005-2012
	John Harer	Duke	Mathematics	2002-2004
	Douglas Kelly	UNC	Statistics	2002-2004
	Jon Kettenring	Telecordia	Mathematics	2002-2004
	James Landwehr	Avaya Labs & NISS	Statistics	2009-2010
	Tom Manteuffel	U of Colorado (SIAM Rep)	Applied Math	2006-2008
	Vijay Nair	NISS Trustees Chair	Statistics	2004-2006
	John Simon	Duke, Asst. Provost	Chemistry	2004-2010
National Advisory Committee	Mary Ellen Bock	Purdue	Statistics	2002-2006
	Peter Bickel	UC Berkeley	Statistics	2002-2005
	Lawrence Brown	Pennsylvania State U	Statistics	2002-2006
	Raymond Carroll	Texas A&M	Statistics	2004-2006
	Carlos Castillo-Chavez	Arizona State U	Mathematics	2003-2011
	Rick Durrett	Cornell U	Mathematics	2006-2010
	David Heckerman	Microsoft	CS & Statistics	2002-2004
	Sallie Keller-McNulty	LANL	Statistics	2002-2005
	Nancy Kopell	Boston U	Mathematics	2005-2008
	John Lehoczky	Carnegie Mellon	Probability	2002-2005
	Rod Little	U of Michigan	Biostatistics	2006-2008
	Jun Liu	Harvard U	Statistics	2007-2010
	David Mumford	Brown U	Applied Math	2006-2008
	George Papanicolaou	Stanford U	Mathematics	2002-2003
	Daryl Pregibon	Google, Inc	CS and Statistics	2003-2010
	Adrian Raftery	U Washington	Statistics	2002-2003
	G.W. Stewart	U of Maryland	Computer Science	2003-2008
	Phillippe Tondeur	U Illinois	Mathematics	2003-2005
	Mary Wheeler	U Texas	Math & Eng	2004-2006
	Shmuel Winograd	IBM	Mathematics	2002-2003
	Margaret Wright	NYU	CS	2002-2004
	Bin Yu (Chair)	U of CA, Berkeley	Statistics	2006-2011
Local Development Committee	David Banks	Duke	Statistics	2003-2010
	H.T. Banks	NCSU	Mathematics	2005-2010
	Andrea Bertozzi	Duke	Mathematics	2002-2003
	Lloyd Edwards	UNC	Biostatistics	2003-2010
	Gregory Forest	UNC	Mathematics	2002-2010
	Jean-Pierre Fouque	NCSU	Mathematics	2002-2004

	Montserrat Fuentes	NCSU	Statistics	2004-2010
	Alan Gelfand	Duke	Statistics	2002-2003
	Mark Genton	NCSU	Statistics	2002-2003
	John Harer	Duke	Mathematics	2004-2010
	Jacqueline Hughes-Oliver	NCSU	Statistics	2002-2004
	Joseph Ibrahim	UNC	Biostatistics	2002-2003
	Christopher Jones	UNC	Mathematics	2002-2003
	Thomas Kepler	Duke	Biostatistics	2002-2004
	Sharon Lubkin	NCSU	Mathematics	2004-2010
	Sally Morton	RTI	Statistics	2005-2010
	Andrew Nobel	UNC	Statistics	2002-2004
	David Schaeffer	Duke	Mathematics	2002-2003
	Scott Schmidler	Duke	Mathematics	2002-2003
	Ralph Smith	NCSU	Mathematics	2002-2003
	Richard Smith	UNC	Statistics	2004-2010
	John Trangenstein	Duke	Mathematics	2003-2004
	Butch Tsiatis	NCSU	Statistics	2004-2010
	Mike West	Duke	Bioinfo & Stat	2004-2010
Local Dept Liaisons	Helen Zhang	NCSU	Statistics	2010-2011
Chairs Committee	Jianwen Cai	UNC	Biostatistics	2005-2006
	Clarence E. Davis	UNC	Biostatistics	2003-2005
	Patrick Eberlein	UNC	Mathematics	2005-2010
	Jean-Pierre Fouque	NCSU	Mathematics	2004-2005
	Richard Hain	Duke	Mathematics	2004-2006
	Christopher Jones	UNC	Mathematics	2003-2004
	Thomas Kepler	Duke	Biostatistics	2005-2006
	Vidyadhar Kulkarni	UNC	Statistics	2003-2006
	Bernard Mair	NCSU	Mathematics	2003-2004
	David Morrison	Duke	Mathematics	2003-2004
	Sastry Pantula	NCSU	Statistics	2003-2010
	William Smith	UNC	Mathematics	2004-2005
	Dalene Stagl	Duke	Statistics	2003-2006
	Mark Stern	Duke	Mathematics	2006-2009
	William Wilkerson	Duke	Biostatistics	2003-2005
Education and Outreach Committee	H. Thomas Banks	NCSU	Mathematics	2004-2005
	Negash Begashaw	Benedict College	Mathematics	2004-2010
	Carlos Castillo-Chavez	Arizona State U	Mathematics	2004-2010
	Karen Chiswell	NCSU	Statistics	2004-2010
	Cammey Cole Manning	Meredith College	Math & CS	2004-2010
	Wei Feng	UNC-Wilmington	Math & Statistics	2004-2006
	Marian Hukle	U of Kansas	Biological Sci	2004-2010
	Negash Medhin	NCSU	Mathematics	2004-2006
	Masilamani Sambandham	Morehouse College	Mathematics	2004-2010

Minutes and Report
Mathematics Institutes Directors Meeting
AIM
May 6 -7, 2011

In attendance May 6, 2011:

Hélène Barcelo MSRI

Estelle Basor AIM

Robert Bryant MSRI

Russel Caflisch IPAM

Brian Conrey AIM

David Farmer AIM

Marty Golubitsky MBI

Leslie Hogben AIM

Sally Koutsoliotas AIM

David Levermore BMSA

Jill Pipher ICERM

Fadil Santosa IMA

Richard Smith SAMSI

Thomas Spencer IAS

1 Welcome and Introductions

2 Diversity Committee Report

David Levermore reported on the observations about diversity contained in the DMS's 2010 Committee of Visitors report.

Leslie Hogben gave a review of the Diversity Committee members, general goals, and recent activities, including involvement with SACNAS, the Alliance program, Blackwell Tapia, and the AWM mentor network.

a. **Modern Math Workshop.** The Modern Math Workshop is a pre-activity of SACNAS. It is now a cooperative effort of all institutes, showcasing the activities of the institutes and offering mini-courses and lectures related to the themes of the institutes.

b. **SACNAS.** Institutes representation at SACNAS includes a booth and reception, but perhaps, there should be more involvement of the institutes in the mathematical aspects.

c. **Alliance Postdoctoral Fellowships.** Another recent activity of the Diversity Committee is the Alliance Postdoctoral program, run by Loek Helminck. The goal is to provide mentoring of postdocs at a research university while also spending a year at an institute. A review of the Alliance postdocs and mentors chosen for this past year was given.

d. **Diversity Workshops.** Leslie reported on the joint effort led by Dave Auckly (MSRI) to secure the collective funding for several diversity workshops. This would alleviate the need to seek individual funding for each activity. Workshops include the Modern Math Workshop, the Spring Opportunities for women and minorities workshops, Blackwell-Tapia conference and the workshop at Infinite Possibilities, which rotate amongst the institutes. In Fall, IPAM will organize the Modern Math Workshop, etc.

e. **AWM Mentor Network.** Leslie described the AWM Mentor Network activities. The report from the 2010 AWM Mentor Network is on the AWM site.

Measures to broaden diversity of institute participants. A number of institutes strive for a minimum 20% participation of women at workshops. Various measures have been implemented, but it was noted that the acceptance rate for women is significantly lower than that for men.

Hélène and Fadil described the possible creation of a database that would be useful for all the institutes in identifying a broader diversity of institute participants. to broaden the diversity of institute participants. There is good cooperation with AWM and the institutes over issues like this, but for organizations such as NAM, communication is more problematic. Some institutes ask for lists of women and minorities at the time of application and allocate more resources to the programs that are diverse. Some institutes help with daycare costs (using private funding) to encourage greater participation.

There was a consensus that this issue needs to addressed on a case-by-case basis and that programs and sub-fields that have a larger proportion of women should support a greater percentage of women participants.

Pipeline concerns. There are indications that the percentages of women in mathematics is decreasing, both at the undergraduate and graduate level. MAA and BMSA are planning to propose that NSF fund a pipeline study. It was noted that this is a complicated matter with many and varied factors.

Diversity Committee recommendation. The Diversity Committee recommended that all institute contributions to diversity efforts be consolidated into a single annual contribution. Funds would support the reception tied to the SACNAS booth, as well as prizes, etc. In this way, all the possible contributions would be identified at the beginning of each year. The Diversity Committee would be charged with overseeing the activities.

MID recommendation. We request that the Diversity Committee prepare an annual budget to be presented at each MID meeting for discussion and review. The budget would identify appropriate diversity activities to be supported by the Institutes.

3 Technical Committee Report

David Farmer gave the Technical Committee Report. Last year there were no action items. David reported that the composition of the Technical Committee restricted its efficacy. In many cases, its members were not in a position to allocate resources or make executive decisions. It was recommended that the directors revisit the scope of the Technical Committee, along with its membership.

Video Collection of Institute Presentations. A discussion of institutes' video-taping of talks followed. The desire to collect both video and PDF versions of the talk was noted. There was discussion of a common mechanism to search all institutes' videos. Each institute would need to submit the information in an agreed upon form.

Highlights on the Joint Institutes' Webpage. The question of when highlights should be removed from the Joint Institute page was raised. Their over-abundance is not yet an issue, and IMA was asked to remind institutes to submit an annual highlight.

MID recommendation 1. The Technical Committee should be disbanded. Items that need to be addressed should be taken on a case-by-case basis.

MID recommendation 2. David Farmer will send a preliminary list of fields for video metadata to each institute as examples. Every institute will then prepare a list of properties (fields) related to its video presentations. The lists will be collected in a reasonable time and these will form the basis of the metadata needed for a database of institutes videos.

4 JMM Institutes' Reception

Format of the 2011 reception was not optimal: the physical setup made it difficult to hear presentations. Another concern was that the JMM schedule lists it only as a social event.

Brian Conrey provided a history of the event: social – display tables – short talks – talks on a theme. This led to a discussion of what math institutes may want to communicate to the mathematics community. Brian suggested that the individual institute tables remain, and have an informational slide show about the institutes as a whole as well as aspects of the individual institutes. Another suggestion was to separate the social event from the presentations by using different rooms at different times. It was also noted that many of the people who visited the institute tables came from small colleges.

Fadil suggested the institutes consider the idea of creating a joint short video about 'math in everyday life' for YouTube.

MID Recommendation for the Joint Reception. The JMM reception shall have no speakers; instead, an entertaining (but silent) slide show will be presented. The slides will highlight the institutes in the form of questions, trivia, pictures, programs, etc. Jill, Fadil, and Richard are co-chairs and will collect and assemble materials from all the institutes for the slide show. The slide show will be first presented at the JMM meeting, with the intention of its possible use at other future meetings, such as MathFest. This recommendation will be presented to representatives of the other institutes not present at MID.

5 MPE 2013

Brian Conrey gave a short history of MPE 2013 and the workshop at AIM in March 2011. A non-public website was created for planning purposes. Themed programs have been planned at some of the institutes: IPAM is planning a program in material science and applications, and other topics; MBI plans a program on bio-sustainability in 2013 with topics on green management, agent based modeling, etc.; SASMI is planning a program for 2012/2013 that will include a workshop on sustainability for spring 2013. SASMI is tentatively considering a 2013/2014 program on sustainability that will include contributions statistics, applied math, and other fields. The Newton Institute, the Canadian Institutes, and others are also planning events.

Brian showed the internal website, the link to the committees, and described the various activities, possible topics, resources, etc.

David Levermore suggested that programs for 2013 should be thought of as the beginning of MPE-type programs and that others would happen for many years to come.

There was additional discussion about possible joint activities, for example, a workshop at one institute that complements another long program. We need to publicize activities that are complementary and let participants know. CSCAMM should also be contacted to be included in the Fields' list so that the activities of CSCAMM are listed.

6 MID Meeting

Brian Conrey gave a short history of the annual MID meeting and the past activities, including the postdoc initiative and CDI projects.

Brian posed the question of what this group should do in the future. There is agreement that this meeting, especially with the size and informality, is very useful for learning about the other institutes, best practices, etc.

For the next meeting, it might be helpful to include discussion time after the meeting with the NSF.

7 Other Business

Institutes' Planning. There was useful discussion about sharing information about upcoming programs (not just MPE-related programs) and how far ahead the planning takes place. Several planned programs were described and discussed.

Inclusion of DDs. There was also discussion of including Deputy Directors during the annual MID meeting with possibly separate parallel session. One advantage is that it would help reinforce cooperation between programs and institutes.

Institute Postdocs. We reviewed the Institutes postdocs initiative. There seemed be a high percentage of postdocs that have positions for next year. Of those reported, 15 out of 20 have positions for next year.

8 Saturday, May 7, 2011

In attendance May 7, 2011:

Hlne Barcelo MSRI

Estelle Basor AIM

Robert Bryant MSRI

Russel Caflish IPAM

Haiyan Cai DMS

Brian Conrey AIM

David Farmer AIM

Marty Golubitsky MBI

Leslie Hogben AIM

Sally Koutsoliotas AIM

David Levermore BMSA

Sastry Pantula DMS

Jill Pipher ICERM

Rosemary Renaut DMS

Fadil Santosa IMA

Richard Smith SAMSI

Thomas Spencer IAS

Christopher Stark DMS

Henry Warchall DMS

Brian Zuckerman STPI

9 Introduction from Sastry Pantula

DMS DD Sastry Pantula gave a review of his vision of the role of institutes in mathematical research and emphasized that he appreciated the efforts and cooperation of the institutes. He also outlined budget considerations. A hard copy of his presentation was given to everyone in attendance.

10 Institute Evaluation

Brian Zuckerman discussed his draft report and answered questions. The recommendation is to evaluate the Long Program Convening Model by using case studies. It was noted that many institutes have short, week-long workshops that are not part of the semester-long programs. These meetings are similar in nature to the AIM workshops. A hard copy of his presentation was given to everyone in attendance.

11 Other Business

Brian Conrey highlighted some of the items from the MID meeting on Friday, including some of the diversity activities, the MID's recommendation for the JMM Joint Institutes' reception, the MPE 2013 initiative, the role of the MIDs, the sharing of future programs, the inclusion of Deputy Directors in the annual meeting, and the technical committee. There was also discussion of the proposed joint database.

Hank Warchall spoke about having organizers of workshops and programs create white papers about the future directions of fields. He emphasized the importance of the white papers. DMS is willing to provide additional funds to Institutes to support this activity. There was a request to see good examples of white papers.

He also reminded us about the needs of plans for self-evaluation and how this should be a back and forth activity between the NSF and the Institutes.

12 Date of the next MID meeting

Friday, May 11 - Saturday, May 12, 2012 at ICERM

II. Special Reports: Program Plan

A. Programs for 2010-2011

1. **Uncertainty Quantification**
2. **Nonlocal Continuum Models for Diffusion, Mechanics and Other Applications (Summer Program 2012)**
3. **Computational Advertising (Summer Program 2012)**

II.A.1: 2011-12 Program on Uncertainty Quantification

This is a preliminary report on the main program that took place at SAMSI during 2011-12. A detailed report will be given in next year's annual report.

Program Leaders: Amy Braverman (JPL, Cal. Tech and UCLA), Don Estep (Colorado State), Roger Ghanem (USC), David Higdon (LANL), Christine Shoemaker (Cornell), Jeff Wu (Georgia Tech)

Local Scientific Coordinators: James Nolen (Duke), Jan Hannig (UNC)

Directorate Liaison: Pierre Gremaud (N.C. State)

National Advisory Committee Liaison: Max Gunzburger (Florida State)

Computer simulations of mathematical models have become commonplace in many areas of science and technology. Quantifying the uncertainty in such models involves addressing many issues, ranging from physical limitations of the model to numerical and statistical issues. Addressing such challenges involves mathematics, statistics, probability, operations research, and computer science, as well as expertise in the disciplines for which computer models are developed. The SAMSI program on Uncertainty Quantification addressed a very wide variety of theoretical and practical problems in this field.

The structure of the program was unique among SAMSI programs in that, instead of running two year-long programs in parallel as is usually done, we ran a single UQ program with four sub-programs:

- Methodology Program
- Climate Modeling Program
- Engineering and Renewable Energy Program
- Geosciences Program

Major Activities

SAMSI/Sandia UQ Summer School

The first activity of the program was a Summer School for graduate students and young researchers held on June 20-24, 2011 in Albuquerque, NM. 80 participants attended.

Workshops

Four opening workshops were held, one for each of the sub-programs:

- Opening Workshop and Tutorials: Methodology of Uncertainty Quantification, September 7-10, 2011, held at the Radisson Hotel

- Climate Change Workshop, August 29-31, 2011, co-organized with the Lawrence Livermore National Lab in Pleasanton, CA
- Engineering and Renewable Energy Workshop, September 19-21, 2011, at Radisson Hotel
- Geosciences Applications Workshop, September 21-23, 2011, at Radisson Hotel

These workshop engaged a large a part of the statistics, mathematics, and scientific communities. Further workshops on more specialized themes were organized during the program:

- Scientific Problems for the Smart Grid - October 3-5, 2011, at SAMSI
- High Dimensional Approximation for Uncertainty Quantification - November 10, 2011, at SAMSI
- Observations Workshop- January 17-19, 2012, co-organized with the Cooperative Institute for Climate and Satellites in Asheville, NC
- Multiscale Modeling and Uncertainty Quantification for Nuclear Fuel Performance, January 25-26, 2012, at SAMSI
- Simulation of Rare Events - February 13-14, 2012, at SAMSI
- Models with Complex and Uncertain Domains - March 22-23, 2012, at SAMSI
- Uncertainty Quantification for High-Performance Computing, May 2-4, 2012, co-organized with Oak Ridge National Lab in Oak Ridge, TN
- Transition Workshop - May 21-23, 2012, at the Radisson Hotel.

In addition to the above official activities of the program, program members were heavily involved with the first SIAM Conference on Uncertainty Quantification, which took place on April 2-4, 2012, at the Raleigh Marriott City Center Hotel in Raleigh, NC. The event was sponsored by the SIAM Activity Group on Uncertainty Quantification. Several working groups from the SAMSI UQ program contributed to, and benefitted from, this event in a significant way.

Courses

Two graduate-level courses were offered, on Numerical Methods for Uncertainty Quantification (Parts 1 and 2), during the Fall and Spring semesters.

Working Groups

- Surrogate Models - Methodology UQ
- Inverse Function-based Inference - Methodology UQ
- Approximating Computationally Intensive Functions and Sampling Design in High Dimensions - Methodology UQ
- Multiphysics - Methodology UQ
- Model Validation - Methodology UQ
- Stochastic to Deterministic Models and Back Again - Methodology UQ

- Data Assimilation - Methodology UQ
- Simulation of Rare Events - Methodology U
- Data Assimilation in IPCC Level Models - Climate UQ
- Parallel Computing Issues - Climate UQ
- Statistics of Extremes - Climate and Methodology UQ
- Nuclear Energy - Engineering UQ
- Renewable Energy - Engineering UQ
- Materials - Engineering UQ
- Sustainability - Engineering UQ
- Engineered Systems - Engineering UQ
- Geosciences - Geosciences UQ

One week Summer Program on Nonlocal Continuum Models for Diffusion, Mechanics, and Other Applications

Nonlocal discrete models are common in many applications. For example, molecular dynamics and other particle methods having interactions that extend beyond nearest neighbors are in common use. Less popular are nonlocal continuum models which are very much dominated in their usage by local continuum models, especially differential equation models. However, in recent years, there has been burgeoning interest in mathematical, scientific, and engineering circles in nonlocal continuum models, especially for solid mechanics, diffusion, and wave propagation. This interest is motivated by the desire to model singular or anomalous behavior such as cracks and fracture in solids and, more generally, by the need to develop multiscale models, that is, models that are valid and tractable over a wide range of temporal and spatial scales. As a result, there are by now researchers all over the world investigating diverse aspects of nonlocal continuum models. This Summer workshop will bring together such researchers from the mathematical, statistical, computational, scientific, and engineering communities to talk about their interests and research and establish lasting, synergistic connections that lead to new research. We are also interested in having participants who may not be involved in research into nonlocal continuum models, but that have expertise in areas that will influence future research into those models. One particular direction of that research is into uncertainty quantification which will require researchers in nonlocal continuum models to interact with statisticians so as to develop good models for Uncertainty Quantification.

Organizers

Max Gunzburger (Florida State) and Richard Lehoucq (Sandia)

SAMSI Liaison

Pierre Gremaud

Date and place

SAMSI, June 25-29, 2012

Confirmed participants as of early April 2012

Burak Aksoylu (Tobb U. Turkey), Stephen Bond (Sandia), Nathanial Burch (SAMSI), Xi Chen (Florida State U.), Marta D'Elia (Florida State), Albert Erkip (Sabanci U., Turkey), Richard Fabiano (UNC Greensboro), Greg Forrest (UNC Chapel Hill), Eliot Fried (McGill), John Fricks (Penn State), Joe Goddard (UCSD), Zhen Guan (U. Tennessee, Knoxville), Maila Hallare (U. Kansas), Steven Henke (Florida State), Xiaojie Hou (UNC Wilmington), Zhan Huang (Penn State), Lili Ju

(U. South Carolina), Jeremy Lechman (Sandia), Alan Lenarcic (UNC Asheville), Jiao Li (U. Wisconsin Milwaukee), Xiaofan Li (IIT), Sharon Lubkin (NCSU), Tchavdar Marinov (Southern U.), Rossitza Marinova (Concordia), Scott McKinney (U. Florida), Tadele Mengesha (Penn State), Mark Meerschaert (Michigan State), Pachenko (Washington State U.), Michael Parks (Sandia), Maria Teresa Perez-Llanos (U. Autonoma de Madrid, Spain), Petronale Radu (U. Nebraska), Michael Raghib (LANL), Naveen Ramunigari (U. Texas), Julio Rossi (U. Alicante, Spain), Ekkehard Sachs (U. Trier, Germany), Alla Sikorskii (Michigan State), Pablo Seleson (Texas at Austin), Peter Straka (Michigan State U.), Zuhal Unlu (LSU), Ying Wang (U. Minnesota), Wojbor Woyczynski (Case Western), Dexuan Xie (U. Wisconsin, Milwaukee), XiaoXia Xie (Auburn), Ling Xu (U. New Mexico), Guannan Zhang (Florida State)

We expect between 70 and 90 participants.

An example

As an example of a nonlocal continuum model, consider the equation

$$u_t = -2 \int_\Omega (u(y,t) - u(x,y))\ \gamma(x,y)\, dy \tag{1}$$

for a function $u(x,t)$; here, $\gamma(x,y)$ is a given kernel. This equation is nonlocal because interaction occurs between points x and y that are separated. This is a nonlocal diffusion equation which has applications in research areas such as fractional diffusion and fractional Laplacians, image analyses, machine learning, nonlocal Dirichlet forms, nonlocal heat conduction, dense graphs (graph Laplacians).

The model (1) and its solutions have several important features that makes them interesting. In additional to being a nonlocal, continuum model, (1) is free of spatial derivatives. For certain kernels $\gamma(x,y)$ and for smooth function u, (1) reduces, in an appropriate limit, to the heat equation. If one considers the operator on the right-hand side of (1), or if one considers the steady-state version of (1)) given by

$$-2 \int_\Omega (u(y,t) - u(x,y))\ \gamma(x,y)\, dy = b(x), \tag{2}$$

practical kernels $\gamma(x,y)$ can be defined such the solution u does not possess two more (weak) derivatives compared to the data $b(x)$ as is the case for elliptic partial differential equations. In fact, one can have, for $s \in [0,1)$, that if $b \in H^{-s}(\Omega)$, then $u \in H^s(\Omega)$. Thus, the case of no smoothing is included. Of particular interest are kernels for which $s \in [0,1/2)$ because then, equations such as (1) and (2) admit solutions with jump discontinuities so that generalizations of these equations can be used to model cracks and other discontinuous behaviors. Furthermore, for such kernels, one can discretize these equations using discontinuous finite element spaces without the need for accounting for any fluxes across element boundaries as is the case for, e.g., elliptic partial differential operators. This is just a sampling of the properties and results about nonlocal models such as (1) that have great mathematical interest and, at the same time, have deep implications to many applications.

Note that if u_t in (1) is replaced by u_{tt}, we then have a nonlocal "wave" equation. Such equations have very different character, e.g., with respect to speed of propagation, dispersion, etc., compared the classical wave equation; these differences could have important implications to several application areas. A vector-valued form of (1) with u_t replaced by u_{tt} is equivalent to the peridynamics model for the mechanics of linear materials. Nonlinear version of (1) and (2) are also of interest, especially in the peridynamics setting.

Program topics

Participants in the workshop will discuss modeling, mathematical, statistical, computational, and applications issues such as

- choices for the kernel $\gamma(x, y)$
- connections between nonlocal continuum models and discrete models such as MD
- well posedness of the equations
- limiting behaviors of solutions
- finite element and other discretization methods
- efficient solution methods for discretized systems
- uncertainty quantification
- applications including but not limited to mechanics, image processing, graphs, diffusion, and wave propagation.

Uncertainty quantification through nonlocal models is especially intriguing because of the very different nature of local and nonlocal models with respect to such effects as dispersion, smoothing or the lack thereof, etc

Program format

The format of the program will be novel. Tutorials will be given on the first 1.5 day (Monday and Tuesday morning). This will be followed by poster blitzes on Tuesday afternoon, Wednesday and Thursday (short presentations paired with posters). On Friday morning, a tutorial on available software will be offered. We expect this slightly nonstandard format to foster new discussions and interactions. The workshop will conclude with a session on future directions.

Two-week Summer Program on Computational Advertising (6-17 August 2012)

The program was described in the proposal for the NAC meeting on 19 December 2011. Alternatively please see `http://www.samsi.info/programs/CA12`

Program Organizers: Deepak Agarwal (LinkedIn), Diane Lambert (Google)
Local Organizer: David Banks (Stat, Duke)
SAMSI Liaison: Ilse Ipsen

Schedule

Talks will be 40-50 minute technical presentations

- *Monday:* McShane, Owen, Eckles, Pedersen, Zhang, Lowe
  Evening: Poster session
- *Tuesday:* Lambert, Shi, Sundaresan, Taddy, Yang, Wang
- *Wednesday:* Ravishanker, Huo, Bell, Zheng, Posse
- *Thursday:*
  1. Panel discussion discussion with Kimberly Sellers (Math & Stat, Georgetown University), David Banks (Stat, Duke), an optimizer, and a rep from MaxPoint
  2. Discussion of datasets: Lead by Yang Yang (Yahoo!).
  3. Formation of working groups
     The two main working groups will be lead by David Banks, and Liang Zhang (LinkedIn).
- *Friday through Friday:* Research working groups meet at SAMSI

Invited speakers

Confirmed: Robert Bell (AT&T Labs), Dean Eckles (Facebook), Ming-hui Chen (Google), Xiaoming Huo (School of Industrial & Systems Eng., Georgia Tech), Diane Lambert (Google), Mark Lowe (MaxPoint Interactive), Blakeley McShane (Kellogg School of Management, Northwestern), Art Owen (Stat, Stanford), Jan Pedersen (Microsoft), Christian Posse (LinkedIn), Nalini Ravishanker (Stat, University of Connecticut), Neel Sundaresan (eBay), Matt Taddy (Booth School of Business, Chicago), Chong Wang (CS, Princeton), Hongxia Yang (IBM), Liang Zhang (LinkedIn), Tian Zheng (Stat, Columbia University)

Waiting for reply from: Dilip Abreu (Princeton), Rakesh Agrawal (Microsoft), Minghui Shi (Google)

Partial list of topics

Bootstrapping r-fold tensor data (Art Owen)
Effects of social cue and tie strength in social advertising: Evidence from field experiments (Dean Eckles)
Statistical models in keyword bidding (Xiaoming Huo)
Fast predictive approaches for predictive inference for time correlated data streams (Nalini Ravishanker)
Inverse regression for analysis of sentiment in text (Matt Taddy)
Multi-relational learning via hierarchical nonparametric bayesian collective matrix factorization (Hongxia Yang)
Web page personalization and user profiling (Liang Zhang)

Data sets

The working groups will have access to two types of data sets: movie recommendations and e-commerce. We hope that MaxPoint, a local startup, will provide programming and IT support. Two students from the Duke Statistics Department (Andrew Cron and Jacopo Soriano) will also be available to help.

Movie Recommendations:

1. Y! music dataset `http://kddcup.yahoo.com/datasets.php`

   Classical problem: Estimating user-item interactions with very large web scale datasets.

   This dataset has timestamps, which makes it possible to analyze user session behavior, and cluster users based on their session behavior.

2. Movielens 100k/1m/10M dataset:

   `http://www.grouplens.org/system/files/ml-10m-README.html`

   These datasets are smaller in scale, and easier to compute with, e.g. in R. They have both, user and item covariates.

3. Netflix data: Not the most exciting data set anymore, but still useful.

e-commerce:

1. epinions data: `http://www.trustlet.org/wiki/Downloaded_Epinions_dataset`

   In addition to product ratings, the data also contain pairs with positive trust values.

2. Yahoo! webscope data

   May or may not be available, we will know by June.

Diversity

Women: Diane Lambert, Nalini Ravishanker, Kimberly Sellers, Minghui Shi, Hongxia Yang, Yang Yang, Tian Zheng

African American: Robert Bell, Kimberly Sellers

B. Plans for 2012-13

1. **Statistical and Computational Methodology for Massive Datasets**
2. **Data-Driven Decisions in Health Care**

Statistical and Computational Methodology for Massive Datasets (2012 - 2013)

`http://www.samsi.info/programs/MD12`

Overall program leaders: Michael Jordan (EE + CS, UC Berkeley), Karen Kafadar (Stat, Indiana), Michael Mahoney (Math, Stanford), Steve Sain (NCAR), Jiayang Sun (Stat, Case Western), Alexander Szalay (Physics & Astronomy, Johns Hopkins)

NAC liaison: Bin Yu (UC Berkeley)
Local scientific coordinator: Yufeng Liu (Stat, UNC)
SAMSI directorate liaison: Ilse Ipsen (NCSU)

We describe the progress that has taken place since the last proposal in December 2011, in regard to visitors and workshop organization.

1 Visitors

Confirmed long-term visitors

Naomi Altman (Stat, Penn State), Fall 2012 + Spring 2013
Andreas Artemiou (Math, Michigan Tech. U.), Fall 2012 + Spring 2013
Jogesh Babu (Stat, Penn State), Fall 2012
Tamás Budavári (Physics & Astro, Johns Hopkins), September + October 2012
Guang Chen (Stat, Purdue), Spring 2013
Marco Ferreira (Stat, Missouri), Fall 2012 + Spring 2013
Karen Kafadar (Stat, U. Indiana), September 2012, January + February 2013
Brandon Kelly (Physics, UC Santa Barbara), September 2012
Amy Langville (CS, College of Charleston), regular visits in 2012 + 2013
Tom Loredo (Astro, Cornell), September 2012
Ashish Mahabal (Astro, Caltech), September + December 2012
Xingye Qiao (Stat, SUNY Binghampton), Spring 2013
Jiayang Sun (Stat, Case Western), Fall 2012
Bowei Xi (Stat, Purdue), Fall 2012 + Spring 2013

Confirmed short-term visitors

Prajval Shastri (Indian Institute of Astrophysics, Bangalore)

Under discussion: Tim Chartier (Math, Davidson College), Matthias Katzfuß (Math, U. Heidelberg, Germany), Scott Pratt (Physics, Michigan State), Joseph Richards (Astro, UC Berkeley), Roberto Trotta (Astrophys, Imperial College, UK), Ricardo Vilalta (CS, U. Houston)

Diversity *Women:* Naomi Altman, Karen Kafadar, Amy Langville, Prajval Shastri, Jiayang Sun
Ethnic: Marco Ferreira, Ricardo Vilalta

2 Workshops

We have three definite workshops that have been approved by the SAMSI Directorate, and four other workshops that are in the early planning stages.

2.1 Opening Workshop

Date: 9-12 September 2012, Location: Radisson Hotel
`http://www.samsi.info/MD-opening-workshop`

Organizers

Naomi Altman (Stat, Penn State), Michael Mahoney (Math, Stanford), Jiayang Sun (Stat, Case Western), Dani Ushizima (Lawrence Berkeley Lab)

Tentative schedule

Sunday: Tutorials
Monday, 9am - 12noon, 2-5pm: Topical talks, 6pm: poster session
Tuesday, 9am - 12noon: *Environmental applications*, 2-5pm: Topical talks
Wednesday morning: 8:30am - 12noon: Talks + panel discussion
Wednesday afternoon – Friday: Research working groups form

Tutorial speakers

Tamás Budavári (Physics & Astro, Johns Hopkins)
Statistical methods in astronomy
Petros Drineas (CS, RPI)
Mining massive datasets: A (randomized) linear algebraic perspective
Michael Jordan (EECS & Stat, UC Berkeley)
Resampling methods for massive data
Haesun Park (College of Computing, Georgia Tech)
Visual analytics for knowledge discovery in high dimensional data
Steve Wright (CS, Wisconsin)
Optimization techniques for statistical analysis of large datasets

Confirmed topical speakers so far

Steffen Bass (Physics, Duke), WG[1] *High-Energy Physics*
Brian Caffo (Biostat, Johns Hopkins), WG *Inference*
Kyle Cranmer (Physics, NYU), WG *High-Energy Physics*
Noel Cressie (Stat, Ohio State), WG *Environment & Climate*
Dan Crichton (Jet Propulsion Lab), WG *Environment & Climate*
Inderjit Dhillon (CS, UT Austin), WG *Datamining*
Jianqing Fan (Finance & Stat, Princeton), WG *Imaging*
Maryam Fazel (EE, U. Washington), WG *Online Sketching & Streaming*
Michael Mahoney (Math, Stanford), WG *Online Sketching & Streaming*
Anna Michalak (EESS, Stanford), WG *Environment & Climate*
Jim Nagy (CS, Emory), WG *Imaging*
Jonathan Taylor (Stat, Stanford), WG *Inference*
Bin Yu (Stat, UC Berkeley), WG *Inference*

Diversity *Women:* Naomi Altman, Maryam Fazel, Anna Michalak, Haesun Park, Jiayang Sun, Dani Ushizima, Bin Yu

[1] "WG" denotes "working group topic".

2.2 SAMSI-SAVI Workshop on Astrostatistics

Date: 19-21 September 2012, Location: SAMSI
`http://www.samsi.info/AS12`

This workshop is partly supported by the *Virtual Institute of Mathematical and Statistical Sciences*[2], which is part of the broader NSF initiative *Science Across Virtual Institutes (SAVI)*. The goal of the workshop is to bring together astrophysicists and statisticians to brainstorm on advanced topics in statistical inference in the context of modern empirical astrophysics. The workshop will bring together astrophysicists from diverse sub-disciplines who have used statistical analysis in their research, and statisticians who have experience with statistical issues in astrophysics, in order to create a forum for extensive interactions on state-of-the-art statistical inference as applicable to astrophysical problems.

Tentative list of topics:

- Astronomy:
  Transients in Astrophysics: the search for transients, search for periodicities; large sky surveys, Dark Energy Survey; cosmic microwave background studies; galaxy evolution; exoplanets; current missions: Fermi, SDSS, DES, Planck; future missions: LSST, LIGO
- Improved statistical inference:
  Measurement errors, non-linear data transformation, Gaussian processes, data mining, Bayesian methods

Organizers

Jogesh Babu (Stat, Penn State),
Prajval Shastri (Indian Institute of Astrophysics, Bangalore)

Confirmed speakers so far

Debbie Bard (SLAC, Stanford University)
Jim Berger (Stat, Duke)
Andrew Connolly (Astro, U. Washington)
Eric Ford (Astro, U. Florida)
Fabrizia Guglielmetti (Max Plank Institute, Germany)
Ashish Mahabal (Astro, Caltech)
Bhuvnesh Jain (Physics & Astro, U. Penn)
Brandon Kelly (Physics, UC Santa Barbara)
Ann Lee (Stat, CMU)
Tom Loredo (Astro, Cornell)
Hakeem Oluseyi (Physics & Space Sciences, Florida Inst. Tech.)
Harrison Prosper (Physics, Florida State)
Joseph Richards (Astro, UC Berkeley)
Aneta Siemiginowska (Harvard-Smithsonian Center for Astrophysics)
Risa Wechsler (Physics, Stanford)

Diversity *Women:* Debbie Bard, Fabrizia Guglielmetti, Ann Lee, Prajval Shastri, Aneta Siemiginowska, Risa Wechsler
Ethnic: Hakeem Oluseyi, Harrison Prosper

[2]`http://www.samsi.info/VI-MSS`

2.3 SAMSI-FODAVA Workshop on Interactive Visualization and Analysis of Massive Data

Date: 10-12 December 2012, Location: SAMSI
`http://www.samsi.info/IV12`

This workshop is organized in conjunction with the joint NSF/DHS Initiative on the *Foundations of Data Analysis and Visual Analytics* (FODAVA). Haesun Park, the PI and FODAVA lead is the main workshop organizer.

With the advance in technology, enormous amounts of data are generated on a daily basis virtually in every area including bioinformatics, astrophysics, chemometrics, social network analysis, web mining, text mining, financial analysis, and security. We are faced with significant analytical challenges due to many special characteristics of these data sets, which are of large volume, often unstructured, high dimensional, noisy, incomplete, time-varying, spatial, and originate from different sources. These challenges can be turned into new opportunities and discoveries when the massive data can be transformed into useful knowledge. Recent developments in data and visual analytics show that incorporating interactive capability through visual interfaces with automated data analysis methods can substantially increase our ability to understand the data and find more meaningful solutions.

The primary goal of the workshop is to bring together researchers in Mathematics, Statistics, Computational Science and Engineering, Computer Science, and Visualization to work on massive scale data and visual analytics. Issues to be investigated include the mathematical, statistical, and algorithmic issues in efficient representation and transformation of data, scalable and dynamic algorithms for real time interaction, visual representation in limited screen space, performing evaluations, and applications.

Organizers

Orly Alter (Bio. Eng., U. Utah)
Kwan-Liu Ma (CS, UC Davis)
Mauro Maggioni (Math, Duke)
Haesun Park (College of Computing, Georgia Tech)
Leland Wilkinson (SYSTAT Software, Inc.)

Diversity *Women:* Orly Alter, Haesun Park

2.4 Workshops under discussion

The workshops below are in the early stages of planning, and have not yet been approved by the SAMSI Directorate.

Nonlinear optimization

- 2 1/2 days, during the week of 15 or 22 October 2012
- Organizers:

  Yufeng Liu (Stat, UNC), Ben Recht (CS, Wisconsin), Steve Wright (CS, Wisconsin)
- Estimated attendance:

  $\gg$ 50, judging by recent related workshops, with 130 participants at the IMA (*Machine learning: theory and computation*, March 2012) and 160 at IPAM (*Modern trends in optimization and its application*, October 2010)
- Topics:
  - Convex optimization (sparse modeling and compressed sensing, fast MRI, matrix completion)
  - Online optimization (streaming data, on-line learning, control theory)
  - Distributed optimization (parallel and GPU computation, data fusion)
  - Machine learning and high-dimensional models (concentration phenomena, random matrices)

  Interactions expected with working groups on: Methodology for online streaming and sketching, and environment & climate.

Large environmental data sets: NCAR, Spring 2013
Amy Braverman (JPL), Steve Sain (NCAR), Richard Smith (Stat, UNC)

High-energy physics: Spring 2013
Steffen Bass (Physics, Duke), Karen Kafadar (Stat, U. Indiana)

Inference & imaging: Fall 2012
Naomi Altman (Stat, Penn State), Jiayang Sun (Stat, Case Western), Dani Ushizima (Lawrence Berkeley)

3 Working groups

Data visualization and analytics: High-speed visualization of high-dimensional datasets; data representation, extraction, integration and transformation; real-time visual interaction; spatio-temporal data mining. *Chair: Haesun Park (College of Computing, GA Tech)*

Online streaming and sketching: Algorithm paradigms for massive datasets (streaming, online, randomized); scalability; filtering; anomaly detection; data structures for fast computation of statistics; database enabled machine learning tools; computing environments and programming models (GPU's, cloud computing, custom chips). *Chair: Michael Mahoney (Math, Stanford)*

Large-scale optimization: Convex optimization (sparse modeling and compressed sensing, matrix completion); online optimization (streaming data, on-line learning, control theory); distributed optimization (parallel and GPU computation, data fusion); machine learning; high-dimensional models. *Chair: Yufeng Liu (STOR, UNC)*

Inference: Dimension reduction for high-dimensional data (feature selection, sub-sampling and screening, sparse PCA); predictive inference and multiple testing (false discovery rates, uncertainty in prediction); high-dimensional MCMC methods for posterior inference (particle filters, hybrids with optimization methods). *Chair: Naomi Altman (Stat, Penn State)*

Imaging: Rapid registration and segmentation (GPU's, distributed computing); multiple testing and inference for large-scale imaging data (sky surveys, satellite images, false discovery rate with dependence); dynamic imaging (streaming data, spatio-temporal models). *Chairs: Jiayang Sun (Stat, Case Western) & Dani Ushizima (Lawrence Berkeley Lab)*

Datamining: Clustering and ranking, nonnegative matrix factorization, Aggregation-disaggregation methods, stochastic matrix methods, clustering based on eigenvalues, integer programming and optimization. *Chairs: Amy Langville (CS, College of Charleston) & Carl Meyer (Math, NCSU)*

High-energy physics: Reconstruction and analysis of particle collisions from the LHC; pattern recognition and parameter extraction; simulations to estimate error rates; parameter estimation for large numbers of parameters; maximum likelihood estimators. *Chairs: Karen Kafadar (Stat, Indiana) & Steffen Bass (Physics, Duke), Sven Leyffer & Salman Habib (Argonne National Lab)*

Astronomy: Statistics on remote resources; computations on special purpose architectures and GPUs; communication avoiding methods; randomized and online algorithms; detection and classification of transient events and outliers; Bayesian inference and machine learning; high dimensional models with empirical priors; non-parametric models; visualization of large high-dimensional datasets. *Chairs: Tamas Budavari (Physics & Astro, Johns Hopkins), Jogesh Babu (Stat, Penn State)*

Environment and climate: Production, validation, processing, distribution and integration of data; data fusion and remote sensing; algorithms for large distributed datasets; spatial or spatio-temporal statistics. *Chair: Jessica Matthews (CICS/NOAA)*

Diversity *Women:* Naomi Altman, Karen Kafadar, Amy Langville, Jessica Matthews, Haesun Park, Jiayang Sun, Dani Ushizima c

2012-13 Program on
Data-Driven Decisions in Healthcare (DDDHC)
Summary of Status (April 9, 2012)

Background

Healthcare is a central political, economic and social issue of our times. In the healthcare process, decisions are made at every level from the treatment of individual patients to formulation and evaluation of national policies. At the same time, data generation is increasing dramatically. Electronic medical records are becoming ubiquitous. Tests produce gigabytes of data, including images and biometric samples. The volumes of data are daunting in themselves; concomitant problems such as confidentiality and data quality exacerbate the challenges in producing usable tools that support principled healthcare decisions.

Research Themes

The SAMSI program on DDDHC addresses issues of mathematical and statistical theory and methodology that will improve evidence-based healthcare decision-making. The goals are to:

- Strengthen the link between data and decisions, a path that includes major challenges in mathematical modeling and statistical inference.
- Highlight and increase the role that statistics, applied mathematics and operations research can play in making data-driven healthcare decisions.

The program has two principal—and intersecting—themes:

- *Operations Research Modeling (ORM),* with particular emphasis on resource allocation.
- *Comparative Effectiveness Research (CER)*

Program Leaders: Sheldon Jacobsen (University of Illinois at Urbana-Champaign), Myron Katzoff (National Center for Health Statistics/CDC), Mark Lewis (Cornell University), Avi Mandelbaum (Technion), Marianthi Markatou (IBM), Robert Obenchain (Risk Benefit Analysis), William Shannon (Washington University in St. Louis)

Local Scientific Coordinators: Nilay Argon (University of North Carolina at Chapel Hill), Vidyadhar Kulkarni (University of North Carolina at Chapel Hill), Stanley Young (National Institute of Statistical Sciences)

SAMSI Directorate Liaison: Alan Karr

National Advisory Committee Liaison: Susan Murphy (University of Michigan)

Personnel

Definite Long-Term Visitors: Sheldon Jacobsen (University of Illinois at Urbana-Champaign), Avi Mandelbaum (Technion), Marianthi Markatou (IBM), Robert Obenchain (Risk Benefit Analysis)

Provisional Long- or Short-Term Visitors: Constantine Gatsonis (Brown), Myron Katzoff (National Center for Health Statistics/CDC, retired), Tammy Kemp (Carillon Clinic), Mark Lewis (Cornell University), David Madigan (Columbia University), Sally Morton (University of Pittsburgh), Keith O'Rourke (University of Toronto?), William Shannon (Washington University in St. Louis), Joseph Stanford (University of Utah)

Partner Institution Researchers (Faculty Fellows indicated by *): Nilay Argon (UNC*), Brian Denton (NCSU), Gregory Forest (UNC*), A. Blanton Godfrey (NCSU), Alan Karr (NISS), Vidyadhar Kulkarni (UNC), Eric Laber (NCSU), Stanley Young (NISS), and possibly one Faculty Fellow from the UNC Biostatistics Department

Postdoctoral Fellow: Kenneth Lopiano (Statistics, University of Florida). Potential postdoctoral associates are under consideration.

Graduate Student Fellows: In the process of being designated

Other Triangle Researchers: Alan Menius (GlaxoSmithKline)

Description of Activities

Workshops: The program will start with an *Opening Workshop* on August 26-29, 2012. Confirmed speakers and panelists include Constantine Gatsonis (Brown), Brenda Dietrich (IBM), Sheldon Jacobsen (Illinois), Avi Mandelbaum (Technion), Marianthi Markatou (IBM). Invitees include Stephen Cohen (AHRQ), Steve Goodman (Stanford), John Ioannidis (Stanford), David Madigan (Columbia), Sally Morton (Pittsburgh), William Shannon (Washington U in St. Louis)

The *Transition Workshop* will be in the late spring of 2013.

Mid-program workshops will be planned in support the Working Groups.

Courses: Fall: Operations Research Methods in Health Care; Spring: TBD

Working Groups: Working groups will meet throughout the program to pursue particular research topics articulated in the Opening Workshop, as well as topics identified subsequently by Working Group participants. The Working Groups consist of SAMSI visitors, postdoctoral fellows, graduate students, faculty and scientists from the Research Triangle area. Remote participation is possible via WebEx. Candidate Working Group foci include:

- *ORM:* Patient Flow; Data-Based Design of Healthcare Operations; Management of Transplant Lists; Emergency Response; Service Delivery
- *CER:* Predictive Modeling; Case Mix Adjustment; Data Confidentiality, Quality & Integration; Public Health Surveillance
- *Cross-Cutting:* Computational Issues, including Massive Data; Observational Medical Studies

APPENDIX A – 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
FP – Faculty or Professional
S – Students (Education & Outreach)
A – Faculty (Education & Outreach)

2010 SUMMER PROGRAM

Semiparametric Bayesian Inference in PKPD Analysis
Participant Summary
July 12-23, 2010

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	26	9	0	18	17	31	2	2	27	15
Unsuppted	40	18	1	27	32	43	5	11	17	7
SAMSI	1	1	0	1	1	2	0	0		

Semiparametric Bayesian Inference in PKPD Analysis
Workshop Participant Support
July 12-23, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Bayarri	M.J. (Susie)	Female	University of Valencia	Stat	FP
Boonstra	Phil	Male	University of Michigan	Stat	NRG
Choi	Leena	Female	Vanderbilt University	Stat	FP
Cornebise	Julien	Male	University of British Columbia	Stat	NRG
Dahl	David	Male	Texas A&M	Stat	FP
Ding	Lili	Female	Cincinnati Children's	Stat	NRG

			Hospital Medical Center		
Dutta	Ritabrata	Male	Purdue University	Stat	NRG
Escobar	Michael	Male	University of Toronto	Stat	FP
Fronczyk	Kassie	Female	University of California, Santa Cruz	Stat	NRG
Ghosh	Kaushik	Male	University of Nevada	Stat	FP
Guidani	Michele	Male	University of New Mexico	Stat	FP
Johnson	Wes	Male	University of California, Irvine	Stat	FP
Kottas	Thanasi	Male	University of California, Santa Cruz	Stat	FP
Lewandowski	Andrew	Male	Purdue University	Stat	NRG
Li	Lang	Male	Indiana University	Biostat	FP
MacEachern	Steven	Male	Ohio State University	Stat	FP
Mueller	Peter	Male	University of Texas	Stat	FP
Oluyede	Broderick	Male	Georgia Southern University	Stat	FP
Pararai	Mavis	Female	Indiana University of Pennsylvania	Stat	NRG
Quispe Vargas	Walter	Male	University of Puerto Rico	Stat	NRG
Rodriguez	Abel	Male	University of California, Santa Cruz	Stat	NRG
Rosner	Gary	Male	Johns Hopkins University	Stat	FP
Roura	Jaime	Male	University of Puerto Rico	Stat	NRG
Tadesse	Mahlet	Male	Georgetown	Stat	FP
Telesca	Donatello	Male	University of California, LA	Stat	NRG

Thomas	Duncan	Male	University of Southern California	Other	FP
Tomasetti	Cristian	Male	University of Maryland, College Park	Math	NRG
Torres	David	Male	University of Puerto Rico	Math	NRG
Vincini	Paolo	Male	Pfizer	Life	FP
Wahed	Abdus	Male	University of Pittsburgh	Stat	FP
Wang	Yanpin	Female	University of Florida	Stat	NRG
Wu	Meihua	Male	University of Michigan	Stat	NRG
Xu	Xinyi	Female	Ohio State University	Stat	FP
Yao	Ping	Female	Northern Illinois University	Stat	NRG
Zou	Yuanshu	Female	University of Cincinnati	Stat	NRG

Semiparametric Bayesian Inference in PKPD Analysis

Workshop Participants

July 12-23, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Armagan	Artin	Male	Duke University	Stat	NRG
Banks	David	Male	Duke University	Stat	FP
Barbare	Ralph	Male	North Carolina State University	Stat	NRG
Bayarri	M.J. (Susie)	Female	University of Valencia	Stat	FP
Berger	Jim	Male	SAMSI	Stat	FP
Berrocal	Veronica	Female	SAMSI	Stat	NRG
Bhattacharya	Anirban	Male	Duke University	Stat	NRG
Bonate	Peter	Male	GlaxoSmithKline	Other	FP

Boonstra	Phil	Male	University of Michigan	Stat	NRG
Burgette	Lane	Male	Duke University	Stat	NRG
Canale	Antonio	Male	Duke University	Stat	NRG
Chen	Shuguang	Male	GlaxoSmithKline	Stat	FP
Chiu	Joannellyn	Female	GlaxoSmithKline	BioEng	FP
Choi	Leena	Female	Vanderbilt University	Stat	FP
Conti	David	Male	University of Southern California	Stat	FP
Cornebise	Julien	Male	University of British Columbia	Stat	NRG
Cui	Kai	Male	Duke University	Stat	NRG
Dahl	David	Male	Texas A&M	Stat	FP
D'Argenio	David	Male	University of Southern California	Engg	FP
Davidian	Marie	Female	North Carolina State University	Stat	FP
de Iorio	Maria	Female	Imperial College	Stat	FP
Ding	Lili	Female	Cincinnati Children's Hospital Medical Center	Stat	NRG
Dunson	David	Male	Duke University	Stat	FP
Dutta	Ritabrata	Male	Purdue University	Stat	NRG
Escobar	Michael	Male	University of Toronto	Stat	FP
Feng	Xingdong	Male	NISS	Stat	NRG
Fox	Emily	Female	Duke University	Stat	NRG
Fronczyk	Kassie	Female	University of California, Santa Cruz	Stat	NRG
Gan	Kevin	Male	GlaxoSmithKline	Stat	NRG

Ghosal	Subhashis	Male	North Carolina State University	Stat	FP
Ghosh	Sujit	Male	North Carolina State University	Stat	FP
Ghosh	Joyee	Female	University of North Carolina, Chapel Hill	Stat	NRG
Ghosh	Kaushik	Male	University of Nevada	Stat	FP
Ghosh	Jayanta	Male	Purdue University	Stat	FP
Gillespie	William	Male	Metrum Inst	Other	FP
Gong	Xiaohua	Male	GlaxoSmithKline	Biology	NRG
Guha	Subharup	Male	University of Missouri	Stat	FP
Guidani	Michele	Male	University of New Mexico	Stat	FP
House	Leanna	Female	Virginia Tech	Stat	NRG
Jelliffe	Roger	Male	University of Southern California	Other	FP
Johnson	Wes	Male	University of California, Irvine	Stat	FP
Johnson	Brendan	Male	GlaxoSmithKline	Corp	FP
Korman	Alexander	Male	Duke University	Stat	NRG
Kottas	Thanasi	Male	University of California, Santa Cruz	Stat	FP
Krzyzanski	Wojciech	Male	State University of New York	Other	FP
Leary	Bob	Male	Pharsight	Comp	FP
Lee	Ju Hee	Female	Ohio State University	Stat	NRG
Lewandowski	Andrew	Male	Purdue University	Stat	NRG
Li	Lang	Male	Indiana University	Biostat	FP
Lin	Lin	Female	Duke University	Stat	NRG

Luo	Man (Melody)	Female	GlaxoSmithKline	Stat	NRG
MacEachern	Steven	Male	Ohio State University	Stat	FP
Mueller	Peter	Male	University of Texas	Stat	FP
Neely	Michael	Male	University of Southern California	Math	FP
Oluyede	Broderick	Male	Georgia Southern University	Stat	FP
Omolo	Bernard	Male	University of North Carolina, Chapel Hill	Stat	FP
O'Rourke	Keith	Male	O'Rourke Consulting / Health Canada	Math	FP
Pang	Herbert	Male	Duke University Medical Center	Stat	NRG
Pararai	Mavis	Female	Indiana University of Pennsylvania	Stat	NRG
Pati	Debdeep	Male	Duke University	Stat	NRG
Petralia	Francesca	Female	Duke University	Stat	NRG
Quispe Vargas	Walter	Male	University of Puerto Rico	Stat	NRG
Reed	Mike	Male	Duke University	Math	FP
Rizwan	Ahsan	Male	University of North Carolina, Chapel Hill	Life	NRG
Rodriguez	Abel	Male	University of California, Santa Cruz	Stat	NRG
Rosner	Gary	Male	Johns Hopkins University	Stat	FP
Roura	Jaime	Male	University of Puerto Rico	Stat	NRG
Scarpa	Bruno	Male	University of Padova	Stat	FP
Sivaganesan	Siva	Male	University of Cincinnati	Stat	FP
Tadesse	Mahlet	Male	Georgetown	Stat	FP
Tatarinova	Tatiana	Female	University of Glamorgan	Math	NRG

Telesca	Donatello	Male	University of California, Los Angeles	Stat	NRG
Thall	Peter	Male	MDACC	Stat	FP
Thomas	Duncan	Male	University of Southern California	Other	FP
Tokdar	Surya	Male	Duke University	Stat	NRG
Tomasetti	Cristian	Male	University of Maryland, College Park	Math	NRG
Torres	David	Male	University of Puerto Rico	Math	NRG
Toto	Criselda	Female	Worcester Polytechnic Institute	Stat	NRG
Vincini	Paolo	Male	Pfizer	Life	FP
Wahed	Abdus	Male	University of Pittsburgh	Stat	FP
Wang	Xia	Female	NISS	Stat	NRG
Wang	Xiaojing	Female	Duke University	Stat	NRG
Wang	Jianqiang	Male	NISS	Stat	NRG
Wang	Yanpin	Female	University of Florida	Stat	NRG
Wilson	Melanie	Female	University of Southern California	Biology	FP
Wolpert	Robert	Male	Duke University	Stat	FP
Wu	Meihua	Male	University of Michigan	Stat	NRG
Xia	Jessie	Female	NISS	Stat	NRG
Xing	Chuanhua	Female	Duke University	Stat	NRG
Xu	Xinyi	Female	Ohio State University	Stat	FP
Yang	Xiaowei	Male	University of California, Davis	Stat	FP
Yang	Hongxia	Female	Duke University	Stat	NRG

Yao	Ping		Northern Illinois University	Stat	NRG
Yu	Li	Female	MedImmune	Stat	NRG
Zhang	Jianping	Male	GlaxoSmithKline	Math	NRG
Zhu	Hongxiao	Female	University of Texas M.D. Anderson Cancer	Stat	NRG
Zou	Yuanshu	Female	University of Cincinnati	Stat	NRG
Zou	Jian	Male	NISS	Stat	NRG

2010-11 PROGRAM EVENTS AFTER AUG 1, 2010

- **Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change**

Spatial Transition Workshop
Participant Summary
October 11-13, 2010

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	9	3	0	7	5	9	2	1	10	8
Unsuppted	18	7	1	17	9	23	1	2	14	6
SAMSI	4	2	0	1	5	6	0	0		

Spatial Transition Workshop
Workshop Participant Support
October 11-13, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Calder	Catherine	Female	Ohio State University	Stat	FP
Chakraborty	Avishek	Male	Texas A&M U	Stat	NRG
Cressie	Noel	Male	Ohio State U	Stat	FP

Ferreira	Marco	Male	U of Missouri	Stat	FP
Guindani	Michele	Male	University of New Mexico	Stat	FP
Huser	Raphael	Male	EPFL	Stat	FP
Kramer	Peter	Male	Rensselaer Polytechnic Institute	Math	FP
Nicolis	Orietta	Female	University of Bergamo	Math	NRG
Sanso	Bruno	Male	University of California, Santa Cruz	Stat	FP
Tingley	Martin	Male	NCAR	PHYS	NRG
Wu	Wei-Ying	Male	Michigan State University	Stat	NRG
Xue	Yun	Female	Michigan State University	Stat	NRG

Spatial Transition Workshop

Workshop Participants

October 11-13, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Assuncao	Renato	Male	Universidade Federal de Minas Gerais	STAT	FP
Banerjee	Sudipto	Male	University of Minnesota	STAT	FP
Berrocal	Veronica	Female	SAMSI	STAT	NRG
Calder	Catherine	Female	Ohio State University	STAT	FP
Chakraborty	Avishek	Male	Texas A&M University	STAT	NRG
Chang	Howard	Male	SAMSI	STAT	NRG
Craigmile	Peter	Male	Ohio State University	STAT	FP
Cressie	Noel	Male	Ohio State U	STAT	FP

Dunson	David	Male	Duke U	STAT	FP
Eidsvik	Jo	Male	NTNU	MATH	FP
Falk	Raymond	Male	OptInference LLC	STAT	FP
Ferreira	Marco	Male	University of Missouri	STAT	FP
Fuentes	Montse	Female	NCSU	STAT	FP
Gamble	Jennifer	Female	NCSU	ENG	NRG
Gelfand	Alan	Male	Duke U	STAT	FP
Ghosh	Souparno	Male	Duke University	STAT	NRG
Guindani	Michele	Male	University of New Mexico	STAT	FP
Haran	Murali	Male	Pennsylvania State University	STAT	FP
Hannig	Jan	Male	UNC	STAT	FP
Heaton	Matthew	Male	Duke U	STAT	NRG
Herring	Amy	Female	UNC, Chapel Hill	STAT	FP
Huser	Raphael	Male	EPFL	STAT	FP
Jeon	Soyoung	Female	UNC	STAT	NRG
Kang	Emily	Female	SAMSI	STAT	NRG
Kolovos	Alexander	Male	SAS Institute, Inc.	PHYS	FP
Kramer	Peter	Male	RPI	MATH	FP
Mannshardt-Shamseldin	Elizabeth	Female	Duke University	STAT	NRG
Martin	Tingley	Decline	NCAR	PHYS	NRG
Nail	Amy	Female	Duke University, Nicholas School of the Environment	STAT	NRG

Nicolis	Orietta	Female	University of Bergamo	MATH	NRG
Page	Garritt	Male	Duke University	STAT	NRG
Rajaratnam	Bala	Male	Stanford University	STAT	NRG
Reich	Brian	Male	NCSU	STAT	FP
Sanso	Bruno	Male	University of California, Santa Cruz	STAT	FP
Shaby	Ben	Male	SAMSI	STAT	NRG
Smith	Richard	Male	SAMSI	STAT	FP
Sun	Dongchu	Male	University of Missouri	STAT	FP
Wang	Xia	Female	NISS	STAT	NRG
Wasserman	Stan	Male	Indiana U	STAT	FP
Wolpert	Robert	Male	Duke University	STAT	FP
Wu	Wei-Ying	Male	Michigan state University	STAT	NRG
Xue	Yun	Female	Michigan State University	STAT	NRG
Zhang	Jun	Male	SAMSI	STAT	NRG
Zhang	Zhengjun	Male	University of Wisconsin	STAT	FP

- **Stochastic Dynamics**

Stochastic Dynamics Transition Workshop

Participant Summary
November 17-19, 2010

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	3	2	0	3	2	1	3	1	4	4
Unsuppted	15	7	1	8	15	4	18	1	11	6

SAMSI	1	0	0	0	1	0	1	0		

Stochastic Dynamics Transition Workshop
Workshop Participant Support
November 17-19, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Betterton	Meredith	Female	University of Colorado	Phys	FP
Fricks	John	Male	Pennsylvania State University	Stat	FP
Fuller	Pam	Female	Rensselaer Polytechnic Institute	Math	NRG
Kramer	Peter	Male	Rensselaer Polytechnic Institute	Math	FP
McKinley	Scott	Male	University of Florida	Math	NRG

Stochastic Dynamics Transition Workshop
Workshop Participants
November 17-19, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Athreya	Avanti	Female	Duke University	Math	NRG
Betterton	Meredith	Female	University of Colorado	Phys	FP
Budhiraja	Amarjit	Male	University of North Carolina	Stat	FP
Durrett	Rick	Male	Duke University and SAMSI	Math	FP
Elmohamed	Saleh	Male	Cornell University	Comp	FP
Fricks	John	Male	Pennsylvania State University	Stat	FP
Fuller	Pam	Female	Rensselaer Polytechnic Institute	Math	NRG
Gremaud	Pierre	Male	NCSU and SAMSI	Math	FP

Harlim	John	Male	North Carolina State University	Math	NRG
Hooker	Giles	Male	Cornell University	Stat	NRG
Kang	Min	Female	North Carolina State University	Math	NRG
Kolba	Tiffany	Female	Duke University	Math	NRG
Kotelenez	Peter	Male	Case Western Reserve University	Math	FP
Kramer	Peter	Male	Rensselaer Polytechnic Institute	Math	FP
Li	Junchi	Male	Duke University	Math	NRG
Liu	Xin	Female	University of North Carolina	Stat	NRG
Lu	Huitian		South Dakota State University	Stat	NRG
Mattingly	Jonathan	Male	Duke University	Math	FP
McKinley	Scott	Male	University of Florida	Math	NRG
Mitran	Sorin	Male	University of North Carolina	Math	FP
Mucha	Peter	Male	University of North Carolina	Math	FP
Rogers	Bruce	Male	Duke University	Math	NRG
Srinivasan	Ravi	Male	University of Texas	Math	NRG
Sun	Yi	Male	SAMSI	Math	NRG
Traud	Amanda	Female	North Carolina State University	Math	NRG
Wang	Xueying	Female	Texas A&M University	Math	NRG
Yamada	Richard	Male	University of Michigan	Math	NRG
Young	Jennifer	Female	Rice University	Math	NRG

2010-11 PROGRAM EVENTS THROUGH JULY 2011

❖ Complex Networks

Complex Networks Opening Workshop
Participant Summary
August 29-September 1, 2010

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	43	17	0	27	33	26	18	16	38	23
Unsuppted	80	31	1	46	66	68	29	15	25	7
SAMSI	5	2	0	0	7	2	5	0		

Complex Networks Opening Workshop
Workshop Participant Support
August 29-September 1, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Airoldi	Edoardo	Male	Harvard University	Stat	NRG
Balachandran	Kash	Male	Duke University	Stat	NRG
Bassett	Danielle	Female	University of California, Santa Barbara	Phys	NRG
Bickel	Peter	Male	University of California, Berkeley	Stat	FP
Bortz	David	Male	University of Colorado	Math	FP
Brummitt	Charlie	Male	University of California, Davis	Math	NRG
Caceres	Rajmonda	Female	University of Illinois at Chicago	Comp	NRG
Cintron-Arias	Ariel	Male	East Tennessee State University	Math	NRG

Clauset	Aaron	Male	Santa Fe Institute	Comp	NRG
Cressie	Noel	Male	Ohio State University	Stat	FP
Cucuringu	Mihai	Male	Princeton University	Math	NRG
Degond	Pierre	Male	CNRS	Math	FP
D'Souza	Raissa	Female	University of California, Davis	Phys	FP
Dukic	Vanja	Female	University of Colorado	Stat	FP
Fienberg	Steve	Male	Harvard University	Stat	FP
Gharibans	Linda	Female	University of California, Los Angeles	Stat	NRG
Girvan	Michelle	Female	University of Maryland	Phys	NRG
Gordillo	Luis	Male	University of Puerto Rico	Life	FP
Graham	Fan Chung	Female	University of California, San Diego	Math	FP
Gromoll	Christian	Male	University of Virginia	Math	FP
Hauck	Cory	Male	Oak Ridge National Laboratory	Math	NRG
Joo	Jaewook	Male	University of Tennessee	Phys	FP
Joyner	Michele	Female	East Tennessee State University	Math	FP
Katzoff	Myron	Male	National Center for Health Statistics	Stat	FP
Kolar	Mladen	Male	Carnegie Mellon University	Comp	NRG
Kurganov	Alexander	Male	Tulane University	Math	FP
Levina	Liza	Female	University of Michigan	Stat	FP

Li	Hongzhe	Male	University of Pennsylvania	Stat	FP
Lin	Guang	Male	Pacific Northwest National Laboratory	Math	NRG
Linkletter	Crystal	Female	Brown University	Stat	FP
Mahoney	Michael	Male	Stanford University	Stat	NRG
Makowski	Armand	Male	University of Maryland	Comp	FP
Meza	Rafael	Male	University of British Columbia	Math	NRG
Murrugarra Tomairo	David	Male	Virginia Tech/ Virginia Bioinformatics Institute	Math	NRG
Nesterko	Sergiy	Male	Harvard University	Stat	NRG
Newhall	Katie	Female	Rensselaer Polytechinc Institute	Math	NRG
Newman	Mark	Male	University of Michigan	Math	FP
Poggi-Corradini	Pietro	Male	Kansas State University	Math	FP
Redner	Sidney	Male	Boston University	Phys	FP
Reyes	Perla	Female	University of Wisconsin-Madison	Stat	NRG
Rodriguez	Abel	Male	University of California, Santa Cruz	Stat	NRG
Rohe	Karl	Male	University of California, Berkeley	Stat	NRG
Rombach	Michaela	Female	University of Oxford	Math	NRG
Salganik	Matthew	Male	Princeton Unviersity	Socl	NRG
Shalizi	Cosma	Female	Carnegie Mellon University	Stat	FP
Simpson	Sean	Male	Wake Forest U School of Medicine	Stat	NRG

Snijders	Tom	Male	University of Oxford	Stat	FP
Soufiani	Hossein Azari	Male	Harvard University	Engg	NRG
Spencer	Bruce	Male	Northwestern University	Stat	FP
Thomas	Andrew	Male	Carnegie Mellon University	Stat	NRG
Toroczkai	Zoltan	Male	University of Notre Dame	Phys	FP
Vaughan	Joel	Male	University of Michigan	Stat	NRG
Verella	J. Tipan	Male	University of Virginia	Engg	NRG
Volz	Erik	Male	University of Michigan	Life	NRG
Wang	Xueying	Female	SAMSI	Math	NRG
Westveld	Anton	Male	University of Nevada	Stat	NRG
Williams	Clarisa	Female	Auburn University	Stat	NRG
Xing	Eric	Male	Carnegie Mellon University	Comp	FP
Yang	Xiaolin	Female	Carnegie Mellon University	Stat	NRG
Zhu	Ji	Male	University of Michigan	Stat	FP

Complex Networks Opening Workshop

Workshop Participants
August 29-September 1, 2010

Last Name	**First Name**	**Gender**	**Affiliation**	**Major/Department**	**Status**
Airoldi	Edoardo	Male	Harvard University	Stat	NRG
Athreya	Avanti	Female	Duke University and SAMSI	Math	NRG
Balachandran	Kash	Male	Duke University	Stat	NRG

Banks	David	Male	Duke University	Stats	FP
Barbaro	Alethea	Female	University of California, Los Angeles	Math	NRG
Bassett	Danielle	Female	University of California, Santa Barbara	Physics	NRG
Bhamidi	Shankar	Male	University of North Carolina	Stat	NRG
Bianconi	Ginestra	Female	Northeastern University	Stat	FP
Bickel	Peter	Male	University of California, Berkeley	Stats	FP
Blitzstein	Joe	Male	Harvard University	Stats	NRG
Blocher	Jesse	Male	UNC-Chapel Hill Kenan-Flagler Business School	Other	NRG
Borrett	Stuart	Male	University of North Carolina, Wilmington	Life	FP
Bortz	David	Male	University of Colorado	Math	FP
Brummitt	Charlie	Male	University of California, Davis	Math	NRG
Caceres	Rajmonda	Female	University of Illinois at Chicago	Comp	NRG
Chari	Manoj	Male	SAS	Stat	FP
Chatterjee	Shirshendu	Male	Cornell University	Stat	NRG
Chen	Jiang	Female	UNC	Stats	NRG
Chertock	Alina	Female	North Carolina State University	Math	FP
Chipman	Hugh	Male	Acadia University	Stat	NRG
Choi	David	Male	Harvard University	Comp	NRG
Cintron-Arias	Ariel	Male	East Tennessee State University	Math	NRG
Clauset	Aaron	Male	Santa Fe Institute	Comp	NRG

Cooper	John	Male	Clemson University	Math	NRG
Costa	Marcelo	Male	Univesidade Federal de Minas Gerais	Stat	FP
Cressie	Noel	Male	Ohio State University	Stats	FP
Cucuringu	Mihai	Male	Princeton University	Math	NRG
Degond	Pierre	Male	CNRS	Math	FP
Degras	David	Male	SAMSI	Stat	NRG
Despatche	Ahmad	Male	NCSU	Stat	NRG
Diaz Espinosa	Oliver	Male	SAMSI	Math	NRG
Dobra	Adrian	Male	University of Washington	Stat	FP
D'Souza	Raissa	Female	University of California, Davis	Phys	FP
Dukic	Vanja	Female	University of Colorado	Stat	FP
Durrett	Rick	Male	Cornell University	Math	FP
Falk	Raymond	Male	OptInference LLC	Stat	FP
Fienberg	Steve	Male	Harvard University	Stats	FP
Fisher	Jake	Male	Duke/Sociology	Socl	NRG
Freeze	Michael	Male	University of North Carolina, Wilmington	Math	FP
Fuller	Pamela	Female	Rensselaer Polytechnic Institute	Math	NRG
Galati	Matthew	Male	SAS	Engg	NRG
Gharibans	Linda	Female	University of California, Los Angeles	Stats	NRG
Girvan	Michelle	Female	University of Maryland	Phys	NRG
Gong	Sam	Male	Duke	Physics	NRG

Gordillo	Luis	Male	University of Puerto Rico	Life	FP
Gordon-Wright	Rachael	Female	North Carolina State University	Math	NRG
Graham	Fan Chung	Female	University of California, San Diego	Math	FP
Gromoll	Christian	Male	University of Virginia	Math	FP
Gross	Justin	Male	University of North Carolina	Socl	NRG
Gutfraind	Alexander	Male	Center for Nonlinear Studies	Math	NRG
Haas	Bertrand	Male	Harvard University	Stat	NRG
Harer	John	Male	Duke University	Biology	FP
Hauck	Cory	Male	Oak Ridge National Laboratory	Math	NRG
Higham	Desmond	Male	University of Strathclyde	Math	FP
Ipsen	Ilse	Female	North Carolina State University	Math	FP
Jin	Jiashun	Male	Carnegie Mellon University	Stats	FP
Joo	Jaewook	Male	University of Tennessee	Phys	FP
Joyner	Michele	Female	East Tennessee State University	Math	FP
Jung	Sung Kyu	Male	University of North Carolina	Stat	NRG
Kabul	Mustafa	Male	SAS	Math	NRG
Kang	Min	Female	North Carolina State University	Math	FP
Kang	Emily	Female	SAMSI	Stat	NRG
Katenka	Natallia	Female	Boston University	Stat	NRG
Katzoff	Myron	Male	National Center for Health Statistics	Stats	FP
Khan	Taufiquar	Male	Clemson University	Math	FP

Kolaczyk	Erick	Male	Boston University	Stat	FP
Kolar	Mladen	Male	Carnegie Mellon University	Comp Sci	NRG
Kolba	Tiffany	Female	Duke	Math	NRG
Kong	Dehan	Male	NCSU	Stats	NRG
Kramer	Peter	Male	Rensselaer Polytechnic Institute	Math	FP
Kurganov	Alexander	Male	Tulane University	Math	FP
Lahiri	Soumen	Male	Texas A & M University	Stats	FP
Lee	Seonjoo	Female	UNC/Student Stat Dept	Stat	NRG
Lenarcic	Alan	Male	UNC/Genetics	Stat	NRG
Levina	Liza	Female	University of Michigan	Stat	FP
Li	Hongzhe	Male	University of Pennsylvania	Stat	FP
Li	Yingbo	Female	Duke University	Stat	NRG
Li	Lexin	Male	NCSU	Stat	FP
Li	Junchi	Male	Duke/Math	Math	NRG
Lin	Ja-an	Female	University of North Carolina	Biostats	NRG
Lin	Guang	Male	Pacific Northwest National Laboratory	Math	NRG
Lin	Xiaodong	Male	Rutgers University	Stat	FP
Linkletter	Crystal	Female	Brown University	Stats	FP
Lioa	Yi	Male	SAS	Stat	FP
Liu	Yufeng	Male	University of North Carolina	Stat	FP
Liu	Xin	Female	University of North Carolina	Stat	NRG

Lloyd	Alun	Male	North Carolina State University	Math	FP
Lynch	Jim	Male	University of South Carolina	Stats	FP
Maggioni	Mauro	Male	Duke University	Stat	FP
Mahoney	Michael	Male	Stanford University	Stat	FP
Makowski	Armand	Male	University of Maryland	Comp	FP
Marron	J. S.	Male	University of North Carolina	Stat	FP
Mattingly	Jonathan	Male	Duke University	Math	FP
McCormick	Tyler	Male	Columbia University	Stats	NRG
McSweeney	John	Male	Concordia University	Math	NRG
Meza	Rafael	Male	University of British Columbia	Math	NRG
Michailidis	George	Male	University of Michigan	Stat	FP
Miller	Joel	Male	Harvard School of Public Health	Math	NRG
Mitra	Ritendranath		MD Anderson Cancer Centre	Stat	FP
Moody	James	Male	Duke University	Socl	FP
Morris	Martina	Female	University of Washington	Stat	FP
Mucha	Peter	Male	University of North Carolina	Math	FP
Murrugarra Tomairo	David	Male	Virginia Tech/ Virginia Bioinformatics Institute	Math	NRG
Nesterko	Sergiy	Male	Harvard University	Stat	NRG
Newhall	Katie	Female	Rensselaer Polytechinc Institute	Math	NRG
Newman	Mark	Male	University of Michigan	Math	FP
Nolen	James	Male	Duke University	Math	FP

Pal Majumber	Abhishek	Male	UNC	Stat	NRG
Pang	Herbert	Male	Duke University	Stat	NRG
Perry	Patrick	Male	Harvard University	Stat	NRG
Poggi-Corradini	Pietro	Male	Kansas State University	Math	FP
Porter	Mason	Male	University of Oxford	Math	FP
Redner	Sidney	Male	Boston University	Phys	FP
Reinhold	Dominik	Male	University of North Carolina	Stat	NRG
Reyes	Perla	Female	University of Wisconsin-Madison	Stat	NRG
Rinaldo	Alessandro	Male	Carnegie Mellon University	Stats	NRG
Robert	Michael	Male	North Carolina State University	Math	NRG
Rodriguez	Abel	Male	University of California, Santa Cruz	Stat	NRG
Rogers	Bruce	Male	SAMSI	Math	NRG
Rohe	Karl	Male	University of California, Berkeley	Stats	NRG
Rombach	Michaela	Female	University of Oxford	Math	NRG
Salganik	Matthew	Male	Princeton Unviersity	Socl	NRG
Sang	Hailin	Male	National Institute of Statistical Sciences	Stats	NRG
Schmidler	Scott	Male	Duke University	Stats	NRG
Scholtens	Denise	Female	Northwestern University Medical School	Stat	FP
Schutt	Rachel	Female	Google Research	Stat	NRG
Shafiei	Mahdi	Male	Acadia University	Comp	NRG
Shalizi	Cosma	Male	Carnegie Mellon University	Stat	FP

Shi	Feng	Male	University of North Carolina	Math	NRG
Shojaie	Ali	Male	University of Michigan	Stat	NRG
Si	Yajuan	Female	Duke/Stat	Stat	NRG
Simpson	Sean	Male	Wake Forest University School of Medicine	Stat	NRG
Singh	Munindar	Male	North Carolina State University	Comp	FP
Sivakoff	David	Male	SAMSI	Math	NRG
Smith	Jeffrey	Male	Duke	Socl	NRG
Snijders	Tom	Male	University of Oxford	Stat	FP
Socolar	Joshua	Male	Duke University	Physics	FP
Soufiani	Hossein Azari	Male	Harvard University	Engg	NRG
Spencer	Bruce	Male	Northwestern University	Stat	FP
Spickenheuer	Anne	Female	IPA	Stat	NRG
Sullivan	Vida	Female	Oak Ridge National Laboratory	Math	NRG
Sun	Yi	Male	SAMSI	Math	NRG
Svihula	Judie	Female	University of North Carolina	Socl	FP
Talih	Makram	Male	CUNY School of Public Health, Hunter College	Stat	FP
Thomas	Andrew	Male	Carnegie Mellon University	Stats	NRG
Toroczkai	Zoltan	Male	University of Notre Dame	Phys	FP
Traud	Amanda	Female	North Carolina State University	Math	NRG
van Haren	Ken	Male	Duke University	Stat	NRG

Vaughan	Joel	Male	University of Michigan	Stat	NRG
Verella	J. Tipan	Male	University of Virginia	Engg	NRG
Vespignani	Alessandro	Male	Indiana University	Stat	FP
Vivar	Juan	Male	Duke University	Stats	NRG
Volz	Erik	Male	University of Michigan	Life	NRG
Wang	Shouqiang	Male	Duke University	Stat	NRG
Wang	Simi	Female	University of North Carolina	Math	NRG
Wang	Xueying	Female	SAMSI	Math	NRG
Wang	Xiaojing	Female	Duke University	Stat	NRG
Wang	Xuan	Male	UNC	Stat	NRG
Wang	Jianqiang	Male	NISS	Stat	NRG
Westveld	Anton	Male	University of Nevada	Stat	NRG
Willett	Rebecca	Female	Duke University	Engg	FP
Williams	Clarisa	Female	Auburn University	Stat	NRG
Wu	Chih-Da	Female	University of North Carolina	Biostats	NRG
Wu	Yichao	Male	North Carolina State University	Stats	NRG
Xing	Eric	Male	Carnegie Mellon University	Comp Sci	FP
Yang	Xiaolin	Female	Carnegie Mellon University	Stats	NRG
Yang	Xongxia	Female	Duke/Stat	Stat	NRG
Zhang	Helen	Female	NCSU/Faculty Stat Dept	Stat	FP
Zhang	Yuan	Male	Duke/Math	Math	NRG

Zhao	Ou	Male	University of South Carolina	Stats	NRG
Zhao	Yunpeng	Male	University of Michigan	Stat	NRG
Zhao	Yi	Female	Duke University	Stat	NRG
Zheng	Tian	Female	Columbia University	Stats	FP
Zheng	Lingling	Female	Duke University	Stat	NRG
Zhou	Hua	Male	NCSU	Stats	NRG
Zhu	Ji	Male	University of Michigan	Stat	FP
Zhu	Hongtu	Male	University of North Carolina	Biostats	FP
Zhu	Hongjie	Male	North Carolina State University	Stat	NRG
Zhuang	Lili	Female	Ohio State University	Stats	NRG

Complex Networks Modeling Workshop
Participant Summary
October 20-22, 2010

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	7	5	0	5	7	10	1	1	10	9
Unsuppted	26	10	0	15	21	24	7	5	13	4
SAMSI	5	1	0	1	5	1	5	0		

Complex Networks Modeling Workshop
Workshop Participants Support
October 20-22, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Fienberg	Stephen	Male	Carnegie Mellon University	Stat	FP
Krivitsky	Pavel	Male	Carnegie Mellon / Instituto Superior	Stat	NRG

			TÃ©cnico		
Matache	Mihaela Teodora (Dora)	Female	University of Nebraska	Math	FP
Michailidis	George	Male	University of Michigan	Stat	FP
Robinson	Lucy	Female	Johns Hopkins University	Stat	NRG
Rohe	Karl	Male	University of California, Berkeley	Stat	NRG
Simpson	Sean	Male	Wake Forest University School of Medicine	Stat	NRG
Singh	Aarti	Female	Carnegie Mellon University	Engg	NRG
Spencer	Bruce	Male	Northwestern University	Stat	FP
Vance	Eric	Male	Virginia Tech	Stat	NRG
Williams	Clarisa	Female	Auburn University	Stat	NRG
Zheng	Tian	Female	Columbia University	Stat	FP

Complex Networks Modeling Workshop

Workshop Participants
October 20-22, 2010

Last Name	**First Name**	**Gender**	**Affiliation**	**Major/Department**	**Status**
Airoldi	Edoardo	Male	Harvard University	Stat	NRG
Austin	Andrea	Female	Brown University	Stat	NRG
Balachandran	Prakash	Male	Duke University	Stat	NRG
Banks	David	Male	Duke University	Stat	FP
Blocher	Jesse	Male	UNC- Kenan-Flagler Business School	Other	NRG

Chintakunta	Harish	Male	North Carolina State Universtiy	Engg	NRG
Diaz	Oliver	Male	SAMSI	Math	NRG
Durrett	Rick	Male	SAMSI / Duke University	Stat	FP
Falk	Raymond	Male	OptInference LLC	Stat	FP
Fienberg	Stephen	Male	Carnegie Mellon University	Stat	FP
Gile	Krista	Female	University of Massachusetts	Stat	NRG
Gremaud	Pierre	Male	SAMSI & NCSU	Math	FP
Harer	John	Male	Duke University	Biology	FP
Kang	Emily	Female	SAMSI & NCSU	Stat	NRG
Khan	Taufiquar	Male	Clemson University	Math	FP
Kolaczyk	Eric	Male	Boston University	Stat	FP
Krivitsky	Pavel	Male	Carnegie Mellon University / Instituto Superior TÃ©cnico	Stat	NRG
Lee	Nam	Male	Johns Hopkins University	Math	NRG
Lee	Seonjoo	Female	University of North Carolina	Stat	NRG
Li	Yingbo	Female	Duke University	Stat	NRG
Linkletter	Crystal	Female	Brown University	Stat	FP
Lloyd	Alun	Male	North Carolina State U	Math	FP
Lynch	Jim	Male	University of South Carolina	Stat	FP
Matache	Mihaela Teodora (Dora)	Female	University of Nebraska	Math	FP
McCormick	Tyler	Male	Columbia University	Stat	NRG
Michailidis	George	Male	University of Michigan	Stat	FP

Mucha	Peter	Male	University of North Carolina	Math	FP
Nettel-Aguirre	Alberto	Male	U of Calgary/Alberta Children's Hospital	Stat	FP
Ratmann	Oliver	Male	Duke University	Stat	NRG
Robert	Michael	Male	SAMSI / North Carolina State University	Math	NRG
Robin	Stephane	Male	Agro Paris Tech	Stat	FP
Robinson	Lucy	Female	Johns Hopkins University	Stat	NRG
Rogers	Bruce	Male	SAMSI	Math	NRG
Rohe	Karl	Male	University of California, Berkeley	Stat	NRG
Sarkar	Purnamitra	Female	University of California, Berkeley	Comp	NRG
Schmidler	Scott	Male	Duke University	Stat	NRG
Scholtens	Denise	Female	Northwestern University Medical School	Stat	FP
Shi	Bill	Male	University of North Carolina	Math	NRG
Simpson	Sean	Male	Wake Forest University School of Medicine	Stat	NRG
Singh	Aarti	Female	Carnegie Mellon University	Engg	NRG
Sivakoff	David	Male	SAMSI/Duke University	Math	NRG
Spencer	Bruce	Male	Northwestern University	Stat	FP
Sun	Yi	Male	SAMSI	Math	NRG
Thomas	Andrew	Male	Carnegie Mellon University	Stat	NRG
Traud	Amanda	Female	North Carolina State University	Math	NRG
Van Haren	Ken	Male	Duke University	Stat	NRG

Vance	Eric	Male	Virginia Tech	Stat	NRG
Viles	Wes	Male	Boston University	Stat	NRG
Wang	Tian	Male	North Carolina State University	Socl	NRG
Williams	Clarisa	Female	Auburn University	Stat	NRG
Yang	Hyuna	Female	Duke University	Stat	FP
Yu	Bin	Female	University of California, Berkeley	Stat	FP
Zhang	Bo	Male	North Carolina State University	Stat	NRG
Zheng	Tian	Female	Columbia University	Stat	FP

Complex Networks Dynamics of Networks Workshop
Participant Summary
January 10-12, 2011

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	10	4	0	4	10	5	3	6	12	7
Unsuppted	16	3	0	12	7	7	7	5	13	5
SAMSI	4	0	0	1	3	0	4	0		

Complex Networks Dynamics of Networks Workshop
Workshop Participants Support
January 10-12, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Albert	Reka	Female	Pennsylvania State University	Life	FP
Bassett	Dani	Female	University of California, Santa Barbara	Phys	NRG
Bender-deMoll	Skye	Male		Socl	FP
Choi	Jihyeok	Male	Iowa State University	Math	NRG

Danson	Rachel	Female	University of California, Los Angeles	Math	NRG
Fienberg	Steve	Male	Cornell University	Stat	FP
Gross	Elizabeth	Female	University of Illinois, Chicago	Math	NRG
Hoff	Peter	Male	University of Washington	Stat	FP
Krivitsky	Pavel	Male	Carnegie Mellon University	Stat	NRG
Lupu	Yonatan	Male	University of California, San Diego	Socl	NRG
Onnela	JP	Male	Harvard University	Comp	NRG
Rolls	David	Male	University of Melbourne	Stat	NRG
Shafiei	Mahdi	Male	Acadia University	Comp	NRG
Sharpnack	James	Male	Carnegie Mellon University	Stat	NRG

Complex Networks Dynamics of Networks Workshop

Workshop Participants
January 10-12, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Albert	Reka	Female	Pennsylvania State University	Life	FP
Bassett	Dani	Female	University of California, Santa Barbara	Phys	NRG
Bender-deMoll	Skye	Male		Socl	FP
Bhamidi	S Shankar	Male	University of North Carolina	Stat	NRG
Bianconi	Ginestra	Female	Northeastern University	Stat	FP
Butts	Carter	Male	University of California, Irvine	Socl	FP
Carvalho	Luis	Male	Boston University	Stat	NRG

Choi	Jihyeok	Male	Iowa State University	Math	NRG
Danson	Rachel	Female	University of California, Los Angeles	Math	NRG
Durrett	Rick	Male	Duke University	Math	FP
Fairweather	Lindon	Male	US Treasury/Comptroller of the Currency	Other	FP
Fienberg	Steve	Male	Cornell University	Stat	FP
Gremaud	Pierre	Male	SAMSI / NCSU	Math	FP
Gross	Elizabeth	Female	University of Illinois, Chicago	Math	NRG
Hoff	Peter	Male	University of Washington	Stat	FP
Kolaczyk	Eric	Male	Boston University	Stat	FP
Krivitsky	Pavel	Male	Carnegie Mellon University	Stat	NRG
Lupu	Yonatan	Male	University of California, San Diego	Socl	NRG
Moody	Jim	Male	Duke University	Socl	FP
Mucha	Peter	Male	University of North Carolina	Math	FP
Murphy	Brendan	Male	University College Dublin	Stat	FP
Onnela	JP	Male	Harvard University	Comp	NRG
Porter	Mason	Male	University of Oxford	Math	FP
Robert	Michael	Male	North Carolina State University / SAMSI	Math	NRG
Rogers	Bruce	Male	SAMSI	Math	NRG
Rolls	David	Male	University of Melbourne	Stat	NRG
Schmutte	Ian	Male	University of Georgia	Socl	NRG

Shafiei	Mahdi	Male	Acadia University	Comp	NRG
Sharpnack	James	Male	Carnegie Mellon University	Stat	NRG
Shi	Feng	Male	University of North Carolina	Math	NRG
Sivakoff	David	Male	SAMSI / Duke	Math	NRG
Spickenheuer	Anne	Female	IPA	Stat	NRG
Sun	Yi	Male	SAMSI	Math	NRG
Traud	Amanda	Female	North Carolina State University	Math	NRG
Ward	Michael	Male	Duke University	Socl	FP
Wiuf	Carsten	Male	Aarhus University	Math	FP
Wu	Hulin	Male	University of Rochester	Stat	FP

Complex Networks Pedestrian Traffic Flow Workshop

Participant Summary
Febraury 14-16, 2011

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	4	0	0	4	2	0	4	0	3	3
Unsuppted	9	1	0	6	2	0	10	0	6	4
SAMSI	1	0	0	0	1	0	1	0		

Complex Networks Pedestrian Traffic Flow Workshop

Workshop Participants Support
February 14-16, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Kurganov	Alexander	Male	Tulane University	Math	FP
Panferov	Vladislav	Male	California State University	Math	NRG

Polizzi	Anthony	Male	Tulane University	Math	NRG
Timofeyev	Ilya	Male	University of Houston	Math	FP

Complex Networks Pedestrian Traffic Flow Workshop

Workshop Participants
February 14-16, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Chertock	Alina	Female	North Carolina State University	Math	FP
Degond	Pierre	Male	CNRS, Universités Paul Sabatier	Math	FP
Hauck	Cory	Male	Oak Ridge Nat'l Laboratory	Math	NRG
Kurganov	Alexander	Male	Tulane University	Math	FP
Liu	Jian-Guo	Male	Duke University	Math	FP
Motsch	Sebastien	Male	University of Maryland	Math	NRG
Panferov	Vladislav	Male	California State University	Math	NRG
Polizzi	Anthony	Male	Tulane University	Math	NRG
Ringhofer	Christian	Male	Arizona State University	Math	FP
Sun	Yi	Male	SAMSI	Math	NRG
Timofeyev	Ilya	Male	University of Houston	Math	FP

Complex Networks Dynamics On Networks Workshop

Participant Summary
March 21-23, 2011

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	9	5	0	7	7	2	7	5	14	6
Unsuppted	21	6	0	15	12	5	14	8	14	10

SAMSI	4	0	0	0	4	0	4	0		

Complex Networks Dynamics On Networks Workshop
Workshop Participants Support
March 21-23, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Almquist	Zack	Male	University of California, Irvine	Socl	NRG
Brummitt	Charlie	Male	University of California, Davis	Math	NRG
Brunson	Jason Cory	Male	Virginia Tech	Math	NRG
Colizza	Vittoria	Female	ISI and INSERM	Life	FP
Gleeson	James	Male	University of Limerick	Math	FP
Gromoll	Christian	Male	University of Virginia	Math	FP
Katenka	Natallia	Female	Boston University	Stat	NRG
Kramer	Peter	Male	Rensselaer Polytechnic Institute	Math	FP
Lanchier	Nicholas	Male	Arizona State University	Math	FP
Leicht	Elizabeth	Female	Oxford University	Phys	NRG
Meng	Xiangxiang	Male	University of Cincinnati	Stat	NRG
Redner	Sid	Male	Boston University	Phys	FP
Schmidt	Deena	Female	Ohio State University	Math	NRG
Solla	Sara	Female	Northwestern University	Life	FP

Complex Networks Dynamics On Networks Workshop
Workshop Participants
March 21-23, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Almquist	Zack	Male	University of California, Irvine	Socl	NRG
Balcan	Duygu	Female	University of Indiana	Phys	NRG
Bhamidi	Shankar	Male	University of North Carolina	Stat	NRG
Brummitt	Charlie	Male	University of California, Davis	Math	NRG
Brunson	Jason Cory	Male	Virginia Tech	Math	NRG
Chatterjee	Shirhsendu	Male	Cornell University	Math	NRG
Colizza	Vittoria	Female	ISI and INSERM	Life	FP
Diaz Espinosa	Oliver	Male	SAMSI-Duke	Math	NRG
Durrett	Rick	Male	SAMSI	Math	FP
Fisher	Jacob	Male	Duke University	Socl	NRG
Gleeson	James	Male	University of Limerick	Math	FP
Gromoll	Christian	Male	University of Virginia	Math	FP
Katenka	Natallia	Female	Boston University	Stat	NRG
Kelly	Daniel	Male	SAS Institute	Comp	FP
Kolaczyk	Eric	Male	Boston University	Stat	FP

Kramer	Peter	Male	Rensselaer Polytechnic Institute	Math	FP
Kranton	Rachel	Female	Duke University	Econ	FP
Lanchier	Nicholas	Male	Arizona State University	Math	FP
Leicht	Elizabeth	Female	Oxford University	Phys	NRG
Li	Junchi	Male	Duke University	Math	NRG
Lloyd	Alun	Male	North Carolina State University	Math	FP
Lynch	Jim	Male	University of South Carolina	Stat	FP
Makowski	Armand	Male	University of Maryland	Comp	FP
McSweeney	John	Male	Concordia University	Math	NRG
Meng	Xiangxiang	Male	University of Cincinnati	Stat	NRG
Michailidis	George	Male	University of Michigan	Stat	FP
Miller	Joel	Male	Harvard University	Math	NRG
Moody	James	Male	Duke University	Socl	FP
Mucha	Peter	Male	University of North Carolina	Math	FP
Porter	Mason	Male	Oxford University	Math	FP
Redner	Sid	Male	Boston University	Phys	FP
Robert	Michael	Male	North Carolina State University / SAMSI	Math	NRG
Rogers	Bruce	Male	SAMSI	Math	NRG
Rombach	Puck	Female	Oxford University	Math	NRG
Schmidt	Deena	Female	Ohio State University	Math	NRG
Sethuraman	Sunder	Male	Iowa State University	Math	FP

Sethuraman	Jayaram	Male	Florida State University	Stat	FP
Shapiro	Michael	Male	Tufts / Duke	Life	FP
Shi	Feng	Male	University of North Carolina	Math	NRG
Sivakoff	David	Male	SAMSI	Math	NRG
Solla	Sara	Female	Northwestern University	Life	FP
Sun	Yi	Male	SAMSI	Math	NRG
Thielbar	Melinda	Female	SAS Institute	Engg	FP
Traud	Amanda	Female	North Carolina State University	Math	NRG
Wang	Simi	Female	University of North Carolina	Math	NRG

Complex Networks Transition Workshop

Participant Summary
June 6-7, 2011

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	10	2	0	0	12	9	1	2	9	8
Unsuppted	17	3	0	8	12	8	9	3	10	6
SAMSI	3	1	0	0	4	0	4	0		

Complex Networks Transition Workshop

Workshop Participants Support
June 6-7, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Azari Soufiani	Hossein	Male	Harvard University	Engg	NRG
Bhattacharyya	Sharmodeep	Male	University of California, Berkeley	Stat	NRG
Caceres	Rajmonda	Female	University of Illinois, Chicago	Comp	NRG

Csardi	Gabor	Male	Harvard University	Stat	NRG
D'Amour	Alexander	Male	Harvard University	Stat	NRG
Franks	Alexander	Male	Harvard University	Stat	NRG
Jin	Ick Hoon	Male	Texas A&M University	Stat	NRG
Kupaev	Peter	Male	University of Notre Dame	Math	NRG
McCormick	Tyler	Male	Columbia University	Stat	NRG
Ramirez	Lilia	Female	Waterloo University	Stat	NRG
Simpson	Sean	Male	Wake Forest University	Stat	NRG
Zhao	Yunpeng	Male	University of Michigan	Stat	NRG

Complex Networks Transition Workshop

Workshop Participants

June 6-7, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Athreya	Avanti	Female	SAMSI	Math	NRG
Azari Soufiani	Hossein	Male	Harvard University	Engg	NRG
Balachandrian	Kash	Male	Duke University	Stat	NRG
Banks	David	Male	Duke University	Stat	FP
Bhattacharyya	Sharmodeep	Male	University of California, Berkeley	Stat	NRG
Blocher	Jesse	Male	University of North Carolina	Other	NRG
Caceres	Rajmonda	Female	University of Illinois, Chicago	Comp	NRG
Csardi	Gabor	Male	Harvard University	Stat	NRG

D'Amour	Alexander	Male	Harvard University	Stat	NRG
Dowling Sullivan	Blair	Female	Oakridge National Laboratories	Math	NRG
Durrett	Rick	Male	Duke University	Math	FP
Franks	Alexander	Male	Harvard University	Stat	NRG
Jin	Ick Hoon	Male	Texas A&M University	Stat	NRG
Kolaczyk	Eric	Male	Boston University	Stat	FP
Kupaev	Peter	Male	University of Notre Dame	Math	NRG
Lenarcic	Alan	Male	University of North Carolina	Stat	NRG
Li	Junchi	Male	Duke University	Math	NRG
Lin	Xiaodong	Male	Rutgers University	Stat	FP
Lloyd	Alun	Male	North Carolina State University	Math	FP
Maggioni	Mauro	Male	Duke University	Stat	FP
Mattingly	Jonathan	Male	Duke U	Math	FP
McCormick	Tyler	Male	Columbia University	Stat	NRG
Milledge	Gaolin	Female	North Carolina Central University	Comp	NRG
Mucha	Peter	Male	University of North Carolina	Math	FP
Ramirez	Lilia	Female	Waterloo University	Stat	NRG
Robert	Michael	Male	North Carolina State University	Math	NRG
Rogers	Bruce	Male	SAMSI	Math	NRG
Shi	Bill	Male	University of North Carolina	Math	NRG
Shojaie	Ali	Male	University of Michigan	Stat	NRG

Simpson	Sean	Male	Wake Forest University	Stat	NRG
Sivakoff	David	Male	SAMSI	Math	NRG
Sun	Yi	Male	SAMSI	Math	NRG
Traud	Amanda	Female	University of North Carolina	Math	NRG
Verella	Tipan	Male	University of Virginia	Engg	NRG
Zhao	Yunpeng	Male	University of Michigan	Stat	NRG
Zhu	Hongjie	Male	North Carolina State University	Stat	NRG

❖ Analysis of Object Data

Analysis of Object Data Opening Workshop
Participant Summary
September 12-15, 2010

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	20	8	1	10	19	27	2	0	21	12
Unsuppted	93	40	6	71	63	111	13	15	47	15
SAMSI	7	2	0	1	8	9	0	0		

Analysis of Object Data Opening Workshop
Workshop Participants Support
September 12-15, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Bharath	Karthik	Male	University of Connecticut	Stat	NRG
Chang	Yaw	Male	University of North Carolina, Wilmington	Math	FP
Chen	Kehui	Female	University of California, Davis	Stat	NRG
Ellingson	Leif	Male	Florida State University	Stat	NRG

Gervini	Daniel	Male	University of Wisconsin, Milwaukee	Stat	FP
Ghosh	Kaushik	Male	University of Nevada, Las Vegas	Stat	FP
Gottlieb	Andrea	Female	University of California, Davis	Stat	NRG
Guillas	Serge	Male	University College London	Stat	FP
He	Jinjiang	Male	University of California, Davis	Stat	NRG
Huzurbazar	Snehalata	Female	University of Wyoming	Stat	FP
Ivanescu	Andrada	Female	East Carolina University	Stat	NRG
Kim	Yongdai	Male	Seoul National University	Stat	FP
Kurtek	Sebastian	Male	Florida State University	Stat	NRG
Liu	Ran	Female	University of Connecticut	Math	NRG
Lu	Tao	Decline	University of Rochester	Stat	NRG
Ma	Junheng	Male	SAMSI	Stat	NRG
Marcy	Peter	Male	University of Wyoming / Los Alamos National Lab	Stat	NRG
Mas	Andre	Male	University Montpellier	Stat	FP
Ogden	Todd	Male	Columbia University	Stat	FP
Oluyede	Broderick	Male	Georgia Southern	Stat	FP
Osborne	Daniel	Male	Florida State University	Stat	NRG
Reiss	Philip	Male	New York University	Stat	NRG
Schwartzman	Armin	Male	Harvard School of Public Health	Stat	NRG

Wang	Lily	Female	University of Georgia	Stat	NRG
Wang	Yishi	Male	University of North Carolina, Wilmington	Stat	NRG
Wang	Yuan	Female	Colorado State University	Stat	NRG
Wu	Shuang	Female	University of Rochester	Stat	NRG
Wu	Ping-Shi	Male	Lehigh University	Stat	FP
Yang	Nuen Tsang	Male	University of California, Davis	Stat	NRG
Zhang	Lingsong	Male	Purdue University	Stat	NRG

Analysis of Object Data Opening Workshop

Workshop Participants

September 12-15, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Aston	John	Male	University of Warwick	Stat	FP
Avery	Matt	Male	North Carolina State University	Stat	NRG
Aydin	Burcu	Female	Hewlett-Packard Laboratories	Stat	NRG
Baladandayuthapani	Veera	Male	University of Texas M.D. Anderson Cancer Center	Stat	FP
Berrocal	Veronica	Female	University of Michigan	Stat	NRG
Bhamidi	Shankar	Male	University of North Carolina, Chapel Hill	Stat	NRG
Bharath	Karthik	Male	University of Connecticut	Stat	NRG
Bian	Xiao	Male	NCSU	Stat	NRG
Boryson	Petro	Male	UNC	Stat	NRG
Bowman	DuBois	Male	Emory University	Stat	FP

Bowman	Adrian	Male	University of Glasgow	Stat	FP
Brunel	Nicolas	Male	ENSIIE	Stat	FP
Cabanski	Chris	Male	UNC	Stat	NRG
Campbell	David	Male	Simon Fraser University	Stat	NRG
Chang	Yaw	Male	University of North Carolina, Wilmington	Math	FP
Chen	Kehui	Female	University of California, Davis	Stat	NRG
Chen	Lu-Hung	Male	UNC	Math	NRG
Chiou	Jeng-Min	Male	Academia Sinica	Stat	FP
Chow	Sy-Miin	Female	University of North Carolina	Psych	FP
Conner	Emil	Male	UNC		
Crainiceanu	Ciprian	Male	Johns Hopkins University	Stat	FP
Damon	James	Male	University of North Carolina	Math	FP
Dass	Sarat	Male	Michigan State University	Stat	FP
Degras	David	Male	SAMSI	Stat	NRG
Dey	Dipak	Male	University of Connecticut	Stat	FP
Di	Chongzhi	Male	Fred Hutchinson Cancer Center	Stat	NRG
Ding	Jimin	Female	Washington Universtiy, St. Louis	Stat	NRG
Dryden	Ian	Male	University of Nottingham	Stat	FP
Dunson	David	Male	Duke University	Stat	FP
Ehintalcunta	Haoish		NCSU		
Ellingson	Leif	Male	Florida State University	Stat	NRG

Feng	Xingdong	Male	NISS	Stat	NRG
Gamble	Jennifer	Female	NCSU	Engg	NRG
Gervini	Daniel	Male	University of Wisconsin, Milwaukee	Stat	FP
Ghosh	Kaushik	Male	University of Nevada, Las Vegas	Stat	FP
Ghoshal	Subhashis	Male	North Carolina State University	Stat	FP
Girolami	Mark	Male	University of Glasgow	Stat	FP
Gottlieb	Andrea	Female	University of California, Davis	Stat	NRG
Guillas	Serge	Male	University College London	Stat	FP
Han	Lu	Female	NCSU	Engg	NRG
Hannig	Jan	Male	University of North Carolina, Chapel Hill	Stat	FP
He	Jinjiang	Male	University of California, Davis	Stat	NRG
Heo	Giseon	Female	University of Alberta	Stat	FP
Holan	Scott	Male	University of Missouri	Stat	FP
Hooker	Giles	Male	Cornell University	Stat	NRG
Hotz	Thomas	Male	Georgia Augusta University of Goettingen	Stat	NRG
Huang	Hanwen	Male	University of North Carolina, Chapel Hill	Stat	NRG
Huang	Xianzheng	Female	University of South Carolina	Stat	NRG
Huay	De		Duke		
Huckemann	Stephan	Male	University of Goettingen	Stat	FP
Huzurbazar	Snehalata	Female	University of Wyoming	Stat	FP

Ivanescu	Andrada	Female	East Carolina University	Stat	NRG
Jiang	Ci-Ren	Male	SAMSI	Stat	NRG
Joshi	Sarang	Male	University of Utah	Engg	FP
Jung	Sungkyu	Male	University of North Carolina, Chapel Hill	Stat	NRG
Kent	John	Male	University of Leeds	Stat	FP
Kim	Peter	Male	University of Guelph	Stat	NRG
Kim	Yongdai	Male	Seoul National University	Stat	FP
King	Aaron	Male	University of Michigan	Math	FP
Kneip	Alois	Male	University of Bonn	Stat	FP
Kogan	Irina	Female	North Carolina Sate University	Math	FP
Kong	Dehan	Male	NCSU	Stat	NRG
Krim	Hamid	Male	North Carolina State University	Engg	FP
Kumna	Amit		UNC		
Kurtek	Sebastian	Male	Florida State University	Stat	NRG
Le	Huiling	Female	University of Nottingham	Stat	FP
Lee	Seonjoo	Female	University of North Carolina, Chapel Hill	Stat	NRG
Lee	Jaeyong	Male	Seoul National University	Stat	FP
Lee	Deokwoo	Male	NCSU	Engg	NRG
Li	Yao	Female	West Virginia University	Stat	NRG
Li	Fan	Female	Duke University	Stat	NRG
Li	Lexin	Male	North Carolina State University	Stat	FP

Li	Xiaoshan	Male	NCSU	Stat	NRG
Lin	Min	Female	Duke	Stat	FP
Lindquist	Martin	Male	Columbia University	Stat	FP
Liu	Yufeng	Male	University of North Carolina, Chapel Hill	Stat	FP
Liu	Ran	Female	University of Connecticut	Math	NRG
Liu	Lan		UNC	Stat	NRG
Lock	Eric	Male	University of North Carolina	Stat	NRG
Lu	Tao	Decline	University of Rochester	Stat	NRG
Lu	Xiaosun	Female	University of North Carolina, Chapel Hill	Stat	NRG
Lu	Wenbin	Male	North Carolina State University	Stat	FP
Ma	Junheng	Male	SAMSI	Stat	NRG
Macaulay	Vincent	Male	University of Glasgow	Stat	FP
Marcy	Peter	Male	University of Wyoming / Los Alamos National Lab	Stat	NRG
Mardia	Kanti	Male	University of Leeds	Stat	FP
Marron	J. S.	Male	University of North Carolina	Stat	FP
Mas	Andre	Male	University Montpellier	Stat	FP
Miller	Ezra	Male	Duke University	Math	FP
Miranda	Michelle	Female	UNC	Stat	NRG
Moriarty	John	Male	University of Manchester	Math	FP
Morris	Jeffrey	Male	University of Texas M. D. Anderson Cancer Center	Stat	FP

Müller	Hans	Male	University of California, Davis	Stat	FP
Munoz Maldonado	Yolanda	Female	Michigan Tech University	Stat	NRG
Ogden	Todd	Male	Columbia University	Stat	FP
Olhede	Sofia Charlotta	Female	University College London	Stat	FP
Oluyede	Broderick	Male	Georgia Southern	Stat	FP
Omolo	Bernard	Male	University of North Carolina	Stat	FP
Osborne	Daniel	Male	Florida State University	Stat	NRG
Owen	Megan	Female	North Carolina State University	Math	NRG
Panaretos	Victor	Male	Swiss Federal Institute of Technology	Stat	FP
Park	Juhyun	Female	Lancaster University	Stat	FP
Patrangenaru	Victor	Male	Florida State University	Stat	FP
Peng	Jie	Female	University of California at Davis	Stat	FP
Petralia	Francesca	Female	Duke	Stat	NRG
Pizer	Stephen	Male	University of North Carolina	Comp	FP
Preston	Simon	Male	University of Nottingham	Stat	NRG
Ramsay	Jim	Male	McGill University	Math/Stat	FP
Randolph	Tim	Male	Fred Hutchinson Cancer Research Center	Math	FP
Reiss	Philip	Male	New York University	Stat	NRG
Ruymgaart	Frits	Male	Texas Tech University	Stat	FP
Samworth	Richard	Male	University of Cambridge	Stat	FP

Sang	Hailin	Male	NISS	Stat	NRG
Schmidler	Scott	Male	Duke University	Stat	FP
Schwartzman	Armin	Male	Harvard School of Public Health	Stat	NRG
Sedransk	Nell	Female	SAMSI / NISS	Stat	FP
Seiler	Christof	Male	University of Bern	Stat	NRG
Senturk	Damla	Female	Pennsylvania State University	Stat	FP
Shaby	Ben	Male	SAMSI	Stat	NRG
Shannon	William	Male	Washington University School of Medicine	Stat	FP
Shen	Dan	Male	University of North Carolina, Chapel Hill	Stat	NRG
Shen	Haipeng	Male	University of North Carolina, Chapel Hill	Stat	FP
Shin	SeungJun		North Carolina State University	Stat	NRG
Si	Yajuan	Female	Duke	Stat	NRG
Skwerer	Sean	Male	University of North Carolina, Chapel Hill	Stat	NRG
Smith	Richard	Male	SAMSI	Stat	FP
Song	Joon Jin	Male	University of Arkansas	Math	NRG
Srivastava	Anuj	Male	Florida State University	Stat	FP
Stadtmueller	Ulrich	Male	Ulm University	Stat	FP
Staicu	Ana-Maria	Female	North Carolina State University	Stat	NRG
Staneva	Valentina	Female	Johns Hopkins University	Math	NRG
Sun	Wei	Male	UNC	Stat	NRG
Tavakoli	Shahin	Male	Swiss Institute of Technology	Stat	NRG

Taylor	Jonathan	Male	Stanford University	Stat	NRG
Tchumtchoua	Sylvie	Female	SAMSI	Stat	NRG
Trouve	Alain	Male	Ecole Normale Supérieure de Cachan	Math	FP
Wang	Haonan	Male	Colorado State University	Stat	FP
Wang	Huixia Judy	Female	NCSU	Stat	NRG
Wang	Jane-Ling	Female	University of California, Davis	Stat	FP
Wang	Jianqiang	Male	NISS	Stat	NRG
Wang	Lily	Female	University of Georgia	Stat	NRG
Wang	Liwei	Female	NCSU	Stat	NRG
Wang	Naisyin	Female	Texas A&M University	Stat	FP
Wang	Tian	Male	NCSU	Socl	NRG
Wang	Yishi	Male	University of North Carolina, Wilmington	Stat	NRG
Wang	Yuan	Female	Colorado State University	Stat	NRG
Wang	Xia	Female	NISS	Stat	NRG
Wang	Xiao	Male	Purdue University	Stat	FP
Wang	Xiaohui	Female	University of Virginia	Stat	FP
Wang	Xiaojing	Female	Duke University	Stat	NRG
Whitaker	Ross	Male	University of Utah	Comp	FP
Wilkinson	Darren	Male	Newcastle University	Stat	FP
Wu	Shuang	Female	University of Rochester	Stat	NRG
Wu	Yichao	Male	North Carolina State University	Stat	NRG

Wu	Yuefeng	Male	SAMSI	Stat	NRG
Wu	Hulin	Male	University of Rochester	Stat	FP
Wu	Ping-Shi	Male	Lehigh University	Stat	FP
Wu	Chih-Da	Male	UNC	Stat	NRG
Xia	Jessie	Female	NISS	Stat	NRG
Xu	Christine	Female	UNC	Comp	NRG
Yang	Nuen Tsang	Male	University of California, Davis	Stat	NRG
Yang	Hongxia	Female	Duke University	Stat	NRG
Yao	Fang	Male	University of Toronto	Stat	FP
Ye	Han	Female	UNC	Stat	NRG
Yi	Sheng	Male	NCSU	Stat	NRG
Younes	Laurent	Male	Johns Hopkins University	Math	FP
Zeng	Donglin	Male	University of North Carolina	Stat	FP
Zhang	Bo	Male	North Carolina State University	Stat	NRG
Zhang	Hao (Helen)	Female	North Carolina State University	Stat	FP
Zhang	Lingsong	Male	Purdue University	Stat	NRG
Zhang	Tingting	Female	University of Virginia	Stat	NRG
Zhang	Jun	Male	SAMSI	Stat	NRG
Zhang	Chong	Male	UNC	Stat	NRG
Zhou	Frank	Male	NISS	Stat	NRG
Zhou	Mipin		Duke		

Zhu	Hongtu	Male	University of North Carolina	Biostats	FP
Zhu	Hongxiao	Female	University of Texas M. D. Anderson Cancer Center	Stat	NRG

Interface Functional and Longitudinal Data Analysis Workshop
Participant Summary
November 8-10, 2010

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	8	5	0	4	9	13	0	0	12	7
Unsuppted	37	19	1	27	30	54	1	2	26	9
SAMSI	4	1	0	0	5	5	0	0		

Interface Functional and Longitudinal Data Analysis Workshop
Workshop Participants Support
November 8-10, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Cao	Jiguo	Male	Simon Fraser University	Stat	FP
Cao	Guanqun	Female	Michigan State University	Stat	NRG
Guo	Wensheng	Male	University of Pennsylvania	Stat	FP
Kurum	Esra	Female	Penn State University	Stat	NRG
Li	Yimei	Female	St. Jude Children's Research Hospital	Stat	NRG
Li	Wei	Female	Rutgers University	Stat	NRG
Liu	Chong	Male	Boston University	Stat	NRG
Tom	Jennifer	Female	University of California, Los Angeles	Stat	NRG
Wu	Hulin	Male	University of Rochester	Stat	FP

Xue	Hongqi	Male	University of Rochester	Stat	FP
Zhang	Lingsong	Male	Purdue University	Stat	NRG
Zhang	Xiaoke	Male	University of California, Davis	Stat	NRG
Zipunnikov	Vadim	Male	Johns Hopkins School of Public Health	Stat	NRG

Interface Functional and Longitudinal Data Analysis Workshop

Workshop Participants
November 8-10, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Altman	Naomi	Female	Penn State University	Stat	FP
Boente	Graciela	Female	Universidad de Buenos Aires and CONICET	Stat	FP
Cao	Jiguo	Male	Simon Fraser University	Stat	FP
Cao	Guanqun	Female	Michigan State University	Stat	NRG
Carrillo-Garcia	Ivan	Male	NISS	Stat	NRG
Carroll	Raymond J.	Male	Texas A&M	Stat	FP
Chakraborty	Sounak	Male	University of Missouri-Columbia	Stat	FP
Degras	David	Male	SAMSI	Stat	NRG
Ding	Jimin	Female	Wash U, St. Louis	Stat	NRG
Du	Pang	Male	Virginia Tech	Stat	NRG
Feng	Xingdong	Male	NISS	Stat	NRG
Ghosh	Samiran	Male	Weill Cornell Medical College	Stat	NRG

Guo	Wensheng	Male	University of Pennsylvania	Stat	FP
Hays	Spencer	Male	University of North Carolina	Stat	NRG
Heo	Giseon	Female	University of Alberta	Stat	FP
Hooker	Giles	Male	Cornell U	Stat	NRG
Hsing	Tailen	Male	University of Michigan	Stat	FP
Janicki	Ryan	Male	U.S. Census Bureau	Stat	NRG
Jiang	Ci-Ren	Male	SAMSI	Stat	NRG
Jung	Sungkyu	Male	University of North Carolina	Stat	NRG
Kong	Dehan	Male	North Carolina State University	Stat	NRG
Kurum	Esra	Female	Penn State University	Stat	NRG
Lee	Seonjoo	Female	University of North Carolina	Stat	NRG
Li	Lexin	Male	North Carolina State University	Stat	FP
Li	Yimei	Female	St. Jude Children's Research Hospital	Stat	NRG
Li	Wei	Female	Rutgers University	Stat	NRG
Lin	Ja-An	Female	University of North Carolina	Stat	NRG
Liu	Chong	Male	Boston University	Stat	NRG
Ma	Junheng	Male	SAMSI	Stat	NRG
Marron	Steve	Male	University of North Carolina	Stat	FP
Morris	Jeff	Male	University of Texas	Stat	FP
Mueller	Hans-Georg	Male	University of California, Davis	Stat	FP
Munoz	Yolanda	Female	Michigan Tech	Stat	NRG

Panaretos	Victor	Male	Swiss Fed	Stat	FP
Peng	Jie	Female	University of California, Davis	Stat	FP
Pourahmadi	Mohsen	Male	Texas A&M University	Stat	FP
Ramsay	Jim	Male	McGill University	Math/Stat	FP
Reiss	Philip	Male	New York University	Stat	NRG
Samworth	Richard	Male	University of Cambridge	Stat	FP
Sang	Hailin	Male	NISS	Stat	NRG
Seier	Edith	Female	East Tennessee State University	Stat	FP
Seiler	Christof	Male	SAMSI	Stat	NRG
Senturk	Damla	Female	Penn State	Stat	FP
Shen	Haipeng	Male	UNC-Stat Faculty	Stat	FP
Shin	Sunyoung	Female	University of North Carolina	Stat	NRG
Staicu	Ana-Maria	Female	North Carolina State University	Stat	NRG
Sun	Jianguo (Tony)	Male	University of MIssouri	Stat	FP
Tchumtchoua	Sylvie	Female	SAMSI	Stat	NRG
Todem	David	Male	Michigan State University	Stat	FP
Tom	Jennifer	Female	University of California, Los Angeles	Stat	NRG
Toto	Criselda	Female	NISS	Stat	NRG
Wang	Jane-Ling	Female	UC Davis	Stat	FP
Wang	Xiaojing	Female	Duke University	Stat	NRG
Wang	Jianqiang	Male	NISS	Stat	NRG

Wang	Haonan	Male	Colorado State University	Stat	FP
Wierzbicki	Michael	Male	University of Pennsylvania	Stat	NRG
Wright	James	Male	Michigan Technological Univ	Stat	NRG
Wu	Yichao	Male	North Carolina State University	Stat	NRG
Wu	Yuefeng	Male	Cornell University	Stat	NRG
Wu	Hulin	Male	University of Rochester	Stat	FP
Wu	Shuang	Female	University of Rochester	Stat	NRG
Wu	Chih-Da	Male	UNC-BIOS Student	Stat	NRG
Xu	Christine	Female	UNC-Computer Sci	Comp	NRG
Xue	Hongqi	Male	University of Rochester	Stat	FP
Yao	Fang	Male	University of Toronto	Stat	FP
Zhang	Hao (Helen)	Female	North Carolina State University	Stat	FP
Zhang	Lingsong	Male	Purdue University	Stat	NRG
Zhang	Bo	Male	North Carolina State University	Stat	NRG
Zhang	Xiaoke	Male	University of California, Davis	Stat	NRG
Zhang	Donghui	Male	sanofi aventis	Comp	FP
Zhu	Hongtu	Male	University of North Carolina	Biostats	FP
Zhu	Hongxiao	Female	U Texas	Stat	NRG
Zhu	Min		SAS Institute	Stat	FP
Zhu	Bin	Male	University of Michigan	Stat	NRG
Zipunnikov	Vadim	Male	Johns Hopkins School of Public Health	Stat	NRG

AOOD Meets Evolutionary Biology Workshop

Participant Summary
April 30-May 2, 2011

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	9	3	0	7	5	6	1	5	9	6
Unsuppted	33	7	0	23	17	24	5	11	11	6
SAMSI	6	0	0	0	6	6	0	0		

AOOD Meets Evolutionary Biology Workshop

Workshop Participant Support
April 30-May 2, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Aston	John	Male	University of Warwick	Stat	FP
Carter	Pat	Male	Washington State University	Life	FP
Du	Pang	Male	Virginia Tech	Stat	NRG
Dworkin	Ian	Male	Michigan State University	Zoology	FP
Gervini	Daniel	Male	University of Wisconsin	Stat	FP
Hadjipantelis	Pantelis	Male	University of Warwick	Stat	NRG
Heckman	Nancy	Female	University of British Columbia	Stat	FP
Houle	David	Male	Florida State University	Biology	FP
Irwin	Kristen	Female	Washington State Univ	Life	NRG
Jamniczky	Heather	Female	University of Calgary	Life	NRG

Mio	Washington	Male	Florida State University	Math	FP
Singh	Sarabdeep	Male	University of Wyoming	Stat	NRG

AOOD Meets Evolutionary Biology Workshop
Workshop Participants
April 30-May 2, 2011

Last Name	**First Name**	**Gender**	**Affiliation**	**Major/Department**	**Status**
Aston	John	Male	University of Warwick	Stat	FP
Beder	Jay	Male	University of Wisconsin	Stat	FP
Carter	Pat	Male	Washington State University	Life	FP
Degras	David	Male	SAMSI	Stat	NRG
Du	Pang	Male	Virginia Tech	Stat	NRG
Dumancas	Gerard	Male	Oklahoma State University	Chem	NRG
Dworkin	Ian	Male	Michigan State University	Zoology	FP
Gervini	Daniel	Male	University of Wisconsin	Stat	FP
Gou	Kun	Male	Texas A&M University	Math	NRG
Hadjipantelis	Pantelis	Male	University of Warwick	Stat	NRG
Heckman	Nancy	Female	University of British Columbia	Stat	FP
Houle	David	Male	Florida State University	Biology	FP
Irwin	Kristen	Female	Washington State Univ	Life	NRG
Jamniczky	Heather	Female	University of Calgary	Life	NRG
Jiang	Ci-Ren	Male	SAMSI	Stat	NRG

Joshi	Sarang	Male	University of Utah	Engg	FP
Kingsolver	Joel	Male	University of North Carolina	Life	FP
Kong	Dehan	Male	North Carolina State University	Stat	NRG
Li	Gen	Male	University of North Carolina	Stat	NRG
Ma	Junheng	Male	SAMSI	Stat	NRG
Marron	J. S.	Male	University of North Carolina	Stat	FP
Miller	Ezra	Male	Duke University	Math	FP
Mio	Washington	Male	Florida State University	Math	FP
Munoz Maldonado	Yolanda	Female	Michigan Tech University / SAMSI	Stat	NRG
Panaretos	Victor	Male	EPFL	Stat	FP
Park	Juhyun	Female	Lancaster University	Stat	FP
Sen	Saunak	Male	University of California, San Francisco	Stat	FP
Singh	Sarabdeep	Male	University of Wyoming	Stat	NRG
Zhang	Bo	Male	North Carolina State University	Stat	NRG
Aston	John	Male	University of Warwick	Stat	FP
Beder	Jay	Male	University of Wisconsin	Stat	FP
Carter	Pat	Male	Washington State University	Life	FP
Degras	David	Male	SAMSI	Stat	NRG
Du	Pang	Male	Virginia Tech	Stat	NRG
Dumancas	Gerard	Male	Oklahoma State University	Chem	NRG
Dworkin	Ian	Male	Michigan State University	Zoology	FP

Gervini	Daniel	Male	University of Wisconsin	Stat	FP
Gou	Kun	Male	Texas A&M University	Math	NRG
Hadjipantelis	Pantelis	Male	University of Warwick	Stat	NRG
Heckman	Nancy	Female	University of British Columbia	Stat	FP
Houle	David	Male	Florida State University	Biology	FP
Irwin	Kristen	Female	Washington State Univ	Life	NRG
Jamniczky	Heather	Female	University of Calgary	Life	NRG
Jiang	Ci-Ren	Male	SAMSI	Stat	NRG
Joshi	Sarang	Male	University of Utah	Engg	FP
Kingsolver	Joel	Male	University of North Carolina	Life	FP
Kong	Dehan	Male	North Carolina State University	Stat	NRG
Li	Gen	Male	University of North Carolina	Stat	NRG
Ma	Junheng	Male	SAMSI	Stat	NRG
Marron	J. S.	Male	University of North Carolina	Stat	FP
Miller	Ezra	Male	Duke University	Math	FP
Mio	Washington	Male	Florida State University	Math	FP
Munoz Maldonado	Yolanda	Female	Michigan Tech University / SAMSI	Stat	NRG
Panaretos	Victor	Male	EPFL	Stat	FP
Park	Juhyun	Female	Lancaster University	Stat	FP
Sen	Saunak	Male	University of California, San Francisco	Stat	FP
Singh	Sarabdeep	Male	University of Wyoming	Stat	NRG

Zhang	Bo	Male	North Carolina State University	Stat	NRG

AOOD Transition Workshop
Participant Summary
June 9-11, 2011

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	9	0	0	3	6	9	0	0	7	6
Unsuppted	20	7	1	11	17	24	2	2	13	6
SAMSI	2	2	0	0	4	4	0	0		

AOOD Transition Workshop
Workshop Participant Support
June 9-11, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Campbell	David	Male	Simon Fraser	Stat	NRG
Di	Chongzhi	Male	Fred Hutchinson Cancer Research Center	Stat	NRG
Ellingson	Leif	Male	Florida State University	Stat	NRG
Jung	Sungkyu	Male	University of North Carolina	Stat	NRG
Kurtek	Sebastien	Male	Florida State University	Stat	NRG
Morris	Jeff	Male	MD Anderson	Stat	FP
Ombao	Hernando	Male	Brown University	Stat	FP
Wu	Hulin	Male	University of Rochester	Stat	FP
Yang	Nuen Tsang	Male	University of California, Davis	Stat	NRG

AOOD Transition Workshop
Workshop Participants
June 9-11, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Aston	John	Male	University of Warwick	Stat	FP
Bolivar	Addy	Female	University of North Carolina	Stat	NRG
Campbell	David	Male	Simon Fraser	Stat	NRG
Cao	Jiguo	Male	Simon Fraser	Stat	FP
Degras	David	Male	SAMSI	Stat	NRG
Di	Chongzhi	Male	Fred Hutchinson Cancer Research Center	Stat	NRG
Dryden	Ian	Male	University of South Carolina	Stat	FP
Dunson	David	Male	Duke University	Stat	FP
Ellingson	Leif	Male	Florida State University	Stat	NRG
Huang	Xinyu	Male	North Carolina Central University	Math	NRG
Jiang	Ci-Ren	Male	SAMSI	Stat	NRG
Jung	Sungkyu	Male	University of North Carolina	Stat	NRG
Kong	Dehan	Male	NCSU	Stat	NRG
Kurtek	Sebastien	Male	Florida State University	Stat	NRG
Lee	Seonjoo	Female	SAMSI	Stat	NRG

Li	Meng		NCSU	Stat	NRG
Marron	Steve	Male	University of North Carolina	Stat	FP
McLean	Mathew	Male	Cornell University	Engg	NRG
Milledge	Gaolin	Female	North Carolina Central University	Comp	NRG
Morris	Jeff	Male	MD Anderson	Stat	FP
Mueller	Hans	Male	University of California, Davis	Stat	FP
Ombao	Hernando	Male	Brown University	Stat	FP
Pati	Debdeep	Male	Duke University	Stat	NRG
Patrangenaru	Victor	Male	Florida State University	Stat	FP
Ramsay	Jim	Male	McGill University	Math/Stat	FP
Shen	Haipeng	Male	University of North Carolina	Stat	FP
Shen	Dan	Male	University of North Carolina	Stat	NRG
Skwerer	Sean	Male	University of North Carolina	Stat	NRG
Su	Jingyong	Male	Florida State University	Stat	NRG
Tchumtchoua	Sylvie	Female	SAMSI	Stat	NRG
Wang	Jane-Ling	Female	University of California, Davis	Stat	FP
Wang	Xia	Female	NISS	Stat	NRG
Wu	Hulin	Male	University of Rochester	Stat	FP
Wu	Yichao	Male	North Carolina State University	Stat	NRG
Yang	Nuen Tsang	Male	University of California, Davis	Stat	NRG
Yaraee	Kate	Female	University of Alberta	Stat	NRG

Zeng	Peng	Male	Auburn University	Stat	FP
Zhang	Tingting	Female	University of Virginia	Stat	NRG
Zhang	Lingsong	Male	Purdue University	Stat	NRG
Zhu	Hongxiao	Female	U Texas	Stat	NRG
Zou	Jian	Male	NISS	Stat	NRG

❖ **Education and Outreach Program**

Undergraduate Two-Day Workshop

Participant Summary

October 29-30, 2010

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Institutions Represented	Number of States Represented
Supported	8	18	0	0	26	25	1	19	13
Unsuppted	11	5	0	12	4	15	1	6	9
SAMSI	5	0	0	5	0	5	0		

Undergraduate Two-Day Workshop

Workshop Participant Support

October 29-30, 2010

Last Name	First Name	Gender	Affiliation	Major/ Department	Status
Brown-Cornell	Lauren	Female	UCSD	Math	S
Cameron	Sharon	Female	E Tenn State U	Math	S
Chandhiramouli	Vaishnavi	Female	UC Berkeley	Stat	S

Chang	Ellen	Female	Arizona State U	Math	S
Choi	Mary	Female	U Georgia	Stat	S
Christensen	Jonathan	Male	Brigham Young	Stat/Math	S
Chu	Michelle	Female	Emory	Math	S
Culiuc	Amalia	Female	Mt. Holyoke	Math	S
Dong	Xiaochen	Female	Wesleyan	Math	S
Fong	Ut (John)	Male	UC Berkeley	Math/Stat	S
Jackson	Ashley	Female	Wesleyan	Math	S
Jacobs	Abigail	Female	Northwestern	Math	S
Knebel	Robert	Male	UNC Willmington	Math	S
Lawton	Elizabeth	Female	UNC Willmington	Math	S
McMurrey	Ian	Male	Midwestern Stat	Comp	S

Messan	Yao	Male	NC A&T	Math	S
Natale	Sara	Female	Reed	Math	S
Orner	Daniel	Male	Emory	Math	S
Rodriguez	Terry	Male	UNC	Math	S
Sauerwald	Natalie	Female	UCSD	Math	S
Schiermeyer	Katie	Female	E Tenn State	Math	S
Strand	Kyle	Male	Colorado Stat	Math	S
Tan	Xin Lu	Female	Purdue	Math/Stat	S
Valeva	Silviya	Female	Mt. Holyoke	Math	S
Williams	Ivana	Female	Winston-Salem	Math	S
Zhu	Xiaoyu	Female	Carnegie Mellon	Stat	S

Undergraduate Two-Day Workshop
Workshop Participants
October 29-30, 2010

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Balachandrian	Kash	Male	Duke U	Stat	A
Banks	David	Male	Duke U	Stat	A
Brown-Cornell	Lauren	Female	UCSD	Math	S
Cameron	Sharon	Female	E Tenn State U	Math	S
Chandhiramouli	Vaishnavi	Female	UC Berkeley	Stat	S
Chang	Ellen	Female	Arizona State U	Math	S
Choi	Mary	Female	U Georgia	Stat	S
Christensen	Jonathan	Male	Brigham Young	Stat/Math	S
Chu	Michelle	Female	Emory	Math	S
Culiuc	Amalia	Female	Mt. Holyoke	Math	S
Diaz	Oliver	Male	SAMSI	Math	A
Dong	Xiaochen	Female	Wesleyan	Math	S
Durrett	Rick	Male	Duke U	Math	A
Fong	Ut (John)	Male	UC Berkeley	Math/Stat	S
Gremaud	Pierre	Male	SAMSI	Math	A
Horn	Corinne	Female	Duke	Engg	S
Jackson	Ashley	Female	Wesleyan	Math	S
Jacobs	Abigail	Female	Northwestern	Math	S
Jung	SungKyu	Male	UNC	Stat	A
Knebel	Robert	Male	UNC Willmington	Math	S

Lawton	Elizabeth	Female	UNC Willmington	Math	S
Li	Yingbo	Female	Duke	Stat	A
Maggioni	Mauro	Male	Duke	Stat	A
McMurrey	Ian	Male	Midwestern Stat	Comp	S
Messan	Yao	Male	NC A&T	Math	S
Natale	Sara	Female	Reed	Math	S
Orner	Daniel	Male	Emory	Math	S
Qian	Kenneth	Male	Arizona State	Math	S
Qian	Zongjin	Male	Duke	Math	S
Robert	Michael	Male	NCSU	Math	A
Rodriguez	Terry	Male	UNC	Math	S
Rogers	Bruce	Male	SAMSI	Math	A
Sauerwald	Natalie	Female	UCSD	Math	S
Schiermeyer	Katie	Female	E Tenn State	Math	S
Shi	Bill	Male	UNC	Math	A
Shojaie	Ali	Male	U Michigan	Stat	A
Singh	Kyra	Female	Carnegie Mellon	Stat	S
Sivakoff	David	Male	SAMSI	Math	A
Strand	Kyle	Male	Colorado Stat	Math	S
Sun	Yi	Male	SAMSI	Math	A
Tan	Xin Lu	Female	Purdue	Math/Stat	S

Traud	Amanda	Female	NCSU	Math	A
Valeva	Silviya	Female	Mt. Holyoke	Math	S
van Haren	Ken	Male	Duke	Stat	A
Williams	Ivana	Female	Winston-Salem	Math	S
Zhu	Xiaoyu	Female	Carnegie Mellon	Stat	S
Zhu	Hongxiao	Female	U Texas	Stat	A

Undergraduate Two-Day Workshop

Participant Summary

February 25-26, 2011

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Institutions Represented	Number of States Represented
Supported	14	12	0	0	26	24	2	24	17
Unsuppted	3	2	0	4	1	5	0	5	3
SAMSI	4	1	0	5	0	5	0		

Undergraduate Two-Day Workshop

Workshop Participant Support

February 25-26, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Adams	Eric	Male	University of Utah	Math	S
Bilo	Pamela	Female	Indiana University	Math	S
Durney	Clinton	Male	Virginia Tech	Math	S
Eng	Alan	Male	University of California, Davis	Stat	S
Evans	Christopher	Male	University of Arkansas, Fayetteville	Math	S
Garg	Anmol	Male	University of Washington, Seattle	Stat	S

Givens	Mark	Male	University of Wyoming	Math	S
Godwin	Jessica	Female	Auburn University	Math	S
Goode	Alexandra	Female	University of North Carolina Asheville	Math, Statistics	S
Hartsock	Leo	Male	University of North Carolina Wilmington	Math	S
Kabila	Ricardo	Male	Connecticut College	Math	S
Li	Ran	Female	Pennsylvania State University, University Park	Stat	S
Manhanga	Mercy	Female	University of Texas at Dallas	Other	S
Ng	Annalyn	Female	University of Michigan, Ann Arbor	Stat	S
Nguyen	Magarit	Female	University of North Carolina - Wilimington	Stat	S
Ni	Tianyi	Male	University of Illinois at Urbana Champaign	Math	S
Ouayoro	Marlene	Female	George Mason University	Math	S
Ray	Ryan	Male	Kent State University	Math	S
Rozzi	Amy	Female	Baylor University	Stat	S
Rutledge	Whitney	Female	Texas A&M University Corpus Christi	Math	S
Schoonover	Joseph	Male	Florida State University	Math	S
Solis Pacheco	Manuel	Male	University of Alberta	Stat	S
Tan	Xin Lu	Female	Purdue University, West Lafayette	Math/Stat	S
Weiss	Sam	Male	carnegie mellon university	Stat	S
Yi	Mengyao	Female	Purdue University	Math/Stat	S
Zhu	Anthony	Male	Cornell University	Engg	S

Undergraduate Two-Day Workshop

Workshop Participants

February 25-26, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Adams	Eric	Male	University of Utah	Math	S
Bilo	Pamela	Female	Indiana University	Math	S
Choi	James	Male	North Carolina State University	Math	S
Degras	David	Male	SAMSI	Stat	A
Durney	Clinton	Male	Virginia Tech	Math	S
Eng	Alan	Male	University of California, Davis	Stat	S
Evans	Christopher	Male	University of Arkansas, Fayetteville	Math	S
Garg	Anmol	Male	University of Washington, Seattle	Stat	S
Givens	Mark	Male	University of Wyoming	Math	S
Godwin	Jessica	Female	Auburn University	Math	S
Goode	Alexandra	Female	University of North Carolina Asheville	Math, Statistics	S
Gremaud	Pierre	Male	SAMSI	Math	A
Hartsock	Leo	Male	University of North Carolina Wilmington	Math	S
Huzurbazar	Snehalata	Female	U Wyoming	Stat	A
Jiang	Ci-Ren	Male	SAMSI	Stat	A

Kabila	Ricardo	Male	Connecticut College	Math	S
Li	Ran	Female	Pennsylvania State University, University Park	Stat	S
Ma	Junheng	Male	SAMSI	Stat	A
Manhanga	Mercy	Female	University of Texas at Dallas	Other	S
Munoz	Yolanda	Female	Michigan Tech	Stat	A
Ng	Annalyn	Female	University of Michigan, Ann Arbor	Stat	S
Nguyen	Magarit	Female	University of North Carolina - Wilimigton	Stat	S
Ni	Tianyi	Male	University of Illinois at Urbana Champaign	Math	S
Ouayoro	Marlene	Female	George Mason University	Math	S
Ramsay	Jim	Male	McGill U	Math/Stat	A
Ray	Ryan	Male	Kent State University	Math	S
Rozzi	Amy	Female	Baylor University	Stat	S
Rutledge	Whitney	Female	Texas A&M University Corpus Christi	Math	S
Schoonover	Joseph	Male	Florida State University	Math	S
Solis Pacheco	Manuel	Male	University of Alberta	Stat	S
Tan	Xin Lu	Female	Purdue University, West Lafayette	Math/Stat	S
Tchumtchoua	Sylvie	Female	SAMSI	Stat	A
Weiss	Sam	Male	carnegie mellon university	Stat	S
Wu	Hulin	Male	U Rochester	Stat	A
Yi	Mengyao	Female	Purdue University	Math/Stat	S

Zhu	Anthony	Male	Cornell University	Engg	S

SAMSI/CRSC Interdisciplinary Workshop for Undergraduates

Participant Summary

May 16-20, 2011

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Institutions Represented	Number of States Represented
Supported	22	18	0	12	28	35	5	28	16
Unsuppted	8	8	0	12	4	15	1	12	6
SAMSI	0	0	0	0	0	0	0		

SAMSI/CRSC Interdisciplinary Workshop for Undergraduates

Workshop Participant Support

May 16-20, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Abgegha	Gerald	M	Johnson C. Smith University	Math	A
Adams	Kevin	M	Furman U	Math	A
Ayvazov	Suliko	M	UCLA	Math	S
Barron	Alexander	M	University of New Mexico	Math	S
Bergsman	Louis	M	Tulane	Math	S
Bernstein	Daniel	M	Davidson College	Math	S
Carey	Jacob	M	Virginia Tech	Math	S
Chao	Matthew	M	Rutgers	Math	S
Chartier	Tim	M	Davidson College	Math	A
Diawara	Norou	M	Old Dominion	Math/Stat	A

Edusei	Kafui	M	Jackson State	Life	S
Fadimba	Koffi	M	University of South Carolina	Math	A
Goode	Alexandra	F	UNC-Asheville	Math/Stat	S
Hunt	Gregory	M	Drew University	Math	S
Hutson	Kevin	M	Furman University	Math	A
Jiang	Haoping	M	Iowa State	Stat	S
Joseph	James	M	Penn State	Stat	S
Jun	Seong-Hwan	M	U British Columbia	Stat	S
Li	Ran	F	Penn State	Stat/Math	S
Llanos	Danielle	F	Georgetown	Math	S
Lloyd	Buffy	F	Arizona State University	Math	S
Lott	Carolanne	F	U Georgia	Stat	S
Naughton	Brian	M	UNC- Charlotte	Math	S
Nicklas	Brent	M	Michigan Tech	Math	S
Novak	Julie	F	McGill	Math	S
Ortega	Omayra	F	Arizona State University	Math	A
Pechenick	Alison	F	U. of Vermont	Comp	A
Price	Teri	F	Tulane	Math	S
Rajkarnikar	Sujana	F	Salem College	Math	S
Rios-Soto	Karen	F	University of Puerto Rico	Math	A
Sanders	Lauren	F	Marymount University	Math	S

Schneider	Max	M	UCLA	Stat	S
Shaefer	Elsa	F	Marymount University	Math	A
Shand	Lyndsay	F	Bucknell	Math	S
Shepard	Emma	F	Georgetown	Life	S
Starkweather	Clara	F	Duke	Life	S
Tang	Pan	F	Iowa State	Stat	S
Turcios	Henry	M	Marymount University	Life	S
Wieman	Bob	M	Virginia State U.	Math	A
Young	Paula	F	Salem College	Math	A

SAMSI/CRSC Interdisciplinary Workshop for Undergraduates

Workshop Participants

May 16-20, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Abgegha	Gerald	M	Johnson C. Smith University	Math	A
Adams	Kevin	M	Furman U	Math	A
Ayvazov	Suliko	M	UCLA	Math	S
Barron	Alexander	M	University of New Mexico	Math	S
Bergsman	Louis	M	Tulane	Math	S
Bernstein	Daniel	M	Davidson College	Math	S
Brooks	Gary	M	Winthrop University	Math	A
Brown	Alexandra	F	Spellman	Math	S
Carey	Jacob	M	Virginia Tech	Math	S

Chao	Matthew	M	Rutgers	Math	S
Chartier	Tim	M	Davidson College	Math	A
Diawara	Norou	M	Old Dominion	Math/Stat	A
Dunbar	Jean	F	Converse College	Math	A
Eberg	Maria	F	McGill	Math	S
Edusei	Kafui	M	Jackson State	Life	S
Fadimba	Koffi	M	University of South Carolina	Math	A
Fernando	Anne	F	Norflok State U.	Math	A
Fitzkee	Thomas	M	Francis Marion University	Math	A
Goode	Alexandra	F	UNC-Asheville	Math/Stat	S
Griffin	Marcia	F	Johnson C. Smith		S
Hamedani	Gholamhos sein	M	Marquette	Math	A
Herron	Andrew	M	East Tennessee	Math	S
Heyer	Laurie	F	Davidson College	Math	A
Hunt	Gregory	M	Drew University	Math	S
Hutson	Kevin	M	Furman University	Math	A
Jiang	Haoping	M	Iowa State	Stat	S
Joseph	James	M	Penn State	Stat	S
Jun	Seong-Hwan	M	U British Columbia	Stat	S
Knotts-Zides	Charlotte	F	Wofford College	Math	A
Lee	Jimin	F	UNC-Asheville	Math	A
Li	Ran	F	Penn State	Stat/Math	S

Llanos	Danielle	F	Georgetown	Math	S
Lloyd	Buffy	F	Arizona State University	Math	S
Lott	Carolanne	F	U Georgia	Stat	S
Naughton	Brian	M	UNC- Charlotte	Math	S
Nicklas	Brent	M	Michigan Tech	Math	S
Novak	Julie	F	McGill	Math	S
Ong	John	M	Mary Baldwin College	Math	A
Ortega	Omayra	F	Arizona State University	Math	A
Pechenick	Alison	F	U. of Vermont	Comp	A
Price	Teri	F	Tulane	Math	S
Rajkarnikar	Sujana	F	Salem College	Math	S
Rios-Soto	Karen	F	University of Puerto Rico	Math	A
Sanders	Lauren	F	Marymount University	Math	S
Schneider	Max	M	UCLA	Stat	S
Shaefer	Elsa	F	Marymount University	Math	A
Shand	Lyndsay	F	Bucknell	Math	S
Shepard	Emma	F	Georgetown	Life	S
Starkweather	Clara	F	Duke	Life	S
Szurley	David	M	Francis Marion University	Math	A
Tang	Pan	F	Iowa State	Stat	S
Turcios	Henry	M	Marymount University	Life	S
Velummylum	Somasundaram	M	Clafin University	Math	A

Wages	Nolan	M	Hampden-Sydney College	Math	A
Wieman	Bob	M	Virginia State U.	Math	A
Young	Paula	F	Salem College	Math	A

SAMSI/CRSC Industrial Mathematical & Statistical Workshop for Graduates

Participant Summary

July 7-15, 2011

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Institutions Represented	Number of States Represented
Supported	20	9	1	1	29	29	1	26	20
Unsuppted	9	5	0	11	3	10	4	9	4
SAMSI	0	0	0	0	0	0	0		

SAMSI/CRSC Industrial Mathematical & Statistical Workshop for Graduates

Workshop Participant Support

July 7-15, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Ashu	Tom	Male	Kent State University	MATH	S
Basu	Treena	Female	University of South Carolina	MATH	S
Basu	Kanadpriya	Male	U South Carolina	MATH	S
Bolotskikh	Alexandra	Female	U Florida	STAT	S
Callaway	Josh	Male	UAMS	STAT	S
Chen	Ming-Chun	Male	CSU East Bay	STAT	S
Cheng	Yansong	Male	Boston University	MATH	S
Eager	Eric	Male	U Nebraska	MATH	S
Fink	Michael	Male	Binghamton University	MATH	S
Gopaladesikan	Mohan	Male	Purdue University	STAT	S

Gregory	Karl	Male	Texas A&M University	STAT	S
Gromenko	Oleksandr	Male	Utah State University	STAT	S
Gu	Yu	Male	Columbia University	MATH	S
Han	Xiao		Penn State U	MATH	S
Jaafari	Fatima	Female	Binghamton University	STAT	S
Jenkins	Eleanor	Female	Clemson University	MATH	A
Khasawinah	Sarah	Female	Johns Hopkins University	STAT	S
Kong	Liang	Male	Auburn University	MATH	S
Kwessi Nyandjou	Eddy Armand	Male	Auburn University	STAT	S
Li	Qi	Female	University of Illinois at Urbana-Champaign	STAT	S
McNichols	Sean	Male	Auburn University	MATH	S
Mess	Raymond	Male	UW-Milwaukee	STAT	S
Ouedraogo	Jerome	Male	Georgia Southern University	STAT	S
Qi	Qi	Male	University of Missouri-Columbia	COMP	S
Silva	Benjamin	Male	Florida State University	STAT	S
VanderPlas	Susan	Female	Iowa State University	STAT	S
Yang	Jingjing	Female	Rice University	STAT	S
Zeng	Yan	Male	University of Rochester	STAT	S
Zhang	Yilong	Male	George Washington University	STAT	S
Zhou	Wenli	Female	U Kentucky	STAT	S

SAMSI/CRSC Industrial Mathematical & Statistical Workshop for Graduates

Workshop Participants

July 7-15, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Ashu	Tom	Male	Kent State University	MATH	S
Banks	H.T.	Male	NCSU	STAT	A
Basu	Treena	Female	University of South Carolina	MATH	S
Basu	Kanadpriya	Male	U South Carolina	MATH	S
Bogacki	Jared	Male	BB&T	OTHER	A
Bolotskikh	Alexandra	Female	U Florida	STAT	S
Calatroni	Agustin	Male	Rho, Inc	STAT	A
Callaway	Josh	Male	UAMS	STAT	S
Chen	Ming-Chun	Male	CSU East Bay	STAT	S
Cheng	Yansong	Male	Boston University	MATH	S
Cole Manning	Cammey	Female	Meredith College	MATH	A
Eager	Eric	Male	U Nebraska	MATH	S
Fink	Michael	Male	Binghamton University	MATH	S
Gopaladesikan	Mohan	Male	Purdue University	STAT	S
Gregory	Karl	Male	Texas A&M University	STAT	S
Gromenko	Oleksandr	Male	Utah State University	STAT	S

Gu	Yu	Male	Columbia University	MATH	S
Han	Xiao		Penn State U	MATH	S
Jaafari	Fatima	Female	Binghamton University	STAT	S
Jenkins	Eleanor	Female	Clemson University	MATH	A
Kees	Christopher	Male	US Army Research	ENGG	A
Khasawinah	Sarah	Female	Johns Hopkins University	STAT	S
Kong	Liang	Male	Auburn University	MATH	S
Kwessi Nyandjou	Eddy Armand	Male	Auburn University	STAT	S
Li	Qi	Female	University of Illinois at Urbana-Champaign	STAT	S
Lu	Qiyi	Female	SUNY	MATH	S
Massad	Jordan	Male	Sandia National Lab	ENGG	A
McNichols	Sean	Male	Auburn University	MATH	S
Mess	Raymond	Male	UW-Milwaukee	STAT	S
Mitchell	Herman	Male	Rho, Inc.	LIFE	A
Ouedraogo	Jerome	Male	Georgia Southern University	STAT	S
Peach	John	Male	MIT Lincoln Lab	MATH	A
Potter	Laura	Female	Syngenta Biotech	MATH	A
Qi	Qi	Male	University of Missouri-Columbia	COMP	S
Rehm	Keri	Female	North Carolina State University	MATH	S
Scroggs	Jeff	Male	NCSU	MATH	A
Silva	Benjamin	Male	Florida State University	STAT	S

Smith	Ralph	Male	NCSU	MATH	A
VanderPlas	Susan	Female	Iowa State University	STAT	S
Wang	Anran	Female	North Carolina State University	STAT	S
Yang	Jingjing	Female	Rice University	STAT	S
Zeng	Yan	Male	University of Rochester	STAT	S
Zhang	Yilong	Male	George Washington University	STAT	S
Zhou	Wenli	Female	U Kentucky	STAT	S

2011-12 PROGRAM EVENTS THROUGH JULY 2011

Uncertainty Quantification Summer School

Participant Summary
June 20-24, 2011

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	35	14	0	16	33	13	18	18	34	17
Unsuppted	22	6	0	26	2	4	12	12	10	6
SAMSI	3	0	0	1	2	0	3	0		

Uncertainty Quantification Summer School

Workshop Participant Support
June 20-24, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Allaire	Douglas	Male	MIT	Other	S
Anitescu	Mihail	Male	ANL	Math	A
Barajas-Solano	David	Male	UCSD	Engg	S
Bryant	Corey	Male	U Texas	Engg	S
Burch	Nate	Male	SAMSI	Math	A

Chowdhary	Kenny	Male	Brown U	Math	S
Cisewski	Jessica	Female	UNC	Stat	S
Cooley	Dan	Male	Colorado State	Stat	A
Cousins	William	Male	NCSU	Math	S
Goh	Joslin	Female	Simon Fraser	Stat	S
Gremaud	Pierre	Male	SAMSI	Math	A
Heller	Miriam	Female	MHITech	Engg	A
Heo	Jaeseok	Male	NCSU	Math	S
Huoh	Yu-Jay	Male	UC Berkeley	Stat	S
Jakeman	John	Male	Purdue	Math	A
Jeon	Soyoung	Female	UNC	Stat	S
Koltakov	Sergey	Male	Stanford	Engg	S
Lee	Ben	Male	Georgia Tech	Engg	S
Lee	Yoonsang	Male	U Texas	Math	S
Li	Jing	Female	Purdue	Math	S
Lieberman	Chad	Male	MIT	Other	S
Lin	Lizhen	Female	U Arizona	Math	S
Mao	Youli	Male	Texas A&M	Math	S
Marcy	Peter	Male	U Wyoming	Stat	S
Morrison	Rebecca	Female	U Texas	Engg	S
Narayan	Akil	Male	Purdue	Math	A

Nychka	Doug	Male	UCAR	Stat	A
Padilla	Dustin	Male	Arizona Stat	Math	S
Presho	Michael	Male	Colorado State	Math	A
Rocklin	Matthew	Male	U Chicago	Comp	S
Rogers	Bruce	Male	SAMSI	Math	A
Roh	Soojin	Female	Texas A&M	Stat	S
Sancier-Barbosa	Flavia	Female	S. Illinois	Math	S
Sandu	Adrian	Male	Virginia Tech	Comp	A
Seleson	Pablo	Male	U Texas	Other	A
Sgouralis	Ioannis	Male	Duke U	Math	S
Stripling	Hayes	Male	Texas A&M	Engg	S
Terejanu	Gabriel	Male	U Texas	Stat	A
Tran	Linda	Female	UC Berekley	Stat	S
van Lier-Walqui	Marcus	Male	U Miami	Phys	S
Varkey	Paul	Male	U Illinois	Comp	S
Villagran	Alejandro	Male	U Connecticut	Stat	A
Wang	Peng	Male	UC San Diego	Engg	S
Wang	ShengYang	Male	U New Mexico	Stat	S
Wei	Jia	Female	Texas A&M	Math	S
Wu	Yanhong	Female	SUNY	Math	S
Xiu	Dongbin	Male	Purdue	Stat	A

Xu	Hongyi	Male	Northwestern	Engg	S
Zamboni	Laura	Female	ANL	Phys	A
Zhang	Guannan	Male	Florida State	Math	S
Zhang	Kai	Male	U Michigan	Life	A
Zhu	Jielin	Female	UBC	Math	S

Uncertainty Quantification Summer School

Workshop Participants
June 20-24, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Abdel-Khalik	Hany	Male	NCSU	Engg	A
Aguilo	Miguel	Male	Sandia	Other	A
Allaire	Douglas	Male	MIT	Other	S
Anitescu	Mihail	Male	ANL	Math	A
Atamturktur	Sez	Female	Clemson	Engg	A
Barajas-Solano	David	Male	UCSD	Engg	S
Barber	Jared	Male	U Pittsburgh	Math	A
Bennun	Sandra	Female	UC Davis	Engg	A
Bryant	Corey	Male	U Texas	Engg	S
Bulaevskaya	Vera	Female	LLNL	Math	A
Burch	Nate	Male	SAMSI	Math	A
Chowdhary	Kenny	Male	Brown U	Math	S

Cisewski	Jessica	Female	UNC	Stat	S
Constantine	Paul	Male	Sandia	Engg	A
Constantinescu	Emil	Male	ANL	Math	A
Cooley	Dan	Male	Colorado State	Stat	A
Cousins	William	Male	NCSU	Math	S
Crespo	Luis	Male	Nat'l Inst. Aerospace	Engg	A
Eldred	Michael	Male	Sandia	Engg	A
Gittelson	Claude	Male	ETH Zurich	Math	A
Goh	Joslin	Female	Simon Fraser	Stat	S
Gremaud	Pierre	Male	SAMSI	Math	A
Han	Xiaoying	Female	Auburn	Math	A
Heller	Miriam	Female	MHITech	Engg	A
Helton	Jon	Male	Sandia	Math	A
Heo	Jaeseok	Male	NCSU	Math	S
Higdon	Dave	Male	LLNL	Stat	A
Hu	Ken	Male	MIT	Engg	A
Huerta	Gabriel	Male	U New Mexico	Math/Stat	A
Huoh	Yu-Jay	Male	UC Berkeley	Stat	S
Hyman	James	Male	Tulane/LANL	Math	A
Jakeman	John	Male	Purdue	Math	A
Jeon	Soyoung	Female	UNC	Stat	S

Jiang	Lijian	Male	LANL	Math	A
Johannesson	Gardar	Male	LLNL	Math	A
Koltakov	Sergey	Male	Stanford	Engg	S
Lee	Ben	Male	Georgia Tech	Engg	S
Lee	Yoonsang	Male	U Texas	Math	S
Lei	Jing	Male	UC Berkeley	Stat	A
Li	Jing	Female	Purdue	Math	S
Lieberman	Chad	Male	MIT	Other	S
Lin	Lizhen	Female	U Arizona	Math	S
Mao	Youli	Male	Texas A&M	Math	S
Marcy	Peter	Male	U Wyoming	Stat	S
May	Elebeoba	Female	Sandia	Engg	A
Morrison	Rebecca	Female	U Texas	Engg	S
Narayan	Akil	Male	Purdue	Math	A
Nychka	Doug	Male	UCAR	Stat	A
Oluwasanmi	Olumuyiwa	Male	U New Mexico	Math	S
Padilla	Dustin	Male	Arizona Stat	Math	S
Presho	Michael	Male	Colorado State	Math	A
Rocklin	Matthew	Male	U Chicago	Comp	S
Rogers	Bruce	Male	SAMSI	Math	A
Roh	Soojin	Female	Texas A&M	Stat	S

Sancier-Barbosa	Flavia	Female	S. Illinois	Math	S
Sandu	Adrian	Male	Virginia Tech	Comp	A
Seleson	Pablo	Male	U Texas	Other	A
Sgouralis	Ioannis	Male	Duke U	Math	S
Stark	Glenn	Male	U New Mexico	Stat	S
Stewart	Jim	Male	Sandia	Comp	A
Stripling	Hayes	Male	Texas A&M	Engg	S
Swiler	Laura	Female	Sandia	Stat	A
Terejanu	Gabriel	Male	U Texas	Stat	A
Tong	Charles	Male	LLNL	Comp	A
Tran	Linda	Female	UC Berekley	Stat	S
Trucano	Timothy	Male	Sandia	Math	A
van Lier-Walqui	Marcus	Male	U Miami	Phys	S
Varkey	Paul	Male	U Illinois	Comp	S
Villagran	Alejandro	Male	U Connecticut	Stat	A
Wang	Peng	Male	UC San Diego	Engg	S
Wang	ShengYang	Male	U New Mexico	Stat	S
Wei	Jia	Female	Texas A&M	Math	S
Weirs	V Gregory	Male	Sandia	Engg	A
Wu	Yanhong	Female	SUNY	Math	S
Xiu	Dongbin	Male	Purdue	Stat	A

Xu	Hongyi	Male	Northwestern	Engg	S
Zamboni	Laura	Female	ANL	Phys	A
Zhang	Guannan	Male	Florida State	Math	S
Zhang	Kai	Male	U Michigan	Life	A
Zhu	Jielin	Female	UBC	Math	S

2012-13 PROGRAM EVENTS THROUGH JULY 2011

Big Data Meeting
Participant Summary
May 20, 2011

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Institutions Represented	Number of States Represented
Supported	5	5	0	8	2	7	1	2	10	8
Unsuppted	12	2	0	12	2	7	1	6	8	3
SAMSI	1	0	0	1	0	1	0	0		

Big Data Meeting
Workshop Participant Support
May 20, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Altman	Naomi	Female	Penn State U	Stat	FP
Budavari	Tamas	Male	Johns Hopkins	Stat	FP
Gray	Alexander	Male	Georgia Tech	Comp	FP
Ipsen	Ilse	Female	NCSU	Math	FP
Kafadar	Karen	Female	Indiana	Stat	FP
Liu	Yufeng	Male	UNC	Stat	FP

Mane	Ketan	Male	RENCI	Comp	NRG
Michalak	Sarah Ellen	Female	LANL	Stat	FP
Sain	Steve	Male	UCAR	Stat	FP
Sun	Jiayang	Female	Case Western	Stat	NRG

Big Data Meeting
Workshop Participants
May 20, 2011

Last Name	First Name	Gender	Affiliation	Major/Department	Status
Altman	Naomi	Female	Penn State U	Stat	FP
Budavari	Tamas	Male	Johns Hopkins	Stat	FP
Crichton	Dan	Male	NASA	Comp	FP
Demmel	Jim	Male	UC Berkeley	Math	FP
Dhillon	Inderjit	Male	U Texas	Comp	FP
El Karoui	Noureddine	Male	UC Berkeley	Stat	FP
Gray	Alexander	Male	Georgia Tech	Comp	FP
Hastie	Trevor	Male	Stanford	Stat	FP
Horta	Arnie	Male	Dept of Defense	Stat	FP
Ipsen	Ilse	Female	NCSU	Math	FP
Jacobsen	Bob	Male	LBL	Phys	FP
Jordan	Michael	Male	UC Berkeley	Stat	FP
Kafadar	Karen	Female	Indiana	Stat	FP

Liu	Yufeng	Male	UNC	Stat	FP
Mahoney	Michael	Male	Stanford	Stat	FP
Mane	Ketan	Male	RENCI	Comp	NRG
Michalak	Sarah Ellen	Female	LANL	Stat	FP
Prabhat	Mr.	Male	LBL	Comp	NRG
Sain	Steve	Male	UCAR	Stat	FP
Smith	Richard	Male	SAMSI	Stat	FP
Sun	Jiayang	Female	Case Western	Stat	NRG
Temple Lang	Duncan	Male	UC Davis	Stat	NRG
Ushizima	Daniela	Female	LBL	Comp	FP
Wu	John	Male	LBL	Comp	FP
Yu	Bin	Female	UC Berkeley	Stat	FP

APPENDIX B – Workshop Programs and Abstracts

#1: 2010 Summer Program on Semiparametric Bayesian Inference: Applications in Pharmacokinetics and Pharmacodynamics. July 12-23, 2010

Schedule:
Monday, July 12, 2010
Radisson Hotel RTP

Tutorials

8:00-8:45	Registration and Continental Breakfast
8:45-9:00	Welcome & Overview
9:00-10:45	Tutorial 1 **Wes Johnson**, U. of California, Irvine and **David Dunson**, Duke University *"Bayesian Nonparametrics: An Overview"*
10:45-11:15	Break
11:15-12:30	Tutorial 1, continued
12:30-1:30	Lunch
1:30-3:00	Tutorial 2 **David D'Argenio**, University of Southern California and **Paolo Vicini**, Pfizer *"Pharmacokinetics (PK) and Parmacodynamics (PD)"*
3:00-3:30	Break
3:30-5:00	Tutorial 2, continued

Tuesday, July 13, 2010
Radisson Hotel RTP

Opening Workshop

8:00-9:00	Registration and Continental Breakfast
	Prior Elicitation and Construction:
9:00-9:45	**Wes Johnson**, University of California, Irvine *"On the Value of Incorporating Scientific Input in Modeling and Data Analysis, and How to Do it Without Pain"*

9:45-10:30	**Peter Thall**, MDACC *"Prior Elicitation in Bayesian Clinical Trial Design"*
10:30-11:00	Break
11:00-11:45	**Siva Sivaganesan**, University of Cincinnati *"Sample Size: How Many Subjects"*
11:45-Noon	Discussion
Noon-1:00	Lunch
1:00-2:30	**Lang Li**, Indiana University and **Paolo Vicini**, Pfizer *"Identifiability and Prior Regularization: A PK-PD Modeling View"*
2:30 – 3:00	Break
3:00 – 3:45	**Robert Leary**, Pharsight *"Computational Challenges of Nonlinear Mixed Effects Analysis for PK/PD Modeling"*
3:45-4:30	**Sujit Ghosh**, North Carolina State University *"Application of Nonlinear Mixed Models Involving ODEs"*
4:30-5:00	Discussion

Wednesday**, July 14, 2010**
Radisson Hotel RTP

8:00-9:00	Registration and Continental Breakfast
	Joint PK and PD:
9:00-9:45	**Wojciech Krzyzanski**, State University of New York *"PK/PD Modeling of Therapeutic Effects of Erythropoietin"*
9:45-10:30	**Michele Guidani**, University of New Mexico *"NP Bayes Functional Regression for a PK/PD Semi-Mechanistic Model"*
10:30-11:00	Break
11:00-11:45	**Thanasis Kottas**, University of California, Santa Cruz *"A Bayesian Nonparametric Modeling Approach for Developmental Toxicology Data"*
11:45-Noon	Discussion
Noon-1:00	Lunch

1:00-2:30	**David Conti**, U. of Southern California; **Duncan Thomas**, U. of Southern California *“Bayesian Hierarchical and PD/PK Modeling for Complex Pathways in Molecular Epidemiology”*
2:30-3:00	**Mike Reed**, Duke University *“Why Metabolism is a Challenge for Statistical Analysis: the Case of Substrate Inhibition”*
3:00-3:30	Break
3:30- 4:15	**Peter Mueller**, MD Anderson *“Nonparametric Bayesian Models for Population PK/PD.”*
4:15-5:00	**Donatello Telesca**, University of California, Los Angeles *“Modeling Pk/PGx Data”*
5:00-5:30	Discussion

Thursday, July 15, 2010
Radisson Hotel RTP

8:00-9:00	Registration and Continental Breakfast
	DDP, Mixture Models and Related Models:
9:00–9:45	**Subharup Guha**, University of Missouri *“Posterior Simulation in Countable Mixture Models for Large Data Sets”*
9:45 – 10:30	**Steve MacEachern**, Ohio State University *“Nonparametric Bayesian Models for Related Subpopulations”*
10:30-11:00	Break
11:00-11:45	**Maria de Iorio**, Imperial College *“Nonparametric Approaches for Heterogeneous Longitudinal Data”*
11:45-12:30	**Tatiana Tatarinova**, University of Glamorgan *“Analysis of Biological Time Series Data in Data-Rich Situations”*
12:30-1:30	Lunch
1:30-2:10	**Roger Jelliffe**, University of Southern California *"Tools for Individualizing Drug Dosage Regimens for Patients"*
2:10-2:50	**William Gillespie**, Metrum Inst.

"Examples of Decison-Focused Modeling and Simulation in Clinical Drug Development and Some Methodology R&D Opportunities They Suggest"

2:50-3:30 **Gary Rosner**, Johns Hopkins University
"Optimal Design with Bayesian Nonparametric Priors"

3:30-4:00 Break

4:00-5:00 **David D'Argenio**, University of Southern California; **William Gillespie**, Metrum Inst.; **Roger Jelliffe**, University of Southern California; **Bob Leary**, Pharsight
"Implementation and Available Software"

5:00-5:30 Poster Advertisement Session: 2 minute ads by each poster presenter

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.

8:00- Informal Debate: *"Parametric PD Models vs. Nonparametric PD Models*

Friday, July 16, 2010
Radisson Hotel RTP

8:00-9:00 Registration and Continental Breakfast

Nonparametric Bayes and Clustering for Random Effects Distributions:

9:00-9:45 **David Dahl**, Texas A&M University
"Strategies for Incorporating Prior Information in Random Partition Models with Examples"

9:45-10:30 **Mahlet Tadesse**, Georgetown University
"Stochastic Partitioning and Model-based Clustering with Variable Selection"

10:30–11:00 Break

11:00 –11:45 **Julien Cornebise**, University of British Columbia
"Inference for Mixed Effects Stochastic Differential PKPD Models with Particle MCMC"

11:45-12:15 **Gary Rosner**, Johns Hopkins University, and **Peter Mueller**, University of Texas
Summary & Discussion

12:15-1:30 Lunch

1:30–2:30 Formation of Working Groups

2:30-3:00 Break

3:00-4:00 Initial discussions

Monday, July 19 – Thursday, July 22, 2010
SAMSI, RTP,
Rooms 104, 150, 201, 203, 259

9:00 – 5:00 Working Groups - TBA

Possible Themes:

1. Identifiability and Prior Regularization
2. Joint PK and PD Modeling
3. PK/PD Simulation for Design and Dose Individualization
4. Pharmacogenomics and Networks
5. Nonparametric Bayes and Clustering for PK and PD Population Analysis

12:00-1:30 Lunch

Friday, July 23, 2010
SAMSI, RTP,
Room 150

9:30-12:00 Presentations by Working Groups and Discussion about Future Projects

12:00-1:30 Lunch

1:30 Adjourn

SPEAKER TITLES/ABSTRACTS

David Conti
University of Southern California
dconti@usc.edu

Duncan Thomas
University of Southern California
dthomas@usc.edu

"Bayesian Hierarchical and PD/PK Modeling for Complex Pathways in Molecular Epidemiology"

Molecular epidemiology is concerned with the study of the joint role of genetic and environmental influences in disease causation. Despite the current enthusiasm for "agnostic"

(hypothesis-free) genome-wide association approaches, there is still a place for biologically-based analyses driven by our current understanding of specific pathways implicated in disease. This talk will focus on statistical approaches to incorporating biological knowledge into the analysis of gene-environment interactions in disease etiology. Two broad frameworks will be introduced, one based on hierarchical modeling, the other based on physiologically-based pharmacokinetic (PBPK) modeling (Thomas et al. 2010). In the former, a standard regression model (such as multiple logistic for a binary disease trait) forms the first level for the study subject's data, with regression coefficients for main effects of various genes, environmental factors, and their interactions; these coefficients are in turn regressed on external "meta-data" about these risk factors, such as indicators for specific pathways involving the various factors, functional data, or expert knowledge organized in formal ontologies. In the PBPK approach, a mechanistic model for the disease process is constructed, typically involving one or more latent variables representing intermediate metabolite concentrations or other unobserved biological processes, each related to each other and to the input variables through a system of ordinary or stochastic differential equations. Both approaches will be illustrated on various epidemiologic studies, such as folate metabolism in colorectal cancer, nicotine dependence, and DNA repair following ionizing radiation exposure in second breast cancers. Either approach typically requires one to deal with the problem of model uncertainty, such as the selection of variables to include in a hierarchical model. We will discuss approaches such as stochastic search variable selection and shrinkage using lasso or ridge priors, as well as novel methods based on Markov chain Monte Carlo searching over the space of all possible pathway models. We will conclude with some thoughts on the prospects for extending these approaches to genome-wide scale.
Reference: Thomas DC, Conti DV, Baurley J, Nijhout F, Reed M, Ulrich CM. Use of pathway information in molecular epidemiology. *Hum Genomics*, 2010;4:21-42.

David Dahl
Texas A&M University
dahl@stat.tamu.edu

"Strategies for Incorporating Prior Information in Random Partition Models with Examples"

Many Bayesian semiparametric and nonparametric models are based on random partitioning of items (e.g., experimental units) into clusters, where items in the same cluster share a common value for a model parameter. Often there is good prior information about the clustering of items, yet traditional random partition models cannot incorporate this information. In this talk, I will review several recently-proposed strategies for incorporating prior clustering information in random partition models. After surveying methods for which the prior clustering information comes in the form of
item-specific covariates, I will discuss methods that instead utilize pairwise distances between items to inform the prior clustering distribution. Applying these methods to a model for protein structure prediction, it is shown that the methods incorporating distance
information substantially improve predictive accuracy.

David Dunson
Duke University
dunson@stat.duke.edu

“Bayesian Nonparametrics: An Overview”

Tutorial Part II: In the second part of this tutorial we provide a brief review and motivation for the use of nonparametric Bayes methods in biostatistical applications. The focus is on methods utilizing random probability measures, with the emphasis on a few approaches that seem particularly useful in addressing the considerable challenges faced in modern biostatistical research. In addition, the emphasis will be entirely on practical applications-motivated considerations, with the goal of encouraging participants to consider related approaches for their own data.

The discussion includes methods for functional data analysis using DP-based methods, approaches for local shrinkage and clustering, methods for hierarchical borrowing of information across studies, centers or exchangeable groups of data, flexible modeling of conditional distributions using priors for collections of random probability measures that evolve with predictors, time and spatial location, some recent applications in bioinformatics, and methods for nonparametric Bayes hypothesis testing.

Maria de Iorio
Imperial College
m.deiorio@imperial.ac.uk

"Nonparametric Approaches for Heterogeneous Longitudinal Data"

We discuss a method for combining different but related longitudinal studies to improve predictive precision. The motivation is to borrow strength across clinical studies in which the same measurements are collected at different frequencies. Key features of the data are heterogeneous populations and an unbalanced design across three studies of interest. The first two studies are phase I studies with very detailed observations on a relatively small number of patients. The third study is a large phase III study with over 1500 enrolled patients, but with relatively few measurements on each patient. Patients receive different doses of several drugs in the studies, with the phase III study containing significantly less toxic treatments. Thus, the main challenges for the analysis are to accommodate heterogeneous population distributions and to formalize borrowing strength across the studies and across the various treatment levels.

We describe a hierarchical extension over suitable semiparametric longitudinal data models to achieve the inferential goal. A nonparametric random-effects model accommodates the heterogeneity of the population of patients. A hierarchical extension allows borrowing strength across different studies and different levels of treatment by introducing dependence across these nonparametric random-effects distributions. Dependence is introduced by building an analysis of variance (ANOVA) like structure over the random-effects distributions for different studies and treatment combinations. Model structure and parameter interpretation are similar to standard ANOVA models. Instead of the unknown normal means as in standard ANOVA models, however, the basic objects of inference are random distributions, namely the unknown population distributions under each study. The analysis is based on a mixture of Dirichlet processes model as the underlying semiparametric model.

Sujit Ghosh
North Carolina State University
sujit_ghosh@ncsu.edu

"Application of Nonlinear Mixed Models Involving ODEs in Biomedical Sciences"

In the field of biomedical applications, data usually consists of repeated measurements on individuals observed under varying experimental conditions. For example, in pharmacokinetics, several blood samples are taken on participating individuals over a period of time, following administration of a drug. The within subject relationships between the measured response and the varying experimental conditions are often nonlinear represented by a system of nonlinear ordinary differential equations (ODEs) and involve unknown subject-specific parameters of interest. These models are useful because they offer a flexible framework where parameters for both individuals and population can be estimated by combining information across all subjects. More often such a system of ODEs does not have any analytical closed form solution, making parameter estimation for these models quite challenging and computationally very demanding. A method based on Euler's approximation is proposed to obtain an approximate likelihood that is analytically tractable and thus making parameter estimation computationally less demanding than other competing methods. These methods will be illustrated using a real data set and simulation studies will be presented to compare the performances of these new methods to other established methods in the literature.

William Gillespie
Metrum Institute
billg@metrumrg.com

"Examples of Decision-Focused Modeling and Simulation in Clinical Drug Development and Some Methodology R&D Opportunities They Suggest"

Modeling and simulation are increasingly used to support decision-making in clinical drug development. Simulations are used to infer the probable range of outcomes for different treatment regimens, patient populations and clinical trials. Simulated results are then used to optimize decisions regarding patient treatment, clinical trial design and analysis, and development strategy, e.g., whether to continue a development program. The modeling and simulation approaches used for such applications vary depending on a variety of factors including the intended purpose, amount and type of prior information, and, of course, the skills and philosophical preferences of the modeler(s).

Two examples will be presented that use very different modeling strategies for different purposes. Example 1 involved the use of a largely empirical Bayesian hierarchical model for efficacy- and safety-related responses as a function of drug exposure and time constructed via model-based meta-analysis of clinical and preclinical data. Simulations based on the resulting model were used to optimize the design of a Phase II clinical trial. Design features explored included adaptive stopping, adapting pruning, and use of informative prior distributions in the trial analyses. In example 2 a multi-scale mechanistic model for calcium homeostasis and bone remodeling was developed. The model was used to explore the impacts of chronic renal failure, parathyroid disorders and discontinuation of denosumab treatment on biomarkers and bone mineral density. It was also used to investigate potential causal mechanisms for long-term bone remodeling effects of estrogen in menopausal and postmenopausal women. The current model is deterministic. An effort is underway to implement the model in a Bayesian framework that incorporates uncertainty in parameter values and inter-individual variation is selected parameters. This will permit probabilistic inference and model updating via MCMC analysis of new data as it becomes available.

Example 1 suggests the following opportunities for method development:

- Modeling relationships between biomarkers and therapeutic outcomes including approaches for integrating models for multiple biomarkers.
- Use of decision analysis to optimize trial design and for analysis of adaptive trials.
- Model-based meta-analysis, particularly the combined analysis of summary and individual data

Example 2 suggests the potential for predictive inferences about clinical response to a drug candidate even before that candidate has been administered to humans. However few such models have been developed and fewer still provide useful probabilistic inference. This suggests the following areas for R&D:

- Development of multi-scale mechanistic models for disease progression and drug effects relevant to other therapeutic areas.
- Bayesian implementation of such models.
- Development of Bayesian modeling methods and software better suited to such complex mechanistic models.

William Gillespie
Metrum Institute
billg@metrumrg.com

"Computational Tools for Bayesian PKPD Modeling"

Markov chain Monte Carlo (MCMC) simulation methods for Bayesian modeling have been recently added to some existing pharmacometric tools, e.g., NONMEM 7 and S-Adapt. Advantages of these programs are the inclusion of built-in PKPD model libraries and model-specification languages, and MCMC and less computationally-demanding algorithms within a single platform. On the other hand those tools are limited to models with two levels of random variation (plus the prior distributions), normally-distributed inter-individual variation and a normal-Wishart prior distribution.

More general-purpose software tools implementing MCMC for Bayesian modeling include the BUGS language programs (WinBUGS, OpenBUGS and JAGS), the SAS MCMC procedure and Microsoft Infer.Net. Some Bayesian modeling packages are also available for R and MatLab though most are limited to specific classes of models. The BUGS language tools have the advantage of being very flexible w.r.t. stochastic model structure, e.g., no limit on levels of variability, choice of many built-in distributions, and one can easily combine sub-models with very different stochastic structures. Their main disadvantages are the lack of specialized support for pharmacometrics applications, less flexible specification of deterministic model components, and the absence of less computationally-demanding algorithms within the same platform.

BUGSModelLibrary is a prototype PKPD model library for use with WinBUGS 1.4.3. The current version includes built-in 1 and 2 compartment models with or without first order absorption, and the ability to specify more general compartmental models in terms of first order ODE's. The models and data format are based on NONMEM conventions. Several test cases were analyzed using both the NONMEM 7 BAYES method and WinBUGS with BUGSModelLibrary. For the classes of models studied:

- MCMC simulations using NONMEM 7 and WinBUGS produced results with comparable accuracy.
- NONMEM 7 more often produced less auto-correlated MCMC samples.
- WinBUGS required much less computation time to produce comparable MCMC results for a mixture model, and about half the computation time for 4 test cases.
- WinBUGS required more time to produce less precise results for 4 other test cases.
- For _ 1=2 of the scenarios NONMEM 7 outperformed WinBUGS w.r.t. computation time adjusted for autocorrelation.

Those results led to the following recommendations:

- NONMEM 7 is a recommended platform for Bayesian modeling when suitable models can be implemented within the limits imposed by NONMEM.
- WinBUGS is a recommended platform when greater flexibility is required w.r.t. stochastic aspects of models.
- Based on the limited testing presented here, WinBUGS appears to perform better with mixture models and models with inter-occasion variability and is the preferred platform for those cases.

Subharap Guha
University of Missouri
guhasu@missouri.edu

"Posterior Simulation in Countable Mixture Models for Large Datasets"

Mixture models, or convex combinations of a countable number of probability distributions, offer an elegant framework for inference when the population of interest can be subdivided into latent clusters having random characteristics that are heterogeneous between, but homogenous within, the clusters. Traditionally, the different kinds of mixture models have been motivated and analyzed from very different perspectives, and their common characteristics have not been fully appreciated. The inferential techniques developed for these models usually necessitate heavy computational burdens that make them difficult, if not impossible, to apply to the massive data sets increasingly encountered in real world studies.

This talk introduces a flexible class of models called generalized Polya urn (GPU) models. Many common mixture models, such as finite mixtures, hidden Markov models, and Dirichlet processes are obtained as special cases of GPU models. Other important special cases include finite dimensional Dirichlet priors, infinite hidden Markov models, analysis of densities models, nested Chinese restaurant processes, hierarchical DP models, nonparametric density models, spatial Dirichlet processes, weighted mixtures of DP priors, and nested Dirichlet processes.

An investigation of the theoretical properties of GPU models offers new insight into asymptotics that form the basis of cost-effective MCMC strategies for large datasets. These MCMC techniques have the advantage that they provide inferences from the exact posterior of interest and are applicable to different mixture models. The versatility and impressive gains of the methodology are demonstrated by simulation studies and by a semiparametric Bayesian analysis of high-resolution comparative genomic hybridization data on lung cancer.

Michele Guindani
University of New Mexico
michele@stat.unm.edu

"NP Bayes Functional Regression for a PK/PD Semi-Mechanistic Model"

In Clinical pharmacology, a major goal is to study the quantitative prediction of drug effects. Complex pharmacokinetic-pharmacodynamic (PK/PD) models with feedback and transition effects have been recently developed, for example to estimate the time course of myelosupression. These models typically involve the presence of covariate dependent parameters. We propose a coherent NP Bayes probabilistic framework for the analysis of these models. Our goal is to use the information from the individuals' time courses and covariates to provide a clustering of patients according to their PK/PD profiles. This information could then be used to predict the PD time course for a patient on the basis of the observed PK profile, as well as deal with missing data. In the talk I will try to outline the features as well as the challenges connected with such implementation. This is a joint project with Peter Mueller, Gary L. Rosner and Lena Friberg.

Roger Jelliffe
University of Southern California
jelliffe@usc.edu

"Tools for Individualizing Drug Dosage Regimens for Patients"

The clinical tools we currently advocate are:

1. Set specific Target Goals (not windows) - Hit them most Precisely. Simply setting a window is not enough, except for knowing a general range of serum concentrations within which most patients, but certainly not all, do well. It does not consider each patient's need for the drug and the risk of toxicity it appears realistic to accept in order to obtain the hoped – for benefits of the drug. It also does not consider the patient's sensitivity to the drug. Using the same clinical judgment, one can set a specific target value for a serum concentration, for example, and then attempt to hit it as precisely as possible. This target may well be outside the so-called therapeutic range, not only below it, but also above it.

2. Describe Assay Errors by the reciprocal of the variance with which each measurement was made. Both for population modeling and for clinical care, the laboratory can do a much better job if it gives up using CV% as a measure of error (they are the only group in the world who do this!) and use instead the reciprocal of the assay variance for each measurement. This is as easy to do as CV%. One can fit a polynomial to the data and store it in the software. Then one has a reliable measure of credibility with which to fit each measurement. In addition, there is no need to censor low data any more. One can drive the HIV PCR down to zero, and document it, and not just report <50 copies, for example. This greatly increases the usefulness of the lab to the community.

3. Determine the remaining Environmental Noise. Having determined the assay error, one can determine the remaining environmental error either as an additive or a multiplicative term. Then one can KNOW how much error is due to the assay the how much to the clinical environment in which the study or the patient's care has taken place. This is important information, which is usually not obtained.

4. Evaluate CHANGING Renal Function. Most methods to estimate creatinine clearance or GFR are based only on a single sample. We much favor a longstanding method based on 2 serum creatinine samples which estimates the creatinine clearance in unstable patients which makes serum creatinine rise or fall from the first to the second sample over a stated time, in a patient of stated age, gender, height and weight.

5. Nonparametric (NP) Population PK/PD Models. We greatly favor the use on nonparametric population models, as they make no assumptions at all about the often genetically polymorphic shape of the parameter distributions. What you see is what you've got. It also leads directly to...

6. "Multiple Model" (MM) Design of Doses. Here the multiple support points of the NP model parameter distributions, when given a candidate dosage regimen, provide many predictions, each weighted by the probability of that support point. It then is easy to compute the weighted squared error of failure of those predictions to hit the target, and then to find the specific regimen which minimizes that error.

7. MM, MAP, Hybrid, and IMM Bayesian analysis and individual models. These four Bayesian methods provide tools to obtain useful Bayesian posterior joint densities for patients, even when the patient's parameter densities may lie quite a distance outside the stated ranges of the population model parameters, and also in very unstable patients whose parameter distributions may be changing significantly during the period of data analysis.

Wes Johnson
University of California, Irvine
wjohnson@uci.edu

"Bayesian Nonparametrics: An Overview"

Tutorial Part I: Nonparametric statistics is a misnomer. The goal of nonparametrics is to provide flexible modeling that does not overly rely on assumptions. In order to accomplish that goal, it is often necessary to expand a parametric family of models to include many more models. A parametric family is indexed by a finite parameter vector. A nonparametric family is indexed by an infinite dimensional parameter. Of course in the real world, we can't quite get to infinity, so we truncate to obtain a family that is manageable. But in this finite approximation to an infinite dimensional index, we are left with what are called richly parametric models, that still allow for much flexibility in modeling. The goal of this talk is to present a few methods of handling the complexity. We discuss the Dirichlet Process Mixture prior, which is used to model data as theoretically infinite random mixtures of parametric distributions like the normal. We also discuss Mixtures of Polya Tree priors that allow the user to embed a parametric family, like the normal, in a broader class of distributions. We also briefly discuss a different kind of nonparametrics that models a regression function in a flexible way. All methods are implemented using Markov chain Monte Carlo methods and thus require some finesse in their implementation. However, they do not require large sample theory for either their development or their implementation. Moreover, once the basic machinery is set in place, virtually any inference is achievable virtually by request through a few lines of code rather than by additional mathematics.

Wes Johnson
University of California, Irvine
wjohnson@uci.edu

"On the Value of Incorporating Scientific Input in Modeling and Data Analysis, and How To Do It Without Pain"

We discuss the value of incorporating scientific information that is converted into a prior distribution for a Bayesian analysis of data. Most of the talk will focus on examples from parametric modeling. For example, imagine a simple binomial example where n individuals have been tested for HIV using an ELISA test. The probability of a test positive outcome is the prevalence of HIV in the population sampled times that sensitivity of the ELISA test plus the prevalence on non-HIV infected individuals times the one minus the specificity of the test. There are three parameters and there is only one data point, the proportion of test positives in the sample. Additional scientific input is required to make further progress in this non-identifiable problem. Moreover, many might place uniform priors on some of these inputs in a misguided effort to be non-informative. But it would be impossible to believe that the accuracies of the ELISA test, or indeed the prevalence, were equally likely to be 0 or 1. I will conclude the talk with a discussion of how to extend informative prior information into non and semiparametric models.

Thanasis Kottas
University of California, Santa Cruz
thanos@soe.ucsc.edu

"A Bayesian Nonparametric Modeling Approach for Developmental Toxicology Data"

We present a Bayesian nonparametric mixture modeling framework for replicated count responses in dose-response settings. We explore the methodology for modeling and risk assessment in developmental toxicity studies, where the primary objective is to determine the relationship between the level of exposure to a toxic chemical and the probability of a physiological or biochemical response, or death. Data from these experiments typically involve features that can not be captured by standard parametric approaches. To provide flexibility in the functional form of both the response distribution and the probability of positive response, the proposed mixture model is built from a dependent Dirichlet process prior, with the dependence of the mixing distributions governed by the dose level. The methodology is tested with a simulation study, which involves comparison with semiparametric Bayesian approaches to highlight the practical utility of the dependent Dirichlet process nonparametric mixture model. Further illustration will be provided through the analysis of data from two developmental toxicity studies.

Joint work with Kassandra Fronczyk, Department of Applied Mathematics and Statistics, University of California, Santa Cruz

Wojciech Krzyzanski
University at Buffalo
wk@buffalo.edu

“PK/PD Modeling of the Therapeutic Effects of Erythropoietin”

Pharmacokinetics describes a relationship between drug dose and drug concentrations at various body sites such as blood and urine. Pharmacodynamics aims at establishing a relationship between drug concentrations and pharmacological effects and links these responses to a clinical outcome. PK/PD modeling relates drug exposure to effect by means of mathematical models. Hematopietic cell populations that include white blood cells, red blood cells, and platelets constitute important therapeutic targets and are used as measures of pharmacodynamic response. Erythropoietin is a major hormone that regulates production of RBCs. Recombinant human erythropoietin (rHuEPO) is a therapeutic agent for treatment of anemia.

This talk will briefly introduce a basic concept of mechanistic PK/PD modeling. The physiology of EPO regulated erythopoietic system will be presented. A PK/PD model addressing receptor mediated disposition of rHuEPO in rats will be described. Lifespan based indirect response models will be exemplified by a PD model of rHuEPO effect on reticulocyte, RBC, and hemoglobin levels.. The talk will conclude with discussion of numerical challenges associated with the stiffness of the PK/PD model, difficulties of implementation of delay differential equations in PK/PD software, and estimability of parameters in high dimension PK/PD systems.

Robert Leary
Pharsight, Corp.
Bob.Leary@certara.com

“Computational Challenges of Nonlinear Mixed Effects Analysis for PK/PD Modeling”

NLME analysis of PK/PD models poses a variety of computational challenges in terms of algorithms, models, data handling, software, and hardware. For example, general purpose NLME software such as the NLME package in R and S+, the PROC NLMIXED procedure in SAS, and the BUGS Bayesian inference package are often not well suited for PK/PD analyses due to lack of domain-specific features for handling the complex dosing, data structures, and models often encountered. In this talk I will give an overview of the current state-of-the-art in computational approaches to NLME analysis for PK/PD models as practiced within the commercial and academic pharmacometric communities. Both parametric and nonparametric methods will be discussed in terms of fundamental algorithmic approaches, specific software implementations, and current limitations.

Lang Li
Indiana University
lali@iupui.edu

"PK Parameter Prior Distribution Construction from the Literature"

In this talk, we will talk about a text mining approach to extract PK parameter numerical data from the literature automatically. The extracted data will then be transformed into compartment models. Both linear and non-linear mixed effect models are used for these analyses.

Steve MacEachern
Ohio State University
snm@stat.osu.edu

"Nonparametric Bayesian Models for Related Subpopulations"

One of the fundamental questions that we face when attempting to analyze data is how to model a collection of related subpopulations. Through time, many approaches have been developed. In this talk, I will describe a natural progression from very simple models to more complex models. The progression takes us from parametric models to nonparametric Bayesian models, with a focus on dependent Dirichlet processes. Attention will be drawn to differences between parametric and nonparametric approaches, both for modeling and for inference.

Peter Mueller
MD Anderson
pmueller@mdanderson.org

"Nonparametric Bayesian Models for Population PK/PD"

We will review some nonparametric Bayesian models and approaches that are useful to model random effects distributions in population PK/PD. We will discuss in particular alternative approaches to incorporating regression on baseline covariates, focusing on approaches based on random partition models.

Bhramar Mukherjee
University of Michigan
bhramar@umich.edu

“Bayesian Analysis of Two-phase Studies of Gene-environment Interaction with an Application to Drug-gene Interaction”

In this talk we consider Bayesian semiparametric analysis of two-phase studies of gene-environment interaction. Phase I data is available on case-control status, basic demographic covariates and medication use (environmental factor, say). Phase II genotype data is collected by stratified sampling on case-control status and enriched for use of medication. The genes are selected on the same pathway of the metabolism of the drug. We propose a Bayesian approach based on a retrospective likelihood that data- adaptively uses the gene-environment independence assumption and can accommodate multiple genetic factors and their interactions through a variable selection procedure. The retrospective likelihood requires flexible modeling of the joint distribution of multiple covariates where we adopt Bayesian nonparametric techniques. The methods are illustrated by analyzing interaction of statins with polymorphisms in the cholesterol synthesis and lipid metabolism pathway in a population-based case-control study of colorectal cancer.

Michael Reed
Duke University
reed@math.duke.edu

“Why Metabolism is a Challenge for Statistical Analysis: the Case of Substrate Inhibition”

Cell metabolism is a complicated network of biochemical reactions that includes many feedback and feedforward interactions that are crucial for regulation and adaptation to the cell’s external environment. In addition, the special properties of enzymes give rise to unusual nonlinear kinetics. It is serious and important challenge to construct statistical procedures that can infer such regulatory mechanisms and unusual kinetics from network input-output maps. A simple example, substrate inhibition, will be discussed.

Gary Rosner
Johns Hopkins University
grosner1@jhmi.edu

“Optimal Design with Bayesian Nonparametric Priors”

In this talk, I review an application that concerned finding an individual’s optimal dose of an anticancer drug. Since there were multiple studies, we used a hierarchical model based on linked Dirichlet processes. We also incorporated patient-specific covariates to help determine the optimal dose. I will review several ways to incorporate covariates and point out how these methods differ. This is joint work with Alejandro Cruz-Marcelo and Peter Mueller.

Co-Presenters: William Gillespie, Metrum Inst.; Roger Jelliffe, University of Southern California

Siva Sivaganesan
University of Cincinnati
sivagas@ucmail.uc.edu

"Sample Size: How Many Subjects"

Sampling design is an important aspect of a PK study. It includes specification of the number of sample per subject, sampling times, and the number of subjects or sample size, among other. In this talk we given an overview of this topic focusing, specifically, on the sample size, and the role of Bayesian approach in this context.

Mahlet Tadesse
Georgetown University
mgt26@georgetown.edu

"Stochastic Partitioning and Model-based Clustering with Variable Selection"

In the analysis of high-dimensional data there is often interest in uncovering cluster structure and identifying discriminating variables. For example, the goal may be to use gene expression data to discover new disease subtypes and to determine genes with different expression levels between classes. Another research question that is receiving increased attention is the problem of relating genomic data sets from various sources. For instance, the goal may be to uncover cluster structures in the data sets and identify groups of associated markers across data sets. I will present some Bayesian methods we have proposed to address these problems in a unified manner.

Tatiana Tatarinova
University of Glamorgan
tvtatari@glam.ac.uk

Alan Schumitzky
University of Southern California
schum@math.usc.edu

"Analysis of Biological Time Series Data in Data-Rich Situations"

In this talk we present two clustering methods: 1) The Kullback-Leibler Markov Chain Monte Carlo (KLMCMC) method, and 2) The Multiple Collapse Clustering (MCC) method for the treatment of biological time series data in a data-rich situation. Assuming a continuous nature of most biological processes, we suggest computing parameters of nonlinear, piecewise continuous functions approximating each cluster. Our approach is based on a combination of methods including Gibbs sampling, Random Permutation Sampling, birth-death MCMC, and Kullback-Leibler distance.

Donatello Telesca
University of California, Los Angeles
donatello.telesca@gmail.com

"Modeling Pk/PGx Data"

We describe a general framework for the exploration of the relationships between pharmacokinetic pathways and plymorphisms in genes associated with the metabolism of a compound of interest. We integrate a population pharmacokinetics model with a simple sampling model of genetic mutation via a conditional dependence prior. Significant interactions are selected allowing the pharmacokinetic parameters to depend on gene sets of variable dimension. We discuss posterior inference and prediction based on MCMC simulation. Our model finds motivation in the study of a chemotherapeutic agent used in the treatment of various solid tumors.

Peter Thall
MD Anderson Cancer Center
rex@mdanderson.org

"Prior Elicitation in Bayesian Clinical Trial Design"

The impact of one's prior on adaptive decisions in a Bayesian clinical trial design can be profound. For example, the prior may have a substantive effect on treatment assignments involving life-and-death outcomes in trials of treatments for rapidly fatal diseases. This issue is at the heart of what may be considered the strongest frequentist criticism of Bayesian methods. Despite this, little attention has been paid to the problem of eliciting and calibrating priors in such settings. In this talk, I will discuss how priors were established in a variety of different Bayesian clinical trials in which I have been involved. As time permits, topics will include Dirichlet priors for discrete 2x2 (response, toxicity) outcomes, the notion of prior effective sample size, logistic regression and generalized logistic regression models for toxicity as a function of either the dose of a single agent or the doses of two agents in phase I trials, using penalized least squares to establish a prior for bivariate (efficacy, toxicity) probabilities as a function of dose, eliciting a hyperprior for a hierarchical model in a phase II activity trial, and eliciting priors for event-free survival time and toxicity as functions of patient covariates in a pediatric brain tumor trial.

Paolo Vicini
Pfizer
Paolo.Vicini@pfizer.com

"Identifiability and Prior Regularization: A PK-PD Modeling View"

Biomedical modeling practitioners have always been aware of the need to characterize their models' identifiability properties. This is because in many instances the models'structure and parameters are indirectly estimated from measurements gathered in one or more accessible compartments (e.g. blood plasma). The remainder (inaccessible portion, e.g. peripheral organs) of the model needs to be reconstructed or inferred from such estimation procedures. This presentation will address various concepts relating to identifiability, including clarification of the diverse terms used in the literature. Procedures to avoid common pitfalls when choosing appropriate parameterizations for PK-PD models will be described. These ideas will be also developed in the context of sparse data population analysis with mixed effects models. Examples from the published literature will be used to reinforce these concepts.

#2: Program on Complex Networks Tutorial & Opening Workshop, August 29 – September 1, 2010.

Schedule:

Sunday, August 29

Radisson Hotel RTP

Overview Tutorials

8:00-8:55	Registration and Continental Breakfast
8:55-9:00	Welcome
9:00-10:30	Tutorial Lecture 1: Eric Kolaczyk, Boston University *"Statistical Analysis of Network Data"*
10:30-11:00	Break
11:00-12:30	Tutorial Lecture 2: Alessandro Vespignani, Indiana University *"Diffusion and Epidemic Processes in Complex Techno-social Networks"*
12:30-1:45	Lunch
1:45-3:15	Tutorial Lecture 3: Rick Durrett, Cornell University *"Some Features of the Spread of Epidemics and Opinions on a Random Graph"*
3:15-3:45	Break
3:45-5:15	Tutorial Lecture 4: Michael Mahoney, Stanford University *"Geometric Tools for Identifying Structure in Large Social and Information Networks"*

Monday, August 30
Radisson Hotel RTP

8:00-8:45	Registration and Continental Breakfast
8:45-9:00	Welcome
9:00-12:10	Session on Sampling and Inference Chair: Eric Kolaczyk, Boston University
9:00-9:40	Peter Bickel, University of California, Berkeley *"Statistical Inference for Unlabelled Graphs"*
9:40-10:20	Mark Newman**, University of Michigan** *"Community Structure and Link Prediction in Networks"*
10:20-10:50	Break
10:50-11:30	**Matt Salganik**, Princeton University *"Improvements to the Network Scale-Up Method for Estimating the Sizes of Hard-to-Count Populations"*
11:30-12:10	Panel Discussion Moderator: Stan Wasserman, Indiana University Liza Levina, University of Michigan Bruce Spencer, Northwestern University
12:10-1:30	Lunch
1:30-3:30	Session on Spectral Analysis and Geometric Algorithms Chair: Michael Mahoney, Stanford University
1:30-2:10	Fan Chung Graham, University of California, San Diego *"PageRank Algorithms with Applications to Graph Sparsification and Partitioning"*
2:10-2:50	Aaron Clauset, Santa Fe Institute *"The Trouble with Community Detection"*
2:50-3:30	Mauro Maggioni, Duke University *"Multiscale Analysis of Random Walks on Graphs"*
3:30-4:00	Break
4:00-5:00	Shorter Talks **Chair:** Alun Lloyd**, North Carolina State University**

4:00-4:20	Crystal Linkletter, Brown University *"Explaining Network Structure: The Importance of Modeling Pair-wise Preferences"*
4:20-4:40	Joe Blitzstein, Harvard University *"Respondent-Driven Sampling: Degrees of Uncertainty with Uncertain Degrees"*
4:40-5:00	Alexander Gutfraind, Los Alamos National Laboratory *"Dark Networks and Vital Infrastructure"*
5:00-5:30	Poster Advertisement (2 minute ads)
5:30-6:00	Break
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.*

Tuesday, August 31
Radisson Hotel RTP

8:15-9:00	Registration and Continental Breakfast
9:00-Noon	Session on Percolation and Diffusion on Networks Chair: Rick Durrett, Duke University
9:00-9:40	Zoltan Toroczkai, University of Notre Dame *"Modeling Functional Networks of the Primate Cortex"*
9:40-10:20	Erik Volz, University of Michigan *"Simple Models for Infectious Disease Epidemics in Complex Networks"*
10:20-10:50	**Break**
10:50-11:30	**Pierre Degond**, CNRS *"Continuum models for Complex Systems"*
11:30-12:10	Panel Discussion Moderator: **James Moody**, Duke University **Ginestra Bianconi**, Northeastern University **Mason Porter**, Oxford University

12:10-1:30	Lunch
1:30-4:40	Session on Dynamics of Networks Chair: Steve Fienberg, Carnegie Mellon University
1:30-2:10	Raissa D'Souza. University of California, Davis *"What are Dynamic Networks?"*
2:10-2:50	Tom Snijders, Oxford University *"Actor-oriented Models for Network Dynamics"*
2:50-3:30	Sidney Redner, Boston University *"Dynamics of Voter Models on Heterogeneous Networks"*
3:30-4:00	**Break**
4:00-4:40	Panel Discussion Moderator: Peter Mucha, University of North Carolina Hugh Chipman, Acadia University Josh Socolar, Duke University
4:40-5:40	**Short Talks** Chair: Peter Mucha, University of North Carolina
4:40-5:00	Natallia Katenka**, Boston University** *"The Impact of Partial Information on Network Inference and Characterization"*
5:00-5:20	Edoardo Airoldi**, Harvard University** ***"***_Integer Polytope Samplers with Applications to Network Analysis"_

Wednesday, September 1
Radisson Hotel RTP

8:15-9:00	Registration and Continental Breakfast
9:00-12:10	Session on Bio Applications Chair: Alex Vespignani, Indiana University
9:00-9:40	Eric Xing, Carnegie Mellon University *"Time-Varying Networks: Reverse Engineering and Analyzing Rewiring Social and Genetic Interactions"*
9:40-10:20	Hongzhe Li**, University of Pennsylvania** *"Statistical Methods for Network-based Analysis of Genomic Data".*

10:20-10:50	Break
10:50-11:30	Michelle Girvan, University of Maryland *"Effects of Network Topology in Simple Models of Gene Regulation"*
11:30-12:10	**Desmond Higham**, University of Strathclyde *"Algorithms for Evolving Networks"*
12:10-1:30	**Lunch**
1:30-3:00	**Working Group Formation and Initial Meeting**
3:00-3:30	**Working Group Report and Scheduling for Thursday and Friday**

SPEAKER TITLES/ABSTRACTS

Edoardo Airoldi
Harvard University
airoldi@fas.harvard.edu

"Integer Polytope Samplers with Applications to Network Analysis"

I will describe three integer polytope samplers and will showcase applications in network tomography, multi-way contingency tables, and sampling networks with given degree distribution.

Peter Bickel
University of California, Berkeley
bickel@stat.berkeley.edu

"Statistical Inference for Unlabelled Graphs"

A great deal of attention has recently been paid to determining sub-communities on the basis of relations, corresponding to edges, between individuals, corresponding to vertices of an unlabelled graph (Newman, SIAM Review 2003; Airoldi et al JMLR 2008; Leskovec & Kleinberg et al SIGKDD 2005). We have developed a nonparametric framework for probabilistic ergodic models of infinite unlabelled graphs (PNAS2009)and made some connections with modularities arising in the physics literature and community models in the social sciences.A fundamental difficulty in implementing these procedures is computational complexity.We develop approaches which bypass these difficulties.

(Joint work with Aiyou Chen and Liza Levina)

Joseph Blitzstein
Harvard University
blitzstein@stat.harvard.edu

"Respondent-Driven Sampling: Degrees of Uncertainty with Uncertain Degrees"

Respondent-driven sampling (RDS) is a network sampling design (introduced by Heckathorn) which is widely and increasingly being used to study properties of individuals in a social network, e.g., HIV prevalance. The standard estimators make many strong assumptions, including that the degree d of an individual in the sample is known exactly and that this individual is equally likely to recruit any of his or her d neighbors in the social network.

Classical issues of model-based vs. design-based inference arise here, with new challenges from having also to account for the network structure. We explore the bias-variance tradeoff encountered in this setting, and show how the uncertainty in the measurement of degrees propagates into uncertainty in the RDS estimates.

Joint work with Sergiy Nesterko.

Aaron Clauset
Santa Fe Institute
aaronc@santafe.edu

"The Trouble with Community Detection"

Although widely used in practice, the performance of the popular module identification technique called "modularity maximization" is not well understood in practical contexts. In this talk, I'll show that when applied to networks with modular structure the modularity function Q exhibits extreme degeneracies in which the global maximum is hidden among an exponential number of high-modularity solutions. Time allowing, using a real-world network as an example, I'll show that these degenerate solutions can also be structurally very dissimilar, implying that any particular partition derived from this approach should be treated with caution. Notably, these results explain why so many heuristics perform well in practice at finding high-modularity partitions and why different heuristics often disagree on the modular composition the same network. I'll conclude with some forward-looking thoughts about the general problem of identifying network modules from connectivity data alone, and the likelihood of circumventing this degeneracy problem.

Pierre Degond
CNRS
pierre.degond@math.univ-toulouse.fr

"Continuum models for Complex Systems"

For complex systems with a large number of interacting agents, scalable models (whose computational cost does not depend on the number of particles) are attractive. Examples of such models are kinetic and fluid models, which describe the system by means of continuous quantities such as average densities. The question of rigorously linking agent-based to kinetic or fluid models is therefore crucial. In a first part of the talk, we will propose the derivation of a hydrodynamic model from an agent-based model describing consensus formation. In a second part, the validity of this model in view of the chaos assumption which states that the particles are required to be nearly independent will be discussed. A few illustrative examples in which the chaos assumption is not true will be given.

Raissa D'Souza
University of California, Davis and Santa Fe Institute
raissa@cse.ucdavis.edu

"What are Dynamic Networks?"

Networks are dynamic in a multitude of ways. Dynamical processes, such as diffusion, synchronization, and epidemic spreading, take place on network structures. Likewise network structures themselves can evolve over time, with constituent nodes and edges arriving, leaving, or participating in an intermittent fashion. Finally, the load and flow on a network is dynamic, for instance consider the changing flows on a gene-regulatory network as different compounds are expressed and inhibited, or flows on information and communication networks. These temporal and spatial variations in the flows can radically alter which nodes and edges should be considered the most vulnerable. Here we will review some well known models of dynamical processes on networks and models of dynamics of network topology, concluding with some open questions that we might explore during the year-long program focus on "dynamic networks" including modeling systems of interdependent networks.

Rick Durrett
Cornell University and SAMSI
rtd1@cornell.edu

"Some Features of the Spread of Epidemics and Opinions on a Random Graph"

Random graphs are useful models of social and technological networks. To date most of the research in this area has concerned geometric properties of the graphs. Here we focus on processes taking place on the network. In particular we are interested in how their behavior on networks differs from that in homogeneously mixing populations or on regular lattices of the type commonly used in ecological models. The talk will cover results in Proc. Natl. Acad. Sci. 107 (2010), 4491-4498

Michelle Girvan
University of Maryland
girvan@umd.edu

"Effects of Network Topology in Simple Models of Gene Regulation"

In recent years, discrete state networks (in particular, Boolean networks) have been studied as potientially useful models of gene regulation. An important aspect of these networks is their dynamical stability to small perturbations. Previous approaches to stability have assumed the simplest type of uncorrelated network structure and usually consider two state (Boolean) variables. Real gene networks depart significantly from these idealizing assumptions; e.g., real networks have nontrivial topology significantly different from the previously used random network paradigm. To more directly address real situations, we present a general method for determining the stability of large discrete state networks of logic updated units, allowing for any specified topology, and a variable number of discrete states. We find that stability is governed by the maximum eigenvalue of a modified adjacency matrix, and we test this result by comparison with numerical simulations. We also discuss possible application of our work to experimentally inferred gene networks.

Fan Chung Graham
University of California, San Diego
fanchunggraham@gmail.com

"PageRank Algorithms with Applications to Graph Sparsification and Partitioning"

PageRank is one of the main ways for determining the ranking of webpages by Web search engines. Based on relations in an interconnected network, PageRank has become a major tool for addressing fundamental problems arising in general graphs, especially for large information networks with hundreds of millions of nodes.

The applications of PageRank have grown far beyond its original scope of ranking webpages. We will discuss an improved algorithm for computing personalized PageRank vectors with tight error bounds which can be as small as $O(n^{-k})$ for any fixed integer k. This sharp PageRank algorithm can then be used to derive improved algorithms for graph sparsification and graph partitioning.

Alexander Gutfraind
Center of Non-Linear Studies
ag362@cornell.edu

"Dark Networks and Vital Infrastructure"

"Dark" networks representing the organization of terrorist groups and secret societies are very different from other social networks. Because dark networks must be resilient while also remaining efficient in order to attack, they typically have a sparse hierarchical structure. In this talk, I will show that quantifying and optimizing the objectives of dark networks gives rise to such a structure. Under most circumstances, the optimal dark network is organized as clandestine cells (spy chiefs and terrorist masterminds suspected this a long time ago). I will show how this model can be applied to the design of vital infrastructure networks and other kinds of legitimate networks because they also need to be resilient. Developing computational methods that solve realistic large-scale versions of this design problem remains an open question.

Desmond Higham
University of Strathclyde
djh@:maths.strath.ac.uk

"Algorithms for Evolving Networks"

Many technological applications generate time ordered sequences of networks, defined over a fixed set of nodes; for example daily `who phoned who' networks. Concepts and algorithms for static networks do not immediately carry through to this dynamic setting. For example suppose A and B communicate on day 1, and then B and C communicate on day 2. Information may then pass from A to C, but not vice versa. This subtlety is lost if we simply summarize using the aggregate network given by the chain A-B-C. I will describe some recent ideas that extend widely use node centrality concepts to the evolving case, reflecting the asymmetry imposed by time's arrow. This will allow us to address questions such as: which nodes are the best broadcasters or the best receivers, which nodes are most valuable intermediaries and when is the network most sensitive? I will illustrate these ideas on synthetic and real data sets and briefly discuss related issues of modelling and inference for evolving networks. This is joint work with Ernesto Estrada (Strathclyde) and Peter Grindord (Reading).

Natallia Katenka
Boston University
nkatenka@bu.edu

"The Impact of Partial Information on Network Inference and Characterization"

Association networks represent systems of interacting elements, where a link between two different elements indicates a sufficient level of similarity between element attributes. We study the impact of partial information on network inference and characterization, when only a subset of attributes is available. Quantifying similarity between nodes using canonical correlation, we examine both theoretically and numerically the implications of the choice and number of node attributes on the ability to detect network links and, more generally, to estimate higher-level network summary statistics, such as node degree, clustering coefficients, and measures of centrality. We consider in detail the case of two attributes and discuss generalization of our findings to more than two attributes. Our work is motivated and illustrated in the context of gene/protein regulatory networks in human cancer cells.

Eric Kolaczyk
Boston University
kolaczyk@bu.edu

“Statistical Analysis of Network Data”

In the past decade, the study of networks has increased dramatically. Researchers from across the sciences -- including biology and bioinformatics, computer science, economics, engineering, mathematics, physics, sociology, and statistics -- are more and more involved with the collection and statistical analysis of network-indexed data. As a result, statistical methods and models are being developed in this area at a furious pace, with contributions coming from a wide spectrum of disciplines. In this tutorial I will present a brief overview of key taxonomy, concepts, and tools relevant to the work being done in this area.

Hongzhe Li
University of Pennsylvania
hongzhe@mail.med.upenn.edu

"Statistical Methods for Network-based Analysis of Genomic Data".

Graphs and networks are common ways of depicting dependency information among a set of random variables. In biology, many different biological processes are represented by graphs, such as regulatory networks, metabolic pathways and protein-protein interaction networks. This kind of a priori use of graphs is a useful supplement to the standard numerical data such as microarray gene expression data. How to effectively incorporate the network/graph information into analysis of high-dimensional genomic data raises interesting statistical questions. In this talk, we will present two approaches
for network-based analysis of genomic data: a hidden Markov random field-based method for differential expression analysis and a graph-constrained regularization procedure for variable and module selection. Both methods aim to take into account the neighborhood information of the variables measured on a graph with a Markov random field prior. We demonstrate these ideas using several real gene expression data sets related to cancer, brain aging and human immune responses. We will also briefly mention some future work and further extensions of these methods.

Crystal Linkletter
Brown University
cdlinkle@stat.brown.edu

“Explaining Network Structure: The Importance of Modeling Pair-wise Preferences”

Social networks are useful schematics for representing complex social interactions. Unfortunately, collecting data on who interacts with whom in an entire school, workplace, or city, for example, is an inherently difficult task, so learning what kinds of people interact is important for making broader inference beyond a sampled network. In this talk, we look at the use of conditional independence models, such as latent position models, for modeling the joint probability distribution for network data. To illustrate the importance of constructing and including pair-wise preference measures, we introduce new exploratory data analysis techniques and "fit" measures that can be used to assess what portion of network topology can be explained by observed individual-level attributes. We discuss how ideas from spatial statistics can be borrowed for modeling network data and for efficiently constructing pair-wise preference measures.

Mauro Maggioni
Duke University
mauro@math.duke.edu

"Multiscale Analysis of Random Walks on Graphs"

We present a technique for viewing weighted undirected graphs at multiple scales by homogenizing the natural random walk at different time scales. This leads to multiscale representations of the graph itself, as well as of functions on the graph. We discuss applications to machine learning, in situations where one tries to predict the values of a function on a graph given the values at a few nodes, and to graphs that vary in time.

Michael Mahoney
Stanford University
mmahoney@theory.Stanford.edu

"Geometric Tools for Identifying Structure in Large Social and Information Networks"

The tutorial will cover recent algorithmic and statistical work on identifying and exploiting "geometric" structure in large informatics graphs such as large social and information networks. Such tools (e.g., Principal Component Analysis, spectral methods, and related non-linear dimensionality reduction methods) are popular in many areas of statistics, machine learning, and data analysis due to their relatively-nice algorithmic properties and their connections with regularization and statistical inference. These tools are not, however, particularly well-suited for many large informatics graphs applications since graphs are more combinatorial objects; due to the noise, size, and sparsity patterns of many real-world networks, etc. Recent theoretical and empirical work has begun to remedy this---this work uses "geometric" properties implicit in scalable worst-case approximation algorithms to test for and identify hypothesized structure of interest---and in doing so it has already elucidated in a very precise manner several surprising and counterintuitive properties of very large networks. Topics include: underlying theoretical ideas; tips to bridge the theory-practice gap; empirical observations; and the usefulness of these tools for such diverse applications as community detection, routing, inference, and visualization.

Mark Newman
University of Michigan
mejn@umich.edu

"Community Structure and Link Prediction in Networks"

We investigate two questions of broad interest in the study of networks: the discovery of "communities" (regions of unusually high density within networks) and link prediction (judgements about which links in a network are mostly likely to be either false negatives or false positives). We approach both problems using likelihood-based fits to appropriate random graph models, including stochastic block models and hierarchical models. The overall conclusion is that the models appear to work well in general, giving useful answers to the questions of interest, but that there are some substantive issues still to be resolved.

Sidney Redner
Boston University
redner@bu.edu

"Dynamics of Voter Models on Heterogeneous Networks"

The dynamics of the voter model and the complementary invasion process are investigated on heterogeneous networks. In the voter model, an individual "imports" its state from a randomly-chosen neighbor. Here the average time to reach consensus on a network of N nodes with an uncorrelated degree distribution scales sublinearly with N when the degree distribution is sufficiently broad. We also identify the conservation law that helps characterize the route by which consensus is reached. Parallel results are derived for the invasion process, in which the state of an agent is "exported" to a random neighbor. The probability for a single fitter mutant located at a node of degree k to overspread the population---the fixation probability---is proportional to k for the voter model and to 1/k for the invasion process.

Matthew Salganik
Princeton University
mjs3@princeton.edu

"Improvements to the Network Scale-Up Method for Estimating the Sizes of Hard-to-Count Populations"

Estimating the sizes of hard-to-count populations is a challenging and important problem that occurs frequently in social science, public health, and public policy. This problem is particularly pressing in HIV/AIDS research because the sizes of the most at-risk populations---drug injectors, men who have sex with men, and sex workers---are both difficult to estimate and of critical importance for designing, evaluating, and funding programs to curb the spread of the disease. A promising approach in this area is the network scale-up method which uses information about the social networks of respondents to make population size estimates. This talk presents two generalizations to the network scale-up estimator that help to relax some of its strong implicit assumptions. Further, the talk introduces a new data collection procedure, "the game of contacts," that turns the laborious task of collecting network data into a game-like activity. Finally, results from a study to estimate the number of heavy drug users in Curitiba, Brazil will be presented. While motivated by the problem of population size estimation, some of the methods developed could be used by network researchers more broadly.

Tom Snijders
University of Oxford
tom.snijders@nuffield.ox.ac.uk

"Actor-oriented Models for Network Dynamics"

Actor-oriented models are probability models for dynamics on directed networks where the set of nodes is fixed and the set of ties is random, changing according to a continuous-time Markov process in which tie variables can change only one at a time. Tie changes are represented as being under the "control" of the node for whom this is an out-tie. Ways for specifying such models will be discussed which are similar to generalized linear models, allowing the same extent of flexibility. Estimation will be discussed for the situation that data are collected according to a panel design, i.e., at two or more discrete time points. Various estimation methods have been developed: estimating functions / method of moments, maximum likelihood, and Bayesian. These are being made available in the R package RSiena. Examples will be discussed and points of current and future development.

Zoltan Toroczkai
University of Notre Dame
toro@nd.edu

"Modeling Functional Networks of the Primate Cortex"

Surprisingly little is known about the network properties of the cortex due to lack of data on direct connectivity and its properties. Using brain-wide, extensive (70 man-years work) retrograde tracing experiments in macaque, our anatomist collaborators from Lyon's Stem Cell and Brain Research Group generated a consistent database of interareal connections comprising projection densities (link weights) and physical lengths. This dense network (66% density) is neither a sparse small-world graph nor scale-free. Local connectivity accounts for 79% of labeled neurons. Link weights are highly characteristic across brains, decay exponentially with distance, and follow a heavy-tailed lognormal distribution over 6 orders of magnitude. Weighted network analysis reveals a trade-off between local and global communication efficiencies. A distance rule predicts the binary features, the global and local communication efficiencies as well as the functionally clustered topography of the graph. These findings underline the importance of weight-based hierarchical layering in cortical architecture and processing.

Alessandro Vespignani
Indiana University
tgholbro@indiana.edu

"Diffusion and Epidemic Processes in Complex Techno-social Networks"

In different contexts we are challenged by multi-scale networks where infrastructures composed by different technological layers are interoperating with the social component that drives their use and development. Examples are provided by the Internet, the social Web, the new WiFi communication technologies and transportation and mobility infrastructures. The multi-scale nature and complexity of these techno-social networks are crucial features in the understanding of these systems and the dynamical processes occurring on top of them.

I will present the development of models based on the network paradigm and the reaction–diffusion processes framework that allows their theoretical understanding. As a foremost example I will focus on a class of epidemic models that allow the analysis of the impact of complex mobility networks on the behavior of emergent disease spreading and the general issue of the predictive power offered by computational approaches. In this framework it is possible to tackle foundational issues by using the particle-network approach and provide new mathematical and computational tools for the study of large scale epidemic processes. In particular I will present recent results concerning the 2009 H1N1 pandemic that exemplify how complex networks, physics and computing provide quantitative models for the spreading of emerging diseases.

Erik Volz
University of Michigan
erikvolz@umich.edu

"Simple Models for Infectious Disease Epidemics in Complex Networks"

I review several special cases of epidemics spreading through random networks which reduce to simple solutions based on ordinary differential equations. This reveals a link between traditional mass action models of epidemics in which contacts are instantaneous and uncorrelated, and networks which have dynamically rearranging ties. I then present several applications to HIV. A simple modification to this class of models allows us to model sero-sorting of HIV positive individuals, which is the tendency to rearrange relationships to those with matching infection status. Other extensions are explored, including diffusion in networks with clustering (transitivity), and heterogeneous susceptibility and infectiousness.

Eric Xing
Carnegie Mellon University
epxing@imap.srv.cs.cmu.edu

"Time-Varying Networks: Reverse Engineering and Analyzing Rewiring Social and Genetic Interactions"

While there is a rich literature in modeling static or temporally invariant networks, until recently, little has been done toward modeling the dynamic processes underlying rewiring networks, and on recovering such networks when they are not observable. In this talk, I will present a new formalism for modeling network evolution over time based on temporal exponential random graphs, and several new algorithms for estimating the structure of time evolving probabilistic graphical models underlying nonstationary time-series of nodal attributes. I will show some promising results on recovering the latent sequence of evolving social networks in the US Senate based it voting history, and the gene networks during the life cycle of fruit fly from microarray time course, at a time resolution only limited by sample frequency.

#3: Program on Analysis of Object Data Tutorial & Opening Workshop, September 12 – 15, 2010

Schedule:

Sunday, September 12
Radisson Hotel RTP

Overview Tutorials

8:00-8:55	Registration and Continental Breakfast
8:55- 9:00	Welcome
9:00-10:30	Tutorial Lecture 1: Functional Data Analysis and Related Topics **Fang Yao**, University of Toronto *"Functional Data Analysis and Related Topics"*
10:30-11:00	Coffee Break
11:00-12:30	Tutorial Lecture 2: Dynamics **Jim Ramsay**, McGill University *"Models for Output-Buffered Systems: An Introduction to Dynamics"*
12:30 –1:45	Lunch
1:45- 2:45	Tutorial Lecture 3: Images **Sarang Joshi**, University of Utah *"Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychol"* **Martin Lindquist**, Columbia University *"The Statistical Analysis of Neuroimaging Data"*
2:45-3:00	Coffee Break
3:00-4:00	Tutorial Lecture 4: Shapes and Manifolds **John Kent**, University of Leeds *"Shape Analysis and Manifold Data"*
4:00-4:15	Coffee Break
4:15-5:-15	Tutorial Lecture 5: Trees **J.S. Marron**, University of North Carolina, Chapel Hill *"Tutorial on Trees as Data"*

Monday, September 13
Radisson Hotel RTP

8:00-8:30	Registration and Continental Breakfast

8:30-8:45	Welcome
8:45-12:15	Session on FDA Chair: **Hans Mueller**, University of California, Davis
8:45-9:30	**Alois Kneip**, University of Bonn *"Challenges for FDA Theory and Methodology"*
9:30-10:15	**Naisyin Wang**, University of Michigan "*Challenges for FDA in Longitudinal Studies*"
10:15-10:40	Break
10:40-11:00	**Shuang Wu**, University of Rochester *"FDA for Non-traditional Data"*
11:00-11:45	**Ciprian Crainiceanu**, Johns Hopkins University *"My first 100 Terabytes of Data: Challenges for FDA Modeling"*
11:45-12:15	Discussion
12:15-1:45	Lunch
1:45-2:30	Minute Madness (1-3 min talks) Moderator: **J.S. Marron**, University of North Carolina
2:30-5:15	Session on Dynamics Chair: **Jim Ramsay,** McGill University
2:30-3:15	**Aaron King**, University of Michigan *"Plug-and-play Inference for Stochastic Dynamical Systems"*
3:15-4:00	Break
4:00-4:45	**Sy-Miin Chow**, University of North Carolina *"Applications of Differential Equation Modeling in the Social Sciences"*
4:45-5:15	Discussion
5:15-5:45	Poster Advertisement Session (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.*

Tuesday, September 14
Radisson Hotel RTP

8:15-8:45	Registration and Continental Breakfast
8:45-12:15	Session on Images Chair: **Jane-Ling Wang**, University of California, Davis
8:45-9:30	**Dubois Bowman**, Emory University *"Statistical Modeling of Brain Imaging Data: An Overview, Challenges, and Future Directions"*
9:30-10:15	**Laurent Younes**, Johns Hopkins University *"Shape Analysis of Diastolic/Systolic Paired Cardiac Images"*
10:15-10:40	Break
10:40- 11:00	**Armin Schwartzman**, Harvard University *"Data objects in Diffusion Tensor Imaging"*
11:00-11:45	**Jonathan Taylor**, Stanford University *"Overview of Statistical Inference, Multiple Comparisons and Gaussian Random Fields for Neuroimaging Data"*
11:45-12:15	Discussion
12:15-1:45	Lunch
1:45-2:30	Minute Madness (1-3 min talks) Moderator: **Jim Ramsay**, McGill University
2:30-5:15	Session on Trees Chair: **J.S. Marron**, University of North Carolina, Chapel Hill
2:30-2:50	**Burcu Aydin**, Hewlett Packard Research *"Principal Component Analyses for Trees"*
2:50-3:10	**Sean Skwerer**, University of North Carolina *"Analysis of Object Data: Averaging Metric Trees"*
3:10-3:40	Break
3:40-4:00	**Shanker Bhamidi**, University of North Carolina *"Dyck Path Correspondence and the Statistical Analysis of Brain Vascular Networks"*
4:00-4:45	**Haonan Wang**, Colorado State University *"Smoothing & Branching Process Inference on Trees"*
4:45-5:15	Discussion

Wednesday, September 15
Radisson Hotel RTP

8:15-8:45	Registration and Continental Breakfast
8:45-12:15	Session on Shapes and Manifolds Chair: **Ian Dryden**, University of Nottingham
8:45-9:30	**Huiling Le**, Nottingham University *"Some Aspects of Statistics on Riemannian Manifolds from the Perspective of Shape Analysis"*
9:30-10:15	**Anuj Srivastava**, Florida State University *"Towards Statistical Modeling of Shapes of Curves and Surfaces"*
10:15-10:40	Break
10:40-11:00	**Victor Panaretos**, EPFL Lausanne *"Statistical Shape and Random Tomography in Structural Biology"*
11:00-11:45	**Alain Trouve**, Ecole Normale Supérieure de Cachan *"From Shape Comparison to Shape Evolution with a Geometrical and Statistical Perspective"*
11:45-12:15	Discussion
12:15-1:45	Lunch
1:45–2:45	Working Group Formation and Initial Meeting
2:45-3:00	Break
3:00-3:45	Working Group Reports and Scheduling for Thursday and Friday

Thursday, September 16, and Friday, September 17

SAMSI: Individual Working Group meetings

SPEAKER TITLES/ABSTRACTS

Burcu Aydin
Hewlett Packard Research
aydin@hp.com

"Principal Component Analyses for Trees"

The active field of Functional Data Analysis (about understanding the variation in a set of curves) has been recently extended to Object Oriented Data Analysis, which considers populations of more general objects. A particularly challenging extension of this set of ideas is to populations of tree-structured objects. We develop analogs of Principal Component Analysis for trees, based on the notion of tree-lines, and propose numerically fast algorithms to solve the resulting problems. The solutions we obtain are used in the analysis of a data set where each data object is a tree of blood vessels in one person's brain. Our analysis revealed a significant relation between the age of the individuals and their brain vessel structure.

Shanker Bhamidi
University of North Carolina
bhamidi@email.unc.edu

"Dyck Path Correspondence and the Statistical Analysis of Brain Vascular Networks"

The Dyck path correspondence is a very powerful technique in combinatorics, allowing one to map trees to paths (i.e. functions as analyzed in FDA) and vice-versa. In the last few years this technique has proven to be extremely powerful in theoretical probability in analyzing many different random tree models. Random tree models arise in a wide array of fields including combinatorics (e.g. spanning trees of various graphs, uniform random trees on labeled vertices and so on), computer science (e.g. search trees) probability (branching process models) and biology (e.g. phylogenetics). Understanding the asymptotics of these models, namely what happens for trees on a large number of vertices is of paramount importance. The Dyck path correspondence provides an extremely powerful technique, allowing us to view these trees as functions and then proving "Brownian motion" type convergence results, and in particular showing the convergence of a wide array of finite random tree models to random fractals such as the continuum random tree. Such results for example allow us to conclude that in many random tree models, the height of the tree, properly rescaled, converges to the height of a standard Brownian excursion. In this talk I will report on preliminary work done in analyzing the structure of brain artery networks from Magnetic Resonance Angiography brain images provided by the CASILab at UNC Chapel Hill. Applying FDA techniques to the Dyck path correspondence version of the data, provides a method of analyzing the contribution of various factors such as age and gender on the variation in the data.

This is joint work with J.S. Marron, Dan Shen and Haipeng Shen at UNC Chapel Hill.

DuBois Bowman
Emory University

dbowma3@sph.emory.edu

"Statistical Methods for Modeling Functional Brain Imaging Data"

Functional neuroimaging studies attempt to identify spatially localized brain regions that drive the execution of experimental tasks targeting, for example, behavior, cognition, or emotion, and sometimes to additionally make links to particular neurological or psychiatric disorders. There are numerous challenges in analyzing functional neuroimaging data. Brain imaging studies using functional magnetic resonance imaging (fMRI) produce massive data sets comprised of 3-D movies of neural activity for each subject, with each 3-D scan (movie frame) containing hundreds of thousands of spatially localized measurement sites (voxels). Also, human brain function, viewed through fMRI,
presents complex patterns of spatial and temporal correlations. In this talk, we consider fMRI data from a study of depression. We discuss a Bayesian hierarchical model that accounts for spatial correlations between voxels, both within and between designated neuroanatomic structures, while also capturing temporal correlations between repeated scans for each subject. Our model provides a framework to detect localized (voxel-level) and regional neural correlates of depression and measures of functional connectivity.

Sy-Miin Chow
University of North Carolina
symiin@email.unc.edu

"Applications of Differential Equation Modeling in the Social Sciences"

The past decade has evidenced the increased prevalence of irregularly spaced longitudinal data in social sciences. Often lacking, however, are practical tools that allow researchers to effectively map a continuous-time model (e.g., ordinary or stochastic differential equations) onto a set of noisy, discrete-time data. A stochastic approximation expectation-maximization algorithm for fitting multivariate linear and nonlinear differential equations with random effects is presented. This approach is used to analyze the ambulatory cardiovascular changes of 182 participants during a typical workday. Pertinent methodological challenges and unresolved issues are discussed.

Ciprian Crainiceanu
Johns Hopkins University
ccrainic@jhsph.edu

"My first 100 Terabytes of Data: Challenges for FDA Modeling"

Several large studies that contain biosignals and imaging data at multiple visits will be described. I will present several methods developed to address the challenges raised by these new data types. In particular, I will describe Multilevel and Longitudinal Functional Principal Component Analysis (M- and LFPCA), Penalized Functional Regression (PFR) and Population Value Decomposition (PVD). The many remaining open problems in this new area of research will be described.

Sarang Joshi

University of Utah
sjoshi@sci.utah.edu

"Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychol"

The purpose of this study is to characterize the neuroanatomical variations observed in neurological disorders such as dementia. We do a global statistical analysis of brain anatomy and identify the relevant shape deformation patterns that explain corresponding variations in clinical neuropsychological measures. The motivation is to model the inherent relation between anatomical shape and clinical measures and evaluate its statistical significance. We use Partial Least Squares for the multivariate statistical analysis of the deformation momenta under the Large Deformation Diffeomorphic framework. The statistical methodology extracts pertinent directions in the momenta space and the clinical response space in terms of latent variables. We report the results of this analysis on 313 subjects from the Mild Cognitive Impairment group in the Alzheimer's Disease Neuroimaging Initiative (ADNI).

John Kent
University of Leeds
j.t.kent@leeds.ac.uk

"Shape Analysis and Manifold Data"

The "shape" of a geometric object in 2 or 3 dimensions consists of the information remaining after removing the effects of translation, rotation and scaling. This talk will focus mainly on landmark-based shape analysis. But it is also of interest to consider outlines, surfaces and solid objects.

Since shapes are equivalence classes of configurations, special methods are needed for their study. From a local perspective, shape analysis is a modified version of multivariate analysis. However, from a more global point of view, shape space can be viewed as a manifold, and ideas from Riemannian geometry become relevant. The study of data on manifolds has a long history in Statistics. Examples include directions and axes on the circle and the sphere, rotation matrices and projection matrices.

Aaron King
University of Michigan
kingaa@umich.edu

"Plug-and-play Inference for Stochastic Dynamical Systems"

An increasingly fruitful approach to ecological inference queries time-series data using stochastic dynamical systems (AKA state-space models, partially-observed Markov processes). Rigorous inference using such models has traditionally been extremely challenging. Most existing methods place severe restrictions on the form of the models that can be entertained. "Plug-and-play" methods, by contrast, require only simulation and are thus free of such

restrictions. These methods accelerate scientific progress by allowing one to entertain and compare multiple competing hypotheses. I will point out several of these methods and describe one, Iterated Filtering, in some detail, using ecological examples to show that one can use it to ask and answer questions previously
unaddressable. Along the way, I will introduce a software package, pomp, that implements a number of plug-and-play methods.

Alois Kneip
University of Bonn
akneip@uni-bonn.de

"Challenges for FDA Theory and Methodology"

The talk first summarizes some basic results in functional principal component analysis which has to be considered as a key technique in functional data analysis. It is then shown that based on high-dimensional factor models some closely related procedures can be useful tools when dealing with more general data structures. The analysis of econometric panel data provides an important example. A final part of the talk will focus on high-dimensional linear regression. Using a factor approach it is possible to combine the points of view of functional regression and commonly used procedures for variable selection like the Lasso.

Huiling Le
Nottingham University
Huiling.Le@nottingham.ac.uk

"Some Aspects of Statistics on Riemannian Manifolds from the Perspective of Shape Analysis"

We present some techniques for statistical analysis of data on manifolds, illustrated by progress made in the context of shapes determined by `landmarks'. We then discuss how concepts such as Euclidean means need to be adapted to account for features of the geometry, such as curvature, and how change of data over time can be modelled.

Martin Lindquist
Columbia University
martin@stat.columbia.edu

"The Statistical Analysis of Neuroimaging Data"

Functional neuroimaging has experienced an explosive growth in recent years. Currently there exist a number of different imaging modalities that allow researchers to study the physiological changes that accompany brain activation. Each of these techniques has advantages and disadvantages and each provides a unique perspective on brain function. In this tutorial we discuss the statistical analysis of neuroimaging data, with a specific focus on functional magnetic resonance imaging (fMRI) data. We discuss the signal and noise properties of the data, as well as the common research objectives of fMRI studies. These include localizing regions of the brain activated by a task, determining distributed networks that correspond to brain function and making predictions about psychological or disease states.

Steve Marron
University of North Carolina
marron@email.unc.edu

"Tutorial on Trees as Data"

Motivation of and introduction to the field of Tree Structured Data Objects.

Victor Panaretos
EPFL Lausanne
victor.panaretos@epfl.ch

"Statistical Shape and Random Tomography in Structural Biology"

We investigate how the notion of shape arises naturally in the electron microscopy of biological macromolecules. In particular, single particle electron microscopy is a modern technique that biophysicists employ to learn the structure of proteins, yielding data that consist of noisy random projections of the protein structure in random directions, with the added complication that the projection angles cannot be observed. This gives rise to an inverse problem that is qualitatively different from the standard problem of tomography, where reconstructions depend crucially on knowledge of the projection angles. One may thus ask what kind of information they can hope to recover given single particle data. Strictly speaking, the problem is unidentifiable and an appropriate reformulation is suggested hinging on ideas from Euclidean Shape Theory. Within this setup, we will explore how a consistent solution to the problem may be derived, without attempting to estimate the unknown angles, provided that the density is assumed to admit a mixture representation. From the practical point of view this can be understood as being able to provide a low-resolution reconstruction, which can already serve as a useful basis for further biophysical investigations.

Anqi Qiu
National University of Singapore
bieqa@nus.edu.sg

"Multi-Manifold Diffeomorphic Brain Mapping"

This talk introduces a variational problem under the setting of large deformation diffeomorphic metric mapping (LDDMM) for whole brain mapping when images, landmarks, gyral/sulcal curves, and cortical surfaces are simultaneously carried from one subject to the other through a flow of diffeomorphisms. Its Euler-Lagrange equation is described in terms of momentum, a linear transformation of the velocity vector field of the diffeomorphic flow. The numerical implementation for solving this variational problem, which involves kernel application in an irregular grid, is made feasible by the introduction of a class of computationally friendly kernels. This algorithm is applied to register 40 magnetic resonance (MR) brain images and evaluated based on anatomical local and global features. We present comparisons of mapping accuracy among LDDMM, FreeSurfer, and CARET.

James Ramsay
McGill University
ramsay@psych.mcgill.ca

"Models for Output-Buffered Systems: An Introduction to Dynamics"

How systems transform streams of input information into one or more output behaviours is a preoccupation over virtually all the sciences, whether pure or applied. In many of these fields, models tend to be defined in terms of systems of differential equations, or dynamic systems. Until recently, statisticians had little to say about how to fit such models to data, or about how to draw inferences from data on such systems.

In this talk, we introduce continuous dynamic systems as a relatively simple variation of a standard regression problem indexed by time. An approach to fitting such systems to real data will be outlined that does not require the system of differential equations to be solved, either analytically or numerically. This approach also leads to optimal parameter estimation and inferential procedures, and is therefore essentially statistical rather than numerical in nature.

Data on a variety of real-world input/output situations are used to illustrate what can be achieved by proposing that inputs are buffered in a variety of ways so as to transfer sharp changes in input to smooth output responses.
The talk will go on to highlight the most common modifications of linear dynamic systems that render them nonlinear, and indicate some of the consequences of nonlinearity.

Sean Skwerer
University of North Carolina
sskwerer@unc.edu

"Analysis of Object Data: Averaging Metric Trees"

The non-positivily-curved tree space introduced in 2001 by Billera et al. gives a geometric distance between trees of non-negative edge length. Generally in NPC spaces, any two points will have a unique geodesic path between them; and a set of points will have a unique mean - a point that minimizes the sum of squared geodesic distances to every point in the set. We are currently working on efficient algorithms for computing the mean tree based on the framework established by Owen, Provan and Miller. Applications include analysis of blood vessel structures in the human body, and phylogenetics.

Anuj Srivastava
Florida State University
asrivastava@fsu.edu

"Towards Statistical Modeling of Shapes of Curves and Surfaces"

Shape analysis of 2D and 3D objects, both static and dynamic, is important in many applications, ranging from graphics, biometrics, and military to medical. Restricting to the boundaries of these objects, we are interested in tools for analyzing shapes of curves and surfaces. In particular we seek probability models that can capture typical variability within shape classes of interest, and can be used for shape estimation, clustering, classification, and tracking. Although parametrized curves and surfaces are efficient in representing objects boundaries, compared to say level sets or embedding in larger spaces, there is an important need to be invariant to parametrization. This is often accomplished using representations and metrics such that the actions of the re-parametrization groups are by isometries. While the techniques for studies of curves are quite developed, similar ideas for surfaces are more recent and deserve attention. I will describe some preliminary results for parametrization-invariant shape analysis of surfaces. In terms of stochastic modeling, the main idea is to use finite-dimensional subspaces of the tangent bundles to define probability models. These models are estimated from training shapes and can be easily sampled from for use in Monte Carlo type inferences. However, for use in more advanced inferences, it is desirable to have analytical expressions for probability density functions on shape spaces. I will discuss some ideas and issues in deriving analytical models on some spherical manifolds associated with shapes of curves.

Jonathan Taylor
Stanford University
jonathan.taylor@stanford.edu

"Overview of Statistical Inference, Multiple Comparisons and Gaussian Random Fields for Neuroimaging Data"

In this talk, we give an overview of some of the approaches to signal detection and population level inference in neuroimaging studies. Some of the approaches involve: Gaussian Random Field Theory (RFT); permutation methods; False Discovery Rate (FDR). These methods are applied to an fMRI neuroeconomics experiment. Time permitting, we will also discuss some aspects of prediction in fMRI experiments.

Alain Trouvé
CMLA ENS-CACHAN
trouve@cmla.ens-cachan.fr

"From Shape Comparison to Shape Evolution with a Geometrical and Statistical Perspective"

During the last decades, the study of shape spaces has been driven by the accelerated development of imaging techniques in biomedical engineering and the emergence of computational anatomy.

We will first explain how the concepts of homogeneous spaces and riemannian manifolds embedded in the large deformation diffeomorphic metric mapping setting (LDDMM) have been a powerful and effective framework to support local statistical analysis in more and more complex shape spaces.

Then we will discuss some important issues about growth modeling and longitudinal studies that are still today major challenges from a statistical perspective. We will explain how the Hamiltonian formulation of geodesic evolution in shape spaces can be the source of an interesting family of deterministic and stochastic second order evolutions extending the concept of splines to shape spaces.

Haonan Wang
Colorado State University
wanghn@stat.colostate.edu

"Some Statistical Analysis when Data Are Tree-structured Objects"

Object oriented data, such as tree-structured data, random graphs, manifold data and curve data, are frequently collected in many scientific studies. Traditional statistical models for multivariate data are built under Euclidean space setting. However, the elements of object oriented data analysis reside in non-Euclidean spaces such as Lie groups, or more complex spaces such as spaces of tree-structured data. For example, two blood vessel systems differ in terms of topological structures and geometric properties (i.e., overall length, number of branches, and branching orientation). A mathematical framework for statistical analysis of object oriented data has been developed. This includes measures of centrality, variability, and regression analysis. Notions of lines and curves are also developed. The methodology is illustrated through applications to the analysis of brain blood vessel data.

Naisyin Wang
University of Michigan
nwangaa@umich.edu

"Challenges for FDA in Longitudinal Studies"

In longitudinal studies, repeated measurements from a selected group of subjects could be recorded at irregular and small to moderate number of time points. When it is assumed that these measurements are measured-with-noise surrogates of smooth underlying process, various approaches that have been developed in the FDA literature provide powerful tools to understand the key features of the longitudinal data. In this talk, we focus on present certain considerations regarding assumptions and choices of approaches that could play a crucial role in the use of FDA methods in analyzing longitudinal data. These considerations would be briefly illustrated from both theoretical considerations and numerical implementation.

Shuang Wu
University of Rochester
Shuang_Wu@URMC.Rochester.edu

"FDA for Non-traditional Data"

The classical theory and methodology of Functional Data Analysis (FDA) has been developed for fully observed or densely sampled functions. In this talk, the focus is on the case of sparsely and irregularly sampled noisy functional data. The method of Functional Principal Component Analysis (FPCA) for the nonparametric analysis of random trajectories in this sparse situation is utilized to extend the classical functional linear regression models to the case of sparsely sampled functional responses, leading to "response-adaptive regression" models. Another example of non-traditional functional data is provided by samples consisting of series of rare events, where for each object in the sample one observes a small number of repeated event times over a fixed time interval. Viewing the underlying intensity functions that generate the event times as random functions leads to a functional FPCA based approach to analyze such data.

Fang Yao
University of Toronto
fyao@utstat.toronto.edu

“Functional Data Analysis and Related Topics”

In this tutorial, I will introduce the concept of functional data analysis (FDA) that has attracted substantial research interest. The data considered in FDA are in the form of repeated measurements over a period of time for a sample of subjects. These data may be considered as observations generated from the realizations of some latent or observed stochastic processes. Following a gentle introduction of the general framework, I will discuss various approaches that have been studied in literature to deal with such complex data structures, such as dimensional reduction through functional principal components, penalized methods using splines, regression models with flexibility to different extent, functional derivatives and etc.

Laurent Younes
Johns Hopkins University
laurent.younes@jhu.edu

“Shape Analysis of Diastolic/Systolic Paired Cardiac Images”

The presentation will describe an approach that deals with grouped images of deformable objects (here, pairs of cardiac images) in two steps, where the first step analyzes shape variations within groups in a local, group-dependent exponential chart, and the second uses parallel transport to normalize these results in a reference chart (template) before running statistical analyzes. The whole approach, including the Riemannian structure on deformable images that specifies exponential charts and parallel transport, is embedded in the "Diffeomorphic Metric Mapping" paradigm that represents shape spaces as homogeneous spaces under the action of diffeomorphisms.

We will present experimental results using this method on a data set consisting of CT cardiac images, where the goal of the analysis is to compare systolic/diastolic paired images across subjects affected by different types of heart disease.

#4: Program on Space-time Analysis for Environmental Mapping, Epidemiology & Climate Change – Spatial Transition Workshop, October 11-13, 2010

Schedule:

Monday, October 11
SAMSI

8:15-8:30	Registration and Continental Breakfast
8:30 -9:00	Welcome **Richard Smith**, SAMSI **Montse Fuentes**, North Carolina State University
9:00- 10:00	**Noel Cressie**, Ohio State University *"Spatio-Temporal Statistical Modeling"*
10:00-10:20	Break
10:20-11:20	Paleoclimate: **Martin Tingley**, NCAR *"Piecing Together the Past: Statistical Insights into Paleoclimatic Reconstructions"* **Elizabeth Mannshardt-Shamseldin**, Duke University *"Paleoclimate Extremes in Proxy Data"*
11:20-12:20	Spatial Health: **Howard Chang**, SAMSI and Duke University *"Impact of Climate Change on Ambient Ozone Level and Mortality in Southeastern United States"* **Catherine Calder**, Ohio State University *"Bayesian Inference for Incomplete Marked Spatial Point Patters: Estimating Individual Activity Spaces"*
12:20-1:40	Lunch
1:40-2:40	Deterministic/Stochastic: **Peter Kramer**, RPI *"Improving Statistical Components of Multi-scale Simulation Schemes"* **Emily Kang**, SAMSI *"Computational Strategies for Spatio-Temporal Filtering in Multiscale Systems"*
2:40-3:40	Computation and Visualization: **Renato Assuncao**, Minas Gerais *"A Point Process Model for Clustering Including Covariates"*

	Bruno Sanso, University of California, Santa Cruz *"Comparing and Blending Regional Climate Model Predictions for the American Southwest"*
3:40-4:00	Break
4:00-5:00	Spatial Extremes: **Raphael Huser**, EPFL *"Space-Time Modelling of Extreme Events and Inference for Max-Stable Processes"*
	Ben Shaby, SAMSI *"Approximate Bayesian Computation for Spatial Extremes via Open-faced Sandwich Adjustment"*

Tuesday, October 12
SAMSI

8:30-9:00	Registration and Continental Breakfast
9:00-10:00	Invited Lecture **Amy Herring**, University of North Carolina *"Air Pollution and Reproductive Outcomes: Opportunities for Increased Research and Translation"*
10:00-11:00	Foundations: **Marco Ferreira**, University of Missouri *"Dynamic Multiscale Spatio-Temporal Models for Areal Data"*
	Garritt Page, Duke University *"The Effect of Spatial Confounding on Covariate Estimation"*
11:00-11:20	Break
11:20-12:20	Geostatistics: **Jo Eidsvik**, NTNU *"Estimation and Prediction for Spatial Data using Composite Likelihood"*
	Brian Reich, North Carolina State University *"Nonparametric Spatial Models for Extremes"*
12:20-1:40	Lunch
1:40-2:40	Point Patterns: **Athanasios Micheas,** University of Missouri *"Bayesian Hierarchical Modeling for Point Processes with Covariate Information using Mixtures"*
	Avishek Chakraborty, Texas A&M University *"Imputing confidential data using point process"*

2:40-3:40 NonGaussian/Nonstationary:
Veronica Berrocal, University of Michigan
"The Predictive Spatial Dirichlet Process with Application to Downscaling"

Michele Guindani, University of Texas
"A Class of Covariate-Dependent Spatiotemporal Covariance Functions for the Modeling of Ozone"

3:40-4:00 Break

4:00-6:00 Reception and Poster Session

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.

Wednesday, October 13
SAMSI

8:30-9:00 Registration and Continental Breakfast

Summary and Future Work by Working Group Leaders (with discussion):

9:00 – 10:40 *"Paleoclimate Reconstructions"*
Bala Rajaratnam, Stanford University

"Spatial Exposures and Health Effects"
Catherine Calder, Ohio State University

"Interaction of Deterministic and Stochastic Models"
Murali Haran, Pennsylvania State University

"Computation and Visualization"
Renato Assuncao, Minas Gerais

"Spatial Extremes"
Richard Smith, SAMSI

10:40-11:00 Break

11:00-12:20 *"Fundamentals"*
Dongchu Sun, University of Missouri

"Geostats"
Sudipto Banerjee, University of Minnesota

"Spatial Point Processes"
Alan Gelfand, Duke University

"Non-Gaussian and Non-Stationary Spatial Models"
David Dunson, Duke University

12:20-12:30 Closing Remarks
Alan Gelfand, Duke University

12:30 Lunch

SPEAKER TITLES/ABSTRACTS

Renato Assuncao
Universidade Federal de Minas Gerais
assuncao@est.ufmg.br

"A Point Process Model for Clustering Including Covariates"

We present a point process model that allows for interaction between the events and for the introduction of covariates. The main aim is to generate patterns with clustering that are due both, enviromental heterogeneity and real interaction between the particles. The covariates can be modelled in a way similar to regression models and therefore parameters can be easily estimated.

Sudipto Banerjee
University of Minnesota
banr009@umn.edu

"Summary of Activities: The GEOSTAT Working Group"

The GEOSTATISTICS working group was led by Sudipto Banerjee (UMN), Brian Reich (NCSU) and Alan E. Gelfand (Duke). This group will explore spatial and spatiotemporal models for geostatistical data analysis with special emphasis on issues relating to preferential sampling. After the opening workshop and a series of weekly meetings, the working group identified some important issues for initial research. Given the larger size and the broader spectrum of topics related to these topics, the working group was further subdivided into smaller subgroups, with each subgroup focusing upon one of the following topics:

a) Preferential sampling: Preferential sampling refers to settings where the process that determines the data locations and the process being modeled are stochastically dependent.

b) Spatial sampling designs: This working group focused upon methodological developments in spatial sampling design including the study of "efficiency estimates" that demonstrate how good sampling designs actually are.

c) Estimation techniques for geostatistical models: Given the increasing presence of large geostatistical datasets in spatial and spatial-temporal analysis, this group devoted time to efficient estimation methods for geostatistical models focusing, in particular, upon

"Integrated Nested Laplace Approximation" (INLA) and other related algorithms to estimate spatial process models.

d) Ozone process models: This group focused upon building spatial-temporal models for ozone processes and devising methods that can help us achieve better estimates and prediction including the development of active models that combine information from different spatial scales.

e) Numerical model validation: This group used geostatistical methods to study the discrepancy between computer model output and observed data in different settings. The objective is to uncover spatial locations, seasons, and meteorological conditions in which the model does and does not perform well.

f) Feedback issue: This group will focus upon combining multiple sources of data: 2-stage modeling approaches versus full Bayesian approaches.

The activities of these subgroups will be summarized and presented.

Veronica Berrocal
University of Michigan
berrocal@umich.edu

“The Predictive Spatial Dirichlet Process with Application to Downscaling"

The spatial Dirichlet process is a flexible mixture model process that allows to represent a spatial process without assuming Gaussianity or stationarity. Prediction of spatial DPs at unobserved locations requires specifying the locations of the sites at which prediction is sought prior to fitting the model. Here, we propose a method to circumvent these difficulties by implementing the predictive process approach of Banerjee et al. (2008). In particular, our predictive spatial DP uses the nodes of the predictive process as the sites associated with the atoms of the spatial DP. Then, prediction at any site is handled deterministically, as in the predictive process framework. As an application, we present our method in the context of downscaling, extending the downscaler of Berrocal et al.
(2010).

Catherine Calder
Ohio State University
calder@stat.osu.edu

“Bayesian Inference for Incomplete Marked Spatial Point Patters: Estimating Individual Activity Spaces”

Co-author: William F. Darnieder, Ohio State University

Geographers use the term activity space to describe the totality of an individual's spatial and temporal day-to-day interaction with his/her environment. Advanced GIS-based measures of an individual's activity space rely on complete activity pattern data (i.e, information about the locations of and time spent on all regular activities). Eliciting complete activity pattern data as part of large-scale sample surveys is typically not feasible due to the costly and time-consuming nature of data collection. Instead, survey participants may be questioned only about the locations of and time spent participating in common activity categories (e.g., working, grocery shopping, visiting relatives). Our research seeks to develop a formal statistical framework for inferring an individual's activity space using survey data that incompletely characterizes participants' activity patterns. We propose a Bayesian statistical model for imputing unobserved activities from the information available about known activities. Our model captures the inherent connection between the spatial and temporal aspects of human activity by assuming activities arise according to a spatial inhomogeneous marked point process, where the marks represent the time spent at a particular location. In addition to providing a general overview of our proposed methodology, we also illustrate our approach using activity pattern data collected as part of the Los Angeles Family and Neighborhood Survey.

Avishek Chakraborty
Texas A&M University
deptaccount@gmail.com

"Imputing Confidential Data Using Point Process"

We address the issue of protecting confidential datasets with location information of the subjects. The usual imputation techniques often interchange the records in an arbitrary way that often makes the released datasets unfit for any modelling or association studies. We suggest a general method based on spatial point pattern to replace the actual dataset with a number of imputed datasets. The 'goodness' of imputed data can be tested with emperical statistics as well as model-based inferences. The confidentiality of the records requires analysis under different degrees of 'information disclosure'. Also highlighted were several tuning parameters present in our model that can be used to control the degree of confidentiality of the released datasets.

Howard Chang
Duke University / SAMSI
howard.chang@duke.edu

"Impact of Climate Change on Ambient Ozone Level and Mortality in Southeastern
United States"

There is a growing interest in quantifying the health impacts of climate change. This paper examines the risks of future ozone levels on non-accidental mortality across 19 urban communities in Southeastern United States. We present a modeling framework that integrates data from climate model outputs, historical meteorology and ozone observations, and a health surveillance database. We first modeled present-day relationships between observed maximum daily 8-hour average ozone concentrations and meteorology measured during the year 2000. Future ozone concentrations for the period 2041 to 2050 were then projected using calibrated climate model output data from the North American Regional Climate Change Assessment Program. Daily community-level mortality counts for the period 1987 to 2000 were obtained

from the National Mortality, Morbidity and Air Pollution Study. Controlling for temperature, dew-point temperature,
and seasonality, relative risks associated with short-term exposure to ambient ozone during the summer months were estimated using a multi-site time series design. We estimated an increase of 0.43 ppb (95% PI: 0.14–0.75) in average ozone concentration during the 2040's compared to 2000 due to climate change alone. This corresponds to a 0.01% increase in mortality rate and 45.2 (95% PI: 3.26–87.1) premature deaths in the study communities attributable to the increase in future ozone level.

This is joint work with Jingwen Zhou (NCSU) and Montserrat Fuentes (NCSU)

Noel Cressie
Ohio State University
ncressie@stat.osu.edu

"Spatio-Temporal Statistical Modeling"

In this talk, spatio-temporal statistical modeling will be looked at through the prism of hierarchical modeling, and the role of dynamical temporal models of spatial fields will be emphasized. A global remote-sensing dataset of aerosol optical depth from the MISR instrument on the Terra satellite will be analyzed, resulting in optimally interpolated and
filtered statistical predictions. This leads naturally to a discussion of data/model/ computing compromises.

Jo Eidsvik
NTNU
joeid@math.ntnu.no

"Estimation and Prediction for Spatial Data using Composite Likelihood"

We suggest composite likelihood models for fast inference and prediction in large Gaussian geostatistical models. The composite likelihood models are built for the joint distribution of pairs of (block) variables, rather than the full joint distribution, which becomes intractable for large datasets. The composite likelihood construction is fast and allows parallel computing. By using the Godambe (sandwich) estimator we calculate the variance of the estimates and predictions. Results are compared on a synthetic study and a real data example.

Marco Ferreira
University of Missouri
ferreiram@missouri.edu

"Dynamic Multiscale Spatio-Temporal Models for Areal Data"

We introduce a new class of dynamic multiscale models for spatio-temporal processes arising from areal data. More specifically, we use nested geographical structures in order to decompose the data using an extension of the Kolaczyk-Huang multiscale decomposition. In our methodology, the original process is decomposed into multiscale coefficients which evolve through time following state-space equations. Our approach naturally accommodates data observed on irregular grids as well as heteroscedasticity. Moreover, we propose a multiscale spatio-temporal clustering algorithm that facilitates construction of the nested geographical multiscale structure. In addition, we propose a singular forward filter backward sampler for efficient Bayesian estimation. Our multiscale spatio-temporal methodology decomposes large data-analysis problems into many smaller pieces and thus leads to scalable and highly efficient computational procedures. Finally, we illustrate the utility and flexibility of our dynamic multiscale Framework through two spatio-temporal applications. The first example examines agricultural production in Espirito Santo State Brazil whereas the second example considers mortality ratios in the state of Missouri. This is joint work with Scott H. Holan and Adelmo I. Bertolde.

Michele Guindani
University of Texas
mguindani@mdanderson.org

"A Class of Covariate-Dependent Spatiotemporal Covariance Functions for the Modeling of Ozone"

In geostatistics, it is common to model spatially distributed phenomena through an under- lying stationary and isotropic spatial process. However, these assumptions are often untenable in practice because of the influence of local effects in the correlation structure. Therefore, it has been of prolonged interest in the literature to provide flexible and effective ways to model non-stationarity in the spatial effects. Arguably, due to the local nature of the problem, we might envision that the correlation structure would be highly dependent on local characteristics of the domain of study, namely the latitude, longitude and altitude of the observation sites, as well as other locally defined covariate information. In this work, we provide a flexible and computationally feasible way for allowing the correlation structure of the underlying processes to depend on local covariate information. We discuss the properties of the induced covariance functions and discuss methods to assess its dependence on local covariate information by means of a simulation study and the analysis of data observed at ozone-monitoring stations in the Southeast United States.

Amy Herring
University of North Carolina
aherring@bios.unc.edu

"Air Pollution and Reproductive Outcomes: Opportunities for Increased Research and Translation"

Exposures to air pollutants have many well-established effects on human health, including increased mortality and cardiopulmonary disease. Fewer studies have examined reproductive outcomes, and the EPA recently indicated that further research was needed. Pregnancy presents a unique opportunity to study health effects of air pollution, as exposure time windows of interest are relatively short, and much is known about the mechanisms of fetal development. We discuss

methods for examining the relationship between PM2.5, ozone, and birth outcomes, with a focus on identifying critical windows of pregnancy in which the developing fetus may be particularly sensitive to exposures. We discuss future work involving extensions of these methods to accommodate multivariate outcomes and to relax parametric assumptions. While the number of epidemiologic studies with spatially correlated data continues to increase rapidly, despite major advances in the analysis of spatio-temporal data, most epidemiologists (and statisticians working with them) rely on simple analytic measures that ignore spatial correlation. We discuss potential reasons for this incomplete translation of statistical methods into standard practice and plans for analysis of several large-scale cohorts.

Raphael Huser
EPFL
raphael.huser@epfl.ch

"Space-Time Modelling of Extreme Events and Inference for Max-Stable Processes"

We consider a precipitation dataset over Switzerland and we focus on analysing the behaviour of extreme events. The cornerstone of extreme value theory (EVT) is the extremal types theorem of Fisher and Tippett, which describes the asymptotic probability distribution of the maximum of a given sample (see [1]). However, univariate EVT is not enough once we want to model extremes spatially and max-stable processes are the natural extension. The characterization of max-stable processes was discussed in [2] and models for them have been proposed by [3]. Let Pi be a Poisson process of rate ds/s^2 on R+ and W_s denote independent copies of a non-negative truncated Gaussian random process with mean 1, variance 1, and isotropic correlation r(h). Then, Z(x) = max_(s in Pi) s W_s(x); x in X defines a max-stable random field on X (see [3]).

This model is asymptotically justified for modelling spatial extremes. However, the full likelihood is not known and therefore, classical inference becomes awkward. The idea is to take advantage of the theory of composite likelihood to make inference, using only the known pairwise margins. We propose a censored threshold-based maximum pairwise likelihood estimator, which turns out to be asymptotically normal under mild conditions, and we discuss inference based on this, including the choice of pairs to be incorporated into the composite likelihood, based on computations for simple time series models.
We consolidate these results by a simulation study in dimension 1 (simple temporal framework) and, if there is time, we will discuss a space-time application.

REFERENCES:
[1] Coles, S. (2001) An Introduction to Statistical Modeling of Extreme Values, Springer Verlag.
[2] de Haan, L. (1984) A spectral representation for max-stable processes The Annals of Probability 12, 1194-1204.
[3] Schlather, M. (2002) Models for stationary max-stable random elds Extremes 5, 33-44.

Emily Kang
SAMSI
lei@samsi.info

“Computational Strategies for Spatio-Temporal Filtering in Multiscale Systems”

Dynamical systems with multiple spatio-temporal scales are common in applied sciences and engineering, such as weather and climate dynamics. In this talk, we investigate fast computational strategies for filtering partially observed systems with multiscale features. We develop filtering algorithms which combine the Heterogeneous Multiscale Methods (HMM) with filtering approaches, such that they do not require the cross covariances between multiscale variables which are typically needed in standard filtering approaches but not computationally feasible in multiscale systems. The proposed filtering algorithms are tested in the framework of the 2-layer Lorenz 96 model. (Joint work with John Harlim)

Peter Kramer
Rensselaer Polytechnic Institute
kramep@rpi.edu

“Improving Statistical Components of Multi-scale Simulation Schemes”

I will describe the investigations of the working group on "Interaction of Deterministic and Stochastic Models" regarding multiscale computing algorithms which have been developing rapidly in the last decade in the applied mathematical and scientific computing community. The statistical aspects of these methods have not received much attention, and the working group has been formulating ideas for potential improvement in efficiency. We have in particular been examining a novel approach by Sorin Mitran (UNC) that involves a richer statistical model than the earlier Heterogeneous Multiscale Methods. The ongoing research program I will describe emerged from a sustained interaction of our working group with the "Numerical Methods for Stochastic Systems" in the contemporaneous Stochastic Dynamics program.

Elizabeth Mannshardt-Shamseldin
Duke University
Elizabeth@Stat.Duke.Edu

“Paleoclimate Extremes in Proxy Data”

There is much debate about the impact of a changing climate on extreme events. Proxy data for temperature series lead to several questions about long-term climate behavior, and how to interpret this behavior given the patterns seen in proxy series. For example, using proxy data to address questions such as \Were the 1990s the warmest decade of the last millennium", and \Is there evidence that the extreme events of recent decades are more extreme than previous decades?" The methodology of extreme value theory has not been widely applied to this problem. This paper looks at what the statistics of extremes has to offer paleoclimate reconstructions through modeling of the original proxy series. We propose to address questions posed concerning the extremes of the climate system through establishing a method for long-term proxy series behavior. This is explored by examining the underlying distributions of past and present proxy series in order to determine whether the distributional properties are changing over time. Several approaches are explored. This includes investigation of temporal trends and possible spatial relationships among the extreme values in proxy data. Once a method is

established, the feasibility of extending the method to apply to the longer, sparser, and lower resolution paleo reconstruction climate data can be investigated. There is also interest in determining how these reconstructions compare to the observed climate. This leads to possible calibration with climate model output, with models run under different greenhouse gas assumptions.
Co-authors: Peter Craigmile, Ohio State University; Martin Tingley, National Center For Atmospheric Research

Garritt Page
Duke University
page@stat.duke.edu

"The Effect of Spatial Confounding on Covariate Estimation"

In modeling contexts that consider a spatially-varying error structure, there often exist (possibly unmeasured) covariates that also vary spatially. In these scenarios it is hard to distinguish the effect of the covariate from that of the residual spatial variation. In a iid normal error setting, it is well know that dependence between a covariate and error produces biased regression coefficient estimates. However, the consequences of this dependence in a spatial setting are not as well understood. In this paper we investigate the effect of spatial confounding on covariate estimation by considering the bias, variance, and mean squared error of the generalized least squares estimator. We provide some analytical results and consequences of spatial confounding in the setting of a simple linear mixed model with a Gaussian spatial error term where the full covariance structure is known.

Bala Rajaratnam
Stanford University
brajarat@stanford.edu

"New Perspecitves on Paleoclimate Reconstructions"

A summary of paleoclimate working group activities will be given. New research avenues and recent advances in paleoclimate reconstructions will be discussed.

Brian Reich
North Carolina State University
reich@stat.ncsu.edu

"Nonparametric Spatial Models for Extremes"

Estimating the probability of extreme temperature events is difficult because of limited records across time and the need to extrapolate the distributions of these events, as opposed to just the mean, to locations where observations are not available. Another related issue is the need to characterize the uncertainty in the estimated probability of extreme events at different locations. Although the tools for statistical modeling of univariate extremes are well-developed, extending these tools to model spatial extreme data is an active area of research. In this paper, in order to

make inference about spatial extreme events, we introduce a new nonparametric model for extremes. We present a Dirichlet-based copula model that is a flexible alternative to parametric copula models such as the normal and t-copula. This presents the most flexible multivariate copula approach in the literature. The proposed modelling approach is fitted using a Bayesian framework that allow us to take into account different sources of uncertainty in the data and models. To characterize the complex dependence structure in the extreme events we use nonstationary (space-dependent) extremalcoefficient functions, and threshold-specific extremal functions. We apply our methods to annual maximum temperature values in the east-south-central United States.

Bruno Sanso
University of California, Santa Cruz
bruno@ams.ucsc.edu

“Comparing and Blending Regional Climate Model Predictions for the American Southwest”

Within the context of NARCCAP we consider output from the regional climate model RegCM3. This model was implemented by the Paleoclimate and Climate Research Group of the University of California Santa Cruz. Simulations are available for two time periods: current climate conditions, covering 1968 to 2000 and future climate
conditions, under the SRES A2 emissions scenario, covering 2038 to 2070. We consider RegCM3 simulations performed using forcings from two different AOGCM: GFDL and CGCM3. We develop a space-time model for annual summer average daily maximum temperature over a region that covers the southwest of the United States. Our model uses observational records to assess the discrepancies of the two different climate model simulations under present day conditions. Those discrepancies are then propagated into the future to obtain blended forecasts of 21st century climate. The model allows for time-varying spatial heterogeneities, providing local comparisons between the two sets of simulations. Additionally, we estimate the different modes of spatial variability, and use the temporally varying coefficients of the spatial factors for comparisons on a global scale. This work is an ongoing project that resulted from a collaboration involving several participants in the 2009-10 SAMSI program on Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change.

Ben Shaby
SAMSI
bshaby@gmail.com

"Approximate Bayesian Computation for Spatial Extremes via Open-faced Sandwich Adjustment"

Analysis of univariate extreme events is often carried out through models built around generalized extreme value distributions. Models that consider dependence among extreme events, however, are more elusive. Max-stable process models explicitly consider spatial dependence of block maxima, and are therefore promising for applications like heat waves or storms. Unfortunately, though a handful of max-stable processes have been proposed, none feature joint likelihoods that can be expressed analytically for dimensions greater than two. The lack of available likelihoods is an obvious stumbling block if one wishes to estimate model parameters in a Bayesian way through MCMC. We propose a method of constructing a sampler based on pairwise likelihoods, which are available. This sampler is motivated by existing asymptotic theory, and results in ``posterior'' inferences that have desirable frequentist properties.

Martin Tingley
NCAR
martin.tingley@gmail.com

"Piecing Together the Past: Statistical Insights into Paleoclimatic Reconstructions"

Reconstructing the spatial pattern of a climate field through time from incomplete instrumental and climate proxy time series poses both scientific and statistical challenges. The societal relevance and scientific challenges posed by this paleoclimate reconstruction problem make it a critical and important problem to the earth science community, while the challenges posed by the data and the space-time complexity of the climate system make it of interest to statisticians. The scope of the challenges, along with the interdisciplinary nature of the reconstruction problem, point to the need for greater cooperation between the earth science and statistics community - a sentiment echoed in recent parliamentary reports.

As a step in this direction, it is prudent to formalize what is meant by (one aspect of) the paleoclimate reconstruction problem using the language and tools of modern statistics. This article considers the challenge of inferring, with uncertainties, a climate field through space and time from overlapping instrumental and climate sensitive proxy time series that are assumed to be well dated. The assumption that the proxy series are well dated is likely only reasonable for reconstructions over at most the last few millennia. Within a unifying modeling framework for this problem, the modeling assumptions made by a number of published methods can be understood as special cases, and the distinction between modeling assumptions and analysis/inference choices becomes more transparent.

Over the last two decades, the statistics community has made major advances in the modeling and analysis of space-time data sets, such as the surface temperature field. In addition, increases in available computational power, advances in statistical computation, and the development of data-reduction techniques now allow rich classes of space-time model to be fit to very large data sets. Many of these advances have not yet been applied to the paleoclimate reconstruction problem, and doing so has the potential to improve understanding of the climate of the past.

The key aims of this article are to 1) establish a general notational and modeling framework for the paleoclimate reconstruction problem that is transparent to both the earth science and statistics communities; 2) outline and distinguish between scientific and statistical challenges and indicate how modern statistical expertise can be brought to bear upon the problem; 3) offer, in broad strokes, some suggestions for model construction and how to perform the required statistical inference; and 4) identify issues that are critically important to both the earth science and applied statistics communities, and encourage greater collaboration between the two.

#5: Program on Complex Networks – Complex Networks Modeling Workshop, October 20-22, 2010

<u>Schedule</u>:

Wednesday, October 20
SAMSI

8:00-8:55	Registration and Continental Breakfast
8:55-9:00	Welcome
9:00-9:40	**Stephen Fienberg**, Carnegie Mellon University *"Statistical Challenges in Network Modeling"*
9:40-10:30	**Edo Airoldi**, Harvard University *"Network Representation"*
10:30-11:00	Break
11:00-11:40	**Tian Zheng**, Columbia University *"Statistical Methods for Studying Social Networks using Aggregated Relational Data"*
11:40-12:30	**Purnamitra Sarkar**, University of California, Berkeley *"Theoretical Justification of Popular Link Prediction Heuristics"*
12:30-2:30	Lunch / Breakout Sessions
2:30-3:10	**Andrew C. Thomas**, Carnegie Mellon University *"Exploring the Limits of Conditionally Independent Dyadic Network Models"*
3:10-3:40	Break
3:40-4:30	**Lucy Robinson**, Johns Hopkins University *"Change Point Detection in Networks"*
4:30-6:30	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, October 21
SAMSI

8:30-9:00	Continental Breakfast
9:00-9:40	**Bruce Spencer**, Northwestern University *"Sampling Research Questions"*
9:40-10:30	**Krista Gile**, University of Massachusetts *"Self-Consistent Network Model-Assisted Prevalence Estimation from Respondent-Driven Sampling Data"*
10:30-11:00	Break
11:00-11:40	**Bin Yu**, University of California, Berkeley *"Spectral Clustering and the High-dimensional Stochastic Block Model"*
11:40-12:30	**Aarti Singh**, Carnegie Mellon University *"Identifying Graph-structured Network Activations"*
12:30-2:30	Lunch / Breakout Sessions
2:30-3:10	**Stephane Robin**, Agro Paris Tech *"Uncovering Structure in Interaction Networks"*
3:10-3:40	Break
3:40-4:30	**Denise Scholtens**, Northwestern University Medical School *"Sequential Sampling Designs for Estimating Local Connectivity in Bait-Prey Graphs"*

Friday, October 22
SAMSI

8:30-9:00	Continental Breakfast
9:00-9:40	**Oliver Ratmann**, Duke University *"Probing the Evolution of Protein Interaction Networks with Approximate Bayesian Computation"*
9:40-10:30	**Eric Kolaczyk**, Boston University *"Looking Ahead: Challenges in Network Sampling, Modeling, and Inference"*
10:30-11:00	Break

11:00-11:40 Closing Summary/Discussion

11:40-1:00 Lunch and Adjourn

SPEAKER TITLES/ABSTRACTS

Edo Airoldi
Harvard University
airoldi@fas.harvard.edu

"Network Representation"

Networks are ubiquitous in science and have become a focal point for discussion in everyday life. Formal statistical models for the analysis of network data have emerged as a major topic of interest in diverse areas of study, and date back to 1959. A number of modern and historical approaches to network modeling focus on the ensemble of paired measurements that express connectivity at the dyad level. However, while the focus on edges is very popular and mathematically tractable, it is not the only approach available, nor the most relevant to science, one could argue. In this talk, we will explore aspects and challenges of the network representation problem.

Stephen Fienberg
Carnegie Mellon University
fienberg@stat.cmu.edu

"Statistical Challenges in Network Modeling"

Through three examples and a brief overview of some statistical models for network data (p1, exponential random graph models, and stochastic blockmodels for mixed membership), I highlight some challenges associated with the statistical analysis of network data, including those associated with sampling, large-sample inference, experiments, and privacy protection.

Krista Gile
University of Massachusetts
gile@math.umass.edu

"Self-Consistent Network Model-Assisted Prevalence Estimation from Respondent-Driven Sampling Data"

Respondent-Driven Sampling is type of link-tracing network sampling used to study hard-to-reach populations. Beginning with a convenience sample, each person sampled is given 2-3 uniquely identified coupons to distribute to other members of the target population, making them eligible for enrolment in the study. This is effective at collecting large diverse samples from many populations.

Current estimation relies on sampling weights estimated by treating the sampling process as a random walk on the underlying network of social relations. These estimates are based on strong assumptions allowing the data to be treated as a probability sample. In particular, existing estimators assume a with-replacement sample with an ideal initial sample. We introduce a new estimator based on fitting a social network model (ERGM), and demonstrate its ability to correct for biases due to the initial sample.

Eric Kolaczyk
Boston University
kolaczyk@bu.edu

"Looking Ahead: Challenges in Network Sampling, Modeling, and Inference"

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 sampling, modeling and inference in practice. In the context of network data, however, such an infrastructure is only beginning to be developed. I will aim to summarize what seem to be some of the major challenges in this area, some of which are being pursued by the SAMSI Working Group on this topic.

Sarkar Purnamrita
University of California, Berkeley
psarkar@eecs.berkeley.edu

"Theoretical Justification of Popular Link Prediction Heuristics"

There are common intuitions about how social graphs are generated (for example, it is common to talk informally about nearby nodes sharing a link). There are also common heuristics for predicting whether two currently unlinked nodes in a graph should be linked (e.g. for suggesting friends in an online social network or movies to customers in a recommendation network). This paper provides what we believe to be the first formal connection between these intuitions and these heuristics. We look at a familiar class of graph generation models in which nodes are associated with locations in a latent metric space and connections are more likely between closer nodes. We also look at popular link-prediction heuristics such as number-of-common-neighbors and its weighted variants (Adamic & Adar, 2003) which have proved successful in predicting missing links, but are not direct derivatives of latent space graph models. We provide theoretical justifications for the success of some measures as compared to others, as reported in previous empirical studies. In particular we present a sequence of formal results that show bounds related to the role that a node's degree plays in its usefulness for link prediction, the relative importance of short paths versus long paths, and the effects of increasing non-determinism in the link generation process on link prediction quality.

Oliver Ratmann
Duke University
oliver.ratmann@duke.edu

"Probing the Evolution of Protein Interaction Networks with Approximate Bayesian Computation"

The evolutionary mechanisms by which protein interaction networks grow and change are beginning to be appreciated as a major factor shaping their present-day structures and properties.

Here, we present a Bayesian approach for testing the adequacy of stochastic models of network evolution, and for fitting these models to observed network topologies. We found that our method, based on Approximate Bayesian Computation, provides a flexible, computationally attractive tool for the analysis of complex evolution models that may account for network incompleteness and incorporate other data. This new method for parameter inference and model checking will be highlighted to ease its transfer to other potential applications in network science.

Examination of three models of network evolution suggests that models of duplication-divergence match the observed protein network topologies of Helicobacter pylori, Treponema pallidum and Campylobacter jenuni best. It transpires that only the simultaneous analysis of several global aspects of protein networks enables credible inference to be made. Parameter inference is sensitive to the choice of measurement model, suggesting that the various forms of measurement error need to be explicitly modeled from the outset. We argue that the value of such analyses is expected to be greatly enhanced as models incorporate more of the details of the biophysical properties of interacting proteins, their phylogenetic relationships, and interface more closely with the primary network data derived from high throughput experiments.

Stéphane Robin
INRA / AgroParisTech
robin@agroparistech.fr

"Uncovering Structure in Interaction Networks"

Networks are a natural way to describe the interaction between biological entities (genes, species, etc…). Flexible random graphs models are needed to describe the wide variety to topologies displayed by such networks. Stochastic block models have become a popular way to describe such topologies. These models are mixture models, the inference of which raises several statistical issues. We will present several advances such as some refinement of variational approximations, a variational Bayes inference or the introduction of covariates to explain the network structure. These results will be illustrated with application to molecular and ecological networks.

Lucy Robinson
Johns Hopkins University
lrobins9@jhu.edu

"Change Point Detection in Networks"

We propose a latent process model for dynamic network data that is amenable to change point detection. Our goal is to detect a subpopulation of vertices exhibiting a change in behavior over time. The dependency structure of the network and distribution of edge attributes are described using a random dot product model. Applications to email and functional neuroimaging data are presented.

Denise Scholtens
Northwestern University Medical School
dscholtens@northwestern.edu

"Sequential Sampling Designs for Estimating Local Connectivity in Bait-Prey Graphs"

Bait-prey technologies that assay cellular protein interactions have recently surged in popularity. The most widely cited applications use steady-state systems such as Saccharomyces cerevisiae under assumedly stable cellular conditions and target global estimation of the cellular 'interactome' (Uetz et al. 2000; Ito et al. 2001; Gavin et al., 2002 and 2006; Krogan et al. 2006). In contrast to genome-wide models, disease-relevant settings often consist of a small set of starting baits with local connectivity among their neighbors being the estimation goal. Experimental design schemes that allow measurement of local connectivity using fewer baits may facilitate more regular use of
bait-prey technologies in smaller lab settings, and could subsequently afford comparison of protein interaction topologies across different experimental conditions, disease types, etc. To this end, we present a collection of sequential experimental design schemes that mirror classic link-tracing approaches but increase coverage of each bait-prey assay and reduce variability of the inferred local regions for small-scale experimental settings.

Arti Singh
Carnegie Mellon University
aarti@cs.cmu.edu

"Identifying Graph-structured Network Activations"

The problem of identifying an activation pattern in a complex, large-scale network that is embedded in very noisy measurements is relevant to several applications, such as detecting traces of a biochemical spread by a sensor network, expression levels of genes, and anomalous activity or congestion in the Internet. Extracting such patterns is a challenging task specially if the network is large (pattern is very high-dimensional) and the noise is so excessive that it masks the activity at any single node.

Most prior work considers situations in which the activation/non-activation of each node is statistically independent. In this case, the signal-to-noise ratio needs to increase as the network size increases, to accommodate for multiple hypothesis testing effects. However, typically there are statistical dependencies in the network activation process that can be leveraged to fuse the measurements of multiple nodes and enable reliable extraction of high-dimensional noisy patterns. In this talk, I will describe a method based on projecting the measurements onto an appropriate basis that is adapted to the structure of the dependency graph. We show that this procedure enables consistent identification of high-dimensional graph-structured patterns even when the signal-to-noise ratio decreases with the network size.

Bruce Spencer
Northwestern University
bspencer@northwestern.edu

"Sampling Research Questions"

Basic questions in network sampling will be posed. Some are similar to questions for traditional sampling, such as differences between the sampled population and the population of inferential interest. Other questions for traditional sampling apply with more force for network samples, such as the effect of measurement error. In link-tracing designs, the very sample itself is affected both by failure to identify links and by detection of links that do not exist. In network samples, inferences about the network often will depend critically on the use of statistical models for the network, and a question is how to design samples that will permit validation of such models. Other questions are posed as well.

Andrew C. Thomas
Carnegie Mellon University
acthomas@stat.cmu.edu

"Exploring the Limits of Conditionally Independent Dyadic Network Models"

Many of us are interested in modelling the formation of network edges with parsimonious models, but the number of approaches makes this all the more difficult to accomplish. Of just as much scientific interest, but with far less attention paid, is the notion that network edges have values or characteristics that are important to the processes that occur on the networks. Because the Generalized Linear Model approach has proven successful at handling many different classes of data, I aim to show the versatility in specifying dyads as independent conditional on both manifest and latent node- and edge- level characteristics -- christened p^G, to follow the naming pattern laid out by p^1, p^2 and p^* -- and how concepts like transitivity and community formation are handled as terms in a time series evolution process, a natural inclusion in a GLM model, rather than explicit higher-order dependence. I conclude by demonstrating how this approach connects to co-evolution models of nodes and edges and discussing the problem of causal inference on node characteristics.

Bin Yu
University of California, Berkeley
binyu@stat.berkeley.edu

"Spectral Clustering and the High-dimensional Stochastic Block Model"

In recent years, network analysis have become the focus of much research in many fields including biology, communication studies, economics, information science, organizational studies, and social psychology. Communities or clusters of highly connected actors form an essential feature in the structure of several empirical networks. Spectral clustering is a popular and computationally feasible method to discover these communities.

The Stochastic Block Model is a social network model with well defined communities. This talk will give conditions for spectral clustering to correctly estimate the community membership of nearly all nodes. These asymptotic results are the first clustering results that allow the number of clusters in the model to grow with the number of nodes, hence the name high-dimensional.

(Based on joint work with Karl Rohe and Sourav Chatterjee at UC Berkeley)

Tian Zheng
Columbia University
tzheng@stat.columbia.edu

"Statistical Methods for Studying Social Networks using Aggregated Relational Data"

Questions of the form "How many X's do you know?" collect aggregated relational data from one's social network and are a common means of learning about populations that are hard to reach or survey directly. McCarty et al. (2001), for example, take X to be individuals who are HIV positive, are homeless, or were recently raped to estimate the size of these traditionally hard-to-count populations. In this talk, we will discuss several recent statistical methodological developments for analyzing ARD to study features of social networks. This is a joint work with Tyler McCormick.

#6: Education and Outreach Program, Two-Day Undergraduate Workshop, October 29-30, 2010

Schedule:

Friday

8:15	Shuttle from Radisson to SAMSI (Group #1)
8:35	Shuttle from Radisson to SAMSI (Group #2)
9:15-9:30	Introduction **Pierre Gremaud**, North Carolina State University and SAMSI
9:30-10:20	*"What are Graphs and Why are they Important?"* **Rick Durrett**, Duke University and SAMSI
10:20-10:50	MATLAB Demo **Yi Sun**, SAMSI
10:50-11:10	Break
11:10-12:00	*"Random Graph"* **David Sivakoff**, SAMSI
12:00-1:00	Lunch
1:00-1:50	"*Graphs, High-Dimensional Data, and Visualization*" **Mauro Maggioni**, Duke University
1:50-2:20	R Demo **Hongziao Zhu**, SAMSI
2:20-2:50	Break
2:50 - 3:40	*"Using R for Scientific Research"* **Joshua Mendelsohn**, Duke University
3:40 - 4:40	R Lab **Joshua Mendelsohn**, Duke University
4:40-5:00	Career Options **Pierre Gremaud**, North Carolina State University and SAMSI
5:00	Shuttle to Radisson (Group #1)
5:20	Shuttle to Radisson (Group #2)
6:00	Dinner @ Radisson

Saturday

8:00	Shuttle from Radisson to SAMSI (Group #1)
8:20	Shuttle from Radisson to SAMSI (Group #2)
9:00-9:50	"*A History of Network Modeling"* **David Banks**, Duke University
9:50-10:40	*"Graph Clustering with Random Walks"* **Bruce Rogers,** SAMSI **Mandi Traud**, North Carolina State University
10:40-11:00	Break
11:00-12:00	MATLAB Related to Rogers and Traud's Presentation
Noon	Adjournment and Departure
12:00	Shuttle from SAMSI to RDU (Group #1)
12:30	Shuttle from SAMSI to RDU (Group #2)

#7: Program on Analysis of Object Data – Workshop on Interface Functional and Longitudinal Data Analysis, November 8-10, 2010

SCHEDULE
Monday, November 8
Radisson Hotel RTP

8:00-8:30	Registration and Continental Breakfast
8:30-8:45	Welcome
8:45-10:00	Multivariate LDA-FDA **Raymond J. Carroll**, Texas A&M University *"Generalized Functional Linear Models with Semiparametric Single-Index Interactions"*
10:00-10:20	**Damla Senturk**, Pennsylvania State University *"Functional Varying Coefficient Models for Longitudinal Data"*
10:20-11:00	Break
11:00-11:20	**Wensheng Guo**, University of Pennsylvania *"Functional Mixed Effects Spectral Analysis"*
11:20 – 12:15	Discussion and Connections to Working Groups
12:15-1:30	Lunch
	Regression Models for LDA-FDA

1:30-2:30	**Tailen Hsing**, University of Michigan *"Uniform Convergence Rates for Principal Component Analysis in Functional/ Longitudinal Data"*
2:30-3:00	**Jeff Morris**, University of Texas *"Adaptive, Robust Functional and Image Regression in Functional Mixed Models"*
3:00-3:30	Break
3:30-4:00	**Fang Yao**, University of Toronto *"Additive Modeling of Functional Regression and its Gradients"*
4:00-4:45	Discussion and Connections to Working Groups
4:45-5:00	Poster Advertisement Session (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.*

Tuesday, November 9
Radisson Hotel RTP

8:00-8:45	Registration and Continental Breakfast
8:45-10:00	Variable Selection for LDA-FDA **Xihong Lin**, Harvard University *"Likelihood Based and Estimating Equation Based Methods for Variable Selection"*
10:00-10:30	**Helen Zhang**, North Carolina State University *"Variance Component Selection in Linear Mixed Models"*
10:30-11:15	Break
11:15-11:45	**Graciela Boente**, Universidad de Buenos Aires and CONICET *"Robust Functional Principal Components: a projection-pursuit approach"*
11:45-12:15	Discussion and Connections to Working Groups
12:15-1:30	Lunch
1:30-2:30	Longitudinal/Functional Models for Dynamics **Jim Ramsay**, McGill University *"Economical Models for Functional Covariation"*
2:30-3:00	**Jie Peng**, University of California, Davis *"Fitting Ordinary Differential Equation Models with Longitudinal Data"*
3:00-3:30	Break

3:30-4:00 **Hulin Wu**, University of Rochester
"Comparing Functional Data Analysis Approach and Nonparametric Mixed-Effects Modeling Approach for Longitudinal Data Analysis"

4:00-5:00 Discussion and Connections to Working Groups

Wednesday, November 10
Radisson Hotel RTP

8:00-8:45 Registration and Continental Breakfast

Multilevel and Joint Models for LDA-FDA
8:45-9:05 **Ana-Maria Staicu**, North Carolina State University
"Skewed Functional Processes and their Applications"

9:05-9:25 **Jimin Ding**, Washington University, St. Louis
"Time-varying Coefficient Cox model with Nonparametric Longitudinal Covariates"

9:25-10:00 Break

10:00-11:45 Discussion and Connections to Working Groups

11:45 – 1:00 Lunch

1:00 Adjournment

SPEAKER TITLES/ABSTRACTS

Graciela Boente
Universidad de Buenos Aires and CONICET
gboente@dm.uba.ar

"Robust Functional Principal Components: a projection-pursuit approach"

When dealing with multivariate data, like classical PCA, robust PCA searches for directions with maximal dispersion of the data projected on it. Instead of using the variance as a measure of dispersion, a robust scale estimator sn may be used in the maximization problem. This approach was introduced by Li and Chen (1985), who proposed estimators based on maximizing (or minimizing) a robust scale, while a maximization algorithm was proposed in Croux and Ruiz-Gazen (1996). Later on, Croux and Ruiz–Gazen (2005) derived the influence functions of the resulting principal components, while their asymptotic distribution was studied in Cui et al. (2003).

Our aim is to adapt the projection pursuit approach to the functional data setting. We focus on functional data that are recorded over a period of time and regarded as realizations of a stochastic process, often assumed to be in the L2 space on a real interval. While principal components analysis (PCA), originally developed for multivariate data, has been successfully extended to accommodate functional data and termed functional PCA, its robust versions are scant in the literature.

Let {X(t) : t 2 I} be a stochastic process defined in (,A, P) with continuous trajectories and finite second moment, where I _ R is a finite interval. Without loss of generality, we may assume that I = [0, 1]. We will denote the covariance function by X(t, s) = cov (X(t),X(s)), and the corresponding covariance operator by X. Moreover, let {_j : j _ 1} and {_j : j _ 1} be, respectively, the eigenfunctions and the eigenvalues of the covariance operator X with _j _ _j+1. Dauxois et al. (1982) derived the asymptotic properties of principal components of functional data defined as the eigenfunctions of the sample covariance operator. On the other hand, Silverman (1996) and Ramsay and Silverman (2005), introduced smooth principal components for functional data, based on roughness penalty methods while Boente and Fraiman (2000) considered a kernelbased approach. More recent work, dealing with estimation of the principal components of the covariance function, includes Gervini (2006), Hall and Hosseini-Nasab (2006), Hall et al. (2006) and Yao and Lee (2006). Up to our knowledge, the first attempt to provide estimators of the principal components less sensitive to anomalous observations was done by Locantore et al.
(1999) who considered the coefficients of a basis expansion. Besides, Gervini (2008) studied a fully functional approach to robust estimation of the principal components by considering a functional version of the spherical principal components defined in Locantore et al. (1999). On the other hand, Hyndman and Ullah (2007) provide a method combining a robust projection–pursuit approach and a smoothing and weighting step to forecast age-specific mortality and fertility rates observed over time.

We will introduce robust estimators of the principal components and give conditions to ensure their consistency. Our approach combines robust projection–pursuit with different smoothing methods.

Co-authors: Lucas Bali, David E. Tyler, Jane–Ling Wang

Raymond J. Carroll
Texas A&M University
carroll@stat.tamu.edu

"Generalized Functional Linear Models with Semiparametric Single-Index Interactions"

We introduce a new class of functional generalized linear models, where the response is a scalar and some of the covariates are functional. We assume that the response depends on multiple covariates, a finite number of latent features in the functional predictor, and interaction between the two. To achieve parsimony, the interaction between the multiple covariates and the functional predictor is modeled semiparametrically with a single-index structure. We propose a two step estimation procedure based on local estimating equations, and investigate two situations: (a) when the basis functions are pre-determined, e.g., Fourier or wavelet basis functions and the functional features of interest
are known; and (b) when the basis functions are data driven, such as with functional principal components. Asymptotic properties are developed. Notably, we show that when the functional

features are data driven, the parameter estimates have an increased asymptotic variance, due to the estimation error of the basis functions. Our methods are illustrated with a simulation study and applied to an empirical data set, where a previously unknown interaction is detected.

Jimin Ding
Washington University, St. Louis
jmding@math.wustl.edu

“Time-varying Coefficient Cox model with Nonparametric Longitudinal Covariates”

In this paper, we propose a time-varying coefficient joint modeling (VJOINT) to extend current joint modeling approach and allow the time-dependent covariate effects. In recent biomedical and clinical studies, time to some interesting events is usually recorded together with longitudinal covariates for each subject. Joint modeling of the longitudinal and survival data has emerged an effective way to investigate the influence of longitudinal covariates on survival time and the pattern of longitudinal process simultaneously. The proposed VJOINT is not only more flexible but also provides a way to test the proportional hazard assumption in the popular joint models. Both time-varying covariate effect and longitudinal process are modeled through basis representation. Under a set of suitable assumptions, we prove asymptotical consistency of the proposed estimator and provide the different consistency rates for the finite and infinite dimensional parameters. We will illustrate the performance of the proposed VJOINT throught simulations. Finally we will apply VJOINT to two liver cirrhosis studies.

Wensheng Guo
University of Pennsylvania
wguo@mail.med.upenn.edu

"Functional Mixed Effects Spectral Analysis",

In many experiments, time series data can be collected from multiple units and multiple time series segments can be collected from the same unit. This article introduces a functional mixed effects spectral model which can be used to estimate the effects of design covariates on the second order power spectrum while accounting for potential correlations among the time series segments collected from the same unit. We propose a mixed effects Cram\'{e}r spectral representation where the transfer function is composed of a deterministic component to account for the population-average effects and a random component to account for the unit-specific deviations. The resultant log-spectrum has a functional mixed effects representation where both the fixed effects and random effects are functions in the frequency domain. It is shown that, when the unit-specific spectra are smooth, the log-periodograms converge to a functional mixed effects model. A data driven iterative estimation procedure is offered for the periodic smoothing spline estimation of the fixed effects, penalized estimation of the functional covariance of the random effects, and unit-specific random effects prediction via the best linear unbiased predictor.

Based on joint work with Robert T. Krafty and Martica Hall

Tailen Hsing
University of Michigan
thsing@umich.edu

"Uniform Convergence Rates for Principal Component Analysis in Functional/ Longitudinal Data"

We consider nonparametric estimation of the mean and covariance functions for functional/longitudinal data. Strong uniform convergence rates are developed for estimators that are local-linear smoothers. Our results are obtained in a unified framework in which the number of observations within each curve/cluster can be of any rate relative to the sample size. We show that the convergence rates for the procedures depend on both the number of sample curves and the number of observations on each curve. For sparse functional data, these rates are equivalent to the optimal rates in nonparametric regression. For dense functional data, root-n rates of convergence can be achieved with proper choices of bandwidths. We further derive almost sure rates of convergence for principal component analysis using the estimated covariance function.

Xihong Lin
Harvard School of Public Health
xlin@hsph.harvard.edu

"Likelihood Based and Estimating Equation Based Methods for Variable Selection"

We first discuss variable selection using likelihood based method using the seamless L0 penalty (SELO), where the penality is constructed to mimic the L0 penalty. We show that the SELO estimator is consistent for model selection and has the orcale property, as the number of observations grows and the number of variables remains fixed. We next study estimating equation based variable methods using weighted Dantzig Selector (WDS). We study the theoretical properties of the DS and WDS estimators, and show the WDS is consistent for model selection if a proper weight function is used. We propose the use of BIC for selecting the toning parameter and show it gives consistent model selection. We evaluate the finite sample performance using simulation studies and illustrate their performance by applying the methods to the HIV drug resistence data.

Jeffrey Morris
University of Texas, Anderson Cancer Center
jefmorris@mdanderson.org

"Adaptive, Robust Functional and Image Regression in Functional Mixed Models"

Many new methods have been developed in recent years for the analysis of functional and data, which are increasingly encountered in many scientific settings. Many of these methods involve extensions of linear regression to functional data such as functional regression and functional mixed models. Existing methods, however, tend to be sensitive to outliers, as no analogs to robust linear regression have been developed to data for the functional setting. Here, we discuss a unified Bayesian method that performs robust functional regression, whereby a functional response of unspecified form is regressed on a set of linear predictors. The method works within the general functional mixed model framework, so is also able to account for between-function correlation induced by the experimental design, and although requiring an MCMC, automatic reference priors are available in the code so that the method can run automatically with no specification of tuning parameters. We demonstrate that this method has outstanding robustness properties, able to do an excellent job estimating functional regression coefficients even in the presence of Cauchy errors and random effects, and yet does not trade off much efficiency even when the true likelihood is Gaussian. We also observed remarkable adaptive smoothing properties in our estimates of the fixed and random effect functions, which arise from an interaction of the robust likelihood and prior structure in the model.

Jie Peng
University of California, Davis
jie@wald.ucdavis.edu

"Fitting Ordinary Differential Equation Models with Longitudinal Data"

In this talk, we discuss dynamical systems which can be described through ordinary differential equations. We focus on the case when the system is nonlinear and/or when there are subject specific random parameters in the model.

We discuss model fitting when the measurements are irregularly and sparsely observed. We compare several fitting strategies including the two-stage method, a Hierarchical likelihood approach and a marginal likelihood approach via Laplace approximation. Their performances are examined by simulation and real data examples.

James Ramsay
McGill University
ramsay@psych.mcgill.ca

"Economical Models for Functional Covariation"

The high dimensionality of functional observations implies that we cannot expect to estimate between-record covariation in the usual way, by estimating the covariances between basis function coefficients. In this context, conventional longitudinal data analysis, and many others, a method is needed which estimates covariance structure using a number of parameters that does not depend on the dimensionality of the data. A method will be discussed using a finite element representation of a covariance surface.

Damla Senturk
Pennsylvania State University
dsenturk@stat.psu.edu

"Functional Varying Coefficient Models for Longitudinal Data"

The proposed functional varying coefficient model provides a versatile and flexible analysis tool for relating longitudinal responses to longitudinal predictors. Specifically, this approach provides a novel representation of varying coefficient functions through suitable auto- and cross-covariances of the underlying stochastic processes, which is particularly advantageous for sparse and irregular designs, as often encountered in longitudinal studies. Existing methodology for varying coefficient models is not adapted
to such data. The proposed approach extends the customary varying coefficient models to a more general setting, in which not only current but also recent past values of the predictor time course may have an impact on the current value of the response time course. The influence of past predictor values is modeled by a smooth history index function, while the effects on the response are described by smooth varying coefficient functions. The resulting estimators for varying coefficient and history index functions are shown to be asymptotically consistent for sparse designs. In addition, prediction of unobserved response trajectories from sparse measurements on a predictor trajectory is obtained, along with asymptotic pointwise confidence bands. The proposed methods perform well in simulations, especially when compared with commonly used local polynomial smoothing methods for varying coefficient models, and are illustrated with longitudinal primary biliary liver cirrhosis data.

Ana-Maria Staicu
North Carolina State University
ana-maria_staicu@ncsu.edu

"Skewed Functional Processes and their Applications"

We propose a novel class of models for the analysis of functional data that display spatially heterogeneous skewness. These models generalize the Karhunen-Loeve expansion, provide an inferential framework for estimating potentially asymmetric marginal quantiles, and are computationally feasible. More importantly, they provide a new set of tools for increasingly complex data collected in medical and public health studies. Our methods were motivated by and applied to a state-of-the-art study of neuronal tracts in multiple sclerosis patients and healthy controls. However our models are general and will be relevant to data sets where the object of inference are functions, which have distributional shape characteristics that vary across point locations.

Hulin Wu
University of Rochester
hwu@bst.rochester.edu

"Comparing Functional Data Analysis Approach and Nonparametric Mixed-Effects
Modeling Approach for Longitudinal Data Analysis"

Sparse curve data can be considered as longitudinal data and the frequent longitudinal data can be treated as functional or curve data. Thus, analysis and modeling technologies from these two research areas can be shared. The most popular nonparametric method for longitudinal or curve data analysis is the nonparametric mixed-effects (NPME) modeling approach such as the mixed-

effects B-splines smoothing and one popular curve or function data analysis approach is the functional principle component analysis (fPCA).

We compare these two nonparametric methods for longitudinal data analysis to estimate the population curve and subject-specific individual curves. We find that, when the between-subject (between-curve) variation is small and individual curves are in a similar pattern, the NPME modeling method performs better and more efficient; while the fPCA approach is better when the between-subject (between-curve) variation is large. Thus, the NPME modeling approach is more appropriate for the longitudinal data that are generated from similar mechanisms with similar patterns, and the NPME modeling technique helps with borrowing information across subjects. On the other hand, the fPCA is better to deal with the longitudinal or curve data that share something in common, but the between-subject (between-curve) variation is large, since both the data-driven PC bases and the corresponding random coefficients (PCA scores) are readily used to capture the large between-subject variation. Hopefully our results provide a good guidance for practitioners to perform functional or longitudinal data analysis. This is a collaboration work with Dr. Shuang Wu.

Fang Yao
University of Toronto
fyao2001@gmail.com

"Additive Modeling of Functional Regression and its Gradients"

In commonly used functional regression models, the regression of a scalar or functional response on the functional predictor is assumed to be linear. We relax the linearity assumption and propose to replace it by an additive structure. This leads to a more widely applicable and flexible framework for functional regression models. The regularization needed for effective estimation of the regression parameter function is implemented through a projection on the eigenbasis of the covariance operator of the functional components in the models. The utilization of functional principal components in an additive rather than linear way leads to substantial broadening of the scope of functional regression models and emerges as a natural approach, as the uncorrelatedness of the functional principal components is shown to lead to a straightforward implementation of the functional additive model, just based on a sequence of one-dimensional smoothing steps and without need for backfitting. We further extend this framework to the estimation of functional gradients which are then defined as functional directional derivatives. Aiming at a model-free approach, we are enabled to study how changes in the predictor function in a specified functional direction are associated with corresponding changes in the response. The proposed modeling framework and estimation methods are illustrated in applications to gene expression time course data and meflies data, respectively.

Helen Zhang
North Carolina State University
hzhang2@stat.ncsu.edu

"Variance Component Selection in Linear Mixed Models"

In many applications, it is common to collect repeated measurements on a subject or take serial observations over time on the same unit, resulting clustered data, longitudinal data, or spatial data. The random effects provide a convenient and effective way of describing the covariance structure of data. The selection of random effects in linear mixed models is an important yet challenging problem in practice. We propose a robust and unified framework for automatically selecting random effects and estimating the covariance components in linear mixed effects models. A moment-based loss function is constructed for the covariance parameters of random effects and thresholding penalties are imposed for the selection of random effects. Compared to other existing approaches, the new estimator does not require any distribution assumption on random effects or the error terms. Large-sample theories show that the new estimator is consistent for both random effects selection and variance components estimation. Due to the nature of variance-covariance matrix of random effects, our procedure involves two challenging computation problems: nonlinear semidefinite programming and nonlinear programming with a linear inequality constraint. Computational strategies are suggested to tackle these issues as well as the model tuning. Simulation studies reveal that a better selection of random effects can yield efficiency gain for the estimation of fixed effects.

#8: Program on Stochastics Dynamics – Stochastics Dynamics Transition Workshop, November 17-19, 2010

SCHEDULE

Wednesday, November 17
SAMSI

8:30-9:00	Registration and Continental Breakfast
9:00-9:15	Welcome **Peter Kramer**, Rensselaer Polytechnic Institute **Jonathan Mattingly**, Duke University **Sorin Mitran**, University of North Carolina
	Session on Numerical Methods for Stochastic Systems
9:15-9:45	"*Continuum-microscopic Modeling of Fibrous Materials"* **Jennifer Young**, Rice University,
9:45-10:15	"*Improving Statistical Components of Multi-scale Simulation Schemes"* **Peter Kramer**, Rensselaer Polytechnic Institute
10:15-10:45	Working group report and discussion (**Sorin Mitran**, U of North Carolina)
10:45-12:00	Coffee break and free discussion
12:00-1:30	Lunch
	Session on Data Assimilation

1:30-2:00	*"Numerical Strategies for Filtering Partially Observed Stiff Stochastic Differential Equations"* **John Harlim**, North Carolina State University
2:00-2:30	*"Control Theory and the Design of Dynamic Experiments"* **Giles Hooker**, Cornell University
2:30-3:00	Working group report and discussion (**Giles Hooker**, Cornell University)
3:00-4:00	Coffee Break
4:00-6:00	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, November 18
SAMSI

8:30-9:00	Registration and Continental Breakfast
	Session on Network Dynamics
9:00-9:30	*"Thresholds for Emergence of Multi-scale Behavior of Epidemics on Networks"* **Bruce Rogers**, Duke University,
9:30-10:00	"*Pairwise Closure Approximations in Epidemic Models on Regular Networks"* **Xueying Wang**, Texas A&M University
10:00-10:30	Coffee Break
10:30-11:00	*"Cascade Induced Synchronous Dynamics on Different Network Topologies"* **Amanda Traud**, North Carolina State University,
11:00-11:30	Working group report and discussion (**Peter Mucha**, U of North Carolina)
11:30-1:00	Lunch
	Session on Biological Stochastic Dynamics
1:00-1:30	*"Theory of Motor Proteins That Change the Length of Filaments"* **Meredith Betterton**, University of Colorado

1:30-2:00	*"Modeling Neck Linker Extension in Kinesin"* **John Fricks**, Pennsylvania State University
2:00-2:30	Coffee Break
2:30-3:00	*"Collective Dynamics of Processive Molecular Motors"* **Scott McKinley**, University of Florida
3:00-3:30	Working group report and discussion (**Peter Kramer**, Rensselaer Polytechnic Institute)

Friday, November 19
SAMSI

8:30-9:00	Registration and Continental Breakfast
	Session on Qualitative Behavior of Stochastic Dynamic Systems
9:00-9:30	*"The Hahn-Jordan Decomposition for Signed Measure Valued Stochastic Partial Differential Equations"* **Peter Kotelenez**, Case Western University
9:30-10:00	*"Ergodicity of Systems with Singular Interaction Terms"* **Jonathan Mattingly,** Duke University
10:00-10:30	Coffee break
10:30-11:00	"*Stochastic Coalescence, Shock Clustering, and Burgers Turbulence"* **Ravi Srinivasan**, University of Texas
11:00-11:30	Working group report and discussion (**Jonathan Mattingly**, Duke University)
11:30-12:00	Closing remarks **Peter Kramer**, Rensselaer Polytechnic Institute **Jonathan Mattingly**, Duke University **Sorin Mitran**, University of North Carolina
12:00-1:00	Lunch

SPEAKER TITLES/ABSTRACTS

Meredith Betterton
University of Colorado, Boulder
mdb@colorado.edu

"Theory of Motor Proteins That Change the Length of Filaments"

In cells, multiple proteins alter the length of filaments called microtubules. This work develops the theory of motor proteins that bind to and move along microtubules, while also altering microtubule polymerization dynamics and therby changing the length of the
microtubule. The theory describes crowding and collective effects in the motor motion as well as the coupled dynamics of motors and the shortening microtubules. The goal is to build the connection between single-molecule motor properties and the collective effects of multiple motors acting on the filament. Under the right conditions, coupled motor motion and microtubule shortening can act as a microtubule length sensor. Important applications of this model are to microtubule shortening by kinesin-8 proteins in mitotsis and the control of antiparallel microtubule overlaps by PCR1 and Xklp1 in anaphase.

John Fricks
Pennsylvania State University
fricks@stat.psu.edu

"Modeling Neck Linker Extension in Kinesin"

The kinesin molecular motor family takes a single 8 nanometer step forward for each ATP hydrolyzed except in rare cases. Recent experiments have demonstrated mulitiple steps including frequent back steps may be possible if the necklinker connecting the heads of the kinesin are extended. This talk will present a detailed intra-step model of kinesin stepping which allows for multiple steps and show that asymptotic quantities can be caclulated using a combination of limit theorems for semi-Markov processes and matrix analytic techniques for Markov chains.

John Harlim
North Carolina State University
jharlim@ncsu.edu

"Numerical Strategies for Filtering Partially Observed Stiff Stochastic Differential Equations"

In this talk, I will describe a fast numerical strategy for filtering stochastic differential equations with multiscale features. This method is designed such that it does not violate the practical linear observability condition and, more importantly, it does not require the expensive cross correlation statistics between multiscale variables that are typically needed in standard filtering approach. The proposed filtering algorithm comprises of a ``macro-filter'' that borrows ideas from the Heterogeneous Multiscale Methods and a ``micro-filter'' that reinitializes the fast microscopic variables to statistically reflect the unbiased slow macroscopic estimate obtained from the macro-filter and observations of the macroscopic variables at asynchronous times. The proposed micro-filter is essentially equivalent to solving an inverse problem for parameterizing differential equations. Numerically, I will show that this microscopic reinitialization is an important novel feature for accurate filtered solutions, especially when the microscopic dynamics is not mixing at all.

Giles Hooker
Cornell University
gjh27@cornell.edu

"Control Theory and the Design of Dynamic Experiments"

This talk considers the problem of experimental design in which a dynamical system is the subject of the experiment. Common examples include single-neuron patch clamp experiments and lab-based ecological studies. Experimental protocols typically fix system inputs at constant values or call for pre-defined functional forms for input values, usually step or linear changes.

In this talk, we examine the problem of designing experimental protocols that can adapt to system observations in order to optimize the estimation of system parameters of interest. This can be particularly important when certain parameters affect system behavior only in particular regions of state space. In particular, we attempt to choose controls to maximize Fisher Information. This problem can be expressed within the context of control theory and we explore the properties of maximally informative designs. The problem becomes more difficult when system states are only indirectly, and noisily, observed and some approaches to this issue are proposed.

Peter Kotelenez
Case Western Reserve University
pxk4@case.edu

"The Hahn-Jordan Decomposition for Signed Measure Valued Stochastic Partial Differential Equations"

We define several metrics on the space of finite signed Borel measures on $\mathbb{R}^d$. The metrics give special value to signed measures in Hahn-Jordan decomposition. Assuming that the SPDEs are generated by flows of SODEs and that their mass are conserved we show that the solution lives on balls with constant radius if the Hahn-Jordan decomposition of the initial configuration is preserved. A local Lipschitz condition on the SODE implies the validity of our result.

Co-author: Brad Seadler, Case Western Reserve University

Peter Kramer
Rensselaer Polytechnic Institute
kramep@rpi.edu

"Improving Statistical Components of Multi-scale Simulation Schemes"

I will describe an interaction that emerged between members of the ``Numerical Methods for Stochastic Systems" working group in the Stochastic Dynamics program and the "Interaction of Deterministic and Stochastic Models" working group in the contemporaneous Space-Time Analysis for Environmental Mapping, Epidemiology, and Climate Change program. The second group, comprising both applied mathematicians and statisticians, studied multiscale computing approaches, particularly the Heterogeneous Multiscale Method (HMM) and the novel time-parallel Continuum-Kinetic-Molecular (tpCKM) method developed and exposed by Sorin Mitran (UNC) in the Numerical Methods working group. We identified some statistical aspects of each of these computational methods which are currently handled in very simple ways but show significant potential for improvement. I will describe the ideas which we have been discussing and are presently pursuing toward a more sophisticated statistical treatment of the interaction between stochastically simulated data at different levels of the multiscale simulation hierarchy.

Jonathan Mattingly
Duke University
jonm@math.duke.edu

"Ergodicity of Systems with Singular Interaction Terms"

I will consider two examples of physical systems which interact through a singular potential and are subject to thermal excitation: Namely a pair of two beads coupled by spring but sitting in a random Stokes flow and two Hamiltonian particles interacting through a Lenard-Jones type potential. In both cases, we will see that the systems do not fit neatly into the standard theorems used to prove Unique Ergodicity of stochastic differential equations. I will describe the necessary ideas to extend the standard results. The talk will have a pedagogical nature as I will start by reviewing some basic ideas from Harris chains.

Scott McKinley
University of Florida
scott.mckinley@ufl.edu

"Cooperative Dynamics of Processive Molecular Motors"

Internal to biological cells is an intricate transport apparatus, which consists of a cytoskeletal network of thin filaments called microtubules and proteins called processive molecular motors. These motors move in a directed fashion along the microtubules while generating sufficient force to tow biomolecular cargos. While experimentalists and theoreticians have had good success in describing the dynamics of a single motor with an attached cargo, there is considerable debate regarding how multiple motors cooperate and/or interfere with each other when they are attached to the same cargo.

Our team from last year's SAMSI working group on Stochastic Biological Dynamics has proposed an SDE model for this scale of intracellular transport and we present some findings which resolve one of the outstanding mysteries encountered by experimentalists:
common sense says there are multiple motors acting in concert, but certain measurements seem to imply that only one motor is working at a time.

Bruce Rogers
Duke University
bruce@math.duke.edu

"Thresholds for Emergence of Multi-scale Behavior of Epidemics on Networks"

Consider the contact process on a strongly clustered graph, described with a stochastic block model. Suppose there are 2 communities in the graph where the probability of an intra-community edge is p1 and the probability of inter-community edge is p2. We start the contact process with a single infected in one of the communities. If p2<<p1, the contact process respects the "bottle-neck" between the communities, where the with-in group contacts occurs at a much higher rate that between group contacts. In this case, a plot of the number of active sites versus time shows a stair-step where the process jumps from one community to the other. We present a heuristic argument to determine conditions on p1 and p2 where there is a sufficient separation of scales to observe the "stair-step" behavior.

Ravi Srinivasan
University of Texas
rav@math.utexas.edu

"Stochastic Coalescence, Shock Clustering, and Burgers Turbulence"

A remarkable stochastic coalescence model arises from the study of shock statistics in scalar conservation laws with random initial data. While originally rooted in the context of Burgers turbulence, this model has deep connections to statistics, kinetic theory, random matrices, and completely integrable systems. We show that the evolution takes the form of a Lax pair which, in addition to yielding interesting conserved quantities, admits some rather intriguing exact solutions.

Amanda Traud
North Carolina State University
altraud@hotmail.com

"Cascade Induced Synchronous Dynamics on Different Network Topologies"

We start with a current-based Integrate-and-Fire neuron model with instantaneous synaptic coupling and create equivalent Markov chain models by discretizing voltage and/or time, while keeping the mean firing rate of a single neuron constant. We connect these model neurons with real world C. Elegans neuronal network data, as well as, different random network topologies and account for excitatory neuron interactions. We investigate the interactions of the models with each network by comparing network firing rates and synchronizability.

Xueying Wang
Texas A&M University
xueying@math.tamu.edu

"Pairwise Closure Approximations in Epidemic Models on Regular Networks"

Mathematical modeling of contagion dynamics spreading on social networks is of great interest from both theoretical and numerical point of view. Pair approximations for dynamic epidemic models incorporate network structures, including the degree distribution and degree correlations, into the models, one that can capture the spatial correlation of the networks. To close the approximations for epidemic dynamics, various closure approximations have been considered.

In this work, we compare pairwise closure approximation (PAC) for the Susceptible-Infective-Susceptible (SIS) model [K. Eames and M. Keeling, 2002] and that for the Susceptible-Infective-Recovered-Susceptible (SIRS) model [G. Rozhnova and A. Nunes, 2009] on regular graphs. We identify a connection between these works. By analyzing a stochastic PAC SIRS epidemic model on regular networks, we show that in the limiting case, the two dimensional principle component of the stochastic paths, which exhibit sustained oscillations, is essentially an Ornstein-Uhlenbeck process up to a time dependent rotation. This is the joint work with Dr. Priscilla Greenwood.

Jennifer Young
Rice University
jjyoung@rice.edu

“A Continuum-Microscopic Modeling Method for Materials with Dynamic, Heterogeneous Micro-Structures”

Creating accurate, macroscopic scale models of microscopically heterogeneous media is computationally challenging. The difficulty is increased for materials with time-varying micro-structures. This talk will present a new continuum-microscopic (CM) modeling approach aimed at modeling such materials. Fibrous media are chosen as a class of materials upon which to present and test the algorithm. What is novel about this algorithm, compared to other CM methods, is that information from the material's micro-structure is saved over time in the form of probability distribution functions (PDFs). These PDFs are then extrapolated forward in time to predict what the micro-structure will look like in the future. Keeping track of the micro-structure over time allows for a more accurate computation of the local mechanical parameters used in the continuum-level equations. Results show that the mechanical parameters computed with this algorithm are similar to those computed with a fully-microscopic model. Errors for continuum level variables in the 5-10% range are deemed an acceptable trade-off for the savings in computational expense offered by this algorithm.

#9: Program on Complex Networks – Dynamics of Networks Workshop, January 10-12, 2011

SCHEDULE

Monday, January 10
SAMSI

8:30-9:00 Registration and Continental Breakfast

9:00-9:45	*"The "OF" Working Group"* **Peter Mucha**, University of North Carolina
9:45-10:30	*"Modeling Complex Social Interaction Within and Across Settings via Relational Events"* **Carter Butts**, University of California, Irvine
10:30-11:00	Break
11:00-11:45	"*Modeling the Dynamics of Biological Signaling Networks"* **Reka Albert**, Pennsylvania State University (either M, Tues-even + Wed AM)
11:45-12:30	*"Likelihood and Likelihood-free Inference for Certain Growing Network Models"* **Carsten Wiuf**, Aarhus University
12:30-2:30	Lunch and Breakout Sessions
2:30-3:15	*"Classifying Evolving Networks"* **J.P. Onnela**, Harvard University
3:15-3:45	Break
3:45-4:30	Working Group Talk
4:30-6:30	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, January 11
SAMSI

8:30-9:00	Registration and Continental Breakfast
9:00-9:45	*"Temporal Patterns in Cybersecurity Networks"* **Aric Hagberg**, LANL
9:45-10:30	*"Escaping from the Matrix: Storing, Exploring, and Explaining Dynamic Networks"* **Skye Bender-deMoll**

10:30-11:00	Break
11:00-11:45	*"Multiway Array Models for Dynamic Networks and Relational Data "* **Peter Hoff**, University of Washington
11:45-12:30	*"A Separable Model for Dynamic Networks"* **Pavel Krivitsky**, Carnegie Mellon University
12:30-2:30	Lunch & Breakout Sessions
2:30-3:15	"*Dynamics of Social Interactions at Short Timescales"* **Ginestra Bianconi**, Northeastern University
3:15-3:45	Break
3:45-4:30	*"Dynamic Community Structure in Adaptive Systems"* **Dani Bassett**, University of California, Santa Barbara

Wednesday, January 12
SAMSI

8:30-9:00	Continental Breakfast
9:00-9:45	*"Visualizing the Structure and Evolution of Science & Technology"* **Katy Borner**, Indiana University
9:45-10:30	Working group talk
10:30-11:00	Break
11:00-11:45	Brainstorming/Closing
11:45-1:00	Lunch and Adjourn

SPEAKER TITLES/ABSTRACTS

Reka Albert
Pennsylvania State University
ralbert@phys.psu.edu

"Modeling the Dynamics of Biological Signaling Networks"

Signal transduction networks describe the steps of information transfer between extracellular signals and within-cell processes. These networks exhibit a considerable diversity of nodes (including mRNA, proteins, small molecules) and edges (including enzyme-catalyzed reactions, protein-protein interactions and transcriptional regulation). Additionally, the nodes generally represent molecular species and not individuals. Consequently the network needs to be complemented with information on the nodes' states (e.g. abundance or concentration). In this context, network dynamics refers to the description of how the nodes' states vary in time and depending on the interactions they participate in. This talk will outline the methods of modeling signal transduction network dynamics and give examples of the biological applications and implications.

Dani Bassett
University of California, Santa Barbara
dbassett@physics.ucsb.edu

"Dynamic Community Structure in Adaptive Systems"

Modular architecture is thought to provide the necessary system adaptability for learning phenomena to occur. We describe the application of dynamic community detection to a temporally evolving system during a learning paradigm and develop a range of evolving network null models to aid in statistical validation. Our work is illustrated in the context of two complementary data sets gathered during the human acquisition of a simple motor skill. Using human brain functional connectivity networks, we show that temporal flexibility of a node's allegiance to a module can be used to predict learning. In a second study, community detection was applied to human behavioural movement networks and revealed movement ``chunking'', a behavioural phenomenon characterized by the organization of items into manageable units, which was strongly related to underlying neurophysiological activity. Together, these two studies represent important biological applications of dynamic community detection to address long standing questions in human neuroscience.

Skye Bender-deMoll
skyebend@skyeome.net

"Escaping from the Matrix: Storing, Exploring, and Explaining Dynamic Networks"

Network dynamics present new representation and analysis challenges. When edgesets, nodes sets, and node attributes can all change over time, we need new data structures to facilitate representation and analysis. I'll discuss several approaches I've worked with, and some rough categorizations of dynamic network types. I'll also share some ideas about characterizing some transmission-relevant properties of dynamic networks and demonstrate some animations and interactive visualization techniques that can help to communicate with colleagues and broader audiences.

Ginestra Bianconi
Northeastern University
g.bianconi@neu.edu

"Dynamics of Social Interactions at Short Timescales"

Recently new data on face-to-face interactions show that the dynamic of social interactions strongly deviates from a Poisson process. Indeed duration of contacts and inter-contact time intervals are distributed as a power-law, and not as an exponential. These data requires the formulation of new models of social interactions at short time scales. Here we propose a model based on a reinforcement dynamics driven by memory effect that nicely reproduce the observation coming from the data. The origin of this reinforcement dynamics will be discussed and related to a positive feedback occurring at the cognitive level.

Katy Borner
Indiana University
katy@indiana.edu

"Visualizing the Structure and Evolution of Science & Technology"

Well designed visualizations play an important role when making sense and communicating the results of network studies within and across disciplinary boundaries but also to the general public. While the visualization of static networks is supported by diverse commercial and research tools, the visualization of dynamic networks remains an open research problem.

This talk presents general strategies and examples for designing insightful yet aesthetically pleasing visualizations of small, medium, and large scale networks. Specifically, we will present and discuss visualizations of the structure and evolution of science itself, e.g., funding patterns, evolving co-authorship networks, knowledge diffusion networks. Tools like the Network Workbench Tool—designed to support the analysis and visualization of networks—and the Sci^2 Tool—designed for the study of science using scientific means—will be showcased.

Links:

- Cyberinfrastructure for Network Science Center at IU: http://cns.iu.edu
- Network Workbench Tool: http://nwb.cns.iu.edu
- Science of Science Tool: http://sci2.cns.iu.edu
- Mapping Science Exhibit: http://scimaps.org

Carter Butts
University of California, Irvine
buttsc@uci.edu

"Modeling Complex Social Interaction Within and Across Settings via Relational Events"

Social behavior over short time scales is often understood in terms of "actions," which can be thought of as discrete events in which one individual emits a behavior directed at one or more other entities in his or her environment (possibly including him or herself). As many different lines of research have emphasized, the propensity to take a given action can depend upon features of the actor's immediate environment, past history, cognitive state, and ongoing relationships. This poses a particular challenge when trying to understand action in complex settings where multiple behavioral mechanisms are plausibly involved. In this talk, I describe one framework for dealing with this problem, a flexible, event-based modeling scheme that supports likelihood-based inference for interacting behavioral mechanisms with complex dependence. Using a hierarchical generalization of the basic framework, I also show how narrative data from multiple sources -- archival materials, informant accounts, or behavioral traces -- can be employed to identify common behavioral mechanisms across multiple settings. The approach is illustrated with applications to responder communication data and classroom interactions.

Aric Hagberg
LANL
hagberg@lanl.gov

"Temporal Patterns in Cybersecurity Networks"

We collect and analyze data packet flows from multiple vantage points in a large computer network for the purpose of classifying normal and abnormal user activity. From this data we build time-labeled flow graphs of traffic in the network. By decomposing those graphs into temporal subgraphs we build a model that can be used to find unusual graph patterns in the background of normal activity.

Peter Hoff
University of Washington
pdhoff@u.washington.edu

"Multiway Array Models for Dynamic Networks and Relational Data "

Network dynamics on a given set of nodes can be described by a three-way array, having two dimensions representing the nodes and one dimension representing time. Descriptions of the main features of a dynamic network can be obtained by using reduced-rank array approximations. In this talk, we discuss how these reduced-rank arrays can be used to build statistical models of network dynamics, thereby providing a framework for parameter estimation, prediction and statistical inference. We illustrate this multiway modeling approach on several dynamic relational datasets.

Pavel Krivitsky
Carnegie Mellon University
pavel@stat.cmu.edu

“A Separable Model for Dynamic Networks”

Models of dynamic networks — networks that evolve over time — have manifold applications. We develop a discrete-time generative model for social network evolution that inherits the richness and flexibility of the class of exponential-family random graph models. The model facilitates separable modeling of the tie duration distributions and the
structural dynamics of tie formation. We develop likelihood-based inference for the model, and provide computational algorithms for maximum likelihood estimation. We illustrate the interpretability of the model in analyzing a longitudinal network of friendship ties within a school.

Co-Author: Mark S. Handcock

Peter Mucha
University of North Carolina
mucha@unc.edu

“The "OF" Working Group”

Meeting weekly since the opening workshop, the "OF" working group has developed a collection of research projects worked on by subgroups of members. As a means of jumpstarting workshop discussions, some of the themes of those projects are summarized broadly here.

J.P. Onnela
Harvard University
Onnela@hcp.med.harvard.edu

“Classifying Evolving Networks”

We develop a tool for classifying evolving networks. Our approach is based on probing the mesoscopic properties of networks, an important source of heterogeneity for their structure and function. We demonstrate how the framework can be used to study (1) the evolution of voting blocs in time in the US Congress, (2) the different stages of growth of fungal mycelial networks extracted from time series of digitized images of colony growth, and (3) stock-return correlation networks for the NYSE over two decades.

Carsten Wiuf
Aarhus University
carsten.wiuf@gmail.com

"Likelihood and Likelihood-free Inference for Certain Growing Network Models"

In many areas of computational biology, the likelihood of a model is intractable, either because of the large amounts of relevant data or because interesting models are highly complex. This is in particular the case for models of biological networks. In this talk the application is to Protein Interaction Networks (PINs) that typically comprise thousands of proteins and interactions. PINs are the result of long evolutionary histories. Here we adopt simple, qualitative mathematical models that capture essentials of protein evolution and develop statistical methods to estimate evolutionary PIN parameters. Our initial approach is based on Importance Sampling and a recursion for the likelihood, but it becomes computationally intractable for reasonably sized networks. A second approach is based on summary statistics and likelihood-free inference, such as Approximate Bayesian Computation (ABC). We discuss various problems with selection of summaries, convergence, and credibility and apply the methods on H. pylori and P. falciparum data. Additionally, I will discuss how the ABC framework can be used in relation to model criticism. Evolutionary models in biology are simple (and qualitatively) and appropriate tools for critsicims and refinement of models are indespensible for advancing our understanding of the underlying evolutionary processes of biological networks

#10: Program on Education and Outreach Program, Two-Day Undergraduate Workshop, February 25 – 26, 2011

SCHEDULE
Friday

8:30	Shuttle from Radisson to SAMSI (Group #1)
8:45	Shuttle from Radisson to SAMSI (Group #2)
9:15-9:30	**Pierre Gremaud**, NCSU and SAMSI Welcome and Introduction
9:30-10:20	**Jim Ramsay**, McGill University "Introduction to Functional Data Analysis"
10:20-10:50	**Ci-Ren Jiang**, SAMSI MATLAB Demo
10:50-11:10	Break
11:10-12:00	**Snehalata Huzurbazar**, U. of Wyoming "Current Collaborative Research Projects which use AOD"
12:00-1:00	Lunch

1:00-1:50	**Hulin Wu**, U. of Rochester “Dynamics Modeling as a Weapon to Defend Ourselves Against Threats from Infectious Diseases and Bioterrorist Attacks”
1:50-2:20	**Junheng Ma**, SAMSI R Intro
2:20-2:50	Break
2:50-3:40	**Hongxiao Zhu**, SAMSI "Statistical Analysis of Functional Object Data"
3:40-4:40	**Snehalata Huzurbazar**, U. of Wyoming and **Sylvie Tchumtchoua**, SAMSI MATLAB Lab
4:40-5:00	**Pierre Gremaud**, Career Options
5:00 5:20	Shuttle to Radisson (Group #1) Shuttle to Radisson (Group #2)
6:00	Dinner @ Radisson
Saturday	
8:15 8:30	Shuttle from Radisson to SAMSI (Group #1) Shuttle from Radisson to SAMSI (Group #2)
9:00-9:50	**Yolanda Munoz-Maldonado**, Michigan Tech "Sample Size Calculation for Functional Data"
9:50-10:40	**David Degras**, SAMSI “Research Topics in Functional Data Analysis”
10:40-11:00	Break
11:00-12:00	**Yolanda Munoz-Maldonado**, Michigan Tech and **David Degras**, SAMSI R lab
12:00	Adjourn
12:00 12:30	Shuttle from SAMSI to RDU (Group #1) Shuttle from SAMSI to RDU (Group #2)

#11: Program Complex Networks, Dynamics on Networks Workshop, March 21 - 23, 2011

SCHEDULE

Monday, March 21
SAMSI

8:30-9:00	Registration and Continental Breakfast
9:00-9:30	**Vittoria Colizza**, ISI and INSERM *"Human Mobility in an Emerging Epidemic: A Key Aspect for Response Planning"*
9:40-10:10	**Elizabeth Liecht**, Oxford University *"Interacting Networks: Formalism and Significance"*
10:10-10:40	Break
10:40-11:10	**Joel Miller**, Harvard University *"Epidemic Spread in Networks with One Equation"*
11:20-11:50	**Duygu Balcan**, University of Indiana *"Phase Transitions in Contagion Processes Mediated by Recurrent Mobility Patterns"*
12:00-2:30	Lunch
2:30-3	**Charlie Brummitt**, University of California, Davis *"Sandpile Cascades on Interacting Tree-like Networks"*
3:15-3:45	**James Gleeson**, University of Limrick *"Analytical Results for Cascades on Networks"*
4:00-4:30	**Mason Porter**, Oxford University *"Cascades on Networks"*
4:30-6:30	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, March 22
SAMSI

8:30-9:00	Registration and Continental Breakfast

9:00-9:30	Sid Redner, Boston University *"The Role of Reinforcement in Social Dynamics"*
9:40-10:10	Nicholas Lanchier, Arizona State University *"The Axelrod Model for the Dissemination of Culture Revisited"*
10:10-10:40	Break
10:40-11:10	Sara Solla, Northwestern University *"Fast and Slow Dynamics in Neural Networks With Small-World Connectivity"*
11:20-11:50	Deena Schmidt, Ohio State University *"Network Structure and Dynamics of Sleep-Wake Regulation"*
12-2:30	Lunch
2:30-3:00	John McSweeney, Concordia University and SAMSI *"Single-Seed Cascades on Networks with Triangles"*
3:15-3:45	Rachel Kranton, Duke University *"Strategic Interaction and Networks"*
4:00-4:30	Michael Macy, Cornell University *"The Critical Mass in Complex Contagion"*
6ish	Self-organized Dining in Durham

Wednesday, March 23
SAMSI

8:30-9:00	Registration and Continental Breakfast
9:00-9:30	**David Sivakoff**, SAMSI *"Contact Process on Modular Networks"*
9:40-10:10	**Shirhsendu Chatterjee**, Cornell University *"Latent Voter Model on Random Regular Graphs"*
10:10-10:40	Break
10:40-11:10	**Shankar Bhamidi**, University of North Carolina *"Flows, First Passage Percolation and Random Disorder in Networks"*
11:10-Noon	Discussion
Noon-1:00	Lunch and Adjourn

SPEAKER TITLES/ABSTRACTS

Duygu Balcan
Indiana University
balcand@indiana.edu

"Phase Transitions in Contagion Processes Mediated by Recurrent Mobility Patterns"

Spatial spread of infectious diseases, diffusion of rumors, and emergence of consensus are a few examples of the wide range of contagion processes mediated by human mobility and activity patterns. Human mobility patterns have been shown to be dominated by a set of specific locations and recurrent flows, poorly modeled by random diffusive dynamics generally used to study spreading processes. Here we develop a theoretical framework to analyze the behavior of contagion processes on a network of locations where individuals have memory of their location of origin. We find a phase transition between a regime in which the contagion phenomenon affects a macroscopic fraction of the system and a regime in which only a few locations are affected. This threshold cannot be uncovered by continuous models as it is related to the stochastic features of the the contagion process. The phase transition defines an invasion threshold for the contagion process that depends on the mobility parameters, providing guidance on how to control disease and information spread by constraining mobility processes. We recover the threshold behavior in the analysis of diffusion processes mediated by real data from human commuting patterns.

Shankar Bhamidi
University of North Carolina
bhamidi@email.unc.edu

"Flows, First Passage Percolation and Random Disorder in Networks"

Consider a connected network and suppose each edge in the network has a random positive edge weight. Understanding the structure and weight of the shortest path between nodes in the network is one of the most fundamental problems studied in modern probability theory and goes under the name first passage percolation. It arises as a fundamental building block in many interacting particle system models such as the spread of epidemics on networks. To a large extent such problems have been only studied in the context of the n-dimensional lattice. In the modern context these problems take on an additional significance with the minimal weight measuring the cost of sending information while the number of edges on the optimal path (hopcount) representing the actual time for messages to get between vertices in the network. Given general models of random graphs with random edge costs, can one develop techniques to analyze asymptotics of functionals of interest which are robust to the model formulation? The aim of this talk is to describe a heuristic based on continuous time branching processes which gives very easily, a wide array of asymptotic results for random network models in terms of the Malthusian rate of growth and the stable age distribution of associated branching process. These techniques allow us to solve not only first passage percolation problems rigorously but also understand functionals such as the degree distribution of shortest path trees, congestion across edges as well as asymptotics for "betweeness centrality" a concept of crucial interest in social networks, in terms of Cox processes and extreme value distributions. These techniques also allow one to exactly solve models of “weak disorder” in the context of the stochastic mean field model of distance, a model of great interest in probabilistic combinatorial optimization.

Charlie Brummitt
University of California, Davis
cbrummitt@math.ucdavis.edu

“Sandpile Cascades on Interacting Tree-like Networks”

The vulnerability of an isolated network to cascades is fundamentally affected by its interactions with other networks. Motivated by failures cascading among electrical grids, we study the Bak-Tang-Wiesenfeld sandpile model on two sparsely-coupled random regular graphs. By approximating avalanches (cascades) as a multi-type branching process and using a generalization of Lagrange’s expansion to multiple variables, we calculate the distribution of avalanche sizes within each network. Due to coupling, large avalanches in the individual networks are mitigated—in contrast to the conclusion for a simpler model [Newman et al., HICSS 2005]. Yet when compared to uncoupled networks, interdependent networks more frequently suffer avalanches that are large in both networks. Thus sparse connections between networks stabilize them individually but destabilize them jointly, as coupling introduces reservoirs for extra load yet also inflicts new stresses. These results suggest that in practice, to greedily mitigate large avalanches in one network, add connections between networks; conversely, to mitigate avalanches that are large in both networks, remove connections between networks. We also show that when only one network receives load, the largest avalanches in the second network increase in size and in frequency, an effect that is amplified with increased coupling between networks and with increased disparity in total capacity. Our framework is applicable to modular networks as well as to interacting networks and provides building blocks for better prediction of cascading processes on networks in general.

Shirshendu Chatterjee
Cornell University
sc499@cornell.edu

“Latent Voter Model on Random Regular Graphs”

Latent voter model was considered in the mean-field setup by Lambiotte et al. We analyze this model on random regular graphs. In this model, each vertex of the underlying graph has a voter who has one of the two opinions) and 1 and is either active or inactive. When a voter is active, it adopts the opinion of an uniformly chosen neighbor at rate 1, and if this causes to changes its opinion, it enters the inactive phase. A voter stays inactive for an independent exponentially distributed amount of time with mean 1/lambda, during which it doesn't change its opinion. When lambda is large, although the model is a small perturbation of the ordinary voter model, the behavior changes discontinuously as the quasi-stationary density of each opinion tends to 1/2 irrespective of any positive starting level. The idea can be used to analyze many nonlinear voter models on locally tree-like random graphs.

Vittoria Colizza
INSERM - Institut National de la Santé et de la Recherche Médicale
vittoria.colizza@inserm.fr

"Human Mobility in an Emerging Epidemic: A Key Aspect for Response Planning"

Human mobility and interactions represent key ingredients in the spreading dynamics of an infectious disease. The flows of traveling individuals form a network characterized by complex features, such as strong topological and traffic heterogeneities, that unfolds at different temporal and spatial scales, from short ranges to the global scale. Computational models can be developed that integrate detailed network structures based on demographic and mobility data, in order to simulate the spatial evolution of an epidemic. Focusing on the 2009 H1N1 influenza pandemic as a paradigmatic example, these approaches allow quantifying the effects of travel restrictions in delaying and controlling the epidemic spread. In addition, simplified model frameworks can be solved providing the assessment of the interplay between individual mobility and epidemic dynamics, under different conditions of heterogeneities characterizing human mobility.

James Gleeson
University of Limrick
James.Gleeson@ul.ie

"Analytical Results for Cascades on Networks"

Cascade dynamics can occur when the (binary) state of a node is positively affected by the states of its neighbours in the network, so that the number of affected nodes grows over time. Such models have been used to aid understanding of the spread of cultural fads and the diffusion of innovations (Watts 2002), and can be generalized to include a variety of other cascading dynamics on networks (Gleeson 2008). I will review the (tree-based) analytical approaches used to study these dynamics on configuration model and degree-correlated networks, and then present recent results on the spread of cascades on (synthetic) random networks with significant clustering (Newman 2009, Gleeson 2009).

Rachel Kranton
Duke University
rachel.kranton@duke.edu

"Strategic Interaction and Networks"

This paper brings a general network analysis to a wide class of games, including strategic innovation, public goods, investment, and social interactions. The major interest, and challenge, is seeing how network structure shapes outcomes. We have a striking result. Equilibrium conditions depend on a single number: the lowest eigenvalue of a network matrix. When the graph is sufficiently tight (as measured by this eigenvalue), there is a unique equilibrium. When it is loose, stable equilibria always involve extreme play where some agents take no actions at all. We combine tools from potential games, optimization, and spectral graph theory to solve for all Nash and stable equilibria for any graph and for any impact of players' actions.

This paper is the first to uncover the importance of this network measure to social and economic outcomes, and we relate it to different network link patterns.

Nicholas Lanchier
Arizona State University
lanchier@math.asu.edu

"The Axelrod Model for the Dissemination of Culture Revisited"

The Axelrod model is a stochastic process based on the voter model which, in addition to social influence, also accounts for homophily, the tendency to interact more frequently with individuals which are more similar. Each individual is characterized by a set of cultural features, and pairs of neighbors interact at a rate proportional to the number of features they share, which results in the interacting pair having one more cultural feature in common. The Axelrod model has been extensively studied during the past ten years based on numerical simulations and simple mean-field treatments while there is a lack of analytical results for the spatial model. This talk gives rigorous clustering and coexistence results about the one-dimensional system that confirm some of the conjectures formulated by statistical physicists and social scientists.

Elizabeth Leicht
University of Oxford
elizabeth.leicht@wolfson.ox.ac.uk

"Interacting Networks: Formalism and Significance"

The majority of recent work in network theory has dealt with individual networks treated as isolated systems. However, many networks are actually just one component in a much larger complex system; such a system can bring together multiple networks with distinct topologies and functions. For instance, modern critical infrastructure spans assorted electric grids, telecom and computer networks, and transportation networks; likewise, a product development project brings together social networks and technical networks to form a system exhibiting so-called socio-technical congruence. Individual networks are increasingly interdependent and previously neglected or "hidden" inter-network connections can significantly impact our understanding of network structure. We develop a mathematical formalism for interacting networks focusing on both intra- and inter-network connectivity properties. With this formalism we show that the critical threshold for the onset of large-scale network connectivity in one network, i.e., the formation of a giant connected component, can be manipulated by the introduction of a multiple network framework. Applications of this work relate to our understanding of disease spreading across geographic regions and how to engineer minimalist communications networks.

Michael Macy
Cornell University
mwm14@cornell.edu

"The Critical Mass in Complex Contagion"

Why do some contagions “go viral” and others do not? Research on “small world” networks shows how a very small number of long-range ties that bridge between clusters can allow contagions to spread almost as rapidly as on a random network of equal density. Recent research shows how long-range ties that accelerate the spread of information and disease can impede the spread of complex contagions—behaviors, beliefs and preferences that diffuse via social influence and therefore often require contact with multiple adopters. In confirming this result analytically and extending the analysis from small world to power law networks, we discovered that complex contagions require a critical mass of infected nodes that corresponds to a phase transition in the ability of the contagion to take advantage of the ”shortcuts” created by long-range ties. We demonstrate how this critical mass is related to the dynamics of the contagion process and identify implications for modeling behaviors that spread via social influence, such as viral marketing and social movements.

John McSweeney
Concordia University and SAMSI
mcsweeney@samsi.info

“Single-Seed Cascades on Networks with Triangles”

Watts' cascade process on a network is defined as follows: each node is in one of two states, active or inactive. Each node x has a threshold t(x) whereby if the proportion of its neighbors that are active exceeds t(x), then it becomes active itself. This can be viewed as a special case of a wider class of processes that includes site, bond and bootstrap percolation, to name a few. Watts' model nevertheless presents interesting features since it is not monotone, in the sense that denser graphs are not necessarily more likely to promote the spread of the cascade. I will discuss how to combine tree-based techniques of Watts and of Gleeson et al. to analyze the spread of the cascade in a network with a high density of triangles, starting from a single activated seed.

Joel Miller
Harvard University
joel.c.miller@gmail.com

“Epidemic Spread in Networks with One Equation”

The structure of social interactions along which disease spreads can be represented using a network. When we investigate disease spread in networks we find that many mass action assumptions fail. Individuals with many contacts tend to become infected earlier, and in turn infect more individuals, leading to faster initial growth. However, the remaining population has fewer contacts than average, and so the growth rate decreases more rapidly and the epidemic dies out sooner than mass action predicts.

Unfortunately models tend to require (arbitrarily) many equations to correct this. Recent work by Volz (JMB 2008) found a low-dimensional system that exactly captures the dynamics. Work by Miller (JMB 2010) simplified this derivation and also simplified
the equations.

Our more recent unpublished work simplifies the derivation further, and allows easy generalization to a wide range of diseases and population structures, including populations whose contacts change in time. The key simplification comes from focusing our attention on the fraction of edges connecting to susceptible, infected, or recovered individuals rather than the fraction of the population with each status. We show how to derive these systems and compare the resulting predictions with simulation.

Mason Porter
Oxford University
porterm@maths.ox.ac.uk

"Cascades on Networks"

I will discuss ongoing research on cascades on networks that builds on some of the work discussed in James Gleeson's talk. I will first discuss the unreasonable effectiveness of tree-based theories for networks with significant clustering and give some hints as to when such theories might and might not work. I will then consider when mean-field theory is accurate for dynamics on networks constructed from real data. Time permitting, I will discuss in-progress work on multi-step cascades and other cascade dynamics.

Sid Redner
Boston University
redner@bu.edu

"The Role of Reinforcement in Social Dynamics"

In many models of social dynamics, some attribute is transferred by a single interaction between agents that causes an agent to change its state. What happens, however, if an agent requires multiple signals from its neighbors before changing state? The result is quite rich. We first present a simple fad propagation model in which an agent adopts the fad only after receiving K inputs to adopt from other agents. For a population size N, the time until the fad is adopted scales as ln N for k=1, and N^{(K-1)/K} for K>1. For the
Voter model with K inputs needed to change state, a two-time scale approach to consensus is found. There is a manifold of initial conditions that leads the population to almost immediately reach an "apathetic" state before consensus is reached in a time that scales as ln N.

Deena Schmidt
Ohio State University
dschmidt@mbi.osu.edu

"Network Structure and Dynamics of Sleep-Wake Regulation"

Sleep and wake states are each maintained by activity in a corresponding neuronal network, with mutually inhibitory connections between the networks. In infant mammals, the durations of both states are exponentially distributed, whereas in adults, the wake states yield a heavy-tailed distribution. What drives this transformation of the wake distribution? Is it the altered network structure or a change in neuronal dynamics? What properties of the network are necessary for maintenance of neural activity on the network and what mechanisms are involved in transitioning between sleep and wake states? We explore these issues using random graph theory, specifically looking at stochastic processes occurring on random graphs, and also by investigating the accuracy of predictions made by deterministic approximations of stochastic processes on networks.

David Sivakoff
SAMSI / Duke University
djsivy@math.duke.edu

"Contact Process on Modular Networks"

We studied the contact process, or SAS epidemic, on weakly coupled Erdos-Renyi random graphs. Simulations and theoretical results demonstrate that for certain parameter values, the time that it takes for the process spreads to the second component is an order of magnitude larger than the time that it takes the process to reach its quasi-stationary state within the first component.

Sara Solla
Northwestern University
solla@northwestern.edu

"Fast and Slow Dynamics in Neural Networks With Small-World Connectivity"

An understanding of emergent dynamics on complex networks requires investigating the interplay between the intrinsic dynamics of the node elements and the connectivity of the network in which they are embedded. In order to address some of these questions in a specific scenario of relevance to the dynamical states of neural ensembles, we have studied the collective behavior of excitable model neurons in a network with small-world topology. Thesmall-world network has local lattice order, but includes a number of randomly placed connections that may provide connectivity shortcuts. This topology bears a schematic resemblance to the connectivity of the cerebral cortex, in which neurons are most strongly coupled to nearby cells within fifty to a hundred micrometers, but also make projections to cells millimeters away. We find that the dynamics of this small-world network of excitable neurons depend mostly on both the density of shortcuts and the delay associated with neuronal projections. In the regime of low shortcut density, the system exhibits persistent activity in the form of fast propagating waves, which annihilate upon collision and are spawned anew via the reinjection of activity through shortcut connections. As the density of shortcuts reaches a critical value, the system undergoes a transition to failure. The critical shortcut density results from matching the time associated with a recurrent path through the network to an intrinsic recovery time of the individual neurons. Furthermore, if the delay associated with neuronal interactions is sufficiently long, activity reemerges above the critical density of shortcuts. The activity in this slow regime exhibits long, chaotic transients composed of noisy, large-amplitude population bursts.

#12: Program 0n Analysis of Object Data, AOOD Meets Evolutionary Biology Workshop
April 30 – May 2, 2011

SCHEDULE
Saturday, April 30
SAMSI

8:30-9:00	Registration and Continental Breakfast
9:00-10:00	Minute Madness (1-5 min talks) Moderator: **J.S. Marron**, University of North Carolina
10:00-10:40	Discussion and Break
10:40-11:00	<u>Open Problems in FVT:</u> **Joel Kingsolver**, University of North Carolina *"Genetic Variation and Evolution of Function-Valued Traits"*
11:00-11:20	**Jay Beder**, University of Wisconsin, Milwaukee *"Estimating the Selection Gradient"*
11:20-11:40	**David Houle**, Florida State University *"Connecting Phenomic Objects to Genomic Predictors"*
11:40-Noon	Discussion
Noon-2:00	Lunch
2:00-2:20	<u>Open Problems in FVT:</u> **Patrick Carter**, Washington State University *"Evolution of the Integrated Phenotype: A Function Valued Approach"*
2:20-2:40	**Nancy Heckman**, University of British Columbia *"Dependencies in the Analysis of Functional Data"*
2:40-3:00	**Heather Jamniczky**, University of Calgary *"Quantification of Unusual Biological Shapes in Three Dimensions"*
3:00-3:20	**Washington Mio**, Florida State University *"Spectral Methods in Shape Analysis"*
3:20-4:00	Break
4:00-5:30	Informal Group Discussion
5:30-7:30	**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.***

Sunday, May 1
SAMSI

8:30-9:00	Registration and Continental Breakfast
9:00-9:40	**<u>Open Problems in FVT:</u>** Saunak Sen**, University of California, San Francisco** *"Genetic Mapping of Function-valued Traits"*
9:40-10:00	Open Discussion
10:00-10:40	Break
10:40-11:00	**<u>Highlights of SAMSI AOOD:</u>** Daniel Gervini**, University of Wisconsin** *"Semiparametric Curve Registration"*
11:00-11:20	Sarang Joshi**, University of Utah** ***"Computational Anatomy: Simple Statistics on Interesting Spaces for Developing Imaging Biomarkers Analysis"***
11:40-12:00	John Aston**, Warwick University** ***"A Step Towards a Function-Valued Typology for Language"***
11:40-Noon	Discussion
Noon-2:00	Lunch
2:00-2:20	<u>Highlights of SAMSI AOOD:</u> **J. S. Marron**, University of North Carolina *"Report from SAMSI Working Group on Trees as Data Objects"*
2:20-2:40	Ezra Miller**, Duke University** *"Stratified Statistics for Evolutionary Biology"*
2:40-3:20	Discussions and Break
3:20-5:30	Informal Group Discussions

Monday, May 2
SAMSI

8:30-9:00	Continental Breakfast
9:00-10:15	Group Discussions
10:15-10:45	Break
10:45-Noon	Group Discussions

Noon-2:00 Lunch

2:00-5:00 Brainstorming/Closing

SPEAKER TITLES/ABSTRACTS

John Aston
Warwick University
j.a.d.aston@warwick.ac.uk

"A Step Towards a Function-Valued Typology for Language"

Traditionally linguistics has examined the relationships between languages using discrete shifts either in lexicographical entries or phonetic symbolic representations. However, it is becoming increasingly of interest to examine the acoustic phenomena that generate these shifts. However, acoustic changes by their very nature are continuous signals. In this work, we present a new statistical methodology based on functional principal components and mixed effect models to produce a phonetic typology of a language. The fundamental frequency (F0, or commonly pitch) will be characterised by both function-valued components as well as linguistic effects, such a tonal class, sentence position, sentence type, that will affect the shape of the function-valued component. It will be shown that several common functional shapes provide a parsimonious dictionary for describing pitch, and while different languages share common function-valued components, these vary in their importance to specific linguistic effects across the languages. (Joint work with Pantelis Hadjipantelis, University of Warwick, and Jeng-Min Chiou and Jonathan Evans, Academia Sinica)

Jay Beder
University of Wisconsin, Milwaukee
beder@uwm.edu

"Estimating the Selection Gradient"

In classical evolutionary biology, a quantitative trait is a finite-dimensional random vector \bz, whose fitness is modeled by a relative fitness function $w(\bz)$. The selection gradient is defined to be the vector of coefficients of the best linear predictor of w based on \bz. This vector is estimated by regressing observed fitness on observed traits.

In 1989 Kirkpatrick and Heckman introduced the study of infinite-dimensional or function-valued traits. We discuss the modeling of such traits, and propose a method for estimating the selection gradient in this setting. Gaussian processes and reproducing kernel Hilbert spaces play a central role.

Patrick Carter
Washington State University

pacarter@wsu.edu

“Evolution of the Integrated Phenotype: A Function Valued Approach”

Many complex traits important in basic and applied biology are “function-valued” in being naturally described as mathematical functions. Function-valued traits include gene expression profiles, life history patterns, and phenotypic responses to continuous environmental gradients. Many important biological questions for function-valued traits involve their variation, selection, or evolution and can best be assessed using a single integrated statistical framework. This function-valued approach is distinct from other methods because it directly uses the mathematical function that describes the trait as the primary unit of description. In this seminar I will show how the function-valued framework can be considered as an extension of multivariate quantitative genetic approaches, I will describe some of the empirical systems in my lab that examine the evolution of function-valued traits, and I will highlight some of the remaining statistical challenges that must be solved to produce a usable framework for evolutionary biologists.

Daniel Gervini
University of Wisconsin
gervini@uwm.edu

“Semiparametric Curve Registration”

In this talk we will present a semiparametric curve alignment procedure based on interpolating Hermite polynomials. This type of splines are not common in statistics, because in most applications interpolation is explicitly avoided, rather than pursued. But for modeling warping functions, landmark or pseudo-landmark interpolation is a strategy that provides parsimonious, interpretable and computationally simple warping functions. An advantage of this method over traditional landmark registration is that the pseudo-landmarks are automatically estimated, rather than specified by the user on a curve-by-curve basis. We will discuss possible extensions of this method to bivariate problems, and inferential issues related to non-i.i.d. curves, which are common in evolutionary biology.

Nancy Heckman
University of British Columbia
nancy@stat.ubc.ca

“Dependencies in the Analysis of Functional Data”

In functional data analysis, one usually assumes that curves are independent. However, in many applications, curves are dependent. The dependencies could arise from multiple observations on the same individual (eg when measuring lung capacity as a function of time on two separate hospital visits) or from spatial proximity (eg when recording temperature over time at each of many locations across British Columbia). In evolutionary biology and in animal breeding, dependency arises because individuals are related, and the theory of genetics provides us with a very specific model for this dependency. Analysis using this model can be carried out by random regression or a modification of PACE (Principal Analysis by Conditional Expectation, Yao, Muller and Wang 2005 – modification of PACE is being carried out with Yao and Lei). I will discuss this genetic model for dependency and how it can be used in functional data analysis. For discussion: can we extend these ideas to other models of dependency?

David Houle
Florida State University
dhoule@bio.fsu.edu

"Connecting Phenomic Objects to Genomic Predictors"

Organisms are objects made of functions, but are currently represented in our data mostly as scalar or vectors. A key goal of biology is to understand the totality of phenotypic characteristics, the phenome. The limited representation of organisms as vectors is a major impediment to this understanding. Our ability to represent and analyze phenomes lags our ability to characterize genomes. If these limitations can be overcome, phenomic data will allow us to address three key phenomena: the pleiotropic effects of genetic variation, the full range of causes of variation in phenotypic traits, and important interactions that do not have genetic causes. I will present examples of the challenges to phenomic understanding using Drosophila melanogaster as a model organism.

Heather Jamniczky
University of Calgary
hajamnic@ucalgary.ca

"Quantification of Unusual Biological Shapes in Three Dimensions"

Natural selection acts on phenotypic variation, but evolutionary change occurs at the level of the genotype. Genotype and phenotype are linked by organismal development in a complex and highly structured fashion, such that developmental processes have a substantial effect on the evolvability of complex phenotypes. Understanding this developmental basis for evolvability is the central question of evolutionary developmental biology. Our research makes extensive use of the techniques of landmark-based three-dimensional geometric morphometrics as we seek to quantify phenotypic variation in response to genetic and developmental perturbation. These techniques are limited by the ability to place homologous landmarks repeatably and reliably across a large sample of specimens. As we delve deeper into the relationship between genotype and phenotype, we find these methods to be insufficient for the quantification and comparison of a number of structures of interest. Two current avenues of active research interest for our group include the study of branching structures, and the study of gene and protein expression patterns in developing organisms. In the first case, the challenge is to find a means by which shape variation in a branching structure, such as a circulatory tree, may be quantitatively compared to variation in both other branching structures as well as in non-branching structures, such as skulls. In the second case, the challenge is to find a means by which shape variation in an irregular, amorphous patch of gene or protein expression within a tissue can be compared to other such patches, and to the shape of the structure in which it resides. We seek to develop methodologies by which we can begin to explore the developmental basis for evolvability as it pertains to instances such as these.

Sarang Joshi
University of Utah
sjoshi@sci.utah.edu

"Computational Anatomy: Simple Statistics on Interesting Spaces for Developing Imaging Biomarkers Analysis"

A primary goal of Computational Anatomy is the statistical analysis of anatomical variability. Large Deformation Diffeomorphic transformations have been shown to accommodate the geometric variability but performing statistics of Diffeomorphic transformations remains a challenge. In this talk I will develop notions of performing statistical analysis of manifold valued data and apply these concepts for characterizing neuroanatomical variations observed in neurological disorders such as dementia. I will start with the simple concept of defining the "Average Anatomy" and then extend this to the study of regression and co-variation of anatomical shape with independent variables. The motivation is to model the inherent relation between anatomical shape and clinical measures and evaluate its statistical significance. We use Partial Least Squares for the multivariate statistical analysis of the deformation momenta under the Large Deformation Diffeomorphic framework. The statistical methodology extracts pertinent directions in the momenta space and the clinical response space in terms of latent variables. We report the results of this analysis on 313 subjects from the Mild Cognitive Impairment group in the Alzheimer's Disease Neuroimaging Initiative (ADNI).

Joel Kingsolver
University of North Carolina
jgking@bio.unc.edu

"Genetic Variation and Evolution of Function-Valued Traits"

I'll give a brief introduction to genetic and environmental variation of quantitative traits, and how this may lead to evolution.

J.S. Marron
University of North Carolina
marron@email.unc.edu

"Report from SAMSI Working Group on Trees as Data Objects"

This will be an overiew talk about research accomplishments of the SAMSI Trees Working Group. Progress made on the three approaches of Combinatorics, Phylogenetics and Dyck Paths will be reported, including some recent surprising and interesting interactions.

Ezra Miller
Duke University
ezra@math.duke.edu

"Stratified Statistics for Evolutionary Biology"

Branched structures in evolutionary and structural biology, including phylogenetic trees and blood vessels, among many others, give rise to data sets for which traditional methods of statistics on vector spaces do not apply, because the data points live on spaces that are singular or curved, or both. Starting from the basics (What is a phylogenetic tree? How do distributions of them arise in practice? What kind of space is the set of phylogenetic trees?), this talk will present themes from the SAMSI working group on Stratified Spaces, including treatments of fundamental statistical notions such as intrinsic
mean and variance for distributions of phylogenetic trees.

Washington Mio
Florida State University
mio@math.fsu.edu

"Spectral Methods in Shape Analysis"

I will discuss methods to analyze biological shape using techniques from spectral geometry.

Saunak Sen
University of California, San Francisco
sen@biostat.ucsf.edu

"Genetic Mapping of Function-valued Traits"

Many traits of interest in genetic studies such as growth curves, skeletal shape and activity patterns, have temporal or spatial structure. They are better treated as function-valued observations. We will review current methods for genetic analysis of function-valued traits which are mostly likelihood-based, requiring specification of the distribution and error structure. Then we will present an alternative framework based on estimating equations that is robust to misspecification of the covariance structure. The method is quite fast as it uses a two-step least squares algorithm. It is thus well-suited for high volume data generated from automated and high throughput phenotyping technologies. We will conclude with some applications (mouse behavior, growth curves), and discuss open problems.

#13: Program 0n Complex Networks, Complex Networks Transition Workshop June 6 – June 7, 2011

SCHEDULE
Monday, June 6
SAMSI

9:00 – 9:30	Registration and Continental Breakfast
9:30-10:00	**Kash Balachandrian**, Duke University *"Comparison of Local Spectral Clustering Algorithms"*
10:10-10:40	**Alan Lenarcic**, University of North Carolina *"Multiple Latent Trait Model for Interaction in Expander Networks"*
10:50-11:20	Break
11:20-11:50	**Blair Dowling Sullivan**, Oak Ridge National Laboratories *"A Breadth-First Traversal of Tree Decompositions for Complex Networks"*
12:00-2:00	Lunch and Breakout Sessions
2:00-2:30	**David Banks**, Duke University *"Grooming Networks in a Baboon Troop"*
2:40-3:10	**Bruce Rogers**, SAMSI *"Fitting Monkey Models"*
3:20-3:50	Break
3:50-4:20	**Amanda Traud**, University of North Carolina *"Exploring Ant Social Networks: Movement, Models, and Numbered Ants"*
4:30-6:30	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, June 7
SAMSI

9:00–9:30	Registration and Continental Breakfast
9:30–10:00	**Tyler McCormick**, Columbia University *"Surveying Hidden Populations Through Sampled Respondents in a Social Network: A Comparison of Two Strategies"*
10:10-10:40	**Ali Shojaie**, University of Michigan *"Reconstructing Directed Regulatory Networks from Multiple Steady-State and Perturbed Gene Expression Profiles"*
10:50-11:20	Break
11:20-11:50	**Eric Kolaczyk**, Boston University *"Some Progress on Asymptotics for ERGMs"*
12:00-2:00	Lunch
2:00-2:30	**Tipan Verella**, University of Virginia *"The Dual of the Random Intersection Graph"*
2:45--	**Bill Shi**, University of North Carolina **David Sivakoff**, SAMSI *"Robust Scaling Behavior in Dynamic Voter Models"*

(with interruptions by **Peter Mucha**, University of North Carolina and **Rick Durrett**, Duke University)

SPEAKER TITLES/ABSTRACTS

Prakash Balachandrian
Duke University
prakash.balachandran@gmail.com

"Comparison of Local Spectral Clustering Algorithms"

In undirected networks, local spectral clustering algorithms have been popular when one wants to compute a cluster/community in the neighborhood of an input set of seed nodes with low conductance. In this talk we compare two local spectral clustering algorithms PageRank Nibble and LocalCut on a variety of networks with qualitatively different properties. We hope that this empirical work will aid in future research in the identification of core/periphery structures in undirected networks and in inferring overlapping community structure. This work is joint with Peter Mucha, Mason Porter, and Michael Mahoney.

David Banks
Duke University
banks@stat.duke.edu

"Grooming Networks in a Baboon Troop"

Susan Alberts and other primatologists have observed troops of baboons in Kenya for many years, recording various social interactions and information on kinship, genetics, dominance hierarchy, and many other covariates. These data provide an ideal testbed for developing dynamic network models: the social relationships are simple, and the longitudinal data covers decades. The application poses two important questions: Can one predict baboon troop fissions before they occur? And do baboons think about Simmelian ties when forming relationships? This talk summarizes several efforts to address these questions.

Eric Kolaczyk
Boston University
kolaczyk@bu.edu

"Some Progress on Asymptotics for ERGMs"

While mathematical modeling of network graphs is an area with both a long history and much attention of late, their statistical modeling and inference still remains relatively under-developed, particularly when compared to the state of affairs in fields like time series and spatial modeling. Even standard asymptotics for parameter estimation in canonical network models is still lacking. I will present some preliminary work in this direction, in the context of ERGMs with independent dyads. An important issue that enters into the question of asymptotics in this setting appears to be that of how best to calibrate the models with network size. This is joint work with Pavel Krivitsky.

Alan Lenarcic
University of North Carolina
alanjazztenor@gmail.com

"Multiple Latent Trait Model for Interaction in Expander Networks"

We posit a probability-based generating model for social network data, where actors are presumed to hold different affinities within multiple community groups. For certain distributions of affinities among the population, scale-free properties can be attained, particularly for "hairball" networks that are difficult to embed in latent Euclidean space. As a statistical likelihood, the affinities can be learned through inference using an EM model designed to identify matching cooperative events between actors. Scalable through kernel-density approximations as a O((V+E)*T) algorithm, we demonstrate an efficient Cuda-based GPU approach suitable for > 10K vertex problems. We apply this algorithm to Facebook historical networks and Biological co-exclusion pathways as a method to sort nodes between highly multi-trait gregarious actors, and those congregating according to a specific trait.

Tyler McCormick
Columbia University
tyler@stat.columbia.edu

"Surveying Hidden Populations Through Sampled Respondents in a Social Network: A Comparison of Two Strategies"

The sampling frame in most social science surveys excludes members of certain groups, known as hard-to-reach groups. One strategy for learning about these groups, or subpopulations, involves reaching hard-to-reach group members through their social network.

We compare the efficiency of two common methods for subpopulation size estimation using data from standard surveys. These data are examples designs that begin with a randomly sampled set of network members (nodes) and then reach other nodes indirectly through questions asked to the sampled nodes. We begin by viewing these two methods as variants of multi-stage cluster sampling. From this perspective, we derive estimates of sampling variance. These mental link tracing designs cost significantly less than traditional link tracing designs, yet introduce additional sources of potential bias. We examine the influence of one such source of bias using simulation studies. We then demonstrate our findings using data from the 2004 General Social Survey. Additionally, we provide survey design suggestions for practitioners designing future surveys incorporating such designs.

Bruce Rogers
SAMSI
bruce@math.duke.edu

"Fitting Monkey Models"

Baboon troops in Africa exhibit complex social interactions. We will outline an agent based model (ABM) describing the interactions among baboons in a single troop and offer a Bayesian framework for parameter estimation and model selection for ABMs. The ABM simulations output artificial social networks that must be compared with the observed network from field data, so we will offer a set of network statistics and functionals that seem appropriate for the comparison.

Bill Shi
University of North Carolina
bill10@email.unc.edu

David Sivakoff
SAMSI
djsivy@math.duke.edu

"Robust Scaling Behavior in Dynamic Voter Models"

Extensive research on models of opinion formation proceeds in two directions. One allows connections to form only between individuals with similar opinions; the other lets each individual change its opinion based on its neighbors' opinions. Inspired by Holme and Newman [Phys. Rev. E 74, 056108 (2006)], we study the combination of these two dynamic processes with one parameter controlling the coupling between them. We find a phase transition in the system as the parameter varies, from the case where one opinion dominates the network to another one where the density of opinions is conserved. The results point to two types of robust scaling behaviors between the opinion density and the rewiring probability.

Ali Shojaie
University of Michigan
shojaie@umich.edu

"Reconstructing Directed Regulatory Networks from Multiple Steady-State and Perturbed Gene Expression Profiles"

Reconstructing genetic regulatory networks is an important task in functional genomics. Data obtained from experiments that perturb genes by knock-outs or RNA interference contain useful information for addressing the reconstruction problem. However, such data can be limited in size and/or expensive to acquire. On the other hand, observational data of the organism in steady state are more readily available, but their informational content inadequate for the task at hand. We discuss a computational approach to appropriately utilize both data sources for estimating a regulatory network. The proposed method offers significant advantages over existing techniques. The proposed approach includes a three-step algorithm to estimate the underlying directed (cyclic or acyclic) regulatory network that uses as input both perturbation screens and steady state gene expression data. First, causal orderings of genes consistent with the perturbation data are determined. In the second step, for each ordering, a regulatory network is estimated using a penalized likelihood based method, while in the third step a consensus network is constructed from the highest scored ones. We establish the consistency of the estimated network in presence of informative perturbation data and present numerical results that indicate the algorithm performs well in uncovering the underlying network and outperforms competing approaches that rely only on a single data source.

Blair D. Sullivan
Oak Ridge National Laboratory
sullivanb@ornl.gov

"A Breadth-First Traversal of Tree Decompositions for Complex Networks"

This talk offers a forward-looking update on work related to several applications of tree decompositions in complex network analysis. Topics include: progress on the development of scalable and parallel codes for dynamic programming on tree decompositions (particularly to solve weighted independent set) and some of the challenges related to memory usage in these techniques; an update on the notion of tree-decomposition based centrality measures; plans to use tree-decompositions for parameter estimation (especially covariance) in noisy graphical models; future work on the relationship between tree-width and hyperbolicity, a geometric notion of "tree-like-ness".

Amanda Traud
University of North Carolina
altraud@email.unc.edu

"Exploring Ant Social Networks: Movement, Models, and Numbered Ants"

Communication is an integral part of living in a group, especially for animals, such as ants, that need to complete large tasks. Studying the interactions of ants within a single colony can help us to understand the social structure of these small creatures and how this structure may affect information flow and the spread of disease. We study the interactions among individuals by observing one communication type, antennation, and in order to understand these interactions, we compare observed behaviors to those seen in a number of mechanistic models. As a first step towards this, we compared interactions to those generated by a null model, namely Brownian motion, for which meetings between individuals have no impact on their movements. We shall discuss our results in the light of recently published work on degree distributions of empirical ant networks," The effect of individual variation on the structure and function of interaction networks in harvester ants" by Noa Pinter-Wollman, Roy Wollman, Adam Guetz, Susan Holmes and Deborah M. Gordon. Future work will involve comparing our empirical data to other movement models, both interactive and non-interactive, and studying how ant density affects network structure. A longer-term aim is to extend our study to other social animals, such as wolves, giraffes or prairie dogs, and to understand how network structure impacts spread of infection through these different social systems.

Tipan Verella
University of Virginia
tipan.verella@gmail.com

"The Dual of the Random Intersection Graph"

We will present ongoing work on the dual of the generalized random intersection graph. Conceptually, given a set of nodes and a set of attributes, the *dual of the random intersection graph* (DRIG) is a graph on the attributes induced by constructing a random bipartite graph between the nodes and the attributes, and saying that an edge exists between two attributes only when at least one node is connected to both those attributes in the random bipartite graph. We will present some preliminary results on the first two moments of the degree distribution of the attributes, which suggest that the distribution is not Poisson. Also, we will briefly discuss the emergence of the giant component in the graph.

Joint work with S. Bamidhi, H.C. Gromoll

#14: Program 0n Analysis of Object Data, AOOD Transition Workshop June 9 – June 11, 2011

SCHEDULE
Thursday, June 9
SAMSI (Room 150)

8:00-8:45 Registration and Continental Breakfast

8:45-9:00 Welcome

Functional Data: Organizer: **Hans-Georg Müller**, University of California, Davis
Chair: **Jane-Ling Wang**, University of California, Davis

9:00-9:30 **Jeff Morris**, MD Anderson
Hierarchical Modeling of Object Data

9:30-9:50 **Hongxiao Zhu**, SAMSI
Robust Classification of Functional and Quantitative Image Data using Functional Mixed Models

9:50-10:10 **Sylvie Tchumtchou**, SAMSI
Online Variational Bayesian Inference in Hierarchical Models for Correlated High-dimensional Data

10:10-10:40 Discussion and Connections to FDA Working Groups: **Jane-Ling Wang**, University of California, Davis

10:40-11:10 Break

11:10-11:40 **Yichao Wu**, North Carolina State University
Continuously Additive Modeling for Functional Predictors

11:40-12:00 **David Degras,** SAMSI
Longitudinal Survey Methods for Functional Data

12:00-12:30 Discussion and Future Developments: **Steve Marron,** UNC; **Jeff Morris,** MD Anderson; **Jim Ramsay,** McGill University; **Jane-Ling Wang**, University of California, Davis

12:30-2:00 Lunch

Dynamics: Organizer/Chair: **Jim Ramsay**, McGill University

2:00-2:30 **Jiguo Cao**, Simon Fraser University
Quantitative Trait Loci Mapping with Differential Equation Models

2:30-3:00 **David Campbell**, Simon Fraser University
Parameter Estimation from Locally Enforced Differential Equation Models

3:00-3:30 **Hulin Wu**, University of Rochester
High-Dimensional ODEs for Dynamic Gene Regulatory Networks

3:30-4:00 Break

4:00-4:30 **Jim Ramsay**, McGill University
Reflections on Impacts and Issues for Statistical Methodology for Dynamic Models Generated by the AOD Project

4:30-5:30 General discussion, animated by all four speakers

5:30–7:00 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, June 10
SAMSI (Room 150)

8:15-9:00 Registration and Continental Breakfast

Trees: Organizer/Chair: **J.S. Marron,** University of North Carolina

9:00-9:10 **J. S. Marron**, University of North Carolina
Driving Example Background and Research Overview

9:10-9:40 **Sean Skwerer**, University of North Carolina
Phylogenetic Trees

9:40-10:10 **Dan Shen**, University of North Carolina
Dyck Path and Branch Length Analyses

10:10-10:50 Break

10:50-11:10 **Lingsong Zhang**, Purdue University
Non-Negative Matrix Factorization Approach to Tree Analysis

11:10-11:30	**Yongdai Kim**, Seoul National University (presented by J. S. Marron) *Thread Bridging Example: Psuedo-Bayesian Factor Analysis*
11:30-12:00	**John Aston**, Warwick University *Thread Bridging Example: Phylogenetic Trees with Dialects as Leaves*
12:00-12:30	Discussion and Future Directions: **J. S. Marron**, University of North Carolina; **John Aston**, University of Warwick; **Lingsong Zhang**, Purdue University; **Hernando Ombao**, Brown University
12:30-2:00	Lunch
	Manifolds: Organizer/Chair: **Ian Dryden**, University of South Carolina
2.00-2.30	**Victor Patrangenaru**, Florida State University *Statistical Analysis of Object Data.*
2.30-2.50	**Sungkyu Jung**, University of North Carolina *Principal Nested Shape Spaces and an Application to Reduction of Number of Landmarks*
2.50-3.10	**Sebastian Kurtek**, Florida State University *Registration of Functional Data Using Fisher-Rao Metric*
3.10-3.40	Break
3.40-4.00	**Jingyong Su**, Florida State University *Fitting Optimal Curves to Time-Indexed, Noisy Observations of Stochastic Processes on Nonlinear Manifolds*
4.00-4.30	**Ian Dryden**, University of South Carolina *Metrics, Manifolds and Geometric Correspondence*
4.30-5.00	Panel Discussion: **Ian Dryden,** University of South Carolina; **J.S. Marron,** University of North Carolina; **Victor Patrangenaru**, Florida State University

Saturday, June 11
SAMSI (Room 150)

8:15-9:00 Registration and Continental Breakfast

Brain Imaging: Organizer: **Jane-Ling Wang**, University of California, Davis
Chair: **John Aston,** Warwick University

9:00-9:30 **Haipeng Shen**, University of North Carolina
Hemodynamic Response Function Modeling

9:30-9:50 **Tingting Zhang**, University of Virginia
Nonparametric Inference of Hemodynamic Response for fMRI data with Inhomogeneous Variances through Kernel Smoothing

9:50-10:10 **Ci-Ren Jiang**, SAMSI
Nonparametric Response Function Estimation Via FPCA with an Application to Dynamic PET Data

10:10-10:40 Discussion and Connections to Working Groups:
John Aston, University of Warwick

10:40-11:10 Break

Chair: **Hernando Ombao**, Brown University

11:10-11:40 **John Aston**, University of Warwick
Spatial Functiona; Data, Temporal Sequences and Populations of Change Points for fMRI Analysis

11:40-12:00 **Seonjoo Lee**, SAMSI
Independent Component Analysis for Autocorrelated Sources with an Application to Fmri

12:00-12:30 Discussion and Future Developments: **John Aston**, University of Warwick; **Ian Dryden**, University of South Carolina; **Jeff Morris**, MD Anderson; **Hernando Ombao**, Brown University; **Haipeng Shen**, University of North Carolina; **Jane-Ling Wang**, University of California, Davis

SPEAKER TITLES/ABSTRACTS

John Aston
University of Warwick
J.A.D.Aston@warwick.ac.uk

"Spatial Functional; Data, Temporal Sequences and Populations of Change Points for fMRI Analysis"

Change point detection in sequences of functional data is examined where the spatial functional observations are temporally dependent and where the distributions of change points from multiple subjects is required. Of particular interest is the case where the change point is an epidemic change (a change occurs and then the observations return to baseline at a later time). The special case where the covariance can be decomposed as a tensor product is considered with particular attention to the power analysis for detection. This is of interest in the application to functional magnetic resonance imaging (fMRI), where the estimation of a full covariance structure for the three-dimensional image is not computationally feasible. It is found that use of basis projections such as principal components for detection of the change points can be optimal in situations where PCA is traditionally thought to perform badly. [Joint work with Claudia Kirch, Karlsruhe Institute of Technology]

John Aston
University of Warwick
J.A.D.Aston@warwick.ac.uk

"Phylogenetic Trees with Dialects as Leaves"

In this talk, preliminary work on the analysis of phonetic data to build linguistic phylogenies will be presented. It will be shown that dialects can be represented as functional data and variation between speakers accounted for in a variety of ways to allow meaningful comparisons across languages. These will then form the first stage of an analysis to infer a linguistic phylogeny from the dialectal data, allowing the possibility of connecting not only ancestral written languages but also speech.

David Campbell
Simon Fraser University
dac5@sfu.ca

"Parameter Estimation from Locally Enforced Differential Equation Models"

In this talk we present a generalization of local polynomial regression for modelling complex behaviours semi-parameterically via locally enforced models. In addition we present a method for using this locally enforced model to obtain globally used parameter estimates describing the differential equation behaviour exhibited in the data.

While the data fitting is non-parameteric in style, effectively a kernel weight function and a differential equation defined by model parameters (theta) is used to build the basis used to smooth the data. Conditional on a set of parameters theta, the smoothing uses a nonlinear basis expansion parametrized by the initial states of the ODE. We then use the data fitting criterion to estimate the ODE model parameters associated with the basis.

Differential equation models describe the changes in system states occurring in a small time interval. We apply this logic in model fitting as we require only that the model fit the data in short time intervals by locally enforcing the model. However we estimate the parameters of the differential equation by examining the global data fit, thereby pairing locally enforced dynamics with an overall relaxation from a the requirement of having a single differential equation solution.

This talk outlines a collaboration with Nicolas Brunel, Jiguo Cao and David Degras that originated while we were all in residence at SAMSI during the AOOD program.

Jiguo Cao
Simon Fraser University
jca76@sfu.ca

"Quantitative Trait Loci Mapping with Differential Equation Models"

Genetic mapping, attributing a phenotypic trait to its underlying genes, known as quantitative trait loci (QTLs), has been proven powerful for constructing the genotype-phenotype relationship. However, the traditional methods often neglect the biological principles underlying the dynamic interactions among different components in a complex biological system. In order to take into account these biological principles, we develop a conceptual model, called systems mapping. A group of ordinary differential equations (ODE) are proposed to quantity how alterations of different components lead to the global change of the biological system under the regulations of specific QTLs. The ODE parameters are estimated from the noisy time-course trait data in a framework of functional mixture models. Through testing genotype-specific differences in ODE parameters, system mapping can identify the genetic effects of QTLs on component-component interactions. System mapping should enable geneticists to shed light on the genetic complexity of the biological system and predict its physiological and pathological states.

David Degras
SAMSI
ddegras@samsi.info

"Longitudinal Survey Methods for Functional Data"

When a phenomenon is measured by machines in continuous time (for example sensor networks, satellites, or medical scanners), there is great exibility in the design of sampling methods to collect and analyze data. Longitudinal survey methods can be applied to build time-varying samples that efficiently exploit transversal and longitudinal information from a population. Due to their efficient sample use, time- varying sampling schemes may prove advantageous over usual time-invariant samples. For a given sample size, they can produce more accurate statistical results whereas, when the collection and analysis of data incurs significant costs, they can achieve equal performances with smaller sample size. In this talk we present two novel time-varying stratified sampling schemes in the context of survey estimation of a mean function based on functional data. At given times, these sampling schemes allow to re-allocate the sample across strata and partly or fully renew the sample within strata. Considering a classical Horvitz-Thompson estimator, we show that the proposed time-varying samples yield equal or smaller expected estimation error than time-invariant samples. After introducing the notion of renewal rate of the sample, we give an asymptotic expression of the variance of the estimation error which depends on the renewal rate, the number of resampling times, and the sampling rates in the strata. This result indicates that increasing the renewal rate and/or the frequency of resampling massively reduces the variability of the estimation error. The effectiveness of the time-varying sampling schemes under study will also be illustrated with real and simulated data sets.

Ian Dryden
University of South Carolina
dryden@stat.sc.edu

"Metrics, Manifolds and Geometric Correspondence"

Three of the SAMSI working groups have discussed a wide range of topics on manifolds and their connections with other themes. The earlier talks in this session demonstrate the wide range of collaborative work from participants. Major themes have been registration and analysis of curves and functions; stratified manifolds; projective shape analysis; splines and curve fitting on manfolds; metric choice; and analysis of symmetric positive definite matrices. This presentation will focus on the specific activites of the Metrics on Shape Spaces working group, outline further collaborative projects and signal future developments.

Ci-Ren Jiang
SAMSI
cirenjiang@gmail.com

"Nonparametric Response Function Estimation Via FPCA with an Application to Dynamic PET Data"

In dynamic PET data analysis, injected radioactive tracer concentrations are measured over time to help understand functional processes in the body. Traditionally, parametric forms are assumed for the implied impulse response functions while estimating the concentration; however, these parametric assumptions are very difficult to verify and may well not hold. Therefore, we propose a nonparametric approach to estimate the response functions and thus the concentration. First, we employ FPCA with a multiplicative structure to represent the signal function for each voxel using a Karhunen-Loeve decomposition. As convolution can be viewed as a linear operator, we secondly apply deconvolution to the mean and eigenfunctions of the voxel signals. Then, the response function for each voxel can be represented as a linear combination of deconvolved mean function and deconvolved eigenfunctions where the linear coefficients are identical to the multiplicative coefficients and principal component scores in the first step. Therefore, the integral of the concentration in a finite time interval (a quantity of particular interest) can be obtained easily. This approach is demonstrated with simulation studies and real data analysis.

Sungkyu Jung

University of North Carolina
sungkyu@email.unc.edu

"Principal Nested Shape Spaces and an Application to Reduction of Number of Landmarks"

This paper discusses a successive dimension reduction of Kendall's shape space. Instead of taking the popular principal component analysis approach, we reduce the number of landmarks by removing a landmark and projecting the rest of landmarks onto a proper shape space with fewer landmarks. We also introduce a scree plot based on the total variances and mean squared deviance that can be used as a rule-of-thumb for effective number of landmarks. Several real and simulated planar shape data are examined.

Yongdai Kim (Presented by J.S. Marron)
Seoul National University
ydkim0903@gmail.com

"Thread Bridging Example: Psuedo-Bayesian Factor Analysis"

An overview of a new approach to the problem of "flat spots" in the Dyck Path approach to trees as data. This has potential to connect to the Bayesian Working Group through a Bayesian Factor Analysis approach.

Sebastian Kurtek
Florida State University
skurtek@stat.fsu.edu

"Registration of Functional Data Using Fisher-Rao Metric"

We introduce a novel geometric framework for separating the phase and the amplitude variability in functional data of the type frequently studied in growth curve analysis. This framework uses the Fisher-Rao Riemannian metric to derive a proper distance on the quotient space of functions modulo the time-warping group. A convenient square-root velocity function (SRVF) representation transforms the Fisher-Rao metric into the standard L^2 metric, simplifying the computations. This distance is then used to define a Karcher mean template and warp the individual functions to align them with the Karcher mean template. The strength of this framework is demonstrated by deriving a consistent estimator of a signal observed under random warping and scaling. These ideas are demonstrated using both simulated and real data from different application domains: the Berkeley growth study and handwritten signature curves. The proposed method is empirically shown to be superior in performance to several recently published methods for functional alignment.

Authors: A. Srivastava[1], W. Wu[1], S. Kurtek[1] (presenting author), E. Klassen[2], J.S. Marron[3]
[1] Department of Statistics, Florida State University
[2] Department of Mathematics, Florida State University
[3] Department of Statistics, University of North Carolina

Seonjoo Lee

University of North Carolina
slee@samsi.info

“Independent Component Analysis for Autocorrelated Sources with an Application to Fmri”

Independent component analysis (ICA) is an effective data-driven method for blind source separation. It has been successfully applied to separate source signals of interest from their mixtures. Most existing ICA procedures are carried out by relying solely on the estimation of the marginal density functions, either parametrically or nonparametrically. In many applications, correlation structures within each source also play an important role besides the marginal distributions. One important example is functional magnetic resonance imaging (fMRI) analysis where the brain-function-related signals are temporally correlated. In this talk, we consider a novel approach to ICA that fully exploits the correlation structures within the source signals. Specifically, we propose to estimate the spectral density functions of the source signals instead of their marginal density functions. This is made possible by virtue of the intrinsic relationship between the (unobserved) sources and the (observed) mixed signals. Our methodology is described and implemented using spectral density functions from frequently used time series models such as autoregressive moving average (ARMA) processes. The time series parameters and the mixing matrix are estimated via maximizing the Whittle likelihood function. We illustrate the performance of the proposed method through extensive simulation studies and a real fMRI application. The numerical results indicate that our approach outperforms several popular methods including the most widely used fast ICA algorithm.

J.S. Marron
University of North Carolina
marron@email.unc.edu

“Driving Example Background and Research Overview”

Introduction to the driving example of a population of brain artery trees. Also, a large scale overview of SAMSI research in the area of trees as data.

Jeffrey Morris
The University of Texas M.D. Anderson Cancer Center
jefmorris@mdanderson.org

“Hierarchical Modeling of Object Data”

The hierarchical modeling of object data working group has focused on developing hierarchical modeling approaches for object data, including functions, images, and more general structures like shapes. The goal has been to develop inferential methodology motivated by specific applications yielding complex, structured data. The idea of hierarchical modeling implies flexible, unified models that can simultaneously take into account variability and structure from multiple sources in the data set, within and between objects, and induced by the design or other measured covariates. Both Bayesian and frequentist approaches have been considered, and connections and distinctions among existing Bayesian and frequentist approaches in the literature have been explored. In this talk, I will summarize working group activities, briefly summarizing a number of ongoing and completed projects by working group members, and concluding with a summary of some of my own work in the area.

Victor Patrangenaru
Florida State University
vic@stat.fsu.edu

"Statistical Analysis of Object Data"

Analysis of object data is the more traditional name for Data Analysis on Sample Spaces with a Manifold Stratification. It includes Multivariate Analysis, Directional Data Analysis, Projective Shape Analysis as well as classical Shape Analysis, Diffusion Tensor Imaging, Functional Data Analysis, Analysis of Phylogenetic Trees Data; pretty much any non-categorical statistical problem can be formulated as object data analysis problem. Much of the standard nonparametric methodology extends from the multivariate case, in the generic case when the Frechet mean of a random object (r.o.) is at a regular point. In practice there are situations when a r.o. has a mean located on the singular part of the stratified sample space, and the manifold CLT based technique break down. Our goal is to understand the asymptotic behavior of the estimators of the Frechet mean of an arbitrary random object, and to develop nonparametric methodologies and fast inference techniques in applications.

James Ramsay
McGill University
ramsay@psych.mcgill.ca

"Reflections on Impacts and Issues for Statistical Methodology for Dynamic Models Generated by the AOD Project"

Some informal reactions to how the AOD Project may impact dynamic modeling in statistics, as well as other aspects of the Project, are offered. The weekly Dynamics and Inference workgroup conferences as well as the face-to-face exchanges during our times at SAMSI seemed to highlight several issues. I think we all agree that the need and opportunities for modeling data with dynamic systems is widespread and expanding rapidly; but relatively few statisticians are exposed to working with this class of models in either their training or career. We need to work hard at making this mathematical technology approachable and interesting to our colleagues. My attempt to teach this topic at the second graduate seminar taught me a great deal about how this might proceed, and certainly highlighted the need for textbook and workshop material. Much our exchange seemed to be around software development, and some emerging packages such as pomp and CollocInfer promise a great deal, but are evidently still in the earlier stages of development. The relative advantages of deterministic versus stochastic models was a fascinating topic, and the Project and our exchanges reshaped my thinking about this issue in many ways.

Dan Shen
University of North Carolina
dshen@email.unc.edu

"Dyck Path and Branch Length Analyses"

Statistical analysis, including PCA, of data on non-Euclidean spaces, such as tree space, can be challenging. An approach is to build a bridge between trees space (a non-Euclidean space) and curve space (standard Euclidean space). Then, we can exploit the power of functional data analysis to explore statistical properties of tree data sets. The Dyck path and branch length representation provide such a bridge. In addition, we also present the pruning tree idea to deal with missing branches in individual trees.

Haipeng Shen
University of North Carolina
haipeng@email.unc.edu

"Hemodynamic Response Function Modeling"

Hemodynamic response function (HRF) has played an important role in many recent functional magnetic resonance imaging (fMRI) based brain studies, where the main focus is to investigate the relationship between experimental stimuli and the neural activity. Standard statistical analysis of fMRI data usually calls for a "canonical" model of HRF, but it is uncertain how well this fits the actual data. Most existing methods have been developed in the time domain. We present a nonparametric frequency-domain method for estimating the HRF. Our method works for both event-related and block experiment designs with multiple stimuli. Its usefulness will be illustrated and compared with existing methods using both simulated and real human brain data. Some asymptotic properties of the method will also be discussed.

Sean Skwerer
University of North Carolina
password@email.unc.edu

"Phylogenetic Trees"

Phylogenetic trees as data objects is an exciting active area of research. The talk starts with an introduction to the space of phylogenetic trees as described by Billera, Holmes and Vogtman in "Geometry of the Space of Phylogenetic Trees". The metric introduced in Billera et al. gives the space of phylogenetic trees nonpositive curvature (NPC). NPC leads to unique shortest paths between points, and unique means for probability distributions. A law of large numbers for NPC metric spaces was established in "Probability Measures on Metric Spaces of NPC" by Sturm; however the central limit theorem has not been generalized to such spaces. SAMSI researchers started work on a central limit theorem for Tree Space, and have discovered that sample means exhibit *Stickiness*. Although it is known that Stickiness will play an important role, a central limit theorem for Tree Space is an active research topic. The talk then moves onto tools for analyzing populations of phylogenetic trees using "A Fast Algorithm for Computing Geodesic Distances in Tree Space" by Owen and Provan. Multidimensional scaling of the distance matrix can be used to create a two-dimensional visualization of a phylogenetic tree population. An iterated pairwise smoothing method can using only geodesic midpoints can be used to smooth time series of phylogenetic trees. Both techniques are used for exploratory analysis of brain artery trees.

Jingyong Su
Florida State University
jingyong@stat.fsu.edu

"Fitting Optimal Curves to Time-Indexed, Noisy Observations of Stochastic Processes on Nonlinear Manifolds"

We address the problem of estimating optimal curves for interpolation, smoothing, and prediction of values along partially-observed stochastic processes. In particular, we focus on processes that evolve on certain nonlinear manifolds of importance in computer vision applications. The observations are given as a set of time-indexed points on manifolds denoting noisy observations of the process at those times. Fitted curves on manifolds amount to geometrically meaningful and efficiently computable splines on manifolds. We adopt the framework developed in Samir et al. (2010) that develops a Palais metric-based steepest-decent algorithm applied to the weighted sum of a fitting-related and a regularity-related cost function. Using the rotation group, the space of positive-definite matrices, and Kendall's shape space as three representative manifolds, we develop the proposed algorithm for curve fitting. This algorithm requires expressions for exponential maps, inverse exponential maps, parallel transport of tangents, and curvature tensors on the chosen manifolds. These ideas are illustrated using a large number of experimental results on both simulated and real data.

Sylvie Tchumtchoua
SAMSI
sylvie.tchumtchoua@gmail.com

"Online Variational Bayesian Inference in Hierarchical Models for Correlated High-dimensional Data"

High-dimensional data with hundreds of thousands of observations are becoming commonplace in many disciplines. The analysis of such data poses many computational challenges, especially when the observations are correlated over time and/or across space. In this paper we propose flexible hierarchical regression models for analyzing such data that accommodate serial and/or spatial correlation. We address the computational challenges involved in fitting these models by adopting an approximate inference framework. We develop an online variational Bayes algorithm that is fast and produces good approximations to the exact posterior distributions. The performance of the method is assessed through simulation studies. We applied the methodology to analyze signal intensity in MRI images of subjects with knee osteoarthritis.

Hulin Wu

University of Rochester School of Medicine and Dentistry
hwu@bst.rochester.edu

“High-Dimensional ODEs for Dynamic Gene Regulatory Networks”

Gene regulation is a complicated process. The interaction of many genes and their products forms an intricate biological network. Identification of this dynamic network will help us understand the biological process in a systematic way. However, the construction of such a dynamic network is very challenging for a high-dimensional system. We propose to use a set of ordinary differential equations (ODE), coupled with dimensional reduction by clustering and mixed-effects modeling techniques, to model the dynamic gene regulatory network (GRN). The ODE models allow us to quantify both positive and negative gene regulations as well as feedback effects of one set of genes in a functional module on the dynamic expression changes of the genes in another functional module, which results in a directed graph network. A six-step procedure, Screening, Clustering, Smoothing, regulation Identification, parameter Estimates refining and Function enrichment analysis (SCSIEF) is developed to identify the ODE-based dynamic GRN. In the proposed CSIEF procedure, a series of cutting-edge statistical methods and techniques are employed. We apply the proposed method to identify the dynamic GRN for yeast cell cycle progression data and immune response to influenza infection. We are able to annotate the identified modules through function enrichment analyses. Some interesting biological findings are discussed. The proposed procedure is a promising tool for constructing a general dynamic GRN and more complicated dynamic networks.

Yichao Wu
North Carolina State University
wu@stat.ncsu.edu

“Continuously Additive Modeling for Functional Predictors”

We propose continuously additive models (CAM), which are an extension of additive regression models to the case of infinite-dimensional predictors, corresponding to random functions, coupled with scalar responses. In the limit, as the number of predictor times grows larger, properly scaled additive models converge to a model where the additivity is conveyed by through an integral, leading to a new functional regression model. We explore this convergence and study estimation of the limiting functional regression model based on basis expansions. The resulting CAM estimators are shown to be consistent and to outperform existing functional regression approaches in simulations and data applications. Based on joint work with Hans-Georg Müller and Fang Yao.

Lingsong Zhang
Purdue University
lingsong@purdue.edu

“Non-Negative Matrix Factorization Approach to Tree Analysis”

We apply nonnegative matrix factorization method to a population of trees. The node length presentation for the set of binary trees provides a nice framework for analysis of trees. At this setting, different statistical methods in Euclidean space can be applied. However, methods such as principal component analysis may generate non-interpretable projections. The nonnegative matrix factorization method provides a useful set of projections, and they have easy interpretations. This provides an insightful exploration tool for analysis of trees.

This is joint work with J. S. Marron and Dan Shen.

Tingting Zhang
University of Virginia
tz3b@virginia.edu

“Nonparametric Inference of Hemodynamic Response for fMRI Data with Inhomogeneous Variances through Kernel Smoothing”

There is vast literature in functional magnetic resonance imaging(fMRI) data analysis on the estimation of the Hemodynamic Response Function (HRF), through both parametric and nonparametric approaches. However, existing methods are generally for analyzing population averaged fMRI data under the single-stimulus design, which has much smaller variance than individual fMRI data. Motivated from real examples, we propose a nonparametric kernel smoothing method for estimating HRFs under multi-stimulus designs with heterogeneous variances across different subjects. To cope with the large variance inherent to individual fMRI data, we introduce a variance-reduction method though Tikhonov regularization. We show that bandwidth selection and the choice of regularization parameter can be conducted easily with computational time linear of the data size. The proposed method is compared with several existing ones and appear to enjoy both merits of fast computation and small errors in inferring brain activities under multiple stimuli.

Hongxiao Zhu
SAMSI
hzhu@samsi.info

"Robust Classification of Functional and Quantitative Image Data using Functional Mixed Models"
We introduce classification of complex, high dimensional functional data in the functional mixed model (FMM) framework. The FMM relates a functional response to a set of scalar predictors through functional fixed and random effects, and therefore is able to account for various factors that affect the functions and induce correlations. Classification is performed through training the data by treating the class as one of the fixed effects, and then predicting on the test data using posterior predictive probabilities. Through a Bayesian scheme, we are able to incorporate not only all factors that influence the functions, but also factors that directly affect class designation. While this classification method is general for all FMM methods, we provide details for two specific Bayesian approaches, the Gaussian, wavelet-based functional mixed model (G-WFMM) and the robust, wavelet-based functional mixed model (R-WFMM). Both methods perform modeling in the wavelet space, which yields parsimonious representations for the functions, and can naturally adapt to local features, and accommodate various nonstationarities. The R-WFMM has the additional advantage of allowing potentially heavier tails for features of the functions indexed by particular wavelet coefficients, leading to a down-weighting of outliers that makes the method robust to outlying functions or regions of functions. The models are applied to a real mass spectroscopy dataset in pancreatic cancer research. Our results show improved classification when comparing

FMM with other typical functional data classification methods and the ad hoc methods that are based on detected spectral peaks.

#15: Program 0n Uncertainty Quantification, SAMSI/Sandia Summer School of UQ, June 20 – June 24, 2011

SCHEDULE
Monday, June 20
Sheraton Albuquerque Uptown

8:00-8:45	Registration and Continental Breakfast
8:45-9:00	Introduction
9:00-12:00	"*Sensitivity Analysis and Polynomial Chaos for Differential Equations"* **Dongbin Xiu** (Purdue), Part 1
12:00-1:30	Lunch
1:30-4:15	"*Sensitivity Analysis and Polynomial Chaos for Differential Equations"* **Dongbin Xiu** (Purdue), Part 2
4:15-4:45	Break
4:45-5:45	"Nonintrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Analysis and Design" **Mike Eldred** (Sandia)
6:00-8:00	Reception

Tuesday, June 21
Sheraton Albuquerque Uptown

8:30-9:00	**Registration and Continental Breakfast**
9:00-12:00	***"Data Assimilation and Applications in Climate Modeling"*** Doug Nychka **(NCAR), Part 1**
12:00-1:30	**Lunch**
1:30-4:15	***"Data Assimilation and Applications in Climate Modeling"*** Doug Nychka **(NCAR), Part 2**
4:15-4:45	**Break**
4:45-5:45	"The Impact of Parameter Uncertainty on the Community Atmospheric

	Model (CAM)"
	Gardar Johannesson **(LLNL)**

Wednesday, June 22
Sheraton Albuquerque Uptown

8:30-9:00	**Registration and Continental Breakfast**
9:00-12:00	***"Statistical Analysis of Rare Events"*** Dan Cooley **(Colorado State), Part 1**
12:00-1:30	**Lunch**
1:30-4:15	***"Statistical Analysis of Rare Events"*** Dan Cooley **(Colorado State), Part 2**
4:15-4:45	**Break**
4:45-5:45	"Methods to Address Mixed Epistemic-Aleatory Uncertainty Quantification" Laura Swiler **(Sandia)**

Thursday, June 23
Sheraton Albuquerque Uptown

8:30-9:00	**Registration and Continental Breakfast**
9:00-12:00	***"Data Assimilation"*** Adrian Sandu **(Virginia Tech), Part 1**
12:00-1:30	**Lunch**
1:30-4:15	***"Data Assimilation"*** Adrian Sandu **(Virginia Tech), Part 2**
4:15-4:45	**Break**
4:45-5:45	*"Gradient-Enhanced Uncertainty Propagation"* Mihai Anitescu **(ANL)**

Friday, June 24
Sheraton Albuquerque Uptown

8:00-8:30	**Registration and Continental Breakfast**

8:30-9:30 ***"Sampling-Based Methods for Uncertainty and Sensitivity Analysis"***
Jon Helton **(Sandia)**

9:30-9:45 **Break**

9:45-10:45 *"Practical Uncertainty Quantification Methods and Methodologies for Multi-physics Applications"*
Charles Tong **(LLNL)**

10:45-11:00 **Break**

11:00- 12:00 *"Inference Using Computer Models: a Survey of Applications at Los Alamos National Laboratory"*
Dave Higdon **(LANL)**

12:00-1:00 **Lunch**

1:00 **Adjourn**

SPEAKER TITLES/ABSTRACTS

Mihai Anitescu
ANL
anitescu@mcs.anl.gov

"Gradient-Enhanced Uncertainty Propagation"

In this work we discuss an approach for uncertainty propagation through computationally expensive physics simulation codes. Our approach incorporates gradient information information to provide a higher quality surrogate with fewer simulation results compared with derivative-free approaches.

We use this information in two ways: we fit a polynomial or Gaussian process model ("surrogate") of the system response. In a third approach we hybridize the techniques where a Gaussian process with polynomial mean is fit resulting in an improvement of both techniques. The surrogate coupled with input uncertainty information provides a complete uncertainty approach when the physics simulation code can be run at only a small number of times. We discuss various algorithmic choices such as polynomial basis and covariance kernel. We demonstrate our findings on synthetic functions as well as nuclear reactor models.

Dan Cooley
Colorado State University
cooleyd@stat.colostate.edu

"Statistical Analysis of Rare Events"

This short course will introduce the current statistical practice for analyzing extreme events. Statistical practice relies on fitting distributions suggested by asymptotic theory to a subset of data considered to be extreme. Both block maximum and threshold exceedance approaches will be presented for both the univariate and multivariate cases.

Mike Eldred
Sandia
mseldre@sandia.gov

"Nonintrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Analysis and Design"

Uncertainty quantification (UQ) is a key enabling technology for assessing the predictive accuracy of computational models and for enabling risk-informed decision making. This presentation will highlight how R&D in stochastic expansion methods, including nonintrusive polynomial chaos expansion (PCE) and stochastic collocation (SC), is addressing critical robustness and efficiency issues that occur in current production UQ approaches. I will describe current research efforts in PCE and SC, including anisotropy detection, uniform and adaptive h-/p-refinement, gradient-enhancement, and extensions to nonprobabilistic domains (i.e., design, epistemic). Finally, these capabilities will be employed as a foundation for enabling higher level analyses, such as design under uncertainty and mixed aleatory-epistemic UQ.

Jon Helton
Arizona State University
jchelto@sandia.gov

"Sampling-Based Methods for Uncertainty and Sensitivity Analysis"

Sampling-based methods for uncertainty and sensitivity analysis are reviewed. The following topics are considered: (i) Definition of probability distributions to characterize epistemic uncertainty in analysis inputs, (ii) Generation of samples from uncertain analysis inputs, (iii) Propagation of sampled inputs through an analysis, (iv) Presentation of uncertainty analysis results, and (v) Determination of sensitivity analysis results.

Dave Higdon
Los Alamos National Laboratory
dhigdon@lanl.gov

"Inference Using Computer Models: a Survey of Applications at Los Alamos National Laboratory"

Large-scale computational models are used in a wide variety of applications at Los Alamos National Laboratory, nearly all requiring some form of inference - predictions with uncertainty, estimation of parameters, statements about safety, etc. This talk will survey a selection of applications and describe Bayesian statistical methods used for analysis. These methods include designing simulation campaigns, modeling simulation output, estimation - or calibration - of key simulation model parameters, and accounting for major sources of uncertainty. Various response surface models will be discussed, as will model formulations for combining the various sources of information.

Co-authors: Brian Williams, Jim Gattiker, Charlie Nakhleh, Salman Habib, Katrin Heitmann

Gardar Johannesson
Lawrence Livermore National Laboratory
johannesson1@llnl.gov

"The Impact of Parameter Uncertainty on the Community Atmospheric Model (CAM)"

Projections from global climate models are uncertain due to numerous factors. Here we report on ongoing research effort at the Lawrence Livermore National Laboratory to advance the science of uncertainty quantification (UQ) and the application of UQ to a climate models. In the first phase of this effort, we have created multiple ensembles of 12-year long simulations of past climate using the Community Atmospheric Model (CAM) by "perturbing" a large collection of important and uncertain input parameters and by varying initial conditions. We present the use of multiple exploratory and statistical methods to analyze the impact of parameter uncertainty on various output quantities of interest and conclude with which parameters are the main uncertainty drivers for a given output of interest. We contrast and compare the CAM ensembles to selected set of observations, using standard comparison metrics that have been developed and used within the climate community, and discuss the challenge of using multiple sets of spatial-temporal observations to carry out statistical calibration of climate models.

Doug Nychka
NCAR
nychka@ucar.edu

"Data Assimilation and Applications in Climate Modeling"

Climate prediction and modeling do not incorporate geophysical data in the sequential manner as weather forecasting and comparison to data is typically based on accumulated statistics, such as averages. This arises because a climate model matches the state of the Earth's atmosphere and ocean "on the average" and so one would not expect the detailed weather fluctuations to be similar between a model and the real system. An emerging area for climate model validation and improvement is the use of data assimilation to scrutinize the physical processes in a model using observations on shorter time scales. The idea is to find a match between the state of the climate model and observed data that is particular to the observed weather. In this way one can check whether short time physical processes such as cloud formation or dynamics of the atmosphere are consistent with what is observed.

Adrian Sandu
Virigina Tech
sandu@cs.vt.edu

"Variational Data Assimilation"

This lecture will discuss the fundamentals of three- and four-dimensional variational data assimilation techniques.

Topics include: - three dimensional variational (3D-var) data assimilation: formulation of the problem, construction of covariance matrices, observation operators, numerical optimization, and analysis of error impact; - four dimensional variational (4D-var) data assimilation: formulation of the problem; - adjoint sensitivity analysis for systems governed by ODEs and PDEs; - discrete versus continuous adjoint models: properties and implementation; automatic differentiation; - adjoint operators and uncertainty quantification; - computational issues and efficient implementation aspects; - applications and examples.

Hands-on examples will include the construction of a 4D-var system for a small test problem, in an idealized setting.

Laura Swiler
Sandia
lpswile@sandia.gov

"Methods to Address Mixed Epistemic-Aleatory Uncertainty Quantification"

Engineering applications and regulatory requirements often specify the explicit separation and quantification of the contributions of epistemic and aleatory uncertainty. Epistemic (lack of knowledge) uncertainty may be represented by several approaches such as intervals, Dempster-Shafer belief structures, or probability distributions. Aleatory (inherent variability, randomness) is usually represented with probability distributions. This talk will present three main classes of algorithms: interval-valued probability, Dempster-Shafer theory of evidence, and second-order probability. Each of these approaches can be implemented using sampling methods, but more efficient methods are available. This talk will discuss the role of optimization and stochastic expansions in making mixed UQ algorithms more accurate and more affordable.

Charles Tong
Lawrence Livermore National Laboratory
tong10@llnl.gov

"Practical Uncertainty Quantification Methods and Methodologies for Multi-physics Applications"

In this talk I will discuss some of our experiences with developing and using uncertainty quantification (UQ) methods and methodologies for large scale multi-physics simulation models. Details of the presentation will include general UQ methodologies, a survey of methods for dealing with high dimensional uncertain parameter space, methods for constructing surrogate models, methods for uncertainty and sensitivity propagation, and finally methods for multistaged data integration. Small scale examples will be given to illustrate some of the techniques.

Dongbin Xiu
Purdue University
dxiu@purdue.edu

“Sensitivity Analysis and Polynomial Chaos for Differential Equations”
This lecture will focus on numerical algorithms for stochastic simulations, with an emphasis on the methods based on generalized polynomial chaos methodology. Both the mathematical framework and the technical details will be examined, along with performance comparisons and implementation issues for practical complex systems.

The main lectures will be supplemented by discussion sessions and by presentations from UQ practitioners from both the Sandia and Los Alamos National Laboratories.

Appendix C – 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 during the year and broken into two categories: Program Workshops and Education and Outreach (E&O) Workshops.

The evaluation form is as follows:

SAMSI Evaluation

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 http://legacy.samsi.info/PartInfo/200910/participantinformationform09101.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
e. Local Transportation	1	2	3	4	5

2a. What were the positive aspects of the organization and running of this workshop?

2b. What parts of the organization and running need improvement?

3. Please comment on the Scientific Quality:

4. **Additional comments on any other aspects of the workshop**

5. **An important goal of SAMSI is to create synergies between disciplines. How well did this workshop further this goal?**

6. **How did you learn of this workshop?**

7. **Please suggest ideas / contacts for future SAMSI activities**

F.2 Evaluation of Scientific Content

Greater than 90% of the respondents rated the scientific content as Very Good or Excellent for the Program Workshops. The outcomes for the E&O Workshops were similar.

F.2.1 Program Workshops (12 events)

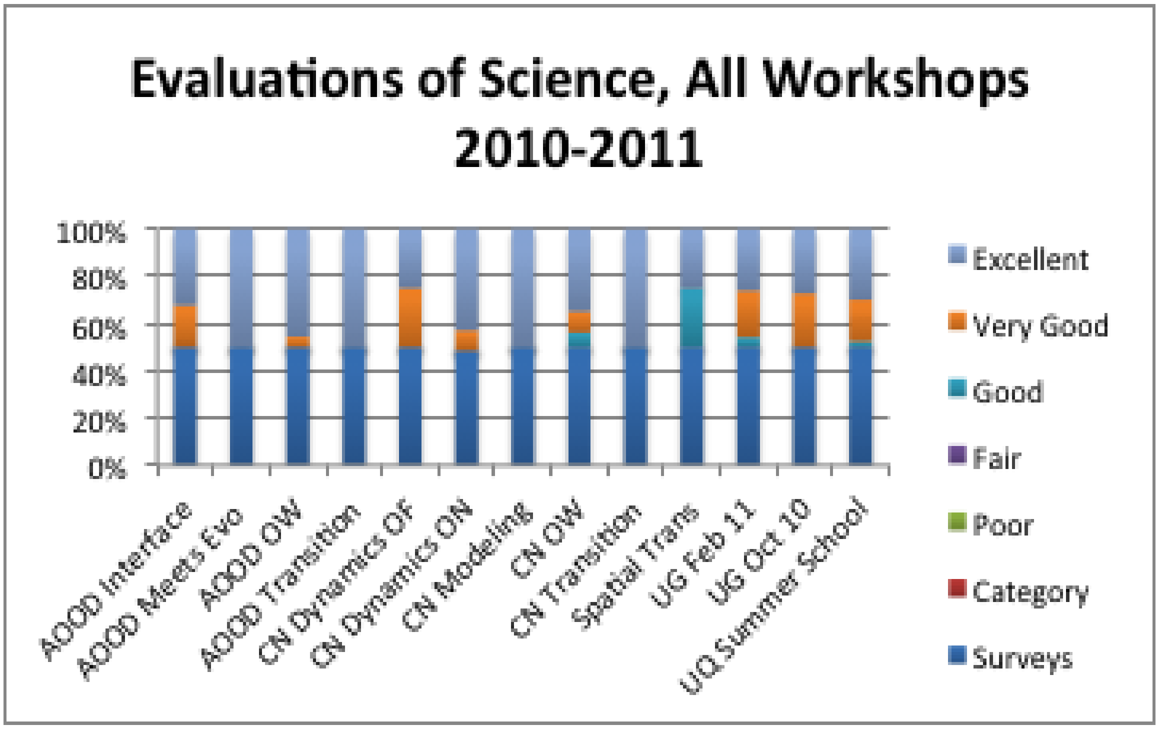
AOD
Opening Workshop
Interface Functional and Longitudinal Data Analysis
Evolutionary Biology
Transition Workshop

CN
Opening Workshop
Modeling
Dynamics Of Networks
Dynamics On Networks
Transition Workshop

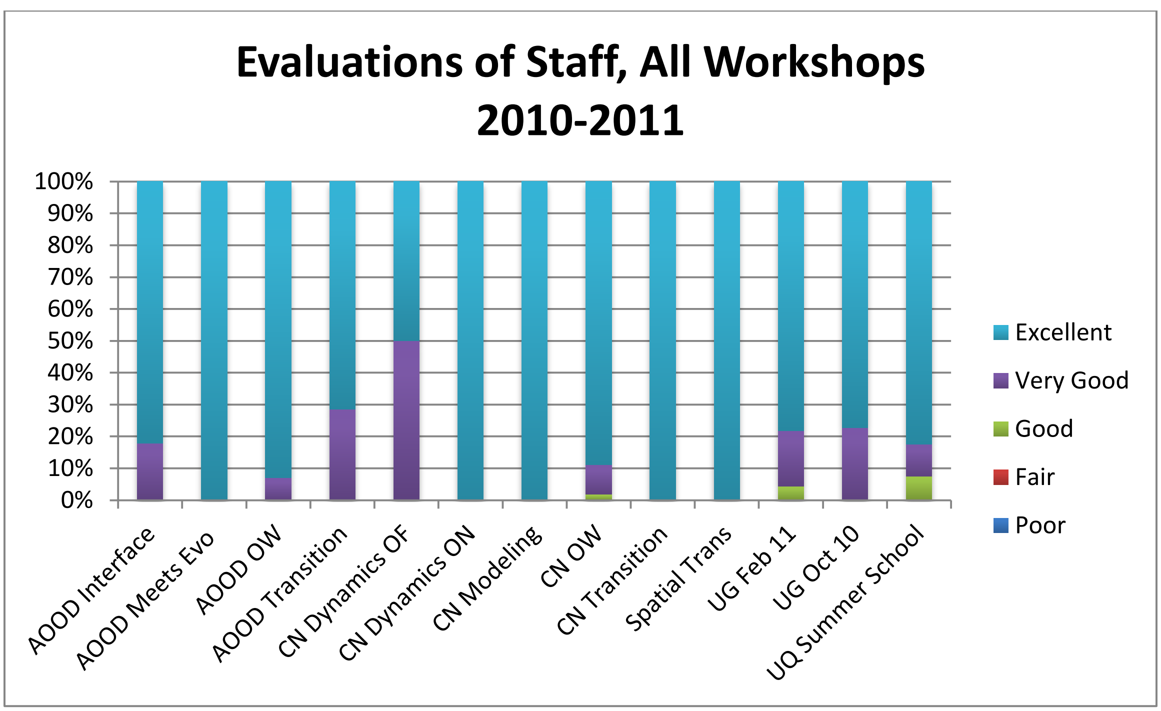
UQ: Summer School 2011

F.2.2 Education and Outreach Workshops

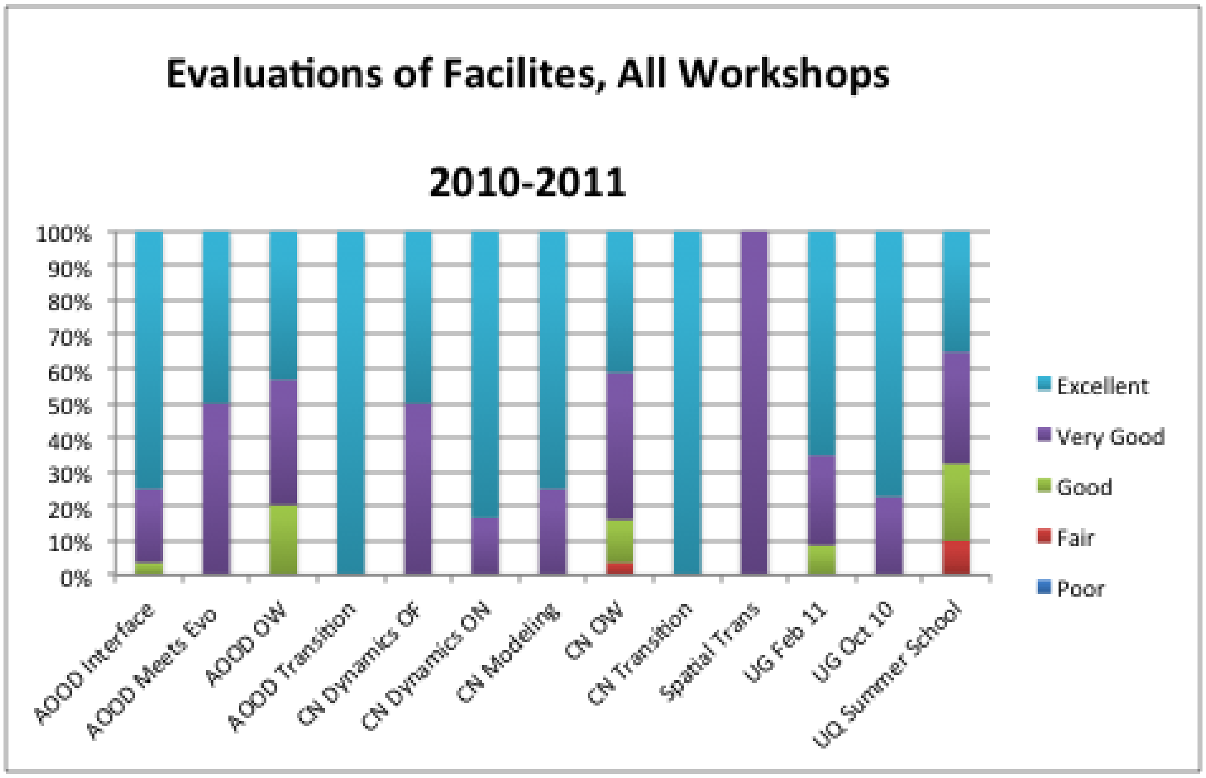
UG: Two Day Undergraduate Workshop October 2010
UG: Two Day Undergraduate Workshop February 2011

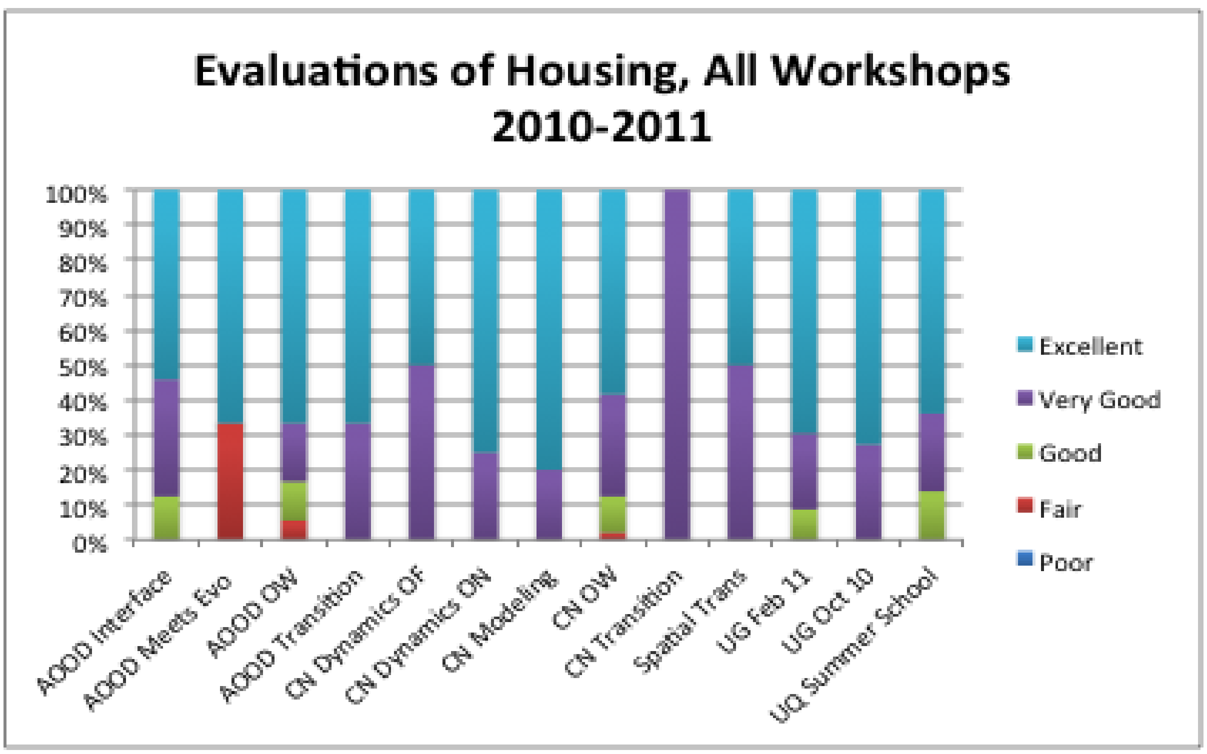


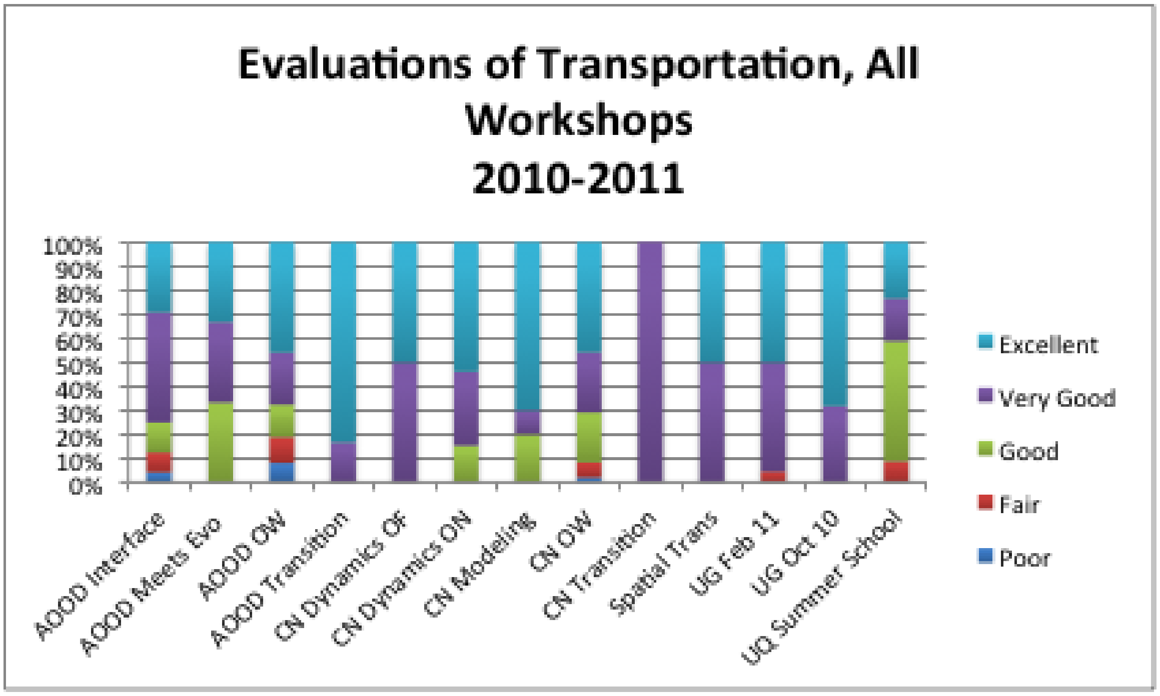
Evaluations of Science, All Workshops 2010-2011
100%
80%
60%
40%
20%
0%
AOOD Interface
AOOD Meets Evo
AOOD OW
AOOD Transition
CN Dynamics OF
CN Dynamics ON
CN Modeling
CN OW
CN Transition
Spatial Trans
UG Feb 11
UG Oct 10
UQ Summer School
Excellent
Very Good
Good
Fair
Poor
Category
Surveys



Evaluations of Staff, All Workshops 2010-2011
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
AOOD Interface
AOOD Meets Evo
AOOD OW
AOOD Transition
CN Dynamics OF
CN Dynamics ON
CN Modeling
CN OW
CN Transition
Spatial Trans
UG Feb 11
UG Oct 10
UQ Summer School
Excellent
Very Good
Good
Fair
Poor


Evaluations of Facilites, All Workshops
2010-2011
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
AOOD Interface
AOOD Meets Evo
AOOD OW
AOOD Transition
CN Dynamics OF
CN Dynamics ON
CN Modeling
CN OW
CN Transition
Spatial Trans
UG Feb 11
UG Oct 10
UQ Summer School
Excellent
Very Good
Good
Fair
Poor


Evaluations of Housing, All Workshops
2010-2011
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
AOOD Interface
AOOD Meets Evo
AOOD OW
AOOD Transition
CN Dynamics OF
CN Dynamics ON
CN Modeling
CN OW
CN Transition
Spatial Trans
UG Feb 11
UG Oct 10
UQ Summer School
Excellent
Very Good
Good
Fair
Poor


Evaluations of Transportation, All Workshops
2010-2011
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
AOOD Interface
AOOD Meets Evo
AOOD OW
AOOD Transition
CN Dynamics OF
CN Dynamics ON
CN Modeling
CN OW
CN Transition
Spatial Trans
UG Feb 11
UG Oct 10
UQ Summer School
Excellent
Very Good
Good
Fair
Poor