samsi

NSF ● Duke ● NCSU ● UNC ● NISS

Annual Scientific Report

2005-2006

May 1, 2006

SAMSI Annual Scientific Report for 2005-2006

This report is a version of the SAMSI Annual Report to the National Science Foundation, with sensitive financial data and personal information removed. It covers the period of SAMSI activities from July 1, 2005 – June 30, 2006. Past and future activities of SAMSI are also discussed.

0. Executive Summary

Executive Summary contains

A. Outline of Activities and Initiatives

1. Fourth Year Programs and Activities

Regular Programs

- Financial Mathematics, Statistics and Econometrics
  - Planning Workshop (joint with CRM, 6/1/05-6/5/05)
  - Opening Tutorials and Workshop (9/18/05-9/21/05)
  - Credit Risk Workshop (10/31/05-11/2/05)
  - Model Uncertainty Workshop (1/27/06)
  - Transition Workshop and Symposium (2/27/06-2/28/06)
- National Defense and Homeland Security (Fall 2005-Spring 2006)
  - Tutorials and Opening Workshop (9/11/05-9/14/05)
  - Anomaly Detection Mid-Year Workshop (2/3/06 at CDC)
  - Social Networks Mid-Year Workshop (3/2/06 at Carnegie-Mellon U.)
  - Data Confidentiality Mid-Year Workshop (3/13/06 at NCHS)
- Astrostatistics (Spring 2006)
  - Planning Workshop (7/14/05-7/15/05 at NASA AMES)
  - Tutorials
    - Bayesian Methods in Astrostatistics (1/18/05-1/20/05)
    - Astronomy for Statisticians (1/21/05-1/22/05)
    - Nonparametric Methods for Astronomers (1/21/05-1/22/05)
  - Opening Workshop (1/23/06-1/25/06)
  - Intensive Research Session in Stellar Evolution (2/20/06-2/23/06)
  - Intensive Research Session in Particle Physics (3/6/06-3/16/06)
  - Transition Workshop (at Penn State, 6/12/06-6/15/06)

Education and Outreach

- Summer School on Data Assimilation: "Fusing Models with Data: From Theory to Practice to Theory" (6/13/05-6/17/05 at NCAR)
- Industrial Mathematical and Statistical Modeling Workshop for Graduate Students (7/25/05-8/2/05)
- Two-Day Workshop for Undergraduates (11/18/05-11/19/05)
- PREP Short Course for College Teachers (1/12/06-1/14/06)
- Two-Day Workshop for Undergraduates (3-3/06-3/4/06)
- Undergraduate Interdisciplinary Workshop (5/22/06-5/26/06)
- Industrial Mathematical and Statistical Modeling Workshop for Graduate Students (7/24/06-8/1/06)
- Graduate Courses at SAMSI
    - Advanced Topics in Financial Econometrics, Fall 2005
    - Special Topics in Financial Mathematics, Fall 2005
    - NDHS Seminar, Full Year
    - Topics in Astrostatistics, Spring 2006

Planning, Hot Topic, Technology Transfer, and Closing Workshops

- Random Graphs and Stochastic Computation (6/13/05-6/14/05)
- Data Mining Technology Transfer Workshop (6/20/05-6/24/05)
- Data Assimilation for Geophysical Systems Transition Workshop (10/5/05)
- Collaborations in the Mathematical Geosciences (10/6/05-10/7/05)
- Latent Variable Models in the Social Sciences Transition Workshop (11/10/05-11/11/05)
- Experimental Analysis of Algorithms: Interfaces between Statistics and Computer Science (3/4/06 at DIMACS)
- 12th Annual Conference for African-American Researchers in the Mathematical Sciences (6/20/06-6/23/06)

2. Fifth Year Program Schedule

- Summer Program on Multiplicity and Reproducibility in Scientific Studies (7/10/06-7/28/06)
    - Opening Workshop (7/10/06-7/12/06)
    - Transition Workshop (7/27/06-7/28/06)
- High Dimensional Inference and Random Matrices (Fall 2006)
    - Opening Workshop and Tutorials (9/17/06-9/20/06)
    - Mid-Program Workshops (TBA, joint with NCAR)
    - Transition Workshop (4/10/07-4/13/07 at AIM)
- Development, Assessment and Utilization of Complex Computer Models (Fall 2006, Spring 2007)
    - Summer School on the Design and Analysis of Computer Experiments (8/11/06-8/16/06 at Simon Fraser U.)
    - Opening Workshop (9/10/06-9/14/06)
    - Mid-Program Workshop in Engineering Models (TBA)

- Mid-Program Workshop in Granular Flow Models (TBA)
- Mid-Program Workshop in Ecological Models (TBA)
- Mid-Program Workshop in Environmental Models (TBA)
- Mid-Program Workshop in Systems Biology Models (TBA)
- Transition Workshop (5/14/07-5/17/07)

Education and Outreach

- Summer School on the Design and Analysis of Computer Experiments (8/11/06-8/16/06 at Simon Fraser U.)
- Two 2-Day Workshops for Undergraduates During the Academic Year (at SAMSI)
- One 2-Day Workshop for Undergraduates at a non-local location
- Interdisciplinary Workshop for Undergraduates (5/21/07-5/25/07)
- The Industrial Mathematical and Statistical Modeling Workshop for Graduate Students (late July, 2007)
- Graduate Courses at SAMSI
    - Computer Modeling, Fall 2006 and Spring 2007
    - High Dimensional Inference and Random Matrices, Fall 2006

Tentative Programs for 2007-2008

- Summer Program on Dynamic Treatment Regimes and Multi-Stage Decision Making
- Risk Analysis, Extreme Events and Decision Theory
- Random Media
- Environmental Sensor Networks

3. Developments and Initiatives

Fourth-Year Developments

- Technology-transfer workshops were tried as a mechanism for dissemination of research from SAMSI programs.
- Continuing efforts were made to optimize the operation of workshops; in particular, directorate liaisons served as workshop facilitators.
- Recognition of the centrality and success of working groups caused us to extend planned one-semester programs by several months.
- National accessibility to working groups was greatly enhanced through initiatives discussed in section C of the Executive Summary.
- Summer schools were instituted (last summer in Data Assimilation, and next summer in Computer Modeling).
- A NAC Scientific Liaison Board was created to provide input into the initiation and development of interdisciplinary SAMSI programs. See section E of the Executive Summary.
- The opportunities for external graduate students to participate in SAMSI research programs were greatly enhanced. See section E of the Executive Summary.
- Databases are being updated:

    - A financial database (not dependent on Partner universities) was installed.
    - Participant and scheduling databases (from IMA) are being installed.
    - Data-reporting to NSF will be standardized among institutes.
- The E&O committee was reconfigured to allow more extensive national input.
- Creation of a weekly announcement of upcoming activities 'SAMSI at a Glance'
- Changes in cooperation among DMS institutes.
    - Yearly directors meeting were instituted
    - Broadening participation, including diversity:
        - Best practices are being exchanged, including diversity databases.
        - At activities involving broadening, an institute will provide time for other institutes to report relevant opportunities.
        - Diversity events, such as CAARMS and the Blackwell-Tapia conference, are being coordinated, in part, through the creation of a new Diversity Committee.
    - A common institutes webpage is being created.
- Additional research collaborations with other institutes were expanded to enhance the overall impact of mathematics and statistics, including
    - Activities with the National Center for Atmospheric Research, relating to the Data Assimilation program, including joint postdoctoral appointments and a planned joint summer graduate educational program;
    - A joint workshop with the Centre de Recherches Mathématiques in financial mathematics;
    - A joint planning workshop with DIMACS and NISS on the Experimental Analysis of Algorithms
    - The transition workshop for the Astrostatistics program at the Center for Astrostatistics at Penn State
    - A variety of coordinated activities with the Canadian National Program on Complex Data Structures, including the Summer School on Computer Modeling at Simon Fraser University next August.

Planned Fifth-Year Developments

- Brainstorming sessions are being planned for areas of research in which there is not a clear path forward for creation of a SAMSI program. Initial areas for these sessions are:
    - Nanotechnology (at Purdue University in Fall, 2006)
    - Quantum Computation
    - Algebraic Techniques in Statistics and Systems Biology
- Enhanced external connections to working groups will be available, including formal teleconferencing facilities. This will allow remote nodes for working groups. As part of this effort to enhance remote connections, a technician will be hired to manage the equipment.
- Expansion of space, through an addition to the NISS building, will be planned.
- One of the two-day undergraduate workshops will be repeated at a non-local site, to investigate the feasibility of taking these "on the road."

C. Directorate's Summary of Challenges and Responses

SAMSI has been successful in achieving its' goals. The scientific programs have been of high caliber, and have led to significant new and ongoing research collaborations between, statistics, applied mathematics, and disciplinary sciences. There has been significant human resource development, through the postdoctoral and graduate programs and through involvement of senior researchers in new interdisciplinary areas. Many students across the country have been shown the SAMSI vision through educational outreach programs and courses. We feel that these successes are amply demonstrated throughout the report; some highlights are given in section D of the Executive Summary. This section discusses the challenges that arose in Year 4 and the Directorate's response to these challenges. Additional issues were raised during the recent Third-Year Review of SAMSI; these issues and our response to them are outlined in Section E of the Executive Summary.

Program Initiation: Most of the programs conducted during the first two years had been part of the initial SAMSI grant proposal, and hence had local individuals as leaders or co-leaders. During the Year 3, programs had a roughly 50-50 mix of local and non-local leaders. In Year 4, the programs were 75% driven by non-local leaders: the Financial Mathematics program had a mix of local and non-local scientists, but the National Defense Program and the Astrostatistics Program had solely non-local leadership. The programs in Year 5 are being entirely driven by outside leaders: High Dimensional Inference and Random Matrices is being led by Iain Johnstone, Helene Massam, and Craig Tracy; while Development, Assessment and Utilization of Complex Computer Models is being led by Susie Bayarri, Bruce Pitman, Peter Reichert and Thomas Santner, all non-local scientists (most of whom will be resident at SAMSI for the bulk of their program). Subsequent programs will also have solely non-local leadership. Of course, these programs will still have significant participation of local scientists; indeed, one of the major strengths of SAMSI is its' ability to draw to its programs stellar local talent in applied mathematics, statistics, and disciplinary sciences. To coordinate this local talent, each program will henceforth also have a local scientific coordinator.

Program Development: As mentioned above, the initial SASMI programs were directed by individuals who had been heavily involved in the creation of SAMSI, and so the process of developing the programs (recruiting participants and postdocs, forming working groups, planning workshops, etc.) was well-understood by the program leaders. The upcoming programs, however, are being led by individuals with no or minimal previous connection to SAMSI, and they do not know how SAMSI 'works.' Documents were prepared outlining the process of program and workshop development, but we came to realize that these are not sufficient guidance. We thus instituted other mechanisms to provide more hands-on guidance; including having directorate and NAC liaisons on each program committee and instituting regular meetings or conference calls between the program leaders, directorate, and support staff. Indeed, where appropriate, we designate two liaisons, one statistician and one applied mathematician.

As part of the leadership role for Institutes, SAMSI is undertaking the exploration of possible roles for Statistics and Applied Mathematics as new science research areas emerge. Often, the incorporation of these mathematical sciences as research foci comes well after the other sciences have well-developed agendas that may have involved the application of extant mathematical tools and computational methods. It is not necessarily the case that the tools and methodologies applied are inherently suited to the particular application; moreover these may be utilized in the absence of mathematical formulation of the essential issues. In consequence, SAMSI is inaugurating a series of small workshops, each with the purpose of examining a new area of interdisciplinary research to determine what, if any, critical role the statistical and mathematical sciences might play in the research and also to determine what kinds of new development in the mathematical sciences might be required. These are discussed further in part 5 of Section E (Program Activities).

Program Operation: Working Groups are the heart of SAMSI programs, meeting weekly or bi-weekly during a program to advance their research agenda. The core of most working groups is formed by the long-term visitors to SAMSI, the postdocs, and interested local participants. Because of the enormous success of the working group concept, it is clearly desirable to enable participation by individuals who cannot be present over a long period.

This year all of the SAMSI working groups have had significant participation by non-local individuals, as the following mechanisms were instituted:

- Technology was made available to all working groups, allowing non-local individuals to participate by teleconferencing; this also involved significant use of working group websites for posting of material and talks. We are continuing to experiment with different technological systems for connection, to maximize the ease of participation. Even though we are still in the experimental phase, the working groups in the NDHS program have had 8 regular remote participants this year, while the working groups in the ASTRO program have had 17 regular remote participants.
- At opening workshops, the opportunity and mechanisms for connection are highlighted, and non-locals who are interested in participating in the working group are 'signed up.'
- Working groups can now be partly based elsewhere; as well as have some participants residing at SAMSI. Indeed, one such remote working group node is being planned for next year at Berkeley, as part of the High Dimensional Inference and Random Matrices Program.
- Mid-program workshops of working groups are now often being held away from SAMSI, in places of significant concentration of external working group participants. For instance, three of the four working groups in the National Defense and Homeland Security Program held their mid-program workshops at external sites.

Program Evaluation: The Evaluation process implemented at SAMSI has several parts. All Program workshops are evaluated individually by the participants, as are the Undergraduate Workshops; and SAMSI graduate courses are evaluated using university course evaluation forms. Postdocs evaluate their experiences both while in residence at SAMSI and afterward; and mentors also provide individual commentaries. In addition, previous Postdocs and Program Leaders are surveyed via email (with telephone follow-up) to try to gain insight into the indirect impact of SAMSI programs through incorporation of ideas and/or actual material into long-term research activities, course syllabi and continuing or subsequent interdisciplinary collaborations. Immediate research impact is evident in publication lists and a compilation of research grants that result from or continue SAMSI research agendas and/or interdisciplinary collaborations. The information gathered is presented in detail in Section B and Appendices E and G.

Program evaluation will be considerably simplified once the new database system is up and running. It will automate maintenance of contact information and tracking of career development of SAMSI Postdocs, new researchers and students. It will also provide immediate access to information about participant composition for workshops, programs and other activities. Planned expansions will include compilations of publications supported by SAMSI, resulting from continuations of SAMSI research or SAMSI collaborations or deriving from ideas first formulated at SAMSI.

Human Resources: We continue to focus on the postdoctoral program. The pool of applicants and eventual hires continues a sharp upward trajectory, as SAMSI has increased in visibility. This year the applicant pool was 99 candidates, most of whom were of very high caliber. Furthermore, we have succeeded in hiring the top applicants; all of the eight offers we made to the top candidates were accepted.

A major threshold seems to have been crossed in terms of long-term visits by leading researchers to SAMSI. We had 11 visitors for at least a semester, and 13 visitors for one to three months. Of these, 8 were new researchers, many of whom were becoming involved in a new area of interdisciplinary research. This significant increase in participation was primarily due to SAMSI becoming widely recognized as having a wonderful research environment, and was also due to a more pro-active role of the directorate; in the past, we had primarily relied on program leaders to contact individuals about participation, but now we personally follow up on many of the suggestions of program leaders, a personal touch that has proven to be very effective.

The large influx of visitors was handled by utilizing shared offices for all postdocs and visitors. The local faculty fellows, who were major participants in the programs, were accommodated by utilization of a large shared office rented from NISS. As this upward trajectory continues, we will become constrained in terms of space, a major reason we are planning a building expansion.

We have developed mechanisms to enable SAMSI Graduate Fellows to be non-local. Next Fall we will have 6 graduate fellows at SAMSI from non-local institutions, and all will be heavily involved in the research programs.

Achieving *diversity* is a never-ending challenge. In addition to the many aspects of diversity involved in our regular Education and Outreach activities, SAMSI will be co-sponsoring the Conference on African-American Mathematical Scientists, with the Department of Mathematics at UNC Chapel Hill, in June, 2006. SAMSI has previously joined other mathematics institutes in sponsoring events at conferences for minorities and has now initiated the forming, with AIM, of a Diversity Coordination Committee that will share information across institutes in the US and Canada about minority opportunities and initiatives.

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

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

1. Research

a) *LATENT VARIABLE MODELING IN THE SOCIAL SCIENCES*: One goal of SAMSI programs is to encourage collaborations between scientists from different disciplines. The Latent Variables in the Social Sciences Program has had great success in this. Six interdisciplinary working groups were formed at the opening workshop and they continued to meet throughout the year.

One of the most active working groups was on model selection. Structural equation modeling is ubiquitous in the social sciences, and the methodology that is currently used for selection from among contending SEM's has recently come under considerable criticism in the social science literatures. Thus, in this working group involving numerous social scientists and statisticians, a natural thrust was to seek to adapt more modern statistical model selection techniques to selection from among SEM's.

One of the most commonly used criteria for selection from among SEM's is BIC, which was designed as an asymptotic model selection technique for a simple setting. The statisticians in the group observed that this was known to be highly inaccurate for complicated settings such as SEM's and a joint effort has created a breakthrough – what seems to be a much more accurate version of BIC (called GBIC) useable for complex models such as SEM's. The working group research has also led to a book, *Model Uncertainty in Random Effects and Latent Variable Models*, being edited by David Dunson and published by Springer.

A sociologist, a business professor, and a statistician from one of the working groups have developed an improved way to estimate models that have categorical dependent variables and independent variables that are measured with error. Scientific research ranging from medicine to economics often involves equations like this, so their research should reach beyond the social sciences.

In the complex survey working group, a statistician and a sociologist have joined together to use latent class analysis to gauge the degree of measurement error in the government's unemployment figures. The goal is to assess their accuracy and to develop better unemployment estimates. Their research was recently presented at the Methodology Section of the American Sociological Association.

b) *COMPUTATIONAL BIOLOGY OF INFECTIOUS DISEASE:* Several major findings and initiatives resulted from the collaborations in the working groups. A key step in the development of an adaptive immune response to pathogens or vaccines is the binding of short peptides to molecules of the MHC for presentation to T lymphocytes, which are thereby activated and differentiate into effector and memory cells. The rational design of vaccines consists in part in the identification of appropriate peptides to affect this process.

There are several algorithms currently in use for making such predictions, but these are limited to a small number of MHC molecules and have good but imperfect prediction power. The working group in this area undertook an exploration of the power gained by taking advantage of a natural representation of the amino acids in terms of their biochemical properties and using powerful statistical classifiers. In all cases, the encoding by biophysical properties leads to substantially lower misclassification error.

The immune system consists of a multitude of motile cells know as *leukocytes*, that move independently through the blood, lymph, and tissues surveilling the environment for evidence of microbial pathogens, unfamiliar molecular patterns, and damage to the host. Once any of these signals has been received, the cells themselves begin signaling to each other, and thereby alter the trafficking patterns of the population as a whole. Understanding the motions of leukocytes is key to understanding the immune response and to the rational design of immune interventions, and in particular, to vaccine design. While the motion of leukocytes under artificial conditions has been well characterized, the mechanisms of motility under natural conditions have not been elucidated. One working group developed Bayesian methods to study the trajectories of these cells, based on using a Langevin Process prior and computing the posterior mean trajectory, and analyzing the residuals with respect to this mean trajectory. The results are consistent with earlier analyses, but provide more detailed insight, suggesting that lymphocyte motion is saltational, with intervals of rapid and relatively unidirectional motion interrupted by pauses and reorientation of the direction of motion.

Another working group focused on *modeling the immune system.* It has led to the development of new models that can be used to study the impact of drug resistance on acute viral infections. Alun Lloyd is also developing a textbook on Infectious Disease Modeling, based on his lecture notes for the modeling course he taught as part of the program.

The transition workshop was entitled “Collective Computational Biology for Infectious Disease” and involved legal and policy scholars, as well as, statisticians, computer scientists, and biologists. The workshop was the occasion for the newly-formed *Tropical Disease Initiative* (TDI) to make the acquaintance of the Program participants. The TDI is devoted to the development of an open-source community for the discovery of drug targets and therapies for diseases of the developing world, starting with malaria. New collaborations were thus formed, and several GGH participants worked on infrastructure development for the TDI. Tom Kepler has become a member of the TDI, and as a result of the workshop now sits on the board of directors for *The Synaptic Leap,* a non-profit open-source biomedical research development company that hosts the TDI (http://thesynapticleap.org). TDI has now grown to include researchers in Europe, Australia and India as well as at several US sites, and is engaged in negotiations with other research groups to provide results from their research to the research community at large for collective computational discovery of novel antimalarial drugs. These seminal developments were made possible by SAMSI and the workshop that brought together for the first time the principal actors who are now working together to bring new, collective, ways of doing statistics, mathematics, and computational biology to bear on the enormous problem of infectious disease in the developing world.

c) *DATA ASSIMILATION IN GEOPHYSICAL SYSTEMS:* The Spring 2005 program was extended into Fall. This program provided a context in which statisticians and applied mathematicians worked closely together and the mutual influence on each other's thinking was extremely productive. This will be reflected in an upcoming special issue of the journal Physica D, which is a leading applied mathematics journal for research on nonlinear phenomena. The journal will contain papers that reflect a signicant shift in thinking concerning data assimilation. The shift is from a position, predominant in weather prediction, that model error is overshadowed by the chaotic effect of a small change in initial conditions to the idea that drift of the state of the system is primarily due to errors in the modeling. This can begin to be dealt with by a strategy that combines a Bayesian view of state estimation with a dynamics view of ensemble formation.

Two key specific areas emerged that focused the efforts and energies of the researchers in the program. First, the promise of assimilating Lagrangian and quasi-Lagrangian data is of enormous importance in ocean state estimation. This program allowed a group of applied mathematicians and oceanographers, with considerable input from statisticians, to consider a number of different approaches to this all-important problem. These approaches include stochastic DE methods, smoothing, and the use of indistinguishable states. The Lagrangian working group spawned a number of group efforts, involving a variety of combinations of core group members, postdocs and students, to implement these approaches. One fundamental question concerned the direct reconstruction of time varying velocity fields from actual float data. Considerable progress was made on characterizing a quasigeostrophic formulation of this problem.

The Lagrangian data assimilation problem stimulated development of a Langevin sampling technique by Andrew Stuart, Amit Apte (SAMSI postdoc) and co-workers. This is a new Bayesian formulation of the problem that promises to overcome some fundamental obstacles to the fusion of Lagrangian data into models.

Significant culture-changes were effected among both the dynamical systems participants and the statistical participants in the program. For instance, the dynamical systems participants became aware of and began to extensively utilize the power of a more formal Bayesian formulation of data assimilation, while the statisticians came to an understanding that their standard techniques for implementation of Markov Chain Monte Carlo methods would fail for dynamical systems. Joint efforts are underway to develop MCMC methodology that is efficiently useable for dynamical systems, a truly exciting possibility.

d) *NATIONAL DEFENSE AND HOMELAND SECURITY:* As an ongoing program, much of the research has yet to be finalized, but several results have been obtained already. A paper by Banks, Karr, Nguyen and Samuels introduced a new paradigm for characterizing the richness of mathematical models. In this instance, the richness is a function of the size of stochastic disturbance terms: as compared to the deterministic model (variance equal to zero), the stochastic may be essentially the same, richer only in terms of transient behavior, richer in terms of both transient and equilibrium behavior, or dominated by the noise.

Research has also been initiated on Bayesian multiple hypothesis testing for Poisson data as well as some on the use of mixing distributions to analyze potential spurious observations. Key advances were made in finding default prior distributions that are computationally implementable and which provide for an automatic ‘correction for multiplicity.’ The work will be broadly applicable to the many anomaly detection problems where the data is discrete.

e) *FINANCIAL MATHEMATICS, STATISTICS AND ECONOMETRICS:* Much of work of this program will be finalized in the coming months, but major activities and research thrusts are as follows.

Credit markets, with popular products such as credit default swaps (CDS) and collateralized debt obligations (CDO's), have become a major place for investments. Understanding the mechanism of defaults and correlation of defaults is the most important issue for pricing and hedging these products and associated risks. The working group on credit risk developed new approaches to the problem based on the introduction of stochastic volatility and correlation through stochastic volatility as well as jumps. These new techniques are being made available to practitioners.

A new Monte Carlo computational technique to handle optimal stopping problems and stochastic control problems has been studied and implemented by the workgroup on computational issues. This provides a significant step toward the solution of high dimensional problems which cannot be treated by classical partial differential equation (PDE) techniques.

The portfolio optimization working group investigated the development and implementation of backward differential equations for applications arising in portfolio management and stochastic portfolio theory. Aspects of this theory have ties to stochastic control and the manner in which stochastic control techniques may influence financial theory represents one of the synergistics aspects of the program.

Some of the main issues investigated by the working group on model uncertainty focused on how one accounts for uncertainty regarding the data generating process in pricing asset, portfolio choice, statistical decision making, and hypothesis testing pertaining to financial monitoring. The variety of topics touched on the foundations of decision making under uncertainty.

f) *ASTROSTATISTICS:* The program only began in January, and so most of the work has yet to be finalized. Here are the highlights of the activities of the working groups.

Exoplanets: Most of the effort has been devoted to determining which methods of marginal likelihood calculations are appropriate for exoplanet models. (Searching for multiple exoplanets is essentially a model choice problem.) None of the existing methods seemed to be successful, because of the very sharp peaks in the likelihood function for planet periodicities. The working group has developed several highly promising methods for turning MCMC chains (which can explore the periodicity space well) into importance functions that allow for efficient marginal likelihood calculations.

Surveys and Population Studies: This working group is focusing on merging model-based and design-based approaches for analyzing survey data in astronomy. A unique aspect of astronomical surveys is the combined presence of truncation (often random) and measurement error (often heteroscedastic). Model-based approaches can readily handle both complications, but astronomers' techniques rely on overly restrictive models. Design-based approaches handle truncation in a more robust way than the model-based methods, but no such methods known to astronomers can account for measurement error. Thus the group is devising methods that marry the rigorous error handling of modeling with the robustness of product-limit estimators. To deal with the problem of multiple testing, the group is investigating the possibility of combining false discover rate (FDR) control techniques with Bayesian modeling. FDR may be used (with a high rate) to generate a subset of data for a subsequent Bayesian analysis. This both reduces the size of the data set, making the subsequent analysis more computationally tractable, and establishes an objective prior for the subset.

PhyStat: During March, there was focused activity on the interface between high-energy physics and statistics. The topics included determination of upper confidence limits in the presence of nuisance parameters, multivariate methods for the separation of signal and background, and goodness of fit with sparse multi-dimensional data. The collaborations initiated during this activity are particularly crucial because of the advent of the new LHC accelerator at CERN in 2007, and the need for new statistical methods to handle the resulting data. Most likely there will be a further workshop at CERN to further advance this research.

2. Human Resource Development

SAMSI's impact on human resources is fully discussed in sections I.B and I.C, with impact on diversity highlighted in section I.H. The individual program reports also contain significant insight into human resource development. Here we give several illustrations indicating the unique impact that SAMSI has in these areas.

Research plans and paths. Few people have left SAMSI doing exactly the same kind of research as when they arrived. Here are some comments in the final reports from 2004 of the postdoctoral fellows, and some comments from the reports submitted by long-term visitors. More details can be found in sections I.B. and Appendix A.

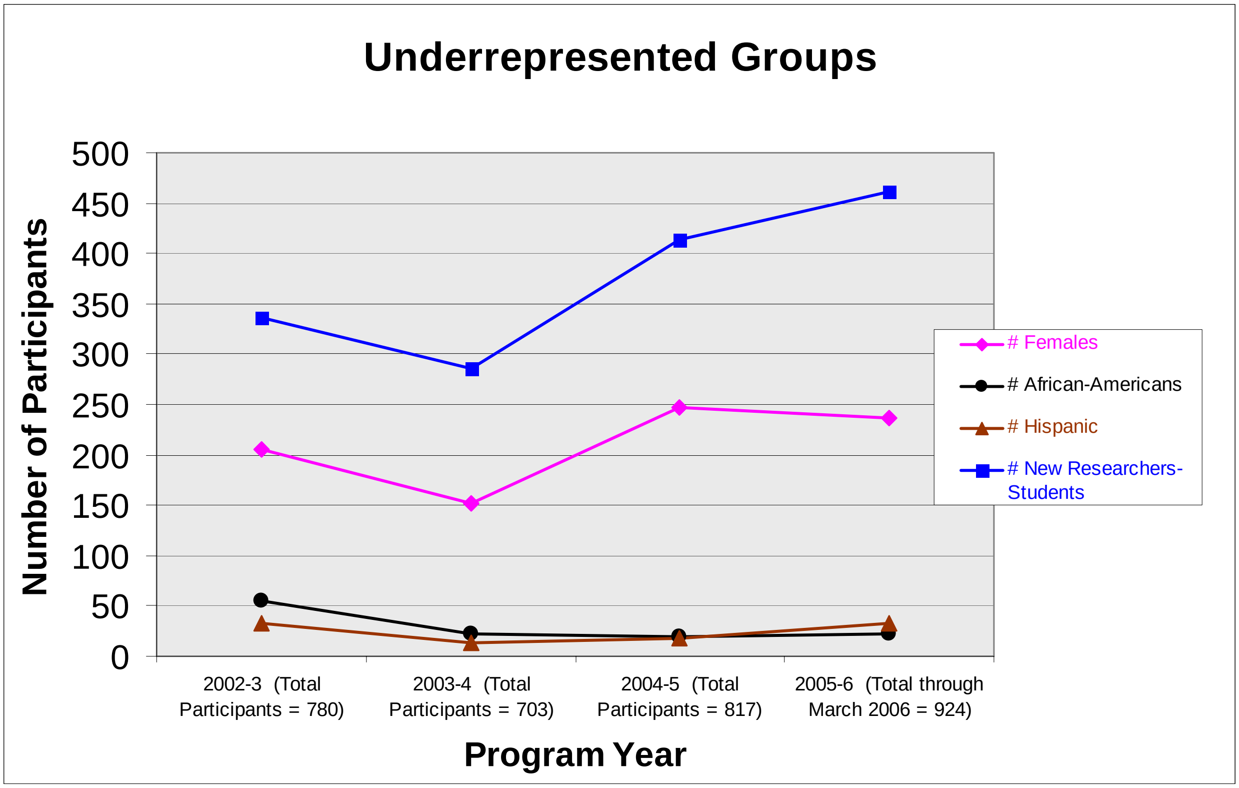
- An INVERSE program postdoc with a degree in applied mathematics, who is now a tenure-track faculty member in statistics at Texas A&M University and says "My current position is a direct result of SAMSI."
- An LVSS program postdoc with a degree in sociology, who is now a tenure track faculty member in sociology at University of Arizona, reports "I am substantially revising my dissertation, which I intend to publish as a book, to incorporate latent variable methods. I give much more attention to measurement error in all my work. The SAMSI affiliation also helped me enjoy considerable success on the job market last year … I am advising several students on use of latent variable methods in their MA and PhD research."
- An NMI program postdoc, with a degree in theoretical statistics, now a faculty member at the University of Georgia, reports "I learned how to organize and led a research group, and how to discuss and share ideas effectively with researchers from other fields. I feel comfortable at working at interdisciplinary research group thanks to the SAMSI experience and currently I am participating in two projects at the University of Georgia."

- A DA program postdoc, with a degree in dynamical systems from physics, has developed an expertise in applying Bayesian statistical techniques and is now pioneering work in infinite-dimensional Langevin sampling, with Andrew Stuart of Warwick, and applying these ideas to Lagrangian data assimilation in ocean models.
- A SAMSI University Fellow, from Oxford University and the London School of Economics, wrote: “My time in SAMSI has already proven of immense value to our research work at CATS and influenced our interactions with ECMWF and other operational weather centres. The SAMSI experience was particularly valuable due to the unique environment that allowed me to view data assimilation in the biggest pictures; being at any particular research university or operational centre naturally leads to a focus on local expertise and interests, while the SAMSI program stimulated vibrant interaction between researchers from a wide range of home institutions. While I have seen this attempted on relative short time scales (a week or so) at locations like the Newton Institute, I have never experienced it done as successfully (in terms of people invited, intensity of discussion, support for research and support for domestic needs) as it was sustained over the duration of this SAMSI program.”
- A University Fellow from the DMML and LVSS, who pursued and received a grant from the Canadian NSERC “based on the proposals for the research I ended up starting at SAMSI” stated “This is the best intellectual/ research environment I've ever had the chance to enjoy. It seemed to me that everyone is keen to bring in the future, teach their fellow travelers and learn in turn, and enjoy the development of new ideas”
- A visiting Professor of Mathematics, Statistics & Computer Science, Northern Michigan University has, as a result of his SAMSI experience, joined an NIH-funded two-year collaborative research program at Texas A&M University. He comments, “The teamwork experience provided by the ‘focus groups’ that are formed after each inaugural conference can be really beneficial to fresh Ph.D.'s.”
- A visiting professor of Mathematics, Louisville University says “I have developed collaborations with NDHS program participants, … resulting in joint presentations at SAMSI undergraduate workshop in March, also a joint invited presentation on Anomaly Detection … at JSM next August. During my first long-term visit to SAMSI, I had an opportunity to organize a seminar series on statistical methods in gene expression analysis. This has created a flow of interesting contacts with a whole bunch of individuals in the summer of 04 including SAMSI postdocs …”
- A postdoc from the London School of Economics, says “(SAMSI has) the atmosphere of genuine collaboration and is well suited to disseminating and gathering knowledge and information.”

Graduate students are also influenced to develop dissertations from the research problems explored by SAMSI Working Groups; and undergraduates are influenced to continue to involve mathematical sciences in their graduate school programs – with some applying to doctoral programs in statistics or applied mathematics.

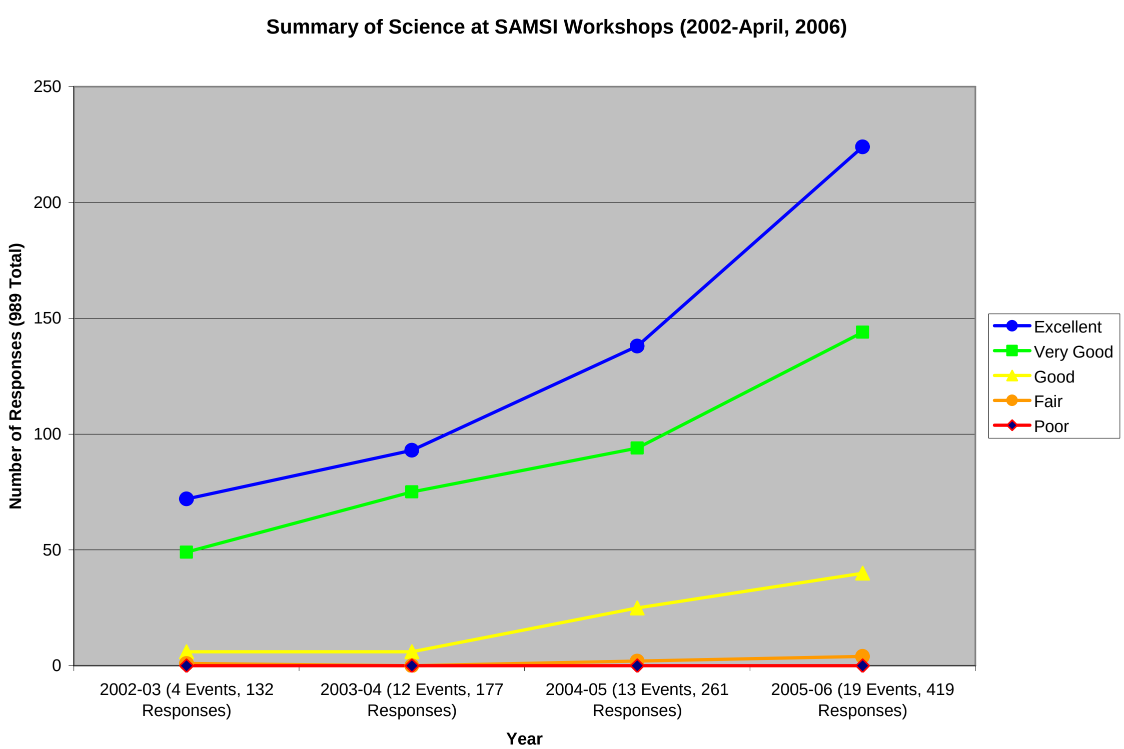
- A *Stochastic Computation Program* graduate student (statistics), Ph.D. September 2005, currently Assistant Professor, Department of Statistics, Ohio State University: “Most of my dissertation research stemmed from work on the SAMSI Program in Stochastic Computation… I would recommend participation in a SAMSI program, primary because SAMSI provides an excellent forum for researchers with different statistical and applied mathematical backgrounds with together on problem areas of interests.”
- A *Latent Variables in Social Sciences Program* graduate student (sociology) Ph.D. expected May 2006, accepted position as Assistant Professor in Crime, Law, and Society Department at the University of California at Irvine: “The work in the group exploring dynamic networks has been influential into the approach used in one chapter of my dissertation. . . . It brings together people from diverse backgrounds, allowing for an interesting cross-fertilization of ideas. … I believe that the insights and ideas that were fostered by the experience will be important over the next several years for the directions my research takes.”
- A *Stochastic Computation Program,* graduate student (statistics); Ph.D., 2003, currently a quantitative trading researcher at a hedge fund in London “The final states and the two chapters of my thesis had an immediate and direct link with my work at SAMSI…My experience at SAMSI's Program on Model Selection in Financial Models was extremely helpful to obtain my First Job at Credit Suisse, and the second one was immediately linked to that one. What I learned at SAMSI has also been quite helpful even in my everyday work.”
- *Undergraduate from Utah State University*: “The two-day undergraduate workshop was excellent in helping me to understand practical applications of my statistical training thus far, and solidified my plans to pursue research in a graduate school setting.”
- *Undergraduate from Meredith College:* “While I didn’t always understand the math …, I found the concepts and questions clear and interesting”
- *Undergraduate from Winona State University:* “SAMSI offered a great insight into real-life applications of using statistics in areas of homeland security. We got to see new and recent research questions: a privileged exposure. The students were really fun and it was a diverse and excited group. I’m glad that I had this opportunity.”

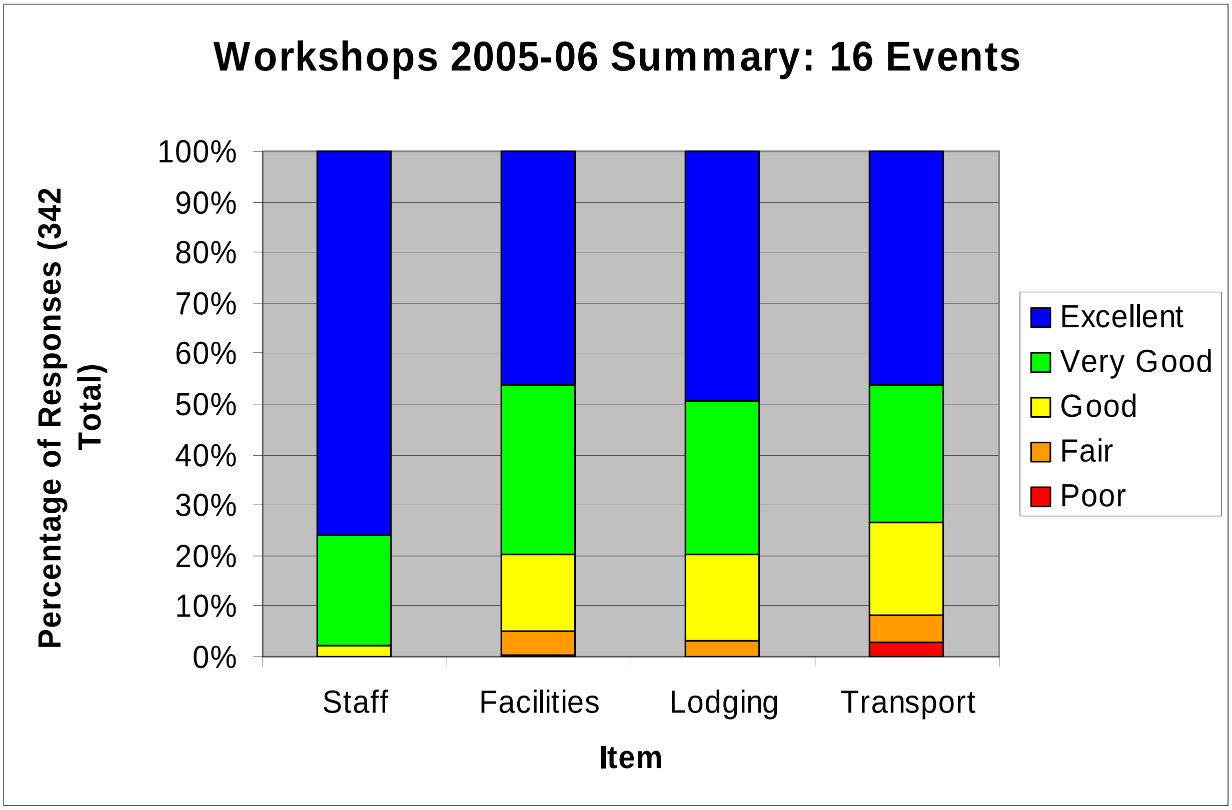
Diversity: SAMSI policy is to give attention to diversity issues throughout all activities, especially in the Postdoc selection process and in the organization and operation of Workshops and Programs. SAMSI’s success in maintaining a diverse community began with the first activities and continues. Numbers for 2005-6 include only the academic year activities through March; summer activities especially include students and new researchers. Because of the vigorous attention given to diversity, underrepresented minorities are not underrepresented at SAMSI.

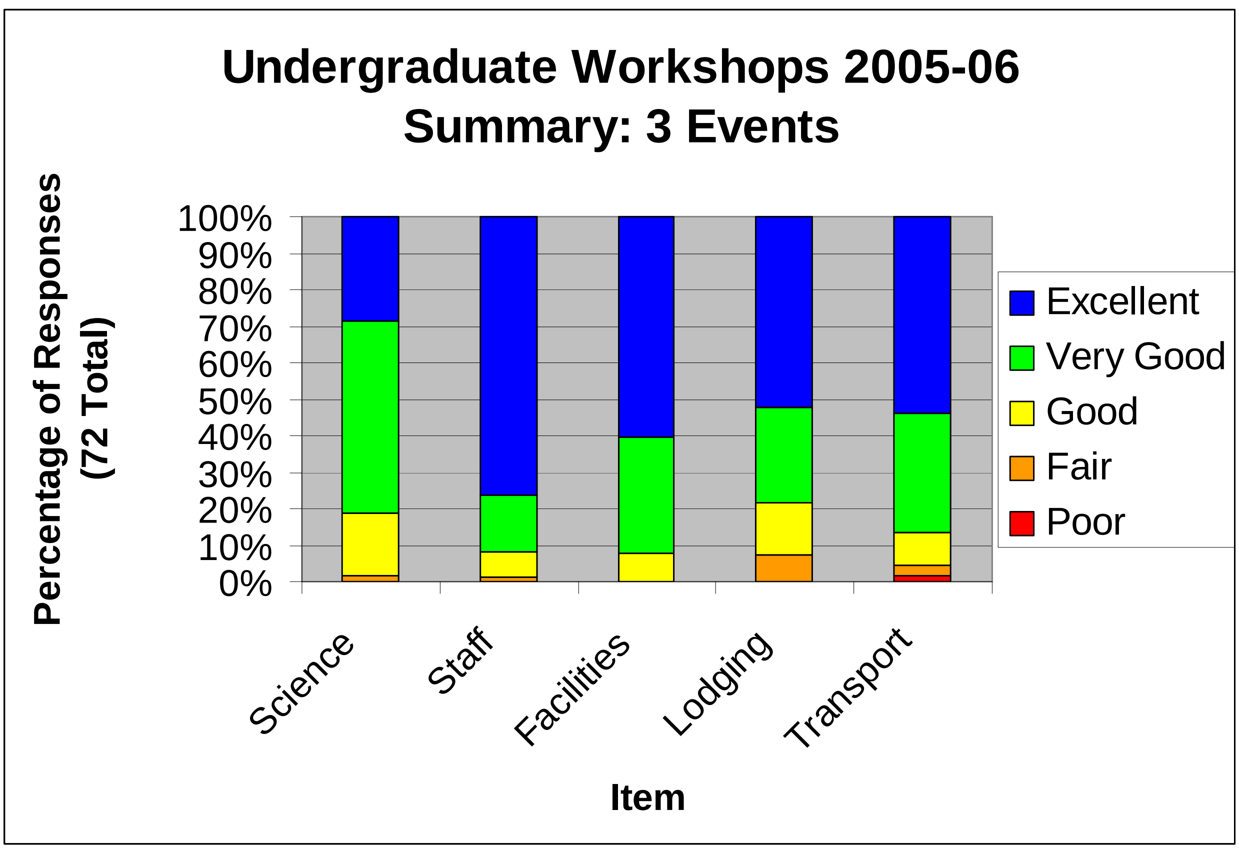


*Individuals may have been participants in several SAMSI activities.

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




Workshops 2005-06 Summary: 16 Events
Percentage of Responses (342 Total)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Staff
Facilities
Lodging
Transport
Item
Excellent
Very Good
Good
Fair
Poor


Undergraduate Workshops 2005-06 Summary: 3 Events
Percentage of Responses (72 Total)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Science
Staff
Facilities
Lodging
Transport
Item
Excellent
Very Good
Good
Fair
Poor

Broadening the DMS research impact: SAMSI's national impact also depends on Institutional Diversity and the inclusion of participants whose home institutions are not already heavily supported by NSF Funding through DMS. Such inclusion develops the national research base by significantly increasing the number of individuals that can engage in cutting edge research.

The SAMSI record in this regard during 2005-06 is excellent, as shown in the following table (for both funded participants and all participants). The 'Other' category primarily includes individuals from other disciplines, governmental agencies or laboratories, and industry.

2005-2006 SAMSI Participation			
Funded Participants	**Home Institution by DMS Funding Level**		
	Top 50 DMS Funded	51-200 DMS Funded	Other
# of Institutions	36	40	83
# of People	130	86	113
% People	39.5%	26.2%	34.3 %
All Participants			
	Top 50 DMS Funded	51-200 DMS Funded	Other
# of Institutions	46	56	149
# of People	239	251	251
% People	32.2%	33.9%	33.9%

3. Education

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

(i) Given the breadth of constituent disciplines coming together in the Astrostatistics Program, the tutorials which initiated the Opening Workshop were expanded significantly from the usual 1 day model used for most programs. The overall goal during the 5 days of tutorials was to introduce statisticians to current trends in astronomy and expose astronomers to modern methodologies in statistics and applied mathematics. Details can be found at http://www.samsi.info/programs/2005astroprogram.shtml. To further disseminate the extensive body of material presented during this period, all tutorials were taped and are being made available to all interested scientists.

(ii) 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 Financial Mathematics, Statistics and Econometrics and National Defense and Homeland Security. 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.

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

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

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

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

To illustrate the scientific level achieved during the workshop, the team investigating a project presented by MIT Lincoln Lab had the following paper accepted as a refereed proceedings: A. Berger, N. Razouk, G. Angelides, A. Bartlett, A. Langville, Z. Li, C. Lipkin, N. Mavinga, E. Perez, E. Tweedy and E. Wheeler, "Locally constrained shortest paths and an application in mission planning," Proceedings of the 44th ACM Southeast Conference, Melbourne, FL, pp.766-767, 2006. The corresponding poster presentation won a 2nd place "Best Poster Award."

E. Responses to the Third-Year Review

On June 8, 2005, we received the report from the Third-Year Review of SAMSI which was very positive and contained numerous recommendations for further enrichment and development of SAMSI. Below, we list each of the issues raised beginning with the summary comment in the NSF letter concerning that issue, along with the 'Weaknesses and Recommendations' contained in the Report itself. (Some of the comments from the report have been repositioned below for better organization of the response.) The changes at SAMSI, current and planned for the future, that bear on these issues are also presented.

1. National Leadership and Presence

An overarching theme of many of the specific issues raised by the review committee (and discussed later) is the strong recommendation that SAMSI should do everything possible to further involve national leadership and to further increase its national presence. Ensuring that SAMSI is focused outward is the result of numerous specific strategies discussed in detail later, but here is a summary.

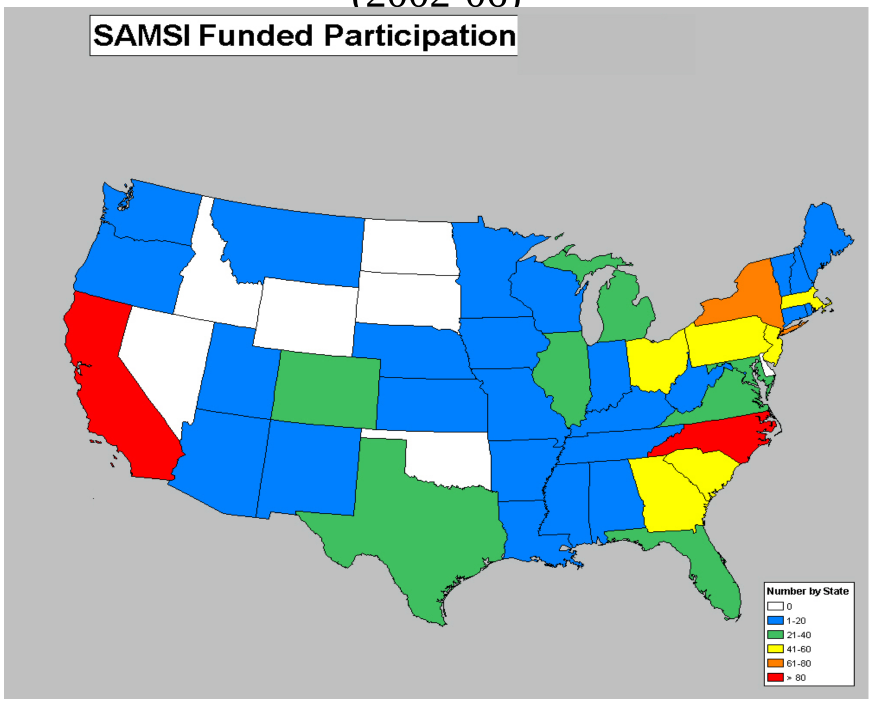
In terms of national scientific leadership:

- As discussed in part C above, *program leaders committees* of all current and future programs are national and international in makeup.
- Two additional external members have been added to the *Governing Board*, appointed by the ASA and SIAM presidents. The governing board thus consists of 3 local individuals (representatives of the administration of each partner university) and 3 external members.
- A *Scientific Liaison Board* of advisory disciplinary scientists is being created to augment the national leadership in the National Advisory Council.
- The *Education and Outreach Committee* now has an extensive national presence.
- In the SAMSI renewal proposal, we will be requesting the creation of a *Deputy Director* position, which will be reserved for two-year rotating visitors, bringing a continual national representation to the Directorate.

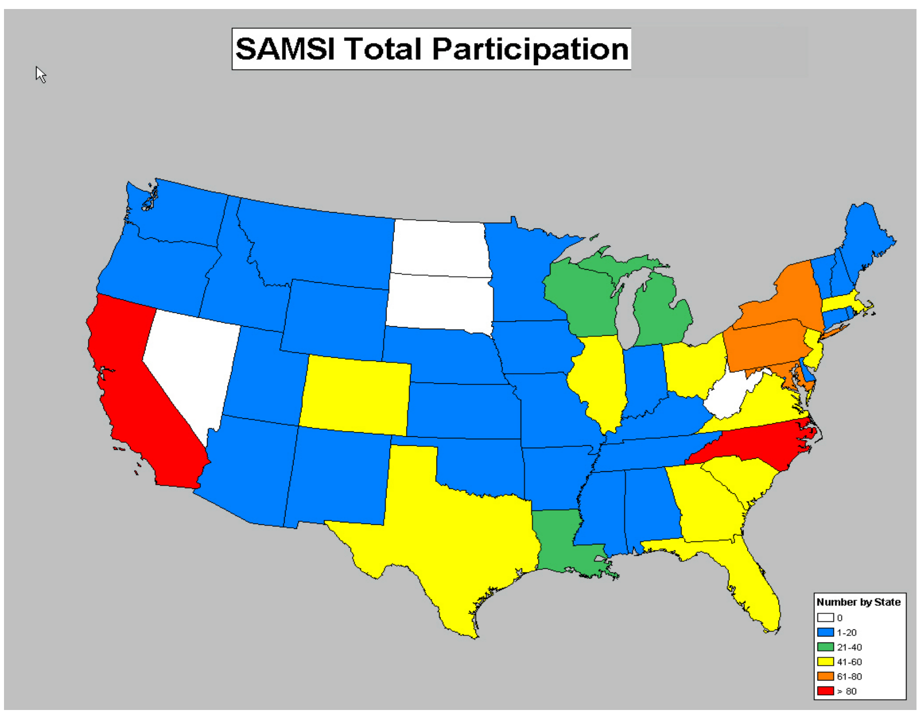
In terms of national and international presence:

- The NSF funding to SAMSI goes almost exclusively to support non-local researchers; the efforts of local researchers are supported by the Partner Institutions of SAMSI. This ensures that the focus of SAMSI is primarily national in terms of the NSF research funding.
- The following two graphs show that workshop participation is highly geographically diverse nationally, both in terms of participation and in terms of funding. These graphs present the cumulative totals over the four years of SAMSI operation.

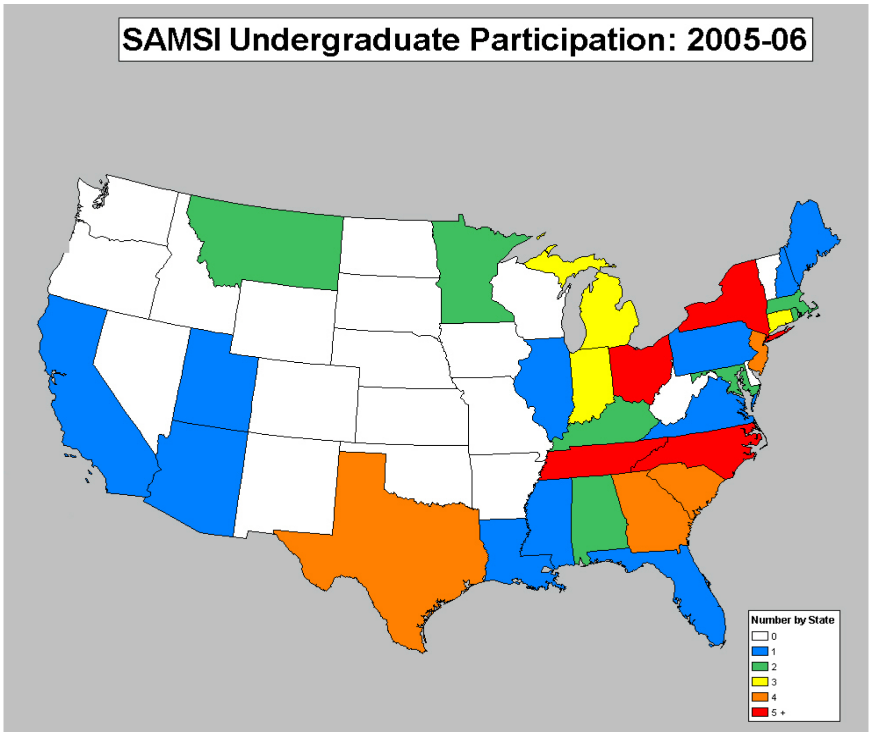
Geographical Distribution of Funded Workshop Participants (2002-06)



Distribution of Workshop Participants (2002-06)



Distribution of Undergraduate Workshop Participants (2005-06)



- SAMSI attracts international visitors as well, although in the areas of statistical sciences and applied mathematical sciences, the concentration of leading research still resides heavily in North America, in contradistinction to other mathematical sciences for which dominant strengths and leading research resides on other continents. It is SAMSI's policy always to attract and support the leading scientists, regardless of nationality; but to focus resources otherwise on domestic participants. This is especially true in selecting Postdocs to be supported by SAMSI. The table below shows that the proportion of non-citizen, non-permanent residents has declined from roughly 35% in SAMSI's early years to a little over 30%, a level that reflects the presence of internationally leading researchers.

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 (to date)	200	95	36	331
TOTAL	787	343	122	1252
Percentage of all funded participants (1249)	63.01%	27.46%	9.77%	

- SAMSI uses multiple strategies to reach and engage the national community, in addition to support for the many visitors to SAMSI and workshop participants, including the following:
  - Participation in working groups is now possible for anyone, through the mechanisms highlighted in Section C above.
  - Last year, over 1/3 of our scientific workshops were held at other locations in North America.
  - The graduate fellows program has been enhanced to enable extensive participation by graduate students from non-local universities.
  - The new website allows easy access of the community to SAMSI.
  - The directorate gives roughly 50 talks around the world each year in which the opportunities at SAMSI are discussed.
  - Annual one-page articles are sent to the newsletters or bulletins of AMS, ASA, IMS, and SIAM (and occasionally elsewhere, e.g. IEEE).
  - Advertisements are placed in the newsletters of ASA, IMS, and SIAM for the postdoctoral program and visiting opportunities.
  - Posters of major activities are sent to all math and stat/biostat departments (and many other places).
  - There are regular announcements of all upcoming activities to the organizations in the affiliate program.
  - Each major event (including E&O events), is announced by mass e-mails to relevant departments and organizations nationally (and internationally). For interdisciplinary events, considerable effort is also spent in trying to identify relevant departments in the other disciplines to be notified. Available discipline-based individual mailing lists are also utilized when appropriate.
  - At the annual Joint Statistical Meetings and SIAM meetings, at least one SAMSI specific session is held, to not only discuss SAMSI research, but also to highlight the next year's (and future) programs. We strongly encourage participation and proposal of new programs. Numerous other groups are typically also addressed at these meetings.
  - At the JSM, there is a SAMSI reception one evening (joint with NISS) and, at the Joint Mathematics Meeting, SAMSI participates with the other institutes in a joint reception/information session.
  - Programs are strongly encouraged (and many do) submit scientific sessions of SAMSI research (identified as such) at a wide range of scientific meetings.

2. Scientific Programs

DMS Letter: The National Advisory Committee consists of statisticians, mathematicians, and computer scientists only. There is no formal mechanism for direct input from other disciplinary sciences for new ideas for scientific programs.

Currently, the development of' the scientific programs seems to rely on the National Advisory Committee. The Director and Associate Directors need to take a more proactive role in identifying scientific programs, particularly in anticipating previously unidentified broad scientific research areas for which the joining of statistics and applied mathematics will play a pivotal role.

Committee Report: Another weakness noted by the SVT is that the NAC consists only of statisticians, mathematicians and computer scientists. There is no current NAC member who can represent the perspective of other scientific disciplines in which SAMSI intends to engage as part of its stated mission Input external to the mathematical sciences would enhance the contributions of the NAC. Include direct input from major figures from the disciplinary sciences in setting the scientific agenda of SAMSI.

Attempt to anticipate previously unidentified broad areas of scientific research in which the joining of statistics and applied mathematics is expected to play a pivotal role.
The SVT feels that SAMSI could still be more ambitious in achieving the goal of' national leadership and scientific impact.

SAMSI Response*:*
(i) A NAC *Scientific Liaison Board* is being created. It will consist of eight individuals, each appointed to 4 year terms. These will be leading individuals in their science (or sciences), or individuals with a broad perspective on science (such as former research directors at a national lab or research deans at a university). Their having some interest in the role of mathematics and statistics is also clearly needed, but they will not be statisticians or applied mathematicians. As a guideline, we would strive to have one individual representing each relevant NSF research directorate, along with someone representing medicine.

Each year we will invite 2 of these individuals to come to the annual NAC meeting, and give a presentation on what they perceive as the most exciting future areas in their science with which SAMSI could potentially be involved. While at the meeting, these leaders would also give advice on the SAMSI programs being discussed that are relevant to their science.

For the six members not in attendance, we would send the relevant proposed SAMSI programs to them in advance of the meeting and either ask for their comments by the time of the meeting to include in our discussion, or ask them to be available for teleconferencing when discussing relevant proposals.

Obtaining meaningful commitments from scientific leaders for endeavors outside their discipline is always difficult. This plan was chosen (among many we considered) as the plan most likely to result in meaningful engagement of such leaders.

It is important to note that this Scientific Liaison Board will just be one of the many mechanisms that SAMSI utilizes to obtain input from other sciences. The SAMSI Directorate, the National Advisory Council, the Local Development Committee, the Chairs Committee and the Governing Board comprise roughly 40 leaders who are heavily connected with other disciplines and routinely seek to engage relevant disciplinary scientists in program development. Of course, those planning interdisciplinary programs also typically have many such contacts themselves.

Leaders in other disciplines are routinely on scientific and organizing committees of SAMSI programs and workshops, and routinely are major participants therein. As but one example, in the current Astrostatistics program, astronomers and physicists were central to the development of the program and four astronomers and one physicist are long-term visitors to the program.

(ii) The Directorate already plays a very pro-active role in the identification of potential programs. In this regard, it is useful to review the development of SAMSI programs:

- *Phase 1:* Ideas come in from a wide variety of sources, including national (and international) individuals, the NAC, and the directorate's own knowledge as to potentially exciting new areas for applied mathematics and statistics. Ideas are given an initial 'credibility screening' by the directorate. In the past ideas that have failed this screening tended to be ideas that are too specialized or too routine.
- *Phase 2:* Ideas that pass the initial screening either proceed to preparation of a formal pre-proposal, or are embedded in other programs – often on a trial basis. Examples of the latter include Causality, Granular flow, and Nanotechnology of Soft Materials that were all embedded in other programs as workshops, to assess their potential for development as a full program. (For various reasons, it was deemed that the time was not right for pursuing these as full programs, although one of them – Granular Flow – is being resurrected as a synergistically fitting subprogram of next year's Computer Modeling program.)
- *Phase 3:* Pre-proposals are brought before the NAC (and the local committees) to assess their potential as SAMSI programs. Those that are approved for development by the NAC enter a developmental stage, leading to a Full Proposal. A Full Proposal must be formally accepted by the directorate and NAC, and includes many of the specifics of the program.
- *Phase 4:* Details of the approved program are planned, including workshops and the research topics for working groups. While there is very considerable discussion of working group research topics in advance – among the long-term program participants and national leaders – the final working groups (and their research topics) are not set until the end of the Opening Workshop of a program. Indeed, the primary purpose of the opening workshop is to obtain input from the assembled national leaders (many from other disciplines) as to the most promising research foci of the working groups; a variety of break-out and discussion sessions at the workshop are oriented towards this result and a member of the SAMSI directorate guides this process.

At each phase, from the initiation of ideas to the final formulation of the specific research working groups, directorate members are heavily involved.

We are, of course, continually seeking new ways to help to identify potential new areas in which statistics and applied mathematics should be involved. Our latest effort in this regard is the initiation of *Brainstorming Meetings,* mentioned in Section C above and extensively discussed in part I.E.5.

3. Governance

DMS Letter: SAMSI has a strong local base and support, which was recognized by the site visit team as a significant factor that helped SAMSI to progress quickly from its infancy to a fully functioning institute. As a national institute, SAMSI should build upon the strong local base to begin incorporating national representations in its management and to ensure national perspectives are an integral part of the SAMSI operations. For example, the pool of candidates for the SAMSI Associate Directors could be enlarged beyond those from the local institutions, and regular evaluations of SAMSI could be conducted by a body consisting of both local and non-local members.

The Governing Board has formal responsibility for the hiring and firing of the SAMSI Director and Associate Directors. It is problematic that the Director, the PI on the award, doesn't have a formal role in the choice of new Associate Directors.

Committee Report: The Governing Board has only local representation, which is highly unusual for a national institute. While SAMSI has to take into account the substantial investment of the local institutions, the investment from NSF represents a commitment from the whole mathematical sciences community and therefore requires accountability at a national level. Concerns over this management structure are heightened by the fact that the only regular evaluation of' SAMSI is now carried out by the local Governing Board.

A mechanism should be found to maintain the support of the local universities and yet expand the pool of candidates for the Associate Directors (ADS), possibly beyond the local community. The goal should be to find an alternative to the current constraint of equal representation from the participating institutions.

Maintain a synergistic and beneficial relationship with NISS while continuing to avoid any potential conflict of interests.

As explained to the SVT, the Governing Board has formal responsibility for the hiring and firing of the Director and the ADS. We see as a problem the fact that the director does not have a formal role in the choice of new AD, someone with whom the Director has to work closely.

SAMSI Response:

(i) The *Governing Board* (GB) has been enlarged by two members, one (George Casella) appointed by the President of the American Statistical Association and another (Tom Manteuffel) appointed by the President of the Society for Industrial and Applied Mathematics. These appointees will serve three year terms. The GB will henceforth consist of three local individuals (representatives from the administrations of the three partner universities) and three external individuals (including the always external Chair of the NISS Board).

The GB currently provides an evaluation of SAMSI at the time of the annual report (included at the end of this executive summary), and will continue to do so. In addition, in its Fall meeting, the GB will formally review some aspect of the SAMSI operation and will include their review in the annual report.

(ii) Under the renewal proposal, we will propose to augment the current directorate by inclusion of a *Deputy Director*. This would be a two-year rotating position, filled by non-local individuals, and bringing continual national presence to SAMSI. In addition, the replacement of the Director is done by a national search and, when replacement of associated directors is necessary, we will strongly consider the possibility of a national search. For instance, Nell Sedransk will be replacing Alan Karr as the Associate Director from NISS, and she was found through a national search. Hence the directorate will continually have a strong national presence.

(iii) The excellent synergistic relationship with NISS is continuing, as evidenced by the willingness of NISS to consider a building expansion to accommodate SAMSI. Everyone is, indeed, very sensitive to conflicts of interest.

(iv) The SAMSI charter is being revised to clarify that the Director of SAMSI is formally involved with all appointments to the Directorate. This was, of course, the reality in the three replacements that have happened, but formalization of this clearly makes sense.

4. Dissemination

Committee Report: In addition, the potential for publicizing the scientific achievements to the general public and political leadership has not yet been exploited. Whereas many elements of the program are very successful, the closing workshops have low attendance and don't seem to be working as they were intended. Finally, the web presence is to date rather weak, although we recognize that efforts are underway to improve the SAMSI website. Recommendations:

- Expedite the upgrading of the web page as soon as possible. Make lectures, tutorials and course materials publicly and easily available via the website.
- If this not already happening, ensure there is a 'champion' for the appropriate disciplinary sciences to help with advertising these opportunities.
- Publicize SAMSI program outcomes more broadly, for example outside scientific community.

SAMSI Response:
(i) The upgrade of the web page is complete, and almost all lectures at SAMSI are now being placed on the web. In addition, the very extensive tutorials for the Astrostatistics program were taped, and we are exploring the most effective means of their dissemination. (The files are large enough that this is an issue.)

(ii) Most programs have disciplinary scientists on their Leaders or Scientific Committees that serve as ‘champions’ for dissemination of information. In addition, the appropriate individuals from the NAC Scientific Liaison Board can help in identifying the proper places for publicity to be sent.

(iii) SAMSI is moving to a varied system of closing events, depending on the circumstances of an individual program. First, closing workshops (now called transition workshops) take place many months later than the previous official ending of programs, to give more time for research to be finalized. Second, opportunities are sought to combine the transition workshop with other events, to increase the awareness of the achievements of a program. For instance, the Astrostatistics program this year is having its transition workshop combined with the Statistical Challenges in Modern Astronomy IV conference in June, which is a very established conference and reaches a wide audience of astronomers and statisticians. The physics side of the Astrostatistics program is having its work discussed in the Statistical Inference Problems in High Energy Physics workshop in July, reaching a wide audience of physicists and statisticians.

Where warranted, we do have more traditional closing workshops. For instance, the just ended Financial Mathematics, Statistics and Econometrics workshop had a closing workshop with an attendance of 70 people, most of which were not affiliated with the program, and thus resulting in considerable transference of program findings to the wider community.

5. Finances and Facilities

Committee Report: The fact that there is a preset, and substantial amount of SAMSI money for local graduate student funding is problematic both in terms of appearance and effective use of limited funds. Further, the budget for senior world class researchers impedes mounting scientific programs at the desired level.

There is a limited capacity for offices, meeting rooms and workshop space. Even medium sized workshops cannot be held at the SAMSI site. Further, there are no university basic services such as libraries, cafeterias or health and exercise facilities within walking distance. Recommendations:

- Seek additional funds for world class visitors from either new sources of money or though re-prioritization of existing funds.
- In the future, consider 1) adding graduate student support for short-term (the length of a working group) visits from other institutions, and 2) opening existing graduate stipends to national competition.
- SAMSI should give serious consideration to external development. We understand the potential conflicts among competing institutions but this might be important for the long term financial stability of SAMSI.
- The NISS plan for new conference facilities appears beneficial to SAMSI.

SAMSI Response:
(i) We have established mechanisms to enable SAMSI Graduate Fellows to be non-local; indeed, the graduate fellows program is now open to all national graduate students. As always, we will prefer applicants for whom mentoring is available (e.g., if the advisor of the student is a visitor at SAMSI or if another visitor is interested in working with the student). Next fall we will have 6 graduate fellows of this type at SAMSI. In addition, we will consider applications from graduate students who simply want to immerse themselves in the courses and working groups at SAMSI, for the educational experience. Such students will have an individual acting as in loco parentis, but will be at SAMSI primarily for the educational side of the research experience. In their application, such students will need to show how the experience will contribute to their academic research.

(ii) We are dealing with the financial issue on two fronts. First, for the renewal, we are reconfiguring the partner university support so that it provides the bulk of the money for graduate students; thus, even if there is significant local graduate student involvement (as we certainly want), it will primarily be funded by the local partner universities.

Indeed, there has been a limited budget for senior researchers; we have managed to involve many leading scholars in SAMSI programs through aggressive negotiation with their university administrations and judicious use of University Fellow positions, but it would be highly desirable to be able to involve a few key individuals without such constraints; this issue will also be addressed in the renewal proposal.

(iii) The enthusiasm for participation in SAMSI programs has caused space limitations to become an issue. The NISS building expansion planned for the renewal is the potential answer to this problem.

(iv) Most visitors are officially appointed through one of the three universities, and hence do have access to university facilities. Of course, these facilities are not within walking distance of SAMSI, but there are also considerable advantages to being centrally located between the three universities in Research Triangle Park. SAMSI is close not only to the universities but also to the many significant research organizations in RTP with very significant interests in the uses of applied mathematics and statistics.

(v) External development is indeed a difficult issue for an institute such as SAMSI, which seeks to be as broad as possible; most development opportunities are for specific research areas. Also, as a very young institute, we have a limited alumni base. As we mature, we will carefully consider the opportunities for development.

6. People and Broadening

Committee Report: The support for the second year of the post-doc may come with additional duties that detract from the post-doc's scientific development. Recommendations:

- Faculty and graduate student working group participation should be extended to the national community in some form.
- If they are not already doing so, the selection of post-docs should involve more input from the program leaders. This selection consideration should also extend to the long term visitors.
- Give some thought to the funding source for the second year for the postdoc with the objective of proposing a long-term model for a nationally oriented institute.

SAMSI Response:
(i) As discussed above in Section C, participation in working groups has been made possible by all non-local faculty and graduate students (as well as the many others in government and industry that engage in SAMSI programs).

(ii) Program leaders are given access to all postdoc applications and their opinions are very much sought in the selection process; indeed, if a program leader is negative about a candidate, the candidate is not appointed. This is also the case for long-term visitors that apply independently of the program leaders (except when we feel that a candidate is trying to move into a new area, and has the potential to do so). We are continually working on mechanisms (beyond simply giving the program leaders access to all the information) to involve the program leaders in these decisions.

(iii) As programs are no longer being led by local scientists, it is indeed somewhat more problematical to arrange for second year funding. We do continue to encourage second year funding by the (non-local) program leaders; we also contact other national organizations and the Affiliates to try to arrange second year funding. For instance, both last year and next year a postdoc is co-supported by the National Center for Atmospheric Research, and next year one of the SAMSI postdocs will be receiving second year support with a non-local program leader. Finally, even when no funding is identified in advance, we do make two-year appointments at SAMSI alone; next year we have two such appointees.

7. Education and Outreach

Committee Report: While the SAMSI Directorate should be commended for their tremendous efforts in maintaining and growing the E&O activity, the current E&O effort seems to be heavily dependent on the leadership and the energy of a tight-knit group of individuals. Further, concerns were expressed that the E&O committee might be too regional in scope. Recommendations:

- Need longitudinal tracking of program impact on career/major choices of students
- Broaden faculty participation in E&O
- Broaden the representation on the E&O committee to ensure that SAMSI's E&O programs will have a national scope consistent with the charter of SAMSI as a national institute.

SAMSI Response:
(i) Tracking of program impact on students in E&O workshops should greatly improve once we have the new data bases in place.

(ii) The scope of Education and Outreach activities is being significantly broadened through several mechanisms. First, other members of the directorate are becoming more involved in E&O activities. The CAARMS conference has Chris Jones as the Directorate Liaison. Under the plan for the renewal, the Deputy Director will potentially be in charge of several E&O activities, such as the 2-day undergraduate workshops.

Secondly, we are increasingly obtaining non-local faculty to serve as mentors at the week-long undergraduate and graduate workshops. For instance, a Meredith College faculty member, Cammey Cole, currently serves as a mentor for the undergraduate workshop. We will expand this model to include multiple non-local mentors for each workshop. This will provide additional mentoring for participants and provide a conduit for disseminating both the workshop concepts and scientific content back to the respective home institutions.

In the present framework, SAMSI graduate students and postdocs play a fundamental role in the 2-day outreach workshops. To broaden the scope of these workshops and dissemination of associated material, each workshop will additionally be presented at a non-local institution by a selected subset of the graduate students, postdocs, and program leaders. We will coordinate the non-local workshop days with members of the national Education and Outreach Committee and faculty at the institutions.

(iii) The Education and Outreach Committee is being reconfigured to strengthen the national base. The national committee will be charged with the task of proposing new E&O activities, providing information to potential participants, and disseminating information regarding the SAMSI E&O Program to the national community.

8. Data and Evaluation

DMS Letter: A formal and efficient data and evaluation mechanism is urgently needed to measure and document the impact of SAMSI, locally, nationally, and internationally.

Committee Report: The Governing Board, as it is presently constituted (with deans/administrators from the three participating institutions), is not an appropriate body to conduct the annual evaluation of SAMSI as a national institute. SAMSI's existing database appears to be inadequate in assessing the national impact of the institute. However, they were able to retrieve much of the information that was asked by the SVT in short order. There does not seem to be any objective mechanism in place for measuring the national impact of SAMSI. Recommendations:

- Keep track of graduate students as a separate subgroup in the grouping of "new researchers and graduate students" (NRGs).
- Measure and document national/international impact in as many ways as possible. For example:
    - Develop "nuggets" to show impact of SAMSI programs and how SAMSI has added value as an institute.
    - Show specific examples of how SAMSI has created timely synergy among the statistical and mathematical sciences and other disciplinary areas of' application which in turn has led to advances that are otherwise not possible.
    - Develop a newsletter to keep contact with participants/visitors.
    - Better organize the statistics showing disciplinary background of participants in order to reveal participation and synergetic interactions between statistical and mathematical scientists and other disciplinary scientists.
    - Follow up formally with participants one, two, or three years after their visit/participation to gauge impact of SAMSI programs on their work, so that effectiveness of the various programs can be traced, e.g., how SAMSI has changed the research direction of participants.
    - Collect statistics on papers published resulting from participation in SAMSI programs to document collaborations that are both interdisciplinary and involve multiple institutions.

SAMSI Response:
(i) The Governing Board has been reconstituted, as described earlier, to become a more appropriate body for evaluation.

(ii) The SAMSI data bases are being upgraded including both expansion and integration, based on adaptation of the IMA Data base. Indeed, the individual who designed and constructed the IMA data base is now in the Triangle area and will be assisting with the adaptation and installation of the IMA data base to meet SAMSI requirements. Among other advantages, this will automate maintenance of contact information and tracking of career development of SAMSI Postdocs, new researchers and students. It will also provide immediate access to information about participant composition for workshops, programs and other activities. Planned expansions will include compilations of publications supported by SAMSI, resulting from continuations of SAMSI research or SAMSI collaborations or deriving from ideas from ideas first formulated at SAMSI.

(iii) Assessment of SAMSI's national impact is being undertaken at the individual level through direct surveying of participants (from Program Leaders through Students) with special attention to accurately capturing the impact of research results (publications, follow-on research grants) and the indirect impact of results (research topics for subsequent students, addition/modification of course content, or applied research protocols and practices), as well as, the dissemination of information about SAMSI accomplishments, activities and opportunities. The 2005-06 Evaluation Survey successfully reached all SAMSI 2004-05 Postdocs; replies are in Appendix A. With telephone follow-up, it also reached about half of previous Program Leaders and Long-time Visitors and a similar proportion of students; these replies are also in that Appendix. Lists of publications from SAMSI research and research deriving from work while at SAMSI gathered from the 2005-6 Evaluation Survey also appear in Appendix A.

With the updated data base of contact information, SAMSI is inaugurating a Newsletter for the SAMSI Family nationwide (and wider). We anticipate a comparably favorable response to that for the SAAG (SAMSI At A Glance) that is published weekly via email for all SAMSI staff and visitors and local subscribers. (A monthly version of SAAG is being contemplated for the SAMSI Family outside the local area.)

Examples of scientific nuggets from SAMSI programs appear in Section D; and personal anecdotes of SAMSI impact also appear earlier in that Section.

F. Evaluation by the SAMSI Governing Board

(Bruce Carney, George Casella, Thomas Manteuffel, Vijay Nair, John Simon, Daniel Solomon – Chair)

The Governing Board provides broad oversight for the Institute's administration, finances, and evaluation, and for relationships among the partnering institutions. As part of the annual evaluation, the Governing Board has elected to address four broad questions. That evaluation follows:

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

The synthesis of applied mathematics, statistics and the disciplinary sciences is a central tenet of the SAMSI mission. There are notable examples of this synthesis in specific SAMSI programs, but the extent varies substantially across the full portfolio.

The data assimilation program provided a context in which statisticians and applied mathematicians worked closely together, and the mutual influence on each other's thinking was extremely productive. This will be reflected in an upcoming special issue of the journal Physica D, which is a leading applied mathematics journal for research on nonlinear phenomena. There will be a paper in that issue by Zoltan Toth of the National Center for Environmental Prediction. Toth was a speaker at both the opening and closing workshops of the program and the shift in his thinking during that time-span is dramatic. Indeed, it is clear in the paper that he has been greatly influenced by the "dynamical-statistical" view of data assimilation that formed the basis for much of the development during the program. The shift is from a position, predominant in weather prediction, that model error is overshadowed by the chaotic effect of a small change in initial conditions to the idea that drift of the state of the system is primarily due to errors in the modeling. This can begin to be dealt with by a strategy that combines a Bayesian view of state estimation with a dynamics view of ensemble formation.

The National Defense and Homeland Security (NDHS) program had other examples of this synthesis. Indeed, the agricultural systems group, comprised of mathematicians, statisticians and veterinary science faculty and students, introduced a new paradigm for characterizing the richness of mathematical models based on the size of stochastic disturbance terms.

The synergy between statistics, mathematics and domain sciences has had a significant impact on participants and postdocs involved with the programs. For instance, Amit Apte was a postdoc in the Data Assimilation program who arrived at SAMSI with a degree in dynamical systems with physical applications. During the program, he developed an expertise in Bayesian statistical techniques and is now investigating techniques involving infinite-dimensional Langevin sampling.

The two 2-day undergraduate workshops associated with the Financial Mathematics, Statistics and Econometrics (FMSE) and NDHS programs had students from both statistics and mathematics participating in lectures and interactive activities having both statistical and applied mathematical components. Likewise the weeklong undergraduate workshop focused on topics that have both stochastic and deterministic components, and students from both backgrounds were chosen to attend. These opportunities provided students with an in-depth exposure to the synergy between the two disciplines.

The 10-Day graduate workshop involved students from both statistics and mathematics who are investigating substantive problems having both stochastic and deterministic components. This is exemplified by the project on "Optimal Mission Planning," which was presented by Greg Angelides, a scientist from MIT Lincoln Laboratory, and involved both deterministic optimization techniques and stochastic models employing random walks. As noted in the Executive Summary, this work was later accepted as a refereed proceedings paper and the corresponding poster presentation won a 2nd place "Best Poster Award."

2) Is the impact of SAMSI on science and human resources growing?

Section D of the Executive Summary highlights some of the developments in science and education that have occurred through SAMSI's programs. Their impact is potentially great on the participants as it offers a redirection of research effort for senior participants and a formative experience for postdocs and other junior scientists. The impact is measured by grants obtained, based on work at SAMSI, and papers stemming from the working groups.

It is clear that SAMSI has "hit a nerve" in the mathematical, and, more broadly, scientific communities. The need to meld statistical, data-based, thinking with the computational models, the purview of applied mathematics, is evident, as reflected in the earlier comments about the Data Assimilation program. Another indication of scientific impact of this program was the comments of … (a SAMSI University Fellow) from Oxford University and the London School of Economics:

> My time in SAMSI has already proven of immense value to our research work at CATS [Committee on Applied and Theoretical Statistics, NAS/NRC] and influenced our interactions with ECMWF [European Center for Medium Range Weather Forecasting] and other operational weather centres. The SAMSI experience was particularly valuable due to the unique environment that allowed me to view data assimilation in the biggest pictures; being at any particular research university or operational centre naturally leads to a focus on local expertise and interests, while the SAMSI program stimulated vibrant interaction between researchers from a wide range of home institutions. While I have seen this attempted on relative short time scales (a week or so) at locations like the Newton Institute, I have never experienced it done as successfully (in terms of

> people invited, intensity of discussion, support for research and support for domestic needs) as it was sustained over the duration of this SAMSI program."

The report on the comp bio program portrays a striking picture of the way that statistical thinking was used to advance understanding in critical issues of the immune system with implications for the design of vaccines. The program on latent variables showed how the introduction of modern statistical ideas and thinking into the social sciences can lead to breakthroughs in sociological modeling and the selection of models.

The lists of refereed publications associated with SAMSI programs (see Section I.G. of the full report) provide another measure of evidence of impact on the mathematical and disciplinary sciences.

SAMSI continues its strong commitment to the development of human resources in the mathematical sciences. Its impacts are discussed in Sections I.B, I.C and I.H (which highlights diversity) of the full report. Indeed some of the effects on early career researchers are described in item 1) above.

Overall, the impact on postdoctoral fellows is reflected by the extremely positive responses coming from postdocs on their experience at SAMSI. For instance, Jane Zaviscka, now at the University of Arizona, was a postdoc in the latent variables program and said of her involvement in the program: "I am substantially revising my dissertation, which I intend to publish as a book, to incorporate latent variable methods. I give much more attention to measurement error in all my work. The SAMSI affiliation also helped me enjoy considerable success on the job market last year."

Other examples of such positive impacts on senior as well as junior scientists, including statements from individual participants, are cited in Section D of the Executive Summary and in the body of the full report.

In the grant support coming out of the SAMSI, over 50% involve participants from beyond the triangle area. The Undergraduate workshop has attracted participants from 30 of the 50 states.

Participation by minorities and women in SAMSI programs has remained stable and high (35-36%) since SAMSI's very first programs. The proportion of young researchers (students and new researchers) has risen from 44% to 50%, with the expectation that the final 2005-6 proportion will be slightly above the 51% of 2004-5.

3) Is the national recognition and respect for SAMSI growing?

SAMSI programs this year have about 75% non-local leadership and future programs have entirely non-local leadership. That leading scientists from outside the local area – e.g., Iain Johnstone (NAS), Craig Tracy and Helene Massam for the High Dimensional

Inference and Random Matrices Program – are eager to lead SAMSI programs is a strong indicator of SAMSI's national stature.

The geographic distribution of participants is summarized in Section E of the Executive Summary, while the detailed participant lists for concluded programs provide ample evidence of the national and international draw of SAMSI activities. Other evidence of SAMSI's reach is in the offers of partnerships with other organizations including the National Center for Atmospheric Research, the Centre de Recherches Mathématiques in Montreal, the (Canadian) National Program on Complex Data Structures, DIMACS and the Center for Astrostatistics at Penn State.

Applications to the postdoctoral program were significantly up again this year – totaling 99 applicants – in spite of the fact that statistics Ph.D. students can typically find immediate tenure track jobs and applied mathematics postdoctoral candidates have many other options. Even more impressive is that the top eight identified candidates all accepted SAMSI postdoc offers (including one who turned down multiple tenure-track offers in order to take a postdoc at SAMSI with a follow-on second and third years at Stanford with the program leader, Iain Johnstone). The directorate observed that the top statistics and probability candidates have been hearing of the considerable benefits of going through a SAMSI postdoctoral experience, while the top applied mathematics candidates are being attracted by a growing recognition of the importance of integrating applied mathematics and statistics.

Long-term visitors to SAMSI also seem to be sharply up, as are applications for visits by outside graduate students; it is impressive that SAMSI will next Fall host six visiting graduate students. The national recognition and respect is also indicated by the excitement that seems to be resulting from the possibility for individuals who are not resident at SAMSI to participate in SAMSI working groups; 8 individuals are currently participating remotely in the NDHS working groups, and 17 in the Astrostatistics working groups. This is in spite of the fact that the telecommunications facilities at SAMSI are not yet fully developed to optimize such involvement.

This continually increasing national presence of SAMSI is also evidenced by the workshops. There appears to be very strong interest in holding SAMSI workshops in other locations nationally, as reflected in the fact that 1/3 of this year's workshops were held at other locations around the country.

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

The directorate model continues to serve SAMSI very well, and transitions in the directorate have gone smoothly, in particular the successful incorporation of new members: Chris Jones of UNC, Nell Sedransk of NISS and Ralph Smith of NCSU. This past year, Alan Karr and Sedransk shared the NISS associate directorship, much to the advantage of SAMSI. Over the next six months, Sedransk will assume the major role, with Karr continuing in an advisory capacity as well as focusing on specific issues such as the building expansion. At this point only one-plus of the four original members of the SAMSI directorate (Berger) remains, but the directorate is arguably functioning more effectively and efficiently than ever.

The directorate is successfully managing the transition to complete national programming, through an evolving delineation of the role of the directorate in facilitating development of SAMSI programs. As SAMSI has moved exclusively to national leadership of programs, the need for strong directorate support has become clear. This support for program development includes contacting program participants (ranging from senior visitors to postdocs to faculty releases) that are suggested by the program leaders; explaining key elements of SAMSI programs, such as working groups, to the leaders and helping in their formulation; and, of course, organization and planning of workshops.

The directorate has been very responsive to suggestions for improvement of SAMSI that arose out of the Third-Year Review. For instance, two additional external members have been added to the Governing Board, appointed by the ASA and SIAM presidents, so that the Governing Board now has extensive national presence; a Scientific Liaison Board of advisory disciplinary scientists is being created to augment the national leadership in the National Advisory Council; accessibility of outside graduate students to SAMSI programs is being greatly enhanced; and, under the renewal, there would be a position for a Deputy Director, which would be reserved for two-year rotating visitors, bringing a continual national representation to the Directorate.

The directorate has also been aggressive in seeking technological solutions to issues, for instance in the continuing integration of technology to enhance working group communication with non-local individuals, and in the planned incorporation of more sophisticated data bases (from the IMA) to enhance the capabilities for planning and evaluation.

The Governing Board Chair and the SAMSI Director have a biweekly telephone conference at which administrative and personnel matters are regularly discussed and issues addressed where they have arisen. There is also excellent cooperation among the partner universities and NISS to ensure that obligations are met and that SAMSI continues to flourish.

Table of Contents

I. Annual Progress Report

The previous annual progress report was complete in all details only through April, 2005. Hence, we also report activities in Year 3 programs that occurred subsequently and were not itemized in the report. These Year 3 programs were *Latent Variable Models in Social Sciences*, *Computational Biology of Infectious Disease*, and *Data Assimilation in Geophysical Systems;* their final reports are in Appendices B, C, and D, respectively.

A. Program Personnel

1. Program and Activity Organizers

Program Organizers

Program	Name	Affiliation	Field
National Defense and Homeland Security	Alicia Carriquiry	Iowa State U	Statistics
	James Crowley	SIAM	Mathematics
2005-06 SAMSI Program	Lawrence Cox (Co-Chair)	Cen for Disease Control	Statistics
	Philip Hanlon	U of Michigan	Mathematics
	Alan F. Karr	NISS	Statistics
	Sallie Keller-McNulty	Los Alamos Nat Lab	Statistics
	Jon Kettenring	Drew U	Statistics
	Carey Priebe	Johns Hopkins U	Mathematics & Statistics
	Fred Roberts	DIMACS	Mathematics and CS
	Stephen Robinson	U of Wisconsin	Industrial Engineering
	Nell Sedransk (Co-Chair)	NISS	Statistics
	Edward Wegman	George Mason U	Computational Statistics
Financial Mathematics, Statistics and	Marco Avellaneda	New York U	Mathmatical Sciences
Econometrics	Ole E. Barndorff-Nielsen	Aarhus U	Math Physics & Stochastics
	Rene Carmona	Princeton U	Op Research & Fin Eng
2005-06 SAMSI Program	Darrell Duffie	Stanford U	Business
	Nicole ElKaroui	Ecole Polytechnique	Applied Math
	Jean-Pierre Fouque (Co-Chair)	North Carolina State U	Mathematics
	Eric Ghysels (Co-Chair)	U of North Carolina	Economics
	Lars Peter Hansen	U of Chicago	Economics
	Robert Jarrow	Cornell U	Business
	John Lehoczky	Carnegie Mellon U	Applied Probability
	Ronnie Sircar	Princeton U	Op Research & Fin Eng
	Ralph Smith	North Carolina State U	Applied Mathematics
	Ruey Tsay	U of Chicago	Business
	Thaleia Zariphopoulou	U of Texas-Austin	Math & Op Management

Astrostatistics *2006 SAMSI Program*	G. Jogesh Babu (Chair) Jim Berger Peter Bickel Alanna Connors Eric Feigelson Tom Loredo Louis Lyons Arlie Petters Donald Richards Larry Wasserman	Penn State U SAMSI Berkeley Eureka Scientific Penn State U Cornell University Oxford University Duke U Penn State U Carnegie Mellon U	Statistics Statistics Statistics Astronomy Astronomy Astronomy Physics Mathematics Statistics Statistics
Education & Outreach Program	Ralph Smith (Chair) Johnny Houston Rachel Levy J. Blair Lyttle Negash Medhin Daniel Teague Wei Feng	North Carolina State U Elizabeth City State North Carolina State U Enloe HS, Raleigh North Carolina State U NC Sch Math & Sci UNC-Wilmington	Applied Math Math and CS Mathematics Statistics Mathematics Mathematics Math and Stat
Latent Variable Models in the Social Sciences *2004-05 SAMSI Program*	Kenneth Bollen (Chair) James Heckman Alan Karr Susan Murphy	U of North Carolina U of Chicago NISS U of Michigan	Sociology Economics Statistics Statistics
Genomes to Global Health: Computational Biology of Infectious Disease *2004-05 SAMSI Program*	Roy Anderson Rustom Antia Carl Bergstrom Arturo Casadevall Carlos Castillo-Chavez Lindsay Cowell Sunetra Gupta Tom Kepler (Chair) Denise Kirschner Jun Liu Alan Perelson Man-Wah Tan	Imperial College Emory University U of Washington Einstein College of Med Arizona State U Duke U Oxford U Duke U U of Michigan Harvard U Los Alamos Nat Lab Stanford U	Disease Epidemiology Biology Biology Microbiology & Immunology Math and Stat Biostats & Bioinformatics Mathematical Biology Biostats & Bioinformatics Microbiology & Immunology Biostatistics Mathematical Biology Genetics
Data Assimilation for Geophysical Sciences *2005 SAMSI Program*	Jeffrey Anderson Mark Berliner Andrew Bennett Craig Bishop Montserrat Fuentes Kayo Ide Christopher Jones (Chair) Eugenia Kalnay Susan Lozier Authur Mariano Ian McKeague Robert Miller Douglas Nychka Juan Restrepo	NCAR Ohio State U Oregon State U Navy Research Lab North Carolina State U UCLA U of North Carolina U of Maryland Duke U U of Miami Florida State U Oregon State U NCAR U of Arizona	Data Assimilation Statistics Atmospheric & Ocean Sci Marine Meterology Statistics Atmospheric & Ocean Sci Mathematics Meterology Earth & Ocean Sciences Marine & Atmospheric Sci Biostatistics Atmospheric & Ocean Sci Geophysical Statistics Mathematics

	Leonard Smith Chris Synder Istvan Szunyogh Olivier Talagrand Keith Thompson Zoltan Toth Francisco Werner Carl Wunsch	Oxford U NCAR U of Maryland Ecole Normale Superier Dalhousie U NCEP U of North Carolina MIT	Industrial & Applied Math Data Assimilation Initiative Physical Science & Tech Meterology & Oceanography Math & Stat Environmental Modeling Marine Sciences Earth, Atmos, Planetary Sci
Data Mining and Machine Learning *2003-2004 SAMSI Program*	David Banks (Co-Chair) Mary Ellen Bock Jerome Friedman Alan F. Karr (Co-Chair) David Madigan William DuMouchel Warren Sarle	Duke U Purdue U, NAC Stanford U NISS Rutgers U AT&T SAS Institute	Statistics Statistics Statistics Statistics CS and Stat Statistics CS and Stat

Activity Organizers

Activity	**Name**
National Defense & Homeland Security (NDHS) Opening Workshop -- *September 11-14, 2005*	James Crowley, Lawrence Cox, Sallie Keller-McNulty, Jon Kettenring, Nell Sedransk
NDHS Mid-Year Meeting on Anomaly Detection (at NCHS) -- *February 3, 2006*	Doug Kelly, Joe Fred Gonzalez, Jr.
NDHS Mid-Year Meeting on Social Networks (at CMU) -- *March 3, 2006*	David Banks, Stephen Fienberg
NDHS Mid-Year Meeting on Data Confidentiality (at NCHS) – March *13, 2006*	Joe Fred Gonzalez, Jr.
NDHS Transition Workshop – *TBD*	James Crowley, Lawrence Cox, Sallie Keller-McNulty, Jon Kettenring, Nell Sedransk
Financial Mathematics Workshop on Stochastic Modeling in Financial Mathematics (U of Montreal) -- *June 1-5, 2005*	Ronnie Sircar, J.P. Fouque
Financial Mathematics, Statistics & Econometrics (FMSE) Opening Workshop -- *September 18-21, 2005*	Eric Ghysels, J.P. Fouque
FMSE Workshop on Credit Risk -- *October 31-November 2, 2005*	Ronnie Sircar, J.P. Fouque

FMSE Workshop on Model Uncertainty -- *January 27, 2006*	Eric Ghysels, Eric Renault
FMSE Transition Workshop -- *February 27-28, 2006*	Eric Ghysels, J.P. Fouque
Astrostatistics Planning Workshop (at NASA) -- *July 14-16, 2005*	J Jogesh Babu, Tom Loredo, Bill Jefferys, Jeff Scargle
Astrostatistics Opening Workshop & Tutorials -- *January 18-25, 2006*	Jogesh Babu, Tom Loredo, Bill Jefferys, Eric Feigelson
Astrostatistics Research Session on Stellar Evolution -- *February 20-24, 2006*	Bill Jefferys
Astrostatistics Research Session on Particle Physics -- *March 6-9, 2006*	Louis Lyons
Astrostatistics Transition Workshop (in conjunction with SCMA VI at Penn State) -- *June 12-15, 2006*	Jogesh Babu, Eric Feigelson
Astrostatistics Transition Workshop (In conjunction with PHYSTAT at BIRS) -- *July 15-20, 2006*	James Linnemann, Louis Lyons, Nancy Reid
MAA Prep Workshop: Mathematics Meets Biology -- *May 25-28, 2005*	Azmy Ackleh
SAMSI Undergraduate Workshop -- *May 30-June 3, 2005*	Ralph Smith, H.T. Banks
Data Assimilation Summer School: Fusing Geophysical Models (at NCAR) -- *June 13-17, 2005*	Doug Nychka, Jeff Anderson, Chris Snyder, Chris Jones, Kayo Ide
SAMSI-CRSC Industrial Mathematical & Statistical Modeling Workshop for Graduates -- *July 24-August 1, 2005*	Ralph Smith
Undergraduate Two-Day Workshop on Financial Mathematics, Statistics & Econometrics -- *November 18-19, 2005*	Ralph Smith
Undergraduate Two-Day Workshop on National Defense & Homeland Security -- *March 3-4, 2006*	Ralph Smith
SAMSI Undergraduate Workshop -- *May 22-26, 2006*	Ralph Smith

Summer School on the Design and Analysis of Computer Experiments (at IRMACS, Simon Fraser U) -- *August 11-16, 2006*	Derek Bingham, David Higdon, Jerry Sacks, Will Welch; Susie Bayarri
Latent Variables in the Social Sciences GLAMM Seminars -- *April 13-15, 2005*	Ken Bollen
Latent Variables in the Social Sciences Working Group Year-End Summaries -- *May 19, 2005*	Ken Bollen
Latent Variables in the Social Sciences Transition Workshop -- *November 10-11, 2005*	Ken Bollen
Computational Biology Transition Workshop -- *May 22-24, 2005*	Tom Kepler, Lindsey Cowell
Data Assimilation Mini-Workshop on Bridging Statistical Approaches -- *June 27, 2005*	Amarjit Budhiraja, Montserrat Fuentes
Data Assimilation Transition Workshop -- *October 5, 2005*	Chris Jones, Kayo Ide
Multiplicity and Reproducibility in Scientific Studies Opening Workshop -- *July 10-12, 2006*	Peter Mueller, Juliet Shaffer, Peter Westfall, Stanley Young
Multiplicity and Reproducibility in Scientific Studies Closing Workshop -- *July 27-28, 2006*	Peter Mueller, Juliet Shaffer, Peter Westfall, Stanley Young
Development, Assessment and Utilization of Complex Computer Models (CompMod) Opening Workshop & Tutorials -- *September 10-13, 2006*	Susie Bayarri, Bruce Pitman, Peter Reichert, Tom Santner, Darren Wilkinson
High Dimensional Inference and Random Matrices (HDIRM) Opening Workshop & Tutorials -- *September 17-20, 2006*	Iain Johnstone, Peter Bickel, Helene Massam, Douglas Nychka, Craig Tracy
Data Mining & Machine Learning Technology Transfer Short Course -- *July 25-29, 2005*	David Banks
NISS-SAMSI Hot Topics Workshop: Collaborations in the Mathematical Geosciences -- *October 6-7, 2005*	Alan Karr
12th Annual Conference for African American Researchers in Mathematical Sciences (CAARMS) -- *June 20-23, 2006*	William Massey, Idris Assani, Patrick Eberlein
Hot Topics Workshop: Random Graphs and Stochastic Computation -- *June 13-14, 2005*	Mike West, Merlise Clyde

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:

D – *Distinguished Lecturer* for the program.
F – *Faculty Fellow***,** defined as an individual from a partner university of SAMSI who is accorded release time for participation in the SAMSI program; the cost-sharing value of this release time is indicated.
FA – *Faculty Associate***,** defined as an individual from a partner university of SAMSI who leads a working group in a SAMSI program but receives no financial support.
G – *Graduate Student*, receiving a research assistantship, in the indicated amount, from SAMSI
N – *New Researcher***,** receiving the indicated support (salary and fringe benefits) from SAMSI
P – *Postdoctoral Fellow***,** receiving the indicated support (salary and fringe benefits) from SAMSI
PA – *Postdoctoral Associate***,** receiving the indicted support (salary and fringe benefits or reimbursement of expenses) from SAMSI
T – *Targeted Expert*, an individual with particular expertise that is felt to be needed for progress in key elements of program research. Such individuals are brought in for shorter intervals of time, for transference of expertise to the program participants.
U – *University Fellow*, a key program participant, visiting for a semester or year, whose primary support is via indicated cost-sharing from a partner university.
V – *Core Visitor*, an individual from outside the Triangle who plays a major role in the program activities, by either a lengthy visit to the program or repeated visits involving ongoing program research.
WG – *Working Group Participant*, an individual, local or non-local, who has significant long-term participation in a working group and does not fit into a category above.

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.

I. National Defense and Homeland Security

Last Name	**First Name**	**Gender**	**Affiliation**	**Department**	**Status**
Airoldi	Edoardo	M	Carnegie Mellon U	Computer Science	V
Apte	Amit	M	SAMSI & U of North Carolina	Mathematics	P
Bai	Ping	F	U of North Carolina	Statistics	G

Banks	David	M	Duke U	Statistics	FA
Banks	H.T.	M	North Carolina State U	CRSC	FA
Bayarri	M.J.	F	U of Valencia	Statistics	V
Blei	David	M	Princeton U	Computer Science	WG
Burkom	Howard	M	Johns Hopkins U	Applied Physics Laboratory	WG
Chipman	Hugh	M	Acadia U	Mathematics and Statistics	T
Clarke	Bertrand	M	U of British Columbia	Statistics	V
Cortez	Ricardo	M	Tulane U	Mathematics	V
Cox	Lawrence	M	National Center for Health Statistics	Statistics	V
Datta	Gauri	M	U of Georgia	Statistics	U
Dediu	Sava	M	SAMSI & North Carolina State U	Mathematics	P
Denogean	Lisa	F	SAMSI & North Carolina State U	Mathematics	P
Dickey	David	M	North Carolina State U	Statistics	F
Fienberg	Stephen	M	Carnegie Mellon U	Statistics	V
Ghosh	Joyee	F	Duke U	Statistics	G
Gonzalez	Joe Fred	M	National Center for Health Statistics	Statistics	WG
Govan	Anjela	F	North Carolina State U	Mathematics	G
Heyward	Shenek	F	North Carolina State U	Statistics	G
Hoff	Peter	M	U of Washington	Statistics and Biostatistics	T
Hong	Chung-Chien	M	North Carolina State U	Mathematics	WG
Katzoff	Myron	M	National Center for Health Statistics	Statistics	WG

Kelly	Douglas	M	U of North Carolina	Statistics	F
Kim	Se Hee	F	U of North Carolina	Biostatistics	G
Lacey	Michelle	F	Tulane U	Mathematics	V
Last	Michael	M	NISS	Statistics	PA
Lynch	James	M	U of South Carolina	Statistics	V
Medhin	Negash	M	North Carolina State U	Mathematics	WG
Mei	Yajun	M	Fred Hutchinson Cancer Center	Statistics	N
Mitra	Robin	F	Duke U	Statistics	WG
Nguyen	Hoan	F	North Carolina State U	Mathematics	PA
Norminton	Ted	M	Carleton University	Statistics	T
Oganyan	Anna	F	NISS	Statistics	PA
Olufsen	Mette	F	North Carolina State U	Mathematics	F
Ozonoff	Al	M	Boston U	Biostatistics	T
Park	Cheolwoo	M	U of Georgia	Statistics	T
Qaqish	Bahjat	M	U of North Carolina	Biostatistics	F
Reiter	Jerome	M	Duke U	Statistics	FA
Rempala	Greg	M	U of Louisville	Mathematics	V
Resnick	Sid	M	Cornell U	Operations Research & Industrial Eng	V
Rolka	Henry	M	CDC	Statistics	WG
Shmueli	Galit	F	U of Maryland	Decision & Info Technlies, Smith Sch of Business	T
Slenning	Barrett	M	North Carolina State U	Veterinary Medicine	WG

Storlie	Curtis	M	SAMSI & North Carolina State U	Statistics	PA
Vance	Eric	M	Duke U	Statistics	WG
Vera	Francisco	M	NISS	Statistics	P
Welsch	Roy	M	MIT	Sch of Management	V
Woo	Mi-Ja	F	NISS	Statistics	PA

II. Financial Mathematics, Statistics and Econometrics

Last Name	**First Name**	**Gender**	**Affiliation**	**Department**	**Status**
de Almeida	Caio	M	IBMEC, Rio de Janeiro	Business School	T
Andreou	Elena	F	University of Cyprus	Economics	V
Bloomfield	Peter	M	North Carolina State U	Statistics	F
Chabi-Yo	Fousseni	M	Bank of Canada	Finance	V
Cont	Rama	M	Ecole Polytechnic	Mathematics	V
Engle	Robert	M	New York U, Stern School of Business	Economics	D
Fan	Yanqin	F	Vanderbilt U	Economics	V
Fouque	Jean-Pierre	M	North Carolina State U	Mathematics	F
Ghysels	Eric	M	U of North Carolina	Economics	F
Hochberg	Kenneth	M	Bar-Ilan U	Mathematics	T
Houdre	Christian	M	Georgia Institute of Technology	Mathematics	T
Hyde	John	M	Duke U	Mathematics	G

Ilhan	Aytac	F	U of Oxford	Mathematics	V
Larsen	Kasper	M	Carnegie Mellon U	Mathematical Sciences	T
Mattingly	Jonathan	M	Duke U	Mathematics	F
Pang	Tao	M	North Carolina State U	Mathematics	F
Pemy	Moustapha	M	SAMSI & North Carolina State U	Mathematics	P
Rodriguez	Jesus	M	SAMSI & North Carolina State U	Mathematics	P
Sinko	Arthur	M	U of North Carolina	Economics	G
Sircar	Ronnie	M	Princeton U	Operations Research and Financial Eng	U
Sloan	Jennifer	F	North Carolina State U	Statistics	G
Solna	Knut	M	U of California, Irvine	Mathematics	T
Tu	Chong	M	Duke U	Statistics	G
Vestal	Doug	M	North Carolina State U	Mathematics	G
Werker	Bas	M	Tilburg U	Economics	U
Wu	Yichao	M	U of North Carolina	Statistics	G
Xu	Mingxin	F	U of North Carolina – Charlotte	Mathematics	N
Zitkovic	Gordan	M	U of Texas	Mathematics	T

III. Astrostatistics

Last Name	First Name	Gender	Affiliation	Department	Status
Arnaud	Keith	M	NASA	Astronomy	T
Babu	Jogesh	M	Pennsylvania State U	Statistics	U

Bhat	Pushpalatha	F	Fermilab	Physics	T
Bullard	Floyd	M	Duke U	Statistics	G
Chernoff	David	M	Cornell U	Astronomy	V
Clyde	Merlise	F	Duke U	Statistics	F
Cranmer	Kyle	M	Brookhaven National Lab	Physics	V
De Gennaro	Steven	M	U of Texas	Astronomy	V
De La Cruz	Pablo	M	U of Valencia	Astronomy	V
Dose	Volker	M	Max Planck Institute	Physics	V
Feigelson	Eric	M	Pennsylvania State U	Astronomy & Astrophysics	T
Ford	Eric	M	U of California, Berkeley	Astronomy	V
Ghosh	Jayanta	M	Purdue U	Statistics	V
Gregory	Phil	M	U of British Columbia	Astronomy	V
Hartigan	John	M	Yale University	Statistics	V
Heinrich	Joel	M	Fermilab	Physics	V
Hendry	Martin	M	U of Glasgow	Astronomy	T
Jang	Woncheol	M	Duke U	Statistics	FA
Jeffery	Elizabeth	F	U of Texas	Astronomy	V
Jefferys	Bill	M	U of Vermont	Astronomy	T
Lee	Hyunsook	M	Pennsylvania State U	Statistics	GA
Lee	Jaeyong	M	Seoul National U	Statistics	V
Linnemann	James	M	Michigan State U	Physics and Astronomy	T

Loh	Ji Meng	M	Columbia U	Statistics	V
Loredo	Tom	M	Cornell U	Astronomy	V
Lyons	Louis	M	U of Oxford	Physics	V
Petters	Arlie	M	Duke U	Mathematics	F
Pilla	Ramani	F	Case Western Reserve U	Statistics	T
Prosper	Harrison	M	Florida State U	Physics	V
Punzi	Giovanni	M	National Institute of Nuclear Physics	Physics	V
Reid	Nancy	F	U of Toronto	Statistics	T
Robbins	Nicholas	M	Duke U	Mathematics	G
Sen	Bodhisattva	M	U of Michigan	Statistics	V
Stein	Nathan	M	U of Texas	Astronomy	V
Van Dyk	David	M	U of California Irvine	Statistics	V
Woodroofe	Michael	M	U of Michigan	Statistics	V
Zhang	Lingsong	M	U of North Carolina	Statistics	G

3. Summary of Activity Participants *

		Underrepresented Groups			
Activity	**# Participants**	**# Female**	**# African-American**	**# Hispanic**	**# New Researcher-Students**
National Defense & Homeland Security (NDHS) Opening Workshop – *September 11-14, 2005*	93	25	5	4	33
NDHS Mid-Year Meeting on Anomaly Detection (at NCHS) -- *February 3, 2006*	35	6	0	2	10

NDHS Mid-Year Meeting on Social Networks (at CMU) -- *March 3, 2006*	22	5	0	1	11
NDHS Mid-Year Meeting on Data Confidentiality (at NCHS) -- March *13, 2006*	35	8	0	3	8
Financial Mathematics Workshop on Stochastic Modeling in Financial Mathematics (U of Montreal) – *June 1-5, 2005*	Co-sponsored Meeting				
Financial Mathematics, Statistics & Econometrics (FMSE) Opening Workshop -- *September 18-21, 2005*	157	31	2	3	83
FMSE Workshop on Credit Risk -- *October 31-November 2, 2005*	36	8	2	1	21
FMSE Workshop on Model Uncertainty -- *January 27, 2006*	27	3	1	0	12
FMSE Transition Workshop -- *February 27-28, 2006*	70	14	2	1	38
Astrostatistics Planning Workshop (at NASA) -- *July 14-16, 2005*	18	4	0	0	2
Astrostatistics Opening Workshop & Tutorials -- *January 18-25, 2006*	74	21	1	6	33
Astrostatistics Research Session on Stellar Evoluation -- *February 20-24, 2006*	Informal meeting; no registration required				
Astrostatistics Research Session on Particle Physics -- *March 6-9, 2006*	Informal meeting; no registration required				
Astrostatistics Transition Workshop (in conjunction with SCMA VI at Penn State) -- *June 12-15, 2006*	to be reported in the 2006-07 Annual Report				
Astrostatistics Transition Workshop (In conjunction with PHYSTAT at BIRS) -- *July 15-20, 2006*	to be reported in the 2006-07 Annual Report				
MAA Prep Workshop: Mathematics Meets Biology -- *May 25-28, 2005*	29	13	0	2	12
SAMSI-CRSC Undergraduate Workshop -- *May 30-June 3, 2005*	36	13	0	0	34

SAMSI-CRSC Industrial Mathematical & Statistical Modeling Workshop for Graduates -- *July 25-August 2, 2005*	60	25	2	2	51
Data Assimilation Summer School: Fusing Geophysical Models (at NCAR) -- *June 13-17, 2005*	40	12	1	0	26
Undergraduate Two-Day Workshop on Financial Mathematics, Statistics & Econometrics -- *November 18-19, 2005*	31	13	3	1	30
Undergraduate Two-Day Workshop on National Defense & Homeland Security -- *March 3-4, 2006*	25	9	1	0	24
SAMSI-CRSC Undergraduate Workshop -- *May 22-26, 2006*	to be reported in the 2006-07 Annual Report				
Latent Variables in the Social Sciences GLAMM Seminars -- *April 13-15, 2005*	15	4	1	0	6
Latent Variable in the Social Sciences Working Group Year-End Summaries -- *May 19, 2005*	29	5	2	1	12
Latent Variable in the Social Sciences Transition Workshop -- *November 10-11, 2005*	75	26	1	3	39
Computational Biology Transition Workshop -- *May 22-24, 2005*	25	6	1	0	8
Data Assimilation Mini-Workshop on Bridging Statistical Approaches -- *June 27, 2005*	27	11	2	1	18
Data Assimilation Transition Workshop -- *October 5, 2005*	39	10	1	1	17
Data Mining & Machine Learning Technology Transfer Short Course -- *July 25-29, 2005*	16	5	0	1	6
12th Annual Conference for African American Researchers in Mathematical Sciences (CAARMS) -- *June 20-23, 2006*	to be reported in the 2006-07 Annual Report				
Hot Topics Workshop: Random Graphs and Stochastic Computation -- *June 13-14, 2005*	36	9	0	0	14
NISS-SAMSI Hot Topics Workshop: Collaborations in the Mathematical Geosciences -- *October 6-7, 2005*	111	23	1	3	43

* Participant lists for workshops are given in Appendix E

B. Postdoctoral Fellows

This section describes the postdoctoral fellow selection and mentoring processes at SAMSI and includes synopses of the activities of the 2004-5 and 2005-6 SAMSI Postdocs from their own perspectives with commentaries by their mentors. Section B.1 describes the SAMSI activities and strategies for effective selection and mentoring; Section B.2 contains the activity reports for SAMSI Postdocs during this grant year; and Section B.3 tracks previous Postdocs.

SAMSI 2005-6 Postdocs and Postdoctoral Associates and their mentors are presented below.

Amit Apte, (Ph.D., Physics, University of Texas at Austin)

SAMSI Programs: *Data Assimilation in Geophysical Sciences*:

- Langevin stochastic differential equations with applications to data assimilation
- Dynamics of inertial particles in fluid flows and Lagrangian data assimilation

National Defense and Homeland Security:

- Agricultural Systems working Group

Research Mentor: Chris Jones
Administrative Mentor: Chris Jones

Sava Dediu, (Ph.D., Applied Mathematics, Rensselaer Polytechnic Institute)

SAMSI Programs: *National Defense and Homeland Security*

- Agricultural Systems and Social Networks working groups
- Sensitivity of dynamical systems to convex metric space parameters

Research Mentor: H. T. Banks
Administrative Mentor: Chris Jones

Lisa Denogean, (Ph.D., Statistics, Cornell University)

SAMSI Programs: *National Defense and Homeland Security*

- Data Confidentiality and Anomaly Detection working groups
- Study of data swapping for categorical variables
- New measures of data utility and risk for data swapping
- Network modeling

Research Mentor: David Dickey
Administrative Mentor: Alan Karr

Moustapha Pemy, (Ph.D., Applied Mathematics, University of Georgia)

SAMSI Programs: *Financial Mathematics, Statistics, and Econometrics*

- Computational Issues, Portfolio Management, and Model Uncertainty working Groups
- Stochastic functional equations

- Theory of viscosity solution

Research Mentor: Tao Pang
Administrative Mentor: Ralph Smith

Jesus Rodriguez, (Ph.D., Probability, Cornell University)

SAMSI Programs: *Financial Mathematics, Statistics, and Econometrics*

- Credit Risk and Portfolio Management working groups
- Multiple default derivatives
- Stochastic portfolio theory
- Pricing issues in energy markets

Research Mentor: J. P. Fouque
Administrative Mentor: Ralph Smith

Curtis Storlie, (Ph.D., Statistics, Colorado State University)

SAMSI Programs: *National Defense and Homeland Security* (participating while on leave at North Carolina State University)

- Anomaly Detection working group
- Method for automated tracking of objects that merge together and/or split apart, such as storms or vortexes
- Nonparametric estimation of the R^2 in a multiple predictor setting

Research Mentor: Jim Berger
Administrative Mentor: Jim Berger

Francisco Vera, (Ph.D., Statistics, University of South Carolina)

SAMSI Programs: *National Defense and Homeland Security*

- Data Confidentiality and Anomaly Detection working groups
- Secure computations software
- Software on Bayesian scan statistics
- Micro aggregation

Research Mentor: Alan Karr
Administrative Mentor: Alan Karr

B.1. Overview of Postdoc Selection, Postdoc Activities and Mentoring Strategies

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

Focus on integrating statistical and applied mathematical aspects in SAMSI programs begins with the Postdoc selection process. During the 2005-6 grant year candidates applied to participate in the 2006-7 SAMSI Programs (*Complex Computer Models* and/or *Random Matrices and High-dimensional Inference*). The recruiting process began with advertisement on the SAMSI web pages and direct contact with university departments, and involved not only SAMSI researchers and the SAMSI Directorate but also the Program Leaders and the Scientific Advisory Committee. The 2006-7 Program Leaders and Scientific Advisory Committee were particularly successful in bringing the SAMSI opportunities to the attention of promising doctoral candidates working in program-relevant areas of research. Additionally, for the 2006-7 *Random Matrices and High-dimensional Inference Program*, the Program Leaders were invited to provide reviews of applications to that program with special attention given to assuring good matches between candidates selected and their possible mentors both during the initial year at SAMSI and the during the second year of fellowship. Final decision continues to rest with the SAMSI Directorate following on-site interview of candidates at SAMSI. However, the careful assessment by the Program Leaders was invaluable and, in this case at least, led to happy consensus decisions.

When Postdocs first arrive at SAMSI they become part of a Postdoc Community that in addition to SAMSI Postdocs includes NISS Postdocs and other young researchers in the NISS-SAMSI complex. This lively Community has monthly Postdoc Lunches with the Directorate where topics often include the practicalities of an academic or a research career (how to interview successfully for a position, how to plan and write a research proposal, how the publication process works in the mathematical sciences from journal selection through interpretation of written reviews to successful revision). A monthly "Postdoc Presentation" seminar with "practice job interview" presentations of research results to the Postdoc Community (and interested graduate students) serves to refine presentation skills at the same time that it serves an interdisciplinary role to inform Postdocs coming from different disciplines and/or working on different SAMSI Programs. Other shared activities within this Community include assisting with the SAMSI Undergraduate workshops, where Postdocs continue to be the most effective presenters for students of this age.

Effective mentoring of Postdocs is an essential part of SAMSI's mission; so that, in principle, each Postdoc acquires two mentors. (Occasionally a single mentor fills both roles.) The first is a Research Mentor, commonly the Working Group Leader of the Postdoc's principal Working Group. The second is a member of the Directorate whose natural role is to be a second, non-technical, pair of ears and a second personality with knowledge of local issues and general SAMSI information. This second mentorship also connects Directorate in a personal, non-evaluative way to Postdoc Life at SAMSI. In their comments, SAMSI Postdocs have continued to report that they feel well-supported by this dual-mentor system and by both particular mentors in their personal evaluations.

B.2. 2005-06 Postdoc Activities

In mid-Spring SAMSI Postdocs summarize their experiences in Activity Reports and they also provide evaluations of their postdoc experiences. In Section B.2.1 each Postdoc has identified the Working Groups (primary and secondary) and other particular activities. The synopsis of the research work follows with accomplishments to date and the longer-term research agenda deriving from this research for each primary Working Group. Finally a list of publications, works in preparation and presentations at conferences, etc., is given. The Research Mentor's commentary on the Postdoc's work follows this self-evaluation. Section B.2.2 contains each current Postdoc's responses to the 10-question evaluation of the SAMSI Postdoc experience.

Amit Apte – Activity Summary

Primary Research

The main focus of my research is on problems in dynamical systems arising from physical applications. I have worked with a broad range of mathematical tools and numerical techniques to study such problems. Specifically, I am working on the following topics.

Current Projects

Langevin stochastic differential equations with applications to data assimilation: We are using the Langevin dynamics to directly attack the main shortcoming of the commonly used data assimilation schemes such as the ensemble Kalman filter (EnKF) - the assumption of a Gaussian prior probability density function (PDF). The Langevin method uses a stochastic differential equation (SDE) whose stationary invariant measure is the posterior PDF that we want to sample. Thus, time-averages using the solutions of this SDE are the same as the averages calculated using the desired distribution. Preliminary calculations using a simple dynamical system have shown a clear improvement over the EnKF and we are applying this method to more realistic problems such as the shallow water dynamics.

Dynamics of inertial particles in fluid flows and Lagrangian data assimilation: Much of the Lagrangian data for the ocean comes from instruments that are not purely Lagrangian, because they follow the dynamics of inertial particles with a small mass. The study of such particles has two aims in the context of data assimilation. (1) It will prove as a test-bed for the assimilation of non-trivial particles. (2) We will study results from the assimilation of data given by inertial particles into a model of Lagrangian particles. This will elucidate how errors in a model affect data assimilation. Using the Langevin dynamics, we can separate model errors from errors in the assimilation scheme. Currently, we are using geometric singular perturbation theory together with Melnikov methods to understand the flow field analytically and to identify trajectories which will behave differently in inertial and Lagrangian fields.

Both these research projects are a direct outcome of the various working group activities associated with the SAMSI program on Data Assimilation in Spring 2005. This year, I have attended the working group on Agricultural systems for the program on National Defense and Homeland Security. I am also involved in projects with my dissertation advisor, continuing my dissertation research. I presented some results from that research at the SIAM conference on dynamical systems at Snowbird, UT and some preliminary ideas about use of Lagrangian data for reconstruction of ocean velocity fields at a workshop in SAMSI. Over the past year, I have refereed articles for the journals Chaos, Physica D, communications in Nonlinear Science and Numerical Simulation, and Discrete and Continuous Dynamical Systems.

Sava Dediu – Activity Summary

<u>Primary Working Group</u>

Agricultural Systems: I am a member of the Agricultural Systems working group which has regular, weekly meetings. In addition, I also maintain its webpage. During Fall Semester of 2005, I gave a talk on the airborne transmission of foot-and-mouth disease during the UK 2001 epidemic. For the Spring Semester of 2006, I have been running numerical simulations for the sensitivity analysis of the Agricultural Network Model and also gave two talks: one on the sensitivity analysis for dynamical systems for the AG working group (joint presentation with Hoan Nguyen and H.T. Banks) and the second one in the Session on Agricultural Systems at the SAMSI Two-Day Undergraduate Workshop.

<u>Secondary Working Group</u>

Social Networks: I am a member of the Social Networks working group which has regular, weekly meetings.

<u>Current Activities</u>

During the Fall Semester of 2005, I attended the kickoff workshop for the NDHS program, I audited the NDHS graduate class taught by Professor Alan Karr at SAMSI and I was actively involved in the Agricultural System and Social Networks working groups. I gave a talk at the SAMSI Postdoc Seminal on the recovery of inhomogeneities in acoustical waveguides using eigensystem decomposition.

<u>Current Projects</u>

Together with Hoan Nguyen and H.T. Banks I am working on a paper about Sensitivity of Dynamical Systems to Convex Metric Space Parameters, which will be submitted to publication soon.

<u>Future Work</u>

Future work includes writing and publishing a paper on the sensitivity analysis of our AG Model in collaboration with other members of the AG working group, give another presentation at the Postdoc seminar and assist in the organization of future workshops for undergraduate and graduate students.

Lisa Denogean – Activity Summary

<u>Primary Working Group</u>

Data Confidentiality: I have presented papers and current research on disclosure limitation during the working group meetings. I am collaborating with Bahjat Qaqish on a new approach to data swapping. We have finished one paper, entitled “A Stochastic Process Approach to the Analysis of Swapping for Categorical Variables”, and are currently working on other case studies with Alan Karr. In March I presented some of our results at the Data Confidentiality Mid-Year Workshop at the National Center for Health Statistics. In addition, I am collaborating with Alan Karr to extend some of this work to develop better measures of data utility and risk. We are currently discussing writing a paper.

Secondary Working Group

Anomaly Detection: The primary direction of this working group has been in the area of syndromic surveillance. We have focused on reviewing research in this area and discussing ways to improve scan statistics for various types of data. I have presented papers on current approaches to syndromic surveillance systems and scan statistics for email network graphs. I am currently collaborating with Sidney Resnick on modeling network traffic data. I also presented an overview of syndromic surveillance for the SAMSI Two-Day Undergraduate Workshop on National Defense and Homeland Security.

Current Projects

A Study of Data Swapping for Categorical Variables: Data swapping can be used by government agencies to protect the confidentiality of publicly released data files. Bahjat Qaqish and I have studied the stochastic process generated by data swapping applied to a data file of categorical variables. The purpose is to understand the effect of swapping and to help the original data owners to determine which variables to swap and how much to swap. We present various utility measures and introduce the idea of measuring distance from the limit rather than from the original file. In addition, we introduce a new type of swapping that we propose is superior to current methods. We are writing this into a paper that we are soon submitting.

New Measures of Data Utility and Risk for Data Swapping: I am working with Alan Karr and Bahjat Qaqish to extend some of the above work. We are attempting to better study the effects of swapping on model building. We are also working on other measures of risk and attempting to formulate an appropriate decision-making model for data owners and data intruders. This is ongoing work with many possible model formulations under consideration, however, we will be presenting some of our current results at the Summer Research Conference in Statistics at Baylor University.

Network Modeling: I am working with Sidney Resnick on building a model of network traffic using real data from the University of North Carolina computer science department. We will attempt to model the timing, size and clustering of network packet communications. Some communications are triggered by human actions, while many are the result of network protocols or are otherwise computer-driven. This work has the application of not only improving network design, but possibly identifying anomalous network traffic or hacker events.

Continuing Thesis Work: I am continuing some work from my Ph.D. thesis that I plan to submit for publication soon. I present methods for forming confidence intervals for the recombination rate in DNA sequences based on various estimators of recombination. In my thesis I compared the Hudson and the Hey and Wakeley estimators, and I am currently studying the behavior of another estimator of interest. In practice, the use of the first-order asymptotic approximation can yield inaccurate confidence intervals due to extremely slow convergence. Due to the computational resources required, there has been little work done on comparing the true behavior of these estimators and their use in forming confidence intervals. I use different approaches to transform and bias correct these estimators and use a bootstrap method to form confidence intervals. I assess the coverage of the resulting intervals under various parameter values.

Presentations

- Joint Statistical Meetings Invited Talk, "New Measures of Data Utility and Risk for Data Swapping (Work in Progress)", July 2006
- Summer Research Conference in Statistics at Baylor University, "New Measures of Data Utility and Risk for Data Swapping (Work in Progress)", June 2006
- College of Charleston Math Seminar Invited Talk, "Confidence Intervals for Mutation and Recombination in Population Genetics", March 2006
- Data Confidentiality Mid-Year Workshop, "A Stochastic Process Approach to the Analysis of Swapping for Categorical Variables", March 2006
- SAMSI Postdoc Seminar, "Confidence Intervals for Mutation and Recombination in Population Genetics", December 2005

Other Activities

- Presenting tutorials and participating in mentoring for the SAMSI/CRSC Undergraduate Workshop, May 2006
- Presented a short talk at the SAMSI Two-Day Undergraduate Workshop on National Defense and Homeland Security, March 2006
- Web design and management for the Anomaly Detection Working Group

Recent Publications

- L. R. Denogean, T. DiCiccio, M. Wells, "Improved Confidence Interval Estimation for the Mutation Rate in the Infinite Sites Model," Submitted to *Statistical Applications in Genetics and Molecular Biology*.
- L. R. Denogean, T. DiCiccio, M. Wells, "Bootstrap Confidence Interval Estimation for the Recombination Rate," *in progress.*
- B. F. Qaqish, L. R. Denogean, A. F. Karr, "A Stochastic Process Approach to the Analysis of Swapping for Categorical Variables," *in progress.*

Moustapha Pemy – Activity Summary

Working Groups:

Computational Issues: This working group was led by Prof. Paul Fackler, I was very active in this group, I presented one of my working paper entitled "Quadratically convergent method", and I received a lot of useful feedbacks and ideas about how to finish the paper, since I still have to figure out how to complete the proof of the main results in this paper. Overall during the semester we went over different methods to solve problems in Financial Mathematics. I have particularly learned a lot about Monte Carlo methods and how they can be used in a nontraditional way.

Portfolio Management: During the group meetings we have devoted a good part of our time to study the theory of Backward Stochastic Differential Equations and how they applied to various problems in finance and stochastic control. Since I was not familiar with this new topic, I have really enjoyed learning this new tool. I will try to incorporate some the ideals I learned in my future research.

Model Uncertainty: This group was led by econometricians, Prof Eric Ghysels and Eric Renault, and the approach they presented to solve various problems in

Asset Pricing and Modeling the Volatility was different from what I have learned in the past, and incite me to think how I can use those ideals together with my functional analysis background to generalize some of their discrete models to continuous functional depending models to forecast the volatility and the asset price. I know this approach may be quite involved and may require a good deal of challenging results in Stochastic Functional Analysis.

Personal Research

My postdoctoral experience has been very positive, in the sense that, it has put in contact with very motivated researchers. My supervisors Prof. Tao Pang and Prof. Harry Chang have introduced to me the area Stochastic Functional Equations, and we have particularly study various control problems in this field. Bringing into play the theory of Viscosity Solution, so far we have submitted in this very short period of time four papers for publication and right now we are working on follow-up papers. In addition, I have continued working with Ph.D. advisor on various Selling Rule problems, we have submitted one paper about the liquidation of a large block of stock when the stock price follows the Geometric Brownian Motion to the Journal of Banking and Finance and right now, we are working on the follow-up paper where we generalize the result to include the case where the underlying model is the Regime Switching Model.

List of Publication and Work in Progress

- *Optimal stock liquidation in a regime switching model with finite time horizon* with Qing Zhang, to appear in the Journal of Mathematical Analysis and Application (available online from September 19, 2005).
- *Optimal Stopping for Stochastic Functional Differential Equations* with Mou-Hsiung Chang and Tao Pang, submitted to SIAM journal of control and optimization Liquidation of a Large Block of Stock with Qing Zhang and George Yin, submitted to the Journal of Banking and Finance.
- *Optimal Control of Functional Stochastic Differential Equations with a Bounded Memory* with Mou-Hsiung Chang and Tao Pang, submitted to the International Journal of Probability and Stochastic Process.
- *Finite Difference Approximations for Stochastic Control Systems with Delay* with Mou-Hsiung Chang and Tao Pang, submitted to Journal of Stochastic Analysis and Application
- *Viscosity Solutions of Infinite Dimensional Black-Scholes Equation and Numerical Approximations,* with Mou-Hsiung Chang and Tao Pang. Submitted to Applied Mathematics and Optimization
- *Optimal Control of a Switching Diffusion* with Tao Pang, working paper
- *Liquidation of a Large Block of Stock under Regime Switching Model* with Qing Zhang and George Yin, working paper.
- *Quadratically Convergent Methods,* working paper.

Jesus Rodriguez – Activity Summary

Primary Working Group

Credit Risk: I have presented papers and current research on modeling credit risk during the working group meetings. I led several discussions on topics of interest, and I am now focusing on modeling correlated multiple default derivatives with

Jean-Pierre Fouque. In March I was invited to present some of our results at the Mathematics Department Seminar Series at Oregon State University.

Secondary Working Group

Portfolio Management: I led several discussions in the Portfolio Management working group, mainly focused on a new area called Stochastic Portfolio Theory. One of the chief contributors to the theory, Adrian Banner from INTECH, attended the Opening Workshop, and I have remained in contact with him throughout the year. I am currently working with Tao Pang on investigating the behavior of stochastic differential equations in this area.

Current Projects

Multiple Default Derivatives: We are considering modeling correlated multiple default derivatives using structural models, which are consistent with short maturity yield spreads. We build on ideas from Stochastic Volatility popularized by Fouque, Sircar, and Papanicolaou, and show how it can be incorporated into complex derivatives in the bond market.

Stochastic Portfolio Theory: I am working with Tao Pang on studying the asymptotic behavior of stochastic differential equations with double reflecting boundaries. We are building on the seminal work done by Banner, Fernholz, and Karatzas on ranked capitalization based portfolios, and removing assumptions in order to work apply the ideas in more generality.

Pricing Issues in Energy Markets: We study models for electricity pricing and derivatives in today's de-regulated markets. We find a class of forward price processes that are consistent with the spot price, and use these forwards to value options on the spot. We first suppose the forward price follows a quite general diffusion process and show a necessary and sufficient condition on the drift term, which is perfectly analogous to the Heath-Jarrow-Morton condition. Using the forwards, we then show how to price European call options with a portfolio that has holdings in forward contracts and bonds. In addition to standard derivatives, we also consider a ``swing" or ``take-or-pay" option which is particular to energy markets. It is these swing options that attract the most interest from market participants, as well as the academic literature since the work of Jaillet, Ronn, and Tompaidis. We show how European calls, along with forward contracts, are used to hedge swing options. We devote special attention to the spot model proposed by M. Barlow, and modify it to prevent arbitrage in our context. A paper based on this work was submitted to *Mathematical Finance*.

Presentations

- *Mathematics Seminar Invited Talk at Oregon State University, "Pricing Issues in the Energy Markets," March 2006*
- *Financial Mathematics Seminar Invited Talk at NC State University, "Forward Models and Options in Electricity Markets", February 2006*
- *SAMSI Postdoc Seminar, "Forward Models and Options in Electricity Markets", October 2005*

Other Activities

- Presenting tutorials and participating in mentoring for the SAMSI/CRSC Undergraduate Workshop, May 2006

- Participated in workshop content and presented a short talk at the SAMSI Two-Day Undergraduate Workshop on Financial Mathematics, Statistics and Econometrics, November 2005
- Web design and management for the Levy Processes and Credit Risk Working Groups

Publications

- J. F. Rodriguez, "Forward Models and Options in Electricity Markets," *submitted to Mathematical Finance.*

Curtis Storlie– Activity Summary

Current Projects

Method for automated tracking of objects that merge together and/or split apart, such as storms or vortexes: This is important for areas such as climate model validation, storm forecasting, turbulence research, and radar/sonar processing. While this is not the first time statistical models have been used for tracking, there are several novel concepts developed here. For example, this is the first time that the merging and splitting of objects has been built explicitly into the model. Secondly, this is the first time anyone has investigated the asymptotic properties of a tracking estimate.

Nonparametric estimation of the R^2 in a multiple predictor setting: Another problem is in the area of nonparametric regression or machine learning. Many applications require an ANOVA or R^2 breakdown of the variability in the response that can be attributed to each predictor variable. The application that directly motivates my work in this area is sensitivity analysis for large scale deterministic computer models. I am interested in nonparametric estimation of the R^2 in a multiple predictor setting. A paper by Doksum gives some consistent estimates of the nonparametric R^2 based on kernel estimators. I have used more flexible estimates such as those achieved by local polynomial regression, recursive partitioning regression, adaptive regression splines, and GAM fitting on real and simulated problems with much success. I am also working to extend the asymptotic results of Doksum to some of these more flexible methods.

Anomaly Detection working group: A third problem coming out of the anomaly detection working group is motivated by the application of syndromic surveillance. We would like to detect outbreaks (natural or terrorist inflicted) early enough to minimize the impact. We would like to improve upon some of the prospective (real-time) methods for anomaly detection such as CUSUM and scan statistics to better meet the needs of syndromic data. The existing methods assume that in the absence of an anomaly, the number of events has Poisson distribution with a background rate that is a deterministic function of time. The complication with syndromic surveillance is that the background rate is not a fixed periodic function from year to year. For example, the flu season changes in temporal location and severity from one year to the next. Thus the rate of hospitalizations with flu-like symptoms or over the counter medication sales cannot be expressed as a function that is the same each year. I am proposing to instead treat the background rate as a random process, which would allow for these year to year differences in the background rate function while still taking advantage of the overall similarities from year to year.

Refereed Journal Articles

- CB Storlie, J Hannig, CM Lee, and D Nychka (2006). The Asymptotic Properties of a Multiple Target Tracker. To be submitted to *Annals of Statistics*.
- CB Storlie, CM Lee, J Hannig, and D Nychka (2006). Tracking of Multiple Merging and Splitting Targets with Application to Convective Systems. Submitted to *Statistica Sinica*.
- CB Storlie and JC Helton (2006). Multiple Predictor Smoothing Methods for Sensitivity Analysis. Submitted to *Reliability Engineering and System Safety*.
- JC Helton, JD Johnson, WL Oberkampf, and CB Storlie (2006). A Sampling-Based Computational Strategy for the Representation of Epistemic Uncertainty in Model Predictions with Evidence Theory. To appear in a special issue of *Computer Methods In Applied Mechanics and Engineering*.
- JC Helton, JD Johnson, CJ Sallaberry, and CB Storlie (2005). Survey of Sampling-Based Methods for Uncertainty and Sensitivity Analysis. *Reliability Engineering and System Safety*.

Conference Proceedings/Lectures

- CB Storlie, CM Lee, J Hannig, D Nychka (2006). Tracking of Multiple Merging and Splitting Targets with Application to Convective Systems. *IMS Annual Meeting with the X Brazilian School of Probability*. (invited)
- CB Storlie, CM Lee, J Hannig, D Nychka (2006). Tracking of Convective Systems for Climate Model Improvement. *Joint Statistical Meetings*. (contributed)
- CB Storlie, CM Lee, D Nychka, B Whichter, C Davis, J Weiss (2005). Identifying and Tracking Turbulence Structures. *38th Asilomar Conference on Signals, Systems, and Computers*. (invited)
- CB Storlie and JC Helton (2005). Multiple Predictor Smoothing Methods for Sensitivity Analysis. *Winter Simulation Conference*. (invited)
- JC Helton, JD Johnson, WL Oberkampf, and CB Storlie (2005). A Sampling-Based Computational Strategy for the Representation of Epistemic Uncertainty in Model Predictions with Evidence Theory. *8th US National Congress on Computational Mechanics*. (invited)

Other Activities

I have also been teaching a Statistics for Engineers course in the Statistics Department at NC State for the past two semesters. I will also be acting as a mentor for the Undergraduate Workshop at SAMSI this coming summer as well. In addition I have been interviewing for tenure track jobs this spring with much success. I received offers from CU Boulder and University of New Mexico. Both of which were very attractive given my collaborations with NCAR and Sandia Laboratories. I have opted to accept the position at UNM starting next spring.

Francisco Vera – Activity Summary

Primary Working Groups:

Data Confidentiality: I have attended weekly meetings, created software for secure computations, give feedback to the group in other areas, and I am currently collaborating with James Lynch in an improvement of micro aggregation techniques.

Anomaly Detection: I have attended weekly meetings, some of which I was the expositor. I have also collaborated with James Lynch and David Dickey on the detection of spurious observations. I was also chosen by the group to talk about our group activities on the undergraduate one-day workshop during spring 2006. I am also collaborating with David Banks and Gauri Datta on a Bayesian version of the scan statistic.

<u>Presentations</u>

As part of the Data Confidentiality group, I have given talks at the working group meetings at SAMSI, at the mid-year workshop on Data Confidentiality at Hyattsville, Maryland and in the Conference on Quantitative Methods & Statistical Applications in Defense and National Security organized by Rand Corporation. These talks were related to the software that I am developing for secure computations.

As part of the Anomaly Detection group, I gave one talk on a paper on spatial criminal prediction at one of the meetings. This talk, along with some ideas for the problem was presented at the mid-year workshop on Anomaly Detection in Hyattsville, Maryland. I also gave a talk with James Lynch on the detection of spurious observation in one of the working group meetings.

At the one-day undergraduate workshop, organized by SAMSI, I gave a talk, together with David Dickey and Lisa Denogean, about the research being done at the working group.

<u>Publications</u>

I am working on a paper about the software that I am developing for secure computations. I am working with James Lynch in the publication of paper from my dissertation with applications in data confidentiality. I am also working with James Lynch and David Dickey on a paper on detection of spurious observations.

<u>Current Projects</u>

Secure Computations Software: I used C#, one of the latest standards in computer programming, to develop software to compute statistics between several parties without revealing individual data between each other or a third party.

The software has currently implemented secure sums, secure regressions and secure data integration. Other secure computations that are being developed include secure maximum likelihood estimation, secure contingency tables and secure generalized linear models. In the future I may develop secure generalized additive models, depending on the progress the group does in this area.

I am also working with David Banks and Gauri Datta to develop software to do Bayesian scan statistic.

I am also working with James Lynch on an improvement of micro aggregation that would allow for better data quality, while preserving data confidentiality.

<u>Activities</u>

I have participated at all of SAMSI workshops on National Defense and Homeland Security, as well as the workshop and tutorials on Astrostatistics.

I have looked also into the current technology being used at SAMSI for working groups and making suggestions on how communications can be improved, in particular for remote participants of the working groups.

C. Graduate Student Participation

I. National Defense and Homeland security

Agricultural Systems Working Group

Ping Bai (Statistics, UNC) My work is focused on the stochastic modeling of the agricultural networks. In our simplified model we consider the stochastic model of a pork food chain with a set of aggregated nodes interacting with each other according to certain realistic assumptions. We model the evolution of the food production network, which currently contains 4 nodes, as a continuous time Markov Chain with discrete state space embedded in R^4 non-negative integer lattice. The standard stochastic simulation algorithm (Gillespie) is used to obtain the long-term behavior of the system. On the other hand, the approximate long term behavior of the appropriately scaled system is also analyzed via the usual macroscopic deterministic rate equations. This deterministic approach makes it possible to carry out the sensitivity analysis of the parameters involved in the model.

Angela Govan (Mathematics, NCSU) I am a graduate fellow participating in the National Defense and Homeland Security SAMSI program. During my fellowship I have been partaking in the Agricultural group which is working on the issue of the agricultural terrorism. Having knowledge of the mathematical modeling of contagious diseases I am responsible for incorporating a model of Foot and Mouth disease into the Agricultural Network model developed by the group. Currently I am modifying an SIR (Susceptible-Infectious-Recovered) disease model to study the Foot and Mouth dynamics in the network.

Anomaly Detection Working Group

Shenek Hayward (Mathematics, NCSUI) was a graduate student working in the anomaly detection working group. She also participated in the seminar course associated with the program. The anomaly detection group focused on the "scan statistic" and Shenek gave a presentation on one of the fundamental papers in the area. As a result of her interest in the topic, she and her SAMSI faculty advisor (D. Dickey) devised a reading course at NCSU in which they are studying a variation called the "double scan statistic"

Data Confidentiality Working Group

Joyee Ghosh (Statistics, Duke) is developing methods for performing nonparametric regression for horizontally partitioned data, i.e. data where three or more agencies possess the same attributes on different records. Her techniques utilize the secure summation protocol, which was demonstrated for linear regression by Karr et al. (2005). Joyee's techniques add flexibility in modeling to the secure computation toolkit, which helps address the difficulty of agencies having to specify models without seeing others' data. Joyee is preparing a manuscript for submission at the end of the semester. She presented the initial stages of her work to the NDHS confidentiality working group. This topic is not likely to lead to a dissertation thesis, although it has been invaluable for Joyee to learn areas of statistics--data confidentiality and nonparametric regression--that she would not normally have studied.

Robin Mitra (Statistics, Duke) is developing methods for modifying survey weights when data are altered to protect confidentiality. This is neglected in the literature on data confidentiality, but it clearly is important: unaltered survey weights could increase the risks of disclosures and could decrease data quality, since the weights are no longer tied to the values of the data. This work will be lead to Robin's preliminary examination as part of his progress toward a PhD at ISDS. He plans to produce a manuscript for submission by the beginning of the next semester. Robin has become very interested in data confidentiality, and he would like to pursue a dissertation in this area.

Saki Kinney (Statistics, Duke) has participated in the working group discussions, which have helped her get a big picture of the issues in data confidentiality. Saki is working on methodology for generating multiply-imputed, synthetic, i.e. simulated, data sets for public release. For her dissertation, she is working on theory of performing large-sample significance tests when multiple imputation is used simultaneously to handle missing data and to replace confidential data.

Social Networks Working Group

Jen-hwa Chu (Statistics, Duke) was involved in building agent-based models of social network dynamics. These models incorporate the latent-variable space approach described by Hoff, Raftery and Handcock (2002), as well as covariate information such as gender and memory of past relationships. The intent is to build rule-sets such that the dynamics of agent behavior mirror the dynamical models being from two other perspectives by other teams in the working group. The comparison of the models is based upon summary statistics from repeated runs, such as the mean and standard deviation of the number of persistent cliques, the first three moments of the in-degrees, the mean and standard deviation of the number of triad completions, and so forth.

John Samuels (Mathematics, NCSU) did research with Alan Karr, Hoan Nguyen (a SAMSI postdoctoral Fellow) and H.T. Banks. During the year they have developed a model for social dynamics based on dynamic agent-specific characteristics and pair-specific attractions. Coupled stochastic differential equations define the evolution of the system. The SDEs are solved by a classical fourth-order Runge-Kutta discretization procedure. The principal focus of the research is on the "richness" of the models as a function of the variances of the stochastic disturbances. Samuels has contributed substantially to the research and methodology and is a co-author on the paper "Sensitivity to noise variance in a social network dynamics model.” He also will play a major leadership role in the SAMSI Undergraduate Workshop to be held May 30-June 3.

Eric Vance (Statistics, Duke) has been studying social network behavior in elephant herds using data collected by ethologists. He has fit the Hoff, Raftery, and Handcock (2002) latent variable version of the dyadic p* model and found that group dynamics change between the wet and dry season, and that genetic relatedness and the social hierarchy play a large role in elephant networks. The inferences are Bayesian, and use Markov chain Monte Carlo to find the posterior distributions of each of these effects. He has written a paper on the social network methodology for and it is submitted to the Journal of Organizational Computation.

Chien-Chung Wong (Mathematics, NCSU), working with Medhin, developed a model where each actor is endowed with a set of dynamic personal attributes, values, and preferences, and a set of statistical information on each of the other actors in the social group. If the social network consists of N actors we construct an NxN matrix of zeros and ones, called sociomatrix. If actor i is friendly toward actor j, then the ij-th entry of the matrix will be one, otherwise zero. The diagonal entries of the matrix are set to zero. If the ij-th entry of the matrix is 1, then we say there is a link from i to j. In the model the ij-th link depends on the maximum of a payoff of an appropriately constructed nonlinear programming problem involving the attributes and values of the actors in the social group. The model can be modified to handle social status as well as general network dependence structure. For example, the link from i to j may not be completely independent of the link from i to k, and/or from j to k. The model also incorporates migration and preferred attributes. The approach developed captures the ideas of the well established P1 model introduced by Holland and Leinhart, and the more recent extension, the P2 model, due to van Duijn, Snijders, and Zijlstra. In particular the model developed reflects reciprocity, and attributes and values of actors i and j play a role in determining whether or not there is linkage between these actors.

II. Financial Mathematics, statistics and econometrics

John Hyde (Mathematics, Duke) was a second year graduate student. In collaboration with J.P. Fouque and Jonathan Mattingly, he explored models of company default. This became one of the major topic in Hyde's preliminary exam to move in to PhD candidacy. John made progress in his understanding of credit risk modeling. He presented and critiqued a number of models to Mattingly. He also explored the model with simulations. He was trying to develop new models of cooperative default which would give insight into how correlations in default time arise. John passed his Exams at the end of the fall. Unfortunately he chose to stop his pursuit of a PhD after passing his exams. Though he has left academic research, he is now perusing employment in the financial industry. His departure has terminated this particular research project, though he hopes to apply what he learned at SAMSI to others in the near future.

Arthur Sinko (Economics, UNC) participated in the activities of the model uncertainty and LÈvy processes workgroups. He is working on three papers and has greatly benefited from the workgroup and the workshops at SAMSI. His research is about MIDAS regressions, and how it relates to volatility modeling. This topic touches on both of the workgroups he attended.

Jennifer Sloan (Statistics, North Carolina State University) was mentored by Peter Bloomfield, who led the discussion in a Working Group session on ``Credit Ratings.'' Sloan jointly led the discussion at another session of the Credit Risk Working Group and participated in the ``Special Topics in Financial Math'' Course. She actively researched Credit Rating Transition and Credit Risk problems, and helped SAMSI in educational outreach activities.

Chong Tu (Statistics, Duke) helped and took part in the Opening Workshop for Financial Mathematics, Statistics and Econometrics from September 18-21, 2005, the transition workshop from February 27-28, 2006 and the Model Uncertainty Workshop. He attended the program courses ``Advanced Topics in Financial Econometrics'' and ``Special Topics in Financial Mathematics.'' He also joined the Model Uncertainty working group which had weekly meetings to discuss frontier papers. He is graduating this year and his SAMSI experience helped considerably in his landing an excellent job in the financial industry.

Doug Vestal (Mathematics, North Carolina State University) participated in a wide range of activities. In the Fall 2005, he took a special topics course in Financial Mathematics that emphasized the growing field of stochastic volatility models, credit risk, and the evaluation of real options. The course placed special emphasis on the current state of research in these fields. In this course, he was also exposed to the ideas that motivated the need for current research in these areas and some of the problems that remained to be solved. He was also a member of the SAMSI Credit Risk working group. In this group, various members presented current research trends in credit risk. He presented a model of recovery rates in a reduced form model along with some of my SAMSI colleagues. Other members gave presentations on the top down approach to credit risk, stochastic volatility and default correlation, credit ratings models, and a model for the unified valuation of equity and credit derivatives. As Doug will be doing his dissertation in credit risk, this working group was very helpful. In addition to a thorough review of the literature, he was able to see the current gaps in research to help develop ideas for his own research. In fact, because of the credit risk workshop, he is working on research to develop a new model of recovery rates. This is in addition to the research he started working on (along with Dr. Fouque and Dr. Carmona, two SAMSI participants) to develop an algorithm that enables the computation of extremely rare events with applications towards intensity based models in credit risk.

For the two-day Undergraduate Workshop in Financial Mathematics at SAMSI in October 2005, Doug wrote a document explaining various types of derivative contracts for the undergraduates. In addition, he explained how to implement, in Matlab, the first passage approach to the pricing of risky bonds. Among other things, this entailed teaching the students about yield curves and how to generate random variables. Dr. Fouque presented the binomial tree method to option evaluation and he showed the undergraduates how to implement it in Matlab to value an Asian option. At the SAMSI Undergraduate Workshop coming up in May 2006, he will be helping to teach, organize and execute a five-day workshop on inverse problems along with the other SAMSI Graduate Fellows and Postdocs. This workshop will place particular emphasis on developing the intuition behind modeling physical processes mathematically, the process of data collection, and the statistical analysis of the data collected for parameter estimation.

Yichao Wu (Statistics and Operations Research, UNC) was a graduate student fellow associated with the SAMSI program in Financial Mathematics, Statistics, and Econometrics (FMSE). Throughout the program, he attended the opening workshop and the transition workshop of FMSE in addition to the opening workshop in the program of National Defense and Homeland Security and one tutorial in the Astrostatistics program. These workshops brought me to the frontiers in the corresponding assorted areas and brought me to a lot of interesting problems in interdisciplinary study. Additionally, I joined the working group on Levy Process led by Prof. George Tauchen. He invited experts on this area to present their recent work and lead discussion. In particular, he invited Prof. Torben Andersen to present ``Jump detection in Finance". I found this is really interesting and try to work on some related problems. Also Prof. Enrique Figueroa led a discussion on ``An Overview of a Nonparametric Estimation Method for Levy Processes". All of these activities improved my understanding how statistics can be applied to other areas and help them, which I think is very important for a graduate student majoring in statistics.

Stephen Zhou (Mathematics, North Carolina State University) participated in the activities of the credit risk workgroup. During the semester he continued his research under J.P. Fouque and defended his dissertation in February 2006. He has greatly benefited from the workgroup and the workshops at SAMSI. His research focuses on modeling the correlation of defaults which is one of the main issues in credit markets.

III. ASTROSTATISTICS

Floyd Bullard (Statistics, Duke) is the SAMSI Graduate Fellow associated with the Exoplanets working group. He is an active member of the group, attending each meeting, maintaining the group’s web page, and coding MCMC, importance sampling and other algorithms for model fitting and model selection. His graduate work is focused on activities of the working group, and he expects to formulate a workable dissertation thesis as a result of the group’s work, possibly on the topic of adaptive design. He has given a presentation in a student seminar series at Duke on the search for exoplanets, and he is planning on another this semester at SAMSI as part of the graduate student and post-doc seminar series. He will be a coauthor on a working group paper on “Model Selection in Astronomy” and on a possible discussion of Skilling’s nested sampling paper.

Pablo de la Cruz (Statistics, University of Valencia) is working on his Ph.D. thesis under the common supervision of Vicent Martinez (Astronomy) and Jose Miguel Bernardo (Statistics) at the University of Valencia. Pablo is working on the analysis of the best available deep three dimensional catalogs of galaxies. He intends to apply novel statistical techniques to characterize the cosmic texture, and to probe the cosmic evolution, since deeper surveys map the universe at earlier cosmic epochs. He is exploring new Bayesian techniques in the studies of the large scale structure of the universe. The SAMSI program in astrostatistics is a great opportunity for Pablo's research projects, since his foundations are quite solid in both statistics and cosmology. His participation in the SAMSI interdisciplinary event will have a huge positive impact in the development of his own ideas for his thesis. The goal is to consolidate several scattered advances in both applied Bayesian statistics and the analysis of the galaxy distribution into a unified body where a set of comprehensible tools and algorithms for the statistical analysis of the galaxy redshift surveys will be developed and made publicly available.

Hyunsook Lee (Statistics, Penn State) is a statistics graduate student with an undergraduate background in astronomy. She attended tutorials and the astrostatistics kickoff workshop. During that time, she presented a poster, titled "Convex Hull Peeling: Nonparametric Multivariate Data Analysis." Some other related results will be presented at Interface 2006, SCMA IV, and JSM2006. After the workshop, she joined various focused working group meetings: exoplanet, source and feature detection, gravitational lensing, particle physics, and survey and population studies. She is in the process of writing papers and a dissertation on model selection with a jackknife method and nonparametric massive data analysis with convex hull peeling. The first topic is of theoretical nature and the latter one focuses on developing algorithms for exploratory data analysis with some supporting theory. She is maintaining the websites for the Survey and Population Studies working group, and for the Particle Physics group.

Nicholas Robbins (Mathematics, Duke) maintains the public web-page for the Gravitational Lensing working group. He is in the early stages of his thesis work with Professor Bray. Topics covered in the lensing session may be integrated in his thesis, but it is too early in the semester to say definitively.

Lingsong Zhang (Statistics, UNC) is interested in multivariate outlier detection and functional data analysis using singular value decomposition. He is currently in charge of maintaining the website for the Source and Feature Detection working group, and he is also an active participant of the discussion. Lingsong had developed visualization tools for functional data, and is currently working on multi-resolution outlier detection methods for detecting outliers in long-range dependent time series, with applications in Internet anomaly detection. He is in the astrostatistics program to look for interesting astronomy applications for which he can apply his visualization tools and outlier detection methods. He is also interested in developing new methodology for challenging astronomy problems.

IV. Latent Variable Models in the Social Sciences

John Hipp (Sociology, UNC) played an active role in several of the working groups. He was a regular participant in the Multilevel and Structural Equation Model group. He regularly attended and participated in the group discussions. John also was a member of the Categorical Observed Variables in Latent Variable Models group. For this group, he provided empirical and simulation examples that enabled us to explore several of the issues that emerged in our discussion. He also used these to demonstrate some of the new analytical results that we developed. John also participated in the social network working group. He is knowledgeable about social network data, theory, and examples and this knowledge was useful to the group. Hipp's dissertation to be completed this spring makes use of latent variable techniques in his analysis of neighborhood satisfaction using survey data collected by the U.S. Census. Finally, John Hipp helped with some of the logistics of making the working groups operate smoothly.

John Samuels (Mathematics, NCSU) has participated in the Social Networks working group that meets every Thursday at SAMSI. He has pursued specific research with Alan Karr, Hoan Nguyen (a SAMSI postdoctoral Fellow) and H.T. Banks. During the year they have developed a model for social dynamics (characteristics associated with a number of agents) of buddy/clique formation. The characteristics (on a continuous time, continuous value scale from -10 to +10) are assigned to each agent. A nonlinear model for degree of connectivity is coupled to a nonlinear stochastic differential equation for the evolution of the characteristics in the agent population. The resulting system of SDE is then solved by a classical fourth-order Runge-Kutta discretization procedure. Samuels has contributed substantially to the research and methodology and will be a co-author on a forthcoming publication this summer. He also played a major leadership role in the SAMSI Undergraduate Workshop held May 30-June 3, 2005.

Jen-hwa Chu (Statistics, Duke) has been involved in building agent-based models of social network dynamics. These models incorporate the latent-variable space approach described by Hoff, Raftery and Handcock (2002), as well as covariate information such as gender and memory of past relationships. The intent is to build rule-sets such that the dynamics of agent behavior mirror the dynamical models being from two other perspectives by other teams in the working group. The comparison of the models is based upon summary statistics from repeated runs, such as the mean and standard deviation of the number of persistent cliques, the first three moments of the in-degrees, the mean and standard deviation of the number of triad completions, and so forth.

Jiezhun Gu (Statistics, NCSU) has also been involved with the program. She has attended all of the weekly meetings, and has expressed interest in working on the asymptotic theory associated with model uncertainty. Subhashis Ghosal is also interested in working on this problem.

Satkartar Kinney (Statistics, Duke) has also made substantial progress on model uncertainty. She has focused on the problem of selecting fixed and random effects in logistic mixed effects models. A number of methods have been proposed in the literature for subset selection in regression, but little has been done for the challenging problem of selecting predictors that vary in their effects for different individuals and hence have associated variance components. Satkartar initially adapted an approach proposed by Chen and Dunson (2003) for the linear mixed model to the probit case. She then further modified this approach to logistic regression using a data augmentation scheme, with slice sampling used in the implementation. Ongoing work focuses on parameter expansion methods, which lead to more efficient computation and improved priors for the variance components.

Eric Vance (Statistics, Duke) has been studying social network behavior in elephant herds using data collected by ethologists. He has fit the Hoff, Raftery, and Handcock (2002) latent variable version of the dyadic p* model and found that group dynamics change between the wet and dry season, and that genetic relatedness and the social hierarchy play a large role in elephant networks. The inferences are Bayesian, and use Markov chain Monte Carlo to find the posterior distributions of each of these effects. He has written a paper on this, submitted to the Journal of Organizational Computation.

Chien-Chung Wong (Mathematics, NCSU), working with Medhin, developed a model where each actor is endowed with a set of dynamic personal attributes, values, and preferences, and a set of statistical information on each of the other actors in the social group. If the social network consists of N actors we construct an NxN matrix of zeros and ones, called sociomatrix. If actor i is friendly toward actor j, then the ij-th entry of the matrix will be one, otherwise zero. The diagonal entries of the matrix are set to zero. If the ij-th entry of the matrix is 1, then we say there is a link from i to j. We have developed a model where the ij-th link depends on the maximum of a payoff of an appropriately constructed nonlinear programming problem involving the attributes and values of the actors in the social group. The model can be modified to handle social status as well as general network dependence structure. For example, the link from i to j may not be completely independent of the link from i to k, and/or from j to k. The model also incorporates migration and preferred attributes. The approach developed captures the ideas of the well established P1 model introduced by Holland and Leinhart, and the more recent extension, P2 model, due to van Duijn, Snijders, and Zijlstra. In particular the model developed reflects reciprocity, and attributes and values of actors i and j play a role in determining whether or not there is linkage between these actors.

V. COMPUTATIONAL BIOLOGY OF INFECTIOUS DISEASE

Ben Cooke (Mathematics, Duke) has participated in the weekly meetings of the working group Mathematical Genomics for Vaccine Design (MagVad). He has been involved in our detailed survey of the literature on mathematical, statistical, and biological aspects of epitope presentation and recognition by the immune system. Throughout the year, Ben has also done work on several aspects of the epitope binding problem. He has produced statistical and graphical analyses of publicly available databases of MHC binding peptide sequences. He has also obtained amino acid "feature sets" for a collaborative effort with SAMSI postdoc Surajit Ray on applying modern classification algorithms (random forests, CART with boosting, SVMs) to the prediction of MHC I peptide binding, which will result in a publication from the working group. Finally he has begun applying multiple structure alignment tools developed in the Schmidler group to examine X-ray crystallographic structures of known MHC binding peptides in complex with MHC I. Preliminary results have suggested new approaches to prediction of peptide/MHC binding by simulation. This will have an impact on Ben's dissertation work (with Schmidler), providing new directions for his work on quantitative peptide and protein simulation. Ben also helped run the audio/visual and computing at the opening program workshop ``Genomes To Global Health: Computational Biology Of Infectious Disease".

Morgan Root (Mathematics, NCSU) has worked on the development of models for the spread of zoonotic infections, such as pneumonic plague, that are considered to pose significant bioterrorism threats. Root is in the early stages of his research career, so much of his time during the year was spent surveying both the biological and modeling literatures. He has developed deterministic and stochastic models for the spread of infection in a heterogeneous community. An example of a question of interest is whether the existence of groups that are underserved by the health system, such as minority communities or individuals of lower socio-economic status, can increase the likelihood or severity of disease outbreaks. Considering a two patch setting (representing the general population and a minority group), he has examined the time lag between outbreaks in the groups. If the underserved group is less closely monitored, then it is possible for the disease to become established there before it is noticed in the general population. The minority group can then become a source of infection for the community at large. Thus, any lag can have a major effect on the severity of an outbreak.

During the course of the year, we started a collaboration with Jay Levine (NCSU College of Veterinary Medicine). Through Levine, Root will be funded by North Carolina's State Department of Health to model bioterrorism scenarios. Beyond this, Root and Lloyd are involved in the preliminary stages of a collaborative project between NCSU's Vet School and Purdue's Homeland Security Center that will develop scenario models examining the impact of and responses to bioterrorism attacks on the nation's agricultural and food supply industry.

Soyoun Park (Statistics/OR, UNC) and **Karl Strohmaier** (Computer Science, UNC) were each funded for one semester, and have been working with mentor Andrew Nobel on problems related to bi-clustering. Bi-clustering refers to a broad class of exploratory data analysis tools that are well suited to genetic, chemometric and microarray type data. The goal of bi-clustering is to identify scientifically interesting interactions between collections of samples (e.g. disease states) and sets of variables (e.g. genes). Bi-clusters correspond to submatrices of a given data matrix. Unlike traditional clustering, bi-clusters can overlap, and they need not cover all the entries of the data matrix.

Park has been working on connections between data mining and classification, with an eye towards producing more interpretable classification schemes for high dimensional data such as microarrays. Specifically, she has been looking at how bi-clustering algorithms can be applied to the classification of high dimensional data.

Strohmaier has been working on bi-clustering for noisy data. Existing bi-clustering algorithms do not account for the presence of noise in a direct way, a fact that may limit their applicability to gene expression and other data, where moderate or large amounts of noise are often present. Strohmaier and others in the group have been working on developing algorithms that will work when noise is present. In particular, he is currently working on the implementation of a heuristic, iterative bi-clustering algorithm. Once preliminary testing is complete, we expect to apply the algorithm to SNP and microarray data.

Abel Rodriguez (Statistics, Duke) has attended two of the program working groups: the Mathematical Genomics for Vaccine Design working group in the fall, and the Modeling of the Immune System in the Spring. Abel also attended Alun Lloyd's course on "Mathematical Modeling of Infectious Diseases". In the MagVad group, Abel contributed to summary analyses of the first group of 86 epitopes we received, and particpated in the reading and presentation of papers on epitope and protein binding prediction. In the Immune System working group, Abel has participated in reading papers on two different lines:† stochastic models for measles and spatial-temporal models for disease spread.

Miriam Nuno (Cornell), **Steven Tennenbaum** (Cornell), **Ariel Cintron** (Cornell), **Alicia Shim** (Arizona State Univ.), **Efrat Barzohar** (Arizona State Univ.), and **Yun Kang** (Arizona State Univ.) were visiting graduate students during Spring, 2005 in the Computational Biology Program. They participated extensively in the SAMSI activities including taking the courses offered by Alun Lloyd (every Monday evening) and Byron Goldstein (every Thursday evening) at SAMSI. In addition they contributed heavily to the success of the working group led by Alun Lloyd on immune response models, attending regularly and discussing papers for the group. At their request, Banks offered a special course on applied mathematical and statistical methods for inverse problems (a version of the course he and Davidian offered in the first year SAMSI program on Inverse Problems). This course was taught at SAMSI on Mondays and had 15 students.

These six students, who were recommended to SAMSI by Carlos Castillo-Chavez after his efforts with them in the Mathematical Biology Theoretical Institute (MBTI) at Cornell and ASU, were each in various stages of their research programs. Nuno finished and defended her Ph.D. thesis during the semester; Tennenbaum finished his in the Fall, 2005: Cintron will finish in Summer, 2006; Shim will finish in two more years; Barzohar and Kang just joined the program at ASU after completing MS degrees and are just getting started on their Ph.D research topics.

The semester went well for all students but especially so for Nuno, Cintron, and Shim. Nuno is now a postdoc at Harvard, while Cintron and Shim have developed a mentoring/working relationship with Banks, Lloyd and some of his postdocs and students. Cintron will return to SAMSI as a SAMSI Fellow in Fall 2006. Several publications have resulted from these interactions.

Banks agreed to act as local scientific mentor/personal advisor to the 6 students. This proved to be quite demanding, but necessary, since their advisor was able to visit SAMSI only periodically during the semester and the students needed extensive scientific assistance and counseling during the course of the semester. While this experiment in outreach and education was a successful one for the students and SAMSI, it reinforced the understanding that such efforts need careful structuring to be successful.

VI. Data Assimilation in Geophysical Systems

Kristen Foley (Statistics, NCSU) is currently working on her dissertation with her advisor Montserrat Fuentes on a data assimilation project in collaboration with a group of meteorologists and oceanographers at NC State. The goal is to improve the prediction of the onshore flooding caused by hurricane force winds along the coastal Carolinas and Georgia. The group, with the help of Foley and Fuentes, has implemented statistical data assimilation methods including the Ensemble Kalman Filter and a 4-D Variational approach to combine observed data with output from a numerical ocean model to predict storm surge associated with hurricanes. They have evaluated the degree that data assimilation can improve forecasting and nowcasting results. They have also investigated the potential improvement of additional observing stations within the domain of interest.

Foley and Fuentes have published a joint paper to appear in the Monthly Weather Review on improving the wind field inputs for storm surge modeling. Foley and Fuentes have submitted (March 2006) a manuscript to the Journal of the Royal Statistical Society, Series C, on a new statistical framework for improving wind field prediction combining physical models with observations (as a request from the journal for a special issue). Foley will graduate in August with her PhD in statistics from NCSU.

Nicole Mich (Geosciences, Duke) was a part of the Lagrangian Data Assimilation working group. She was a graduate student working under the direction of Susan Lozier on a study of the pathways that Labrador Sea water takes from the subpolar gyre to the subtropical gyre. Mich graduated in November of 2005 with a master's degree from Earth and Ocean Sciences at Duke. The overall objective of her work was to understand climate pathways in the deep ocean. Her work was in support of an ongoing observational program that involves the deployment and tracking of Lagrangian floats (RAFOS floats) from the Labrador Sea to the subtropical basin. Her work sets the Eulerian framework for an understanding of Lagrangian data that first came available in July of 2005. In her work, Mich established a link between the Mediterranean Overflow Water pathways, Labrador Sea water pathways and the North Atlantic Oscillation, the dominant mode of decadal variability in the basin.

Stephen Foster (Mathematics, UNC) was a second year student and involved in the program as a younger student with potential interest in this area. He attended various working groups and was an energetic in the SAMSI course. His work focused on Lagrangian data assimilation and, specifically, he worked on a problem aimed at designing smart float placement strategies. The idea is to have optimal float placement for assimilation of the resulting data. In his report for the course, he worked out the ideas on a model problem based on the shallow water equations with wind forcing. He is expected to graduate with a Master’s degree in May, 2006.

Minjung Kyung (Statistics, NCSU) was a student in the program working under the supervision of Sujit Ghsosh, who was a participant from her department. She was also the assigned assistant to the program with note-taking and other support responisbilities. At the Tutorial and Opening Workshop of the DA program in January 2005, she took detailed notes of the lecture on "A Brief Introduction to Bayesian Statistics and Application to Data Assimilation" given by Chris Wikle. She produced careful notes for the Data Assimilation course taught at SAMSI. Her course project was on Bayesian statistical modeling concerning spatial-temporal variability. She participated regularly in the working group on Estimation and Prediction. Kyung is working with Ghosh as a doctoral candidate and her thesis proposal entitled "Generalized Conditionally Autoregressive Models" grew out of the discussion group. The proposed methods will be illustrated with a DA specific data set, which is appropriate with these types of models. She regularly attended the group meetings and the related workshops. She is scheduled to take her doctoral prelim oral exam on April 12th, 2006 and she plans to graduate with PhD in Statistics by end of this year.

Liyan Liu (Mathematics, UNC) is working with Jones. She has been developing Lagrangian data assimilation schemes for two-layer point vortex problems. Her work grew out of the Lagrangian Data Assimilation working group. The idea is to assess the extent to which information gleaned from Lagrangian floats at one level might impact the flow field at other levels.

As a result of the program, she is effectively now jointly supervised by Jones and Kayo Ide (University Fellow). Her work has been greatly influenced by Ide, who defined the problem for her. She has also consulted extensively with Susan Lozier in implementing her ideas in a two-layer quasigeostrophic model. She was deeply involved in the working group and presented a poster at both of the workshops.

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 I.E, the various Program Working Groups, and their members, are discussed.

E. Program Activities

1. Program on National Defense and Homeland Security

1.1 Objectives

The principal purpose of the SAMSI program on National Defense and Homeland Security is to identify research promising paths for the statistical sciences, applied mathematics and decision sciences in problems of National Defense and Homeland Security (NDHS), and to initiate research on them. This effort is especially important to the statistical sciences, applied mathematics and decision sciences because previous efforts by these communities[1], have failed to create a self-sustaining research momentum on NDHS. In addition, there have been few if any research efforts that haves panned the statistical sciences, the applied mathematical sciences *and* the decision sciences.

The SAMSI program is meant fill this gap, in part by providing proof of concept that the necessary collaborations are feasible.

1.2 Working Groups

Four Working Groups have operated throughout the year, whose principal function, as in all SAMSI programs, is to organize the research and ensure communication:

Agricultural Systems, led by Mette Olufsen, Faculty Fellow (Applied Mathematics, NCSU). Barrett Slenning (College of Veterinary Medicine, NCSU) is providing significant leadership as well.

Anomaly Detection, led by David Dickey, Faculty Fellow (Statistics, NCSU), and Douglas Kelly, Faculty Fellow (Statistics, UNC).

Data Confidentiality, led by Lawrence Cox (NCHS), Alan Karr (NISS).

---

[1]Which include an April, 2003 workshop sponsored by the Board of Mathematical Sciences and its Applications Board on Mathematical Sciences and their Applications (2002) a November, 2002 Intelligence Community Workshop hosted by the NSF, which led to a 2003 call for proposals on "Approaches to Combat Terrorism," a December 2002 planning meeting on statistics and national defense and security Karr (2003), which was sponsored by the Los Alamos National Laboratory (LANL) and National Institute of Statistical Sciences (NISS), which led to letters sent by the president of the American Statistical Association (ASA) to officials in the Department of Homeland Security outlining the potential contributions of the statistical sciences to the department, a May, 2003 workshop on Statistical Issues in Counterterrorism sponsored by the ASA Committee on National Defense and Homeland Security and the Washington Statistical Society, creation of the ASA Section on Statistics in Defense and National Security and a January, 2004 meeting in Santa Fe, NM, addressing the interface between statistics and operations research.

Social Networks, led by David Banks (Statistics, Duke).

All of these will function for the entire year.

Each working group has significant external participation. Specifically,

- Greg Rempala and Ryan Gill (University of Louisville) have been occasional participants in the **Agricultural Systems** Working Group, and Michelle Lacey (Tulane University), who was at SAMSI for the fall of 2005, was a regular participant.

- Deepak Agarwal (AT&T Research), Howard S. Burkom (Johns Hopkins University Applied Physics Laboratory), Kevin Ward Drummey (National Security Agency (NSA), Joe Fred Gonzalez, Jr. (National Center for Health Statistics (NCHS)), Myron Katzoff (NCHS), James Lynch (University of South Carolina), Ted Norminton (Carleton University), Cheolwoo Park (University of Georgia), Henry Rolka (Centers for Disease Control and Prevention (CDC)), and Galit Shmueli (University of Maryland) are regular participants in the **Anomaly Detection** Working Group.

- Lawrence Cox (NCHS), Joe Fred Gonzalez (NCHS) and James Lynch (University of South Carolina, who spent the spring semester at SAMSI) are regular participants in the **Data Confidentiality** Working Group.

- Deepak Agarwal (AT&T Research), Edoardo Airoli (Carnegie Mellon University), Hugh Chipman (Arcadia University), Stephen Fienberg (Carnegie Mellon University), Myron Katzoff of the NCHS, and Ted Norminton (Carleton University), are regular participants in the **Social Networks** Working Group.

1.3 Anticipated Outcomes and Measures of Success

1.3.1 Program Level

Outcomes at the program level are:

- Significant research accomplishments by the Working Groups, leading to papers submitted during or shortly after the program year. As noted in §3.2, at least one paper has already been accepted for publication.
- Formation of new collaborations, leading to proposals and research in following years.
- Extremely positive career impact on participants, especially postdoctoral researchers.
- Strong community interest in the program, leading to engagement in the form of research visits and workshop participation.

Each of these has clear measures of success, based on either quantifiable information (e.g., numbers of papers and proposals produced, numbers of visitor) or participant self-assessment (e.g., feedback from database providers, follow-up with postdoctoral fellows). Not all of these measures, however, operate within the one-year time frame of the program, and so post-program follow-up will be necessary.

1.3.2 Working Groups

Each Working Group produced a "sound bite" summary of its plans and identified one or more outcomes that it would consider to be major successes. Not all did so, but each addressed high-level goals. Each Working Group also developed a detailed research agenda that both focuses energy and defines measures of success.

Agricultural Systems. The group set as a major goal and success identification of state of the art model predicting spread of Foot and Mouth Disease (FMD) that can be applied to study of the disease in North Carolina and the US. Work on this model is proceeding.

Participants from SAMSI and the Research Triangle are Amit Apte (postdoc, UNC), Ping Bai (student, UNC), David Banks (Duke), Thomas Banks (NCSU), Ricardo Cortez (Tulane University; currently housed at UNC), Sava Dediu (postdoc, SAMSI), Anjela Govan (student, NCSU), Christopher Jones (SAMSI and UNC), Michelle Lacey (Tulane University; currently housed at Duke), Michael Last (postdoc, NISS), Hoan Nguyen (postdoc, NCSU), Mette Olufsen (NCSU), Barrett Slenning (NCSU) and Ralph Smith (NCSU).

Initial thrusts of the research included:

- Background presentation "Agricultural Disasters, Natural or Not: Risks and Readiness" by Slenning.
- Literature search for mathematical and statistical models for FMD.
- Review and discussion of items identified by the literature search.
- An introduction to Jackson networks.

Anomaly Detection. A central problem in NDHS is the early detection of events. These might include early signs of a disease outbreak (natural or terrorist induced), imminent disastrous weather events, or any form of terrorist attack. As the nation thinks about strategies for dealing with these issues, the Working Group is trying to surveying statistical methods that might be part of a plan to deal with some of these threats.

The group created a web site with descriptions of statistical methods applicable to the kinds of data involved in NDHS issues. Examples would include time series, issues of type I versus type II error tradeoffs, outlier and change point detection.

SAMSI- and Research Triangle-based participants are David Banks (Duke), M. J. Bayarri (SAMSI), Gauri Datta (University of Georgia, visiting SAMSI), Lisa R. Denogean (postdoc, SAMSI), David A. Dickey (NCSU), Joyee Ghosh (student, Duke), Shenek Heyward (student, NCSU), Yajun Mei (Fred Hutchinson Cancer Center, visiting

SAMSI), Bahjat Qaqish (UNC), Curtis Storlie (postdoc, NCSU/SAMSI) and Francisco Vera (postdoc, NISS and SAMSI).

Research thrusts are:

- Study of syndromic surveillance, which is roughly described as statistical analysis of indicators from many sources, for example, admissions to several hospitals for flu-like symptoms as well as sales of over-the-counter products related to the flu. One particular technique, the SCAN statistic, was the focus of some effort and some thinking about how to modify it to take into account background effects.
- Review of related statistical topics, such as time series, outlier detection, sequential testing methods, and multiple testing in a Bayesian framework.

Cross-fertilization of ideas has been very important. To illustrate, Dickey presented a time series of American Airlines stock volume as part of a time series presentation, and Lynch immediately did some additional analyses that were presented to the group.

Data Confidentiality. Because of increasing availability of external databases, construction of public release databases by federal agencies is verging on the impossible. A long-term goal is to produce an integrated suite of software tools that customize disclosure limitation strategies to statistical characteristics of databases, prolonging by years the ability of the agencies to release useful microdata.

SAMSI- and Research Triangle-based participants are Ping Bai (student, UNC), Lisa R. Denogean (postdoc, SAMSI), Joyee Ghosh (student, Duke), Alan Karr (NISS), Robin Mitra (student, Duke), Anna Oganian (postdoc, NISS), Bahjat Qaqish (UNC), Jerome Reiter (Duke), Saki Sarker (student, Duke), Francisco Vera (postdoc, NISS and SAMSI) and Mi-Ja Woo (postdoc, NISS).

Visitors during the spring semester were James Lynch (South Carolina) and Xiaodong Lin (Cincinnati; supported by NISS).

Thrusts of the research:

- Development of techniques that combine two or more methods for statistical disclosure limitation (SDL), resulting in disclosure risk and data utility performance superior to either method alone.
- Construction of new utility measures for numerical microdata, based on propensity scores and clustering.
- Exploration of SDL for data satisfying constraints such as positivity.
- Utility measures that account for analyses of released microdata that accommodate whatever SDL measures were applied.
- Methods and software systems for statistical analysis of distributed databases with actual integration of the data.
- A new formulation of data swapping that randomizes which attributes are swapped as well as which records are selected for swapping.

As a “capstone” project for the year, the group is writing a concept paper that will provide a new decision-theoretic formulation for SDL problems.

Social Networks. Social network statisticians have developed a rich class of static models. But new challenges that have arisen in counterterrorism, communications networks, and disaster response show the need for robust, reliable models that capture network dynamics. A major success would be the development of a body of theory for dynamic network models and the benchmarking of such models against applications from many different fields.

In addition to external participants listed in Section A of the Annual Report, participants based at SAMSI or in the Research Triangle are David Banks (Duke), Thomas Banks (NCSU), Sava Dediu (postdoc, SAMSI), Chung-Chien Hong (student, NCSU), Alan Karr (NISS), Michael Last (postdoc, NISS), Negash Medhin (NCSU), Hoan Nguyen (postdoc, NCSU) and Eric Vance (student, Duke).

Activities of the group include:

- Development of multiple models that introduce dynamics into social networks.
- Organization of and participation in a Board on Mathematical Sciences and their Applications (BMSA) Workshop on Statistics in Networks (Washington, September 26–27) and DARPA Workshop on Virtual Worlds and Social Networks (Washington, October 18–19).
- Review of latent space models and ideas for generalization to dynamic network models, of optimization methods for dynamic networks, of estimation of homophily issues and of dynamic process models.
- Application of network models to Hurricane Katrina evacuee surveys.
- Preparation of a research proposal on network models, collaboratively with AT&T Labs, which will allow models to be tested on real data.
- Planning and preparation of papers for a special issue of *Computational Mathematics andOrganizational Theory*.

Main research directions include:

- Development of a framework for characterizing behavior of dynamic social network models, which has produced a paper that has been accepted for publication (Banks et al., 2005).
- Testing multiple dynamic network models on data from Enron, AT&T and Hurricane Katrina.
- Assessing agent-based models vis-a-vis dynamic versions of p^* models.

1.4 Workshops

1.4.1 KickoffWorkshop

The September 11–14, 2005 kickoff workshop for the program attracted more than 100 attendees, and met the stated goal of informing the composition and activities of theWorking Groups. Details of the program are at www.samsi.info/workshops/2005ndhs-workshop200509.shtml.

1.4.2 Mid-Program Workshops

Three of the Working Groups held off-site mid-program workshops, which were extremely successful in attracting attendees who are not working group participants, as well as bringing together in-Triangle and out-of-Triangle participants. Specifics are as follows:

Anomaly Detection. The workshop was held on February 3, 2006 at the National Center for Health Statistics in Hyattsville, MD. Attendance was approximately thirty. Keynote presentations were made by Donald E. Brown (Systems and Information Engineering, University of Virginia), Howard S. Burkom (National Security Technology Department, Johns Hopkins Applied Physics Laboratory) and Carey E. Priebe (Applied Mathematics and Statistics, Johns Hopkins University). working group participants making presentations were Gauri Datta, David Dickey, Ryan Gill, Michael Last and Francisco Vera. Full details are available at www.samsi.info/200506/ndhs/workinggroup/ad/Workshop.htm.

Data Confidentiality. This workshop was held on March 13, 2006, also at the National Center for Health Statistics in Hyattsville, MD, and attracted more than 45 attendees, including researchers from the Bureau of Labor Statistics, Census Bureau, Energy Information Administration, National Center for Education Statistics and National Center for Health Statistics. Principal presentations were by Lawrence Cox (NCHS), Jay Kim (NCHS) and Aleksandra Slavkovic (Penn State). Presentations on research conducted by the working group were made by Lisa Denogean (SAMSI), Anna Oganian (NISS), Mi-JaWoo (NISS) and Francisco Vera (NISS and SAMSI). James Lynch and Jerome Reiter led discussion sessions. Full details are available at www.samsi.info/200506/ndhs/workinggroup/dc/workshops/midyearworkshop.htm.

Social Networks. The workshop was held on March 2, 2006 at Carnegie Mellon University, the principal location of non-Triangle participants in the working group. Attendance was approximately 25, including Canadian participants Shirley Mills and Ted Norminton. Presentations were made by Alan Karr (NISS), Purnamrita Sarkar (CMU), Anna Goldenberg (CMU), Eric Xing (CMU), Eric Kolaczyk (Boston University), Steve Henneke (CMU) and Eric Vance (Duke University).

1.4.3 Transition Workshop

The Transition Workshop for the NDHS program is planned to be held in Research Triangle Park, NC, in October, 2006, in conjunction with the Army Conference on Applied Statistics.

1.5 Education and Outreach

1.5.1 Education and Outreach

A seminar course on National Defense and Homeland Security took place in the fall of 2006 at SAMSI, led by Alan Karr. A total of eight students were enrolled.

1.5.2 Undergraduate Workshop

An undergraduate workshop based on the National Defense and Homeland Security program took place on March 3–4, 2006. There were approximately 30 attendees, one-half of them women. Presentations were made by David Banks, Negash Medhin, Hoan Hguyen, Eric Vance, Ping Bai, Sava Dediu, Anjela Govan, Lisa Denogean, Francisco Vera, David Dickey, and Alan Karr. Details are available at www.samsi.info/workshops/2005ug-workshop200603.shtml.

References

Banks, H. T., Karr, A. F., Nguyen, H. K., and Samuels, Jr., J. R. (2005). Sensitivity to noise variance in a social network dynamics model. *Q. Appl. Math.* To appear. Available on-line at www.samsi.info/reports/index.shtml.

Board on Mathematical Sciences and their Applications (April 26–27, 2002). Workshop on the The Mathematical Sciences' Role in Homeland Security. Information available on-line at www7.nationalacademies.org/bms/Mathml.

Karr, A. F. (2003). Planning meeting: Statistics and national defense and security. *Amstat News*, 308:15.

2. Financial Mathematics Statistics and Econometrics

2.1 Introduction and Overview

The goal of the SAMSI Program on Financial Mathematics, Statistics and Econometrics (FMSE) was to identify short and long term research directions deemed necessary to achieve both fundamental and practical advances in this rapidly growing field and to initiate collaborative research programs — between mathematicians, statisticians, and economists — focused on the multi-disciplinary and overlapping set of fields which involves disciplines such as: Applied Mathematics, Economics and Finance, Econometrics and Statistics. A prominent theme throughout both the workshop and program was the necessity of exploiting the natural synergy between areas of financial mathematics, statistics and econometrics. The goal of the SAMSI program in Financial Mathematics and Econometrics was to bring together these disciplines and initiate a discussion regarding what is really important and what is missing in three essential tasks. (1) **Modeling:** Model development was considered in domains ranging from financial and energy derivatives to real options. (2) **Data:** The size of financial data can be considerable when looking at high frequency data for large numbers of stocks for example. (3) **Computation:** Once a model has been written and calibrated to data, it remains necessary to compute quantities of interest. These three key themes transpired through the entire program and all its activities, including workshops, courses and the diversity of visitors and participants. In particular the program included the following components and activities.

- **Opening Workshop** (September 18-21, 2005): The goal of the opening workshop was to initiate discussion focused on identifying avenues of research in financial mathematics, statistics and econometrics, including models of risk and uncertainty, extreme events, equity, credit, and energy markets; computational and practitioners issues, portfolio optimization. Leading researchers from a variety of fields were asked to provide overview presentations and make recommendations concerning research directions deemed necessary to make significant advances in the field. The keynote speaker was Robert F. Engle, Nobel Prize Laureate in 2003, for his contributions in the field of financial econometrics. This provided a framework for the research directions pursued in the program. Over 160 participants were present at the Opening Workshop.

- **Tutorials** SAMSI Fellows Ronnie Sircar, Princeton University, and Bas Werker, Tilburg University, provided tutorials on Financial Mathematics and Econometrics, respectively.

- **Workshop on Credit Risk** (October 31 - November 2, 2005): Chair: Jean-Pierre Fouque (NCSU). The goal of this workshop was to bring together researchers from academia and the financial industry to present and discuss recent developments in the area of credit risk. Modeling default, correlation of defaults and the associated structured products were the main topics of the workshop.

- **Workshop on Model Uncertainty** (January 27, 2006): Chair: Eric Ghysels (UNC). In this workshop, participants focused on how to introduce uncertainty about the data generating process into asset pricing models and financial decision making.

- **Transition Workshop** (February 27-28, 2006): Participants in the program outlined progress made in the different fields covered by the various workgroups; that is (1) Computational Issues, (2) Credit Risk, (3) L´evy Processes, (4) Model Uncertainty and (5) Portfolio Management.

- **Distinguished Lecture:** Professor Robert F. Engle (Stern Business School, New York University) - Nobel Laureate 2003, "The Spline GARCH Model for Unconditional Volatility and its Global Macroeconomic Causes"

- **SAMSI University Fellows:** Ronnie Sircar (Princeton University) and Bas Werker (Tilburg University)

- **SAMSI Postdoctoral Fellows:** Jesus Rodriguez and Moustapha Pemy

- **SAMSI Young Research Fellow:** Mingxin Xu (UNC-Charlotte)

- **SAMSI Graduate Fellows:** John Hyde (Mathematics, Duke), Arthur Sinko (Economics, UNC), Jennifer Sloan (Statistics, NCSU), Chong Tu (Statistics, Duke), Doug Vestal (Mathematics, NCSU), Yichao Wu (Statistics and Operations Research, UNC)

- **SAMSI Graduate Associate:** Stephen Zhou (Mathematics, NCSU)

- **Short and Long-Term Visitors:** Evan Anderson (NIU), Elena Andreou (University of Cyprus), Fousseni Chabi-Yo (Bank of Canada), Rama Cont (Ecole Polytechnique), Yanqin Fan (Vanderbilt), Kenneth Hochbery (Bar-Ilan University), Aytac Ilhan (Oxford), Kasper Larsen (CMU), Jennifer Juergens (ASU), Knut Solna (UC Irvine), Gordan Zitkovic (Austin, Texas)

- **Working Groups:** Five formal working groups were formed to organize, pursue, and communicate research investigated during the program.
  - Computational Issues: Led by Paul Fackler (NCSU)
  - Credit Risk: Led by Jean-Pierre Fouque (NCSU) and Ronnie Sircar (SAMSI Fellow from Princeton University)
  - L´evy Processes: Led by George Tauchen (Duke)
  - Model Uncertainty: Led by Eric Ghysels (UNC) and Eric Renault (UNC)
  - Portfolio Management: Led by Tao Pang (NCSU) and Mingxin Xu (UNC Charlotte and SAMSI Young Research Fellow)

- **Courses:** Two courses were taught in conjunction with the program.
    - Advanced Topics in Financial Econometrics, Cross-listed at NCSU, UNC and Duke and co-taught by Eric Ghysels (UNC) and Bas Werker (Tilburg University and SAMSI Fellow).
    - Special Topics in Financial Mathematics, Cross-listed at NCSU, UNC and Duke and co-taught by Jean-Pierre Fouque (NCSU), Paul Fackler (NCSU) and Ronnie Sircar (SAMSI Fellow from Princeton University)

- **Technical Reports, Papers, and Books**

- **Presentations**

Details regarding these activities are provided in subsequent sections.

2.2 Program Objectives, Anticipated Outcomes and Recruitment

2.2.1 Objectives

The following program objectives were identified during the Opening Workshop. Details regarding these activities are provided in the various sections in the context of the working groups.

- **Modeling.** In equity markets there is a profusion of models ranging from local volatility to stochastic volatility, multi factors with and without jumps, based on Brownian motions or Levy processes. The situation is similar in fixed income markets with short rates models, HJM or BGM models to name only a few. There is also a variety of discrete time models as the ARCH family for example. The scope of the program focused on addressing the fundamental question of relevance of these models as well as links between physical measures and pricing measures through market prices of risks. Closely connected are the problems of hedging and portfolio optimization which will also be addressed.

- **Data.** The size of financial data can be considerable when looking at high frequency data for large numbers of stocks for example. Data is essential in the modeling part in at least two ways: writing models which quantify the main effects seen in the data (for example, “are jumps present?”) and calibrating the models with an assessment of the stability of the parameters. A lot has been done in this direction in Statistics and Econometrics and to a lesser extent in Applied Mathematics. The program brought these disciplines together, presented the state of the art and discussed issues on choosing, preparing and using financial data. The program also ensured that statistical software companies were involved.

- **Computation.** Once a model has been written and calibrated to data, it remains necessary to compute quantities of interest. For instance, in option pricing, one has to compute expected values along the trajectories (time evolution) of multidimensional stochastic processes. These quantities are also often obtained as solutions of partial differential equations (or inequalities) with various boundary conditions. The program addressed the question of choosing the most efficient computational method for classes of problems. In particular Monte Carlo methods and numerical methods were discussed, keeping in mind that the computational difficulty has a feedback effect on the modeling and data calibrating parts.

2.2.2 Anticipated Outcomes

Expected outcomes at the program level were the following. The methods used to document each outcome are indicated in brackets. A number of the outcomes were still in progress at the program conclusion and post-program evaluation will be performed in these cases.

- Significant research accomplishments by the working groups leading to collaborations and interdisciplinary research with the goal of starting to write papers and research proposals by the end of the program.

- Significant educational benefits for mathematics, statistics, finance and econometrics graduate students, postdocs, and faculty through the two semester-long courses.

- Significant educational experience for SAMSI graduate students and postdocs as well as visiting students and postdocs. This will strongly enhance their perspectives regarding future research directions considered as important by experts in the field.

- Provide a focal point within the mathematics, statistics, finance and econometrics communities for rigorous analysis of financial theory and applications work and advanced materials through the visitor program and program website.

- Dissemination of research directions and results through presentations at international conferences (e.g., NBER/NSF Time Series conference in Montreal September 2006 and Joint Statistical Meetings (JSM) in August 2006), SIAM meetings (Boston, July 2006), AMS meeting (New Orleans, January 2007), articles in society publications (e.g., Journal of Financial Econometrics, Journal of Econometrics, Econometrica, Review of Financial Studies, Mathematical Finance, Applied Mathematical Finance, Finance and Stochastics, SIAM Journals, Wilmott Magazine, Risk Magazine, Quantitative Finance).

2.2.3 Recruitment and Diversity

Two significant goals of the program were to make it as widely accessible to young people as possible and to recruit a diverse range of participants. Both goals were addressed through aggressive solicitation by the program leaders and committee via personal and research contacts as well as formal symposia and presentations. For example, the majority of participants who attended the Opening Workshop were notified by either the organizers or committee. To illustrate, on the first day of the Opening Workshop there were 78 attendees of whom 20 were women and 2 were African

Americans. On the second day, during the inaugural lecture we had approximately 165 attendees with similar proportions of women and minorities. Similar demographics were observed during the remainder of the workshop as well as at the other workshops. For example, the Credit Risk workshop had 27 attendees which included 5 women.

Several formal presentations have also made by the program chairs, postdoc, and members of the SAMSI directorate. As detailed in Section 3.16, Ghysels gave keynote addresses at the Deutsche Bank Conference in honor of Eugene Fama, at the Second AFF and at the EC2 conference on Financial Econometrics. He has also discussed aspects of the SAMSI program at seminars and colloquia throughout the last year and similar presentations are being made by other SAMSI participants, chairs and directors.

2.2.4 Program Conclusions

One of the primary goals of the program was to ascertain the manner through which synergistic mathematical/statistical/econometrics investigations could significantly improve our understanding of financial asset pricing and risk management. A number of different possibilities were explored by the working groups and conclusions can be broadly summarized as follows.

The SAMSI program on FMSE identified a number of key issues of current research in the area. In many regards the program provided the unique platform that started a closer dialogue between mathematicians, statisticians and econometricians on very key topics in financial modeling. Model uncertainty is one such topic. The foundations of probability theory, risk and uncertainty were discussed at length and several key contributions and open questions were identified, notably in a workshop on model uncertainty that included a number of leaders in the field. Another key topic is the identification of jump processes in financial data. At the opening workshop, experts such as Ole Barndorff-Nielsen (Aarhus, Denmark) made several observations on the topic and a workgroup on the topic identified the recent advances in the area. Credit risk is another theme to which much attention was devoted. The program has given the opportunity to a number of young researchers to enter the field of defaultable instruments and produce significant improvements on the modeling of these instruments. The workgroup on computational issues has concentrated on issues concerning solving stochastic control models inmultiple-dimensions. A number of methods were examined and an attempt was made to develop a practical Monte Carlo based approach. The portfolio optimization group met weekly (some times twice a week) at SAMSI to study backward stochastic differential equation and its applications in portfolio management. Stochastic portfolio theory, a new theory about portfolio management, was also studied by the group.

2.3 Program Activities

2.3.1 Program Leadership

Marco Avellaneda (NYU), Jean-Pierre Fouque (NC State; Co-Chair), Eric Ghysels (UNC; Co-Chair), John Lehoczky (Carnegie-Mellon University; National Advisory Committee Liaison), Ronnie Sircar (Princeton), Ralph Smith (SAMSI; Directorate

Liaison), Ruey Tsay (University of Chicago), and Thaleia Zariphopoulou (University of Texas Austin).

2.3.2 Opening Workshop

The SAMSI Opening Workshop for the Program on Financial Mathematics, Statistics and Econometrics was held on September 18–21, 2005. The goal of the workshop was to initiate discussion focused on identifying avenues of research that would run across the entire program. The inaugural lecture of Nobel Laureate Robert Engle touched on the question of how to link stock market volatility to its fundamental sources. This theme resonated through the rest of the program. Another theme was the use of Levy processes in financial modeling. Neil Shephard in joint work with Ole Barndorff-Nielsen reported on the current state of the art and the outstanding challenges. A prominent theme throughout the workshop and the entire program was the improvement of communication between financial mathematicians, statisticians and econometricians. To that end, there as a panel discussion involving Ole Barndorff-Nielsen (University of Aarhus, leading scholar in the fields of statistics and stochastic process theory) Robert F. Engle (New York University Stern School of Business, 2003 Nobel Laureate for contributions to financial econometrics), Lars Peter Hansen (University of Chicago, leading scholar in the field of financial econometrics and expert on model uncertainty) and Ruey Tsay (University of Chicago, leading scholar in the field of financial statistics).

Tutorials on *Financial Mathematics*, and *Financial Econometrics* were offered on the first day of the workshop with 78 participants in attendance. The program during the remaining three days consisted of six focus sessions on topics pertaining to financial mathematics, statistics and econometrics, and a poster session. Each session consisted of two or three 40 minute overview presentations followed by discussions focused on identifying directions necessary to achieve significant scientific advances in that facet of the field. The open forums were expanded in the final session of the workshop to identify objectives and research directions to be pursued in the program.

A total of 161 individuals registered for the workshop of whom 38 were affiliated with mathematics departments, 34 were in statistics departments, and 33 had affiliations in economics and finance. The demographics for the second day are representative of those for the full workshop and can be summarized as follows. There were 132 attendees of whom 27 were women and 2 were African American.

During the open forum on the final day of the workshop, five focus areas were identified and suggested as topics for working groups: (i) Computational Issues, (ii) Credit Risk, (iii) Levy Processes, (iv) Model Uncertainty and (v) Portfolio Management. The first focuses on computational issues including the calculation of complex derivative prices, dynamic portfolio optimization and simulation design and methods. The second topic studied credit risk issues that appear in term structure models, derivative pricing and consumer credit. It was recommended that the third workgroup would study application of Levy process theory to financial problems. This includes extreme events and jump processes. The fourth group emerged from the presentation of Lars Hansen and Jennifer Juergens at the Opening Workshop, both covering methods to introduce model uncertainty in asset pricing models. Finally, it was recommended that the portfolio management group focuses on optimal portfolio allocation. It was stressed during the forum that groups communicate frequently through joint meetings and overlapping participants and this has been promoted during the program.

Details regarding the working groups are provided in Section I.E.2.3.12 and a summary of the full program is provided in Appendix F.

2.3.3 Workshop on Credit Risk (Jean-Pierre Fouque, Organizer)

The workshop started on Monday, October 31st, with a combination of talks from a leader in the industry (Randy Miller from bank of America) to three young researchers in the field of credit risk: Mirela Predescu from Toronto working with John Hull, Kiseop Lee from Louisville, and Nan Chen from Columbia working with Steve Kou. During the second day, leaders in the field from both academia (Ren´e Carmona from Princeton, David Lando from Copenhagen, JP Fouque from NC State, Mirela Preddescu for a second talk, and Ulrich Horst from UBC) and industry (Pierre Collin-Dufresne from Goldman Sachs), presented various ways to attack the problem of modeling default and correlation of defaults. The structural approach and intensity based models were discussed and compared. Some of the talks were data oriented which gave a good opportunity to the young participants to appreciate the role of data and the real difficulties in handling data. On Wednesday morning a new class of models, "top-down" approach was presented by Lisa Goldberg, senior researcher at Barra. A discussion on credit ratings was led by Peter Bloomfield. The last speaker, Thaleia Zariphopoulou (from Austin Texas) presented recent results on dynamic risk measures.

The meeting was a great success and unique in quality of the talks and discussions. A brief summary of the schedule is included in Appendix F.

2.3.4 Workshop on Model Uncertainty (Eric Ghysels, Organizer)

The meeting was a great success scientifically. It brought together the top two scholars in the field - Lars Peter Hansen from the University of Chicago and Thomas J. Sargent from New York University and Hoover Institute at Stanford. They helped to put a program together that put emphasis not only on the various aspects of the topic but also gather a group of young scholars in the field.

The main issue is how does one account for uncertainty regarding the data generating process in pricing asset, portfolio choice, statistical decision making, hypothesis testing pertaining to financial monitoring. The variety of topics touches on the foundations of decision making under uncertainty. This problem is, of course, far from solved and there was a stimulating variety of viewpoints.

The presentations covered a variety of topics, including how to conduct economic policy, like monetary policy, when policy makers do not know the true model, how can learning be used to address model uncertainty, how do we assess the skills of a fund manager, how can we empirically measure the impact of model uncertainty on asset prices, how do we distinguish risk from uncertainty and measure its impact on expected returns. There was much discussion about the notion of 'ambiguity'. This concept relates to environments where one does not know probability laws. One approach relates to robust control with hidden states, the topic of a presentation by Thomas Sargent. Martin

Schneider, a young scholar from New York University, presented related research on asset pricing with ambiguity. Computational aspects were covered by Evan Anderson, a young scholar from Northern Illinois University. Laura Veldkamp, also a young scholar from New York University, talked about how one would adopt investment strategies when assets were included with unknown or ambiguous return processes. Jennifer Juergens, a young scholar from Arizona State University, presented empirical work on model uncertainty using disagreement among financial analysts.

Yossi Feinberg and Mark Henry, both young scholars at respectively Stanford and Columbia, presented work on statistical decision making with uncertainty.

Each presentation, 25 minutes long, was followed by lengthy and penetrating discussions. The workshop was hailed by many participants as an enormous success. One of its key features was the young scholars had the opportunity to present their work with some of the most senior and leading scholars present and commenting on their work. A brief summary is included in Appendix F.

2.3.5 Transition Workshop

The SAMSI Transition Workshop for the Program on Financial Mathematics, Statistics and Econometrics was held on February 27-28, 2006. Participants in the program outlined progress made in the different fields covered by the various workgroups, that is (1) Computational Issues, (2) Credit Risk, (3) Levy Processes, (4) Model Uncertainty and (5) Portfolio Management. A total of 70 applicants registered for the workshop of whom 18 were affiliated with mathematics departments, 17 were in statistics and 6 had affiliations in mechanical engineering, materials science or engineering. Of the participants, 14 were female and 1 was African-American.

To keep the spirit of the SAMSI program of enhancing discussions between mathematicians, statisticians and econometricians, the transition workshop was structured in morning sessions devoted to financial mathematics and afternoon sessions devoted to financial econometrics and statistics. During the first day of the workshop, participants had the opportunity to hear Nicole El-Karoui, a leading female scholar in the field of financial mathematics. She presented material that was linked to the workgroup on portfolio choice and the general theme of the FMSE program. Her talk was followed by a presentation of a well-known practitioner (Bruno Dupire from Bloomberg, Inc.) who discussed volatility models, a topic that was also the topic of the inaugural lecture of Robert Engle. The last talk of the morning was on the topic of “Executive Stock Options” presented by a young female researcher from Princeton. In the afternoon we had several presentations related to three of the workgroups, namely Levy processes, model uncertainty and portfolio choice. Two of those presentations were by female young scholars.

During the second day of the workshop, one of the leading scholars in financial mathematics covered progress on stochastic control theory, and touched on progress that was discussed in the computational and portfolio choice workgroups. The afternoon contained three presentations, all pertaining to progress made in the areas of computational issues.

A summary of the full program is provided in Appendix F.

2.3.6 Distinguished Lecture

Speaker: Robert F. Engle, 2003 Nobel Laureate, New York University

Date: September 19, 2005

Title: "The Spline GARCH Model for Unconditional Volatility and its Global Macroeconomic Causes"

2.3.7 SAMSI University Fellows

Ronnie Sircar (09/01 - 12/15): Sircar taught half of a weekly class based on research pertaining to the SAMSI program. Preparing this class was integral to a book in preparation (with Jean-Pierre Fouque and others) to be finished in 2006. He also worked directly on research papers with Jean-Pierre Fouque, Aytac Ilhan, Erhan Bayraktar and Knut Solna, who also visited at various times. He had many research-related discussions with Jesus Rodriguez, Doug Vestal, Randy Miller and Peter Bloomfield.

Bas Werker (09/17 - 11/25): Werker worked on market microstructure econometrics with E. Renault and SAMSI visitors. He also worked with SAMSI visitor E. Andreou on residual based testing. Finally, a new research project in the area of structural break testing was started with E. Andreou and E. Ghysels. He also continued his work with M.J. Boes and F. Drost on overnight closure periods of financial markets and its impact on option pricing and another project on risk-neutral volatility distribution estimation. Additionally, he taught half a PhD level course in Financial Econometrics.

2.3.8 SAMSI Postdoctoral Fellows

Jesus Rodriguez

Credit Risk: I have presented papers and current research on modeling credit risk during the working group meetings. I led several discussions on topics of interest, and I am now focusing on modeling correlated multiple default derivatives with Jean-Pierre Fouque. In March I was invited to present some of our results at the Mathematics Department Seminar Series at Oregon State University.

Portfolio Management: I led several discussions in the Portfolio Management working group, mainly focused on a new area called Stochastic Portfolio Theory. One of the chief contributors to the theory, Adrian Banner from INTECH, attended the Opening Workshop, and I have remained in contact with him throughout the year. I am currently working with Tao Pang on investigating the behavior of stochastic differential equations in this area.

Multiple Default Derivatives: We are considering modeling correlated multiple default derivatives using structural models which are consistent with short maturity yield spreads. We built on ideas from Stochastic Volatility popularized by Fouque, Sircar, and Papanicolaou, and show how it can be incorporated into complex derivatives in the bond market.

Stochastic Portfolio Theory: I am working with Tao Pang on studying the asymptotic behavior of stochastic differential equations with double reflecting boundaries. We are building on the seminal work done by Banner, Fernholz, and Karatzas on ranked capitalization based portfolios, and removing assumptions in order to work apply the ideas in more generality.

Pricing Issues in Energy Markets: We study models for electricity pricing and derivatives in today's deregulated markets. We find a class of forward price processes that are consistent with the spot price and use these forwards to value options on the spot. We first suppose the forward price follows a quite general diffusion process and show a necessary and sufficient condition on the drift term, which is perfectly analogous to the Heath-Jarrow-Morton condition. Using the forwards, we then show how to price European call options with a portfolio that has holdings in forward contracts and bonds. In addition to standard derivatives, we also consider a "swing" or "take-or-pay" option which is particular to energy markets. It is these swing options that attract the most interest from market participants, as well as the academic literature since the work of Jaillet, Ronn, and Tompaidis. We show how European calls, along with forward contracts, are used to hedge swing options. We devote special attention to the spot model proposed by M. Barlow, and modify it to prevent arbitrage in our context. A paper based on this work was submitted to Mathematical Finance.

Moustapha Pemy

Computational Issues: I was very active in this working group, led by Paul Fackler, and I presented one of my working papers entitled "Quadratically convergent method." I received a lot of useful feedbacks and ideas about how to finish the paper since I still have to figure out how to complete the proof of the main results in this paper. Overall during the semester, we went over different methods to solve problems in Financial Mathematics. I have particularly learned a lot about Monte Carlo methods and how they can be used in a nontraditional way.

Portfolio Management: During the group meetings, we devoted significant time to study the theory of Backward Stochastic Di_erential Equations and how they applied to various problems in finance and stochastic control. This is a new area for me and I will try to incorporate some the ideas I learned in my future research.

Model Uncertainty: This group was led by econometricians, Prof Eric Ghysels and Eric Renault, and the approach they presented to solve various problems in Asset Pricing and Modeling the Volatility was different from that I have learned in the past. This caused me to think how I can use those ideals together with my functional analysis background to generalize some of their discrete models to continuous functional depending models to forecast the volatility and the asset price. I know this approach may be quite involved and may require a good deal of challenging results in Stochastic Functional Analysis.

Personal Research: My postdoctoral experience has been very positive in the sense that it has put in contact with very motivated researchers. My supervisors Tao Pang and Harry Chang have introduced me to the area Stochastic Functional Equations and we have studied various control problems in this field. Bringing into play the theory of Viscosity Solution, so far we have submitted in this period of time four papers for publication and right now we are working on follow-up papers. In addition, I have continued working with Ph.D. advisor on various Selling Rule problems. We have submitted one paper about the liquidation of a large block of stock when the stock price follows the Geometric Brownian Motion to the Journal of Banking and Finance and right now, we are working on the follow-up paper where we generalize the result to include the case where the underlying model is the Regime Switching Model.

2.3.9 SAMSI Young Research Fellow

Mingxin Xu (09/01 – 12/15): Mingxin Xu gave a talk at the opening workshop and another one at the two-day undergraduate workshop (November 19, 2005). She attended the transition workshop and gave presentations and shared the group-leadership of portfolio management. She studied in the new research area of backward stochastic differential equations and stochastic portfolio theory during this period.

2.3.10 SAMSI Graduate Fellows

Please Refer to Section I.C

2.3.11 Short and Long-Term Visitors

Evan Anderson (NIU), Elena Andreou (University of Cyprus), Fousseni Chabi-Yo (Bank of Canada), Rama Cont (Ecole Polytechnique), Yanqin Fan (Vanderbilt), Kenneth Hochbery (Bar-Ilan University), Aytac Ilhan (Oxford), Kasper Larsen (CMU), Jennifer Juergens (ASU), Knut Solna (UC Irvine), Gordan Zitkovic (Austin, Texas).

2.3.12 Working Groups

Credit Risk: This group identified the following high level objectives: CDO/Copula/Computational, Recovery Rate, Top down approach, Hybrid Models, Stochastic Volatility Structural Models, Dynamic Copulas. The group engaged these issues throughout the program.

Group Members: Bloomfield, Peter (North Carolina State University); Chen, Wei (SAS); Cotton, Peter (Morgan Stanley); Fouque, JP (North Carolina State University); Goldman, Irina (Stevens Institute of Technology); Harmantzis, Fotios (Stevens Institute of Technology); Hoe, SingRu (University of Texas at Arlington); Hyde, John (Duke University); Jabri, Hannah (Rice University); Jeanblanc, Monique (University of Evry); Mathur, Apoorv (North Carolina State University); Mattingly, Jonathan (Duke

University); Miller, Randy (Bank of America); Nyamadi, Tsatsu (North Carolina State University); Pong, Der-Chuang (North Carolina State University); Rodriguez, Jesus (North Carolina State University); Sayit, Hasan (University of Houston); Sircar, Ronnie (Princeton University); Sloan, Jennifer (North Carolina State University); Vestal, Doug (North Carolina State University); Xu, Mingxin (University of North Carolina at Charlotte); Zhou, Stephen (North Carolina State University)

Computational Issues: This group concentrated on issues concerning solving stochastic control models in multipledimensions. A number of methods were examined and an attempt was made to develop a practical Monte Carlo based approach.

Group Members: Paul Fackler(North Carolina State University), Jean Pierre Fouque (North Carolina State University), Ron Gallant (Duke University), Eric Hillebrand (Louisiana State University), Abdul Khaliq (Middle Tennessee State University), Peng Liu (North Carolina State University), Giray Okten (Florida State University), Moustapha Pemy (North Carolina State University), Ricky Rambharat (Duke University), Jesus Rodriguez (North Carolina State University), Jeff Scroggs (North Carolina State University)

Levy Processes: The Levy process group was comprised of more than twenty senior and junior researchers from all over the world. Their areas of expertise were probability, statistics, and econometrics. In view of the vast distances separating us, and the scheduling issues, we felt that long distance/electronic presentations were the only practical way to communicate. Scheduling (and financing) a common group meeting was considered infeasible.

The group held five formal presentations (talks), plus extensive and elaborate e-mail exchanges. The topics covered a wide range of important issues from jump-driven stochastic volatility, hedging jump risk, time changed Levy processes, and practical jump detection issues in observed market indices. The unifying theme, of course, was the role of the jump portion of a Levy process in modeling financial time series.

There was the usual, and expected, outward export flow of information from field experts to others, but there was also some intriguing and very stimulating back flows as well. To cite an example, a probabilist provided the economists with some very useful insights in the complications of hedging jump risk. Sometimes one can gain insights into a topic of one’s own domain from others who view that topic from a completely different angle.

The technology was a bit limited, in that we used a common phone hub and viewed the slides on own computers. It worked alright, except that the speaker never received continuous visual feedback regarding the audiences understanding. Speaking into a phone is far different than lecturing to an audience, so speakers were at times unsure of the audience’s level of understanding. The group leader (George Tauchen, Duke Economics) kept everyone on the same page of the set of slides, but the lack of video interaction made for limitations.

From an intellectual standpoint, the group appeared to be very successful. The communications issues were surmountable, and certainly will become much better as technology progresses. Based on the experience, the workshop leader strongly urged SAMSI and NISS to investigate electronic distance conferencing very intensively.

This area continues to be a pole of attraction in the field of finance. Practitioners and theorists regularly meet to discuss how to model, hedge and identify jumps. The workgroup was instrumental in moving the topic ahead.

Model Uncertainty: The model uncertainty working group consisted of Eric Ghysels, Eric Renault, Moustapha Pemy (SAMSI), Arthur Sinko (UNC), and various graduate students from UNC, Duke and NCSU. This group investigated various aspects of model uncertainty. The highlight was the tremendously successful workshop held at the end of January 2006. Other than that, the group met regularly to discuss current research in this field.

- The group was centered on graduate student participation. Students were asked to present state-of-the-art papers.
- The group held five formal presentations (talks), plus extensive and elaborate e-mail exchanges. Topics ranged from model uncertainty and portfolio choice, asset pricing and model selection.
- The intellectual leadership was maintained by three key participants, (1) Eric Ghysels, (2) Eric Renault and (3) Bas Werker. They provided the guidance to the study group and helped identify the issues that graduate students pursued.
- The meetings were leading up to the Model Uncertainty workshop described in detail in Section 2.3.4

Present and Future Research: The SAMSI program left its mark on the field by its January 2006 workshop on the topic. It is probably the single most important workshop on the topic that took place recently.

Portfolio Management: The activities of this group during the program are summarized as follows.

Sept 27, Sept 29, Oct 4 2005, “Backward stochastics differential equations,” N. El Karoui, S. Peng, M.C. Quenez; “Backward stochastics differential equations in Finance, Mathematical Finance,” Vol. 7 No. 1 (1997), Shige Peng; “Nonlinear Expectations, Nonlinear Evaluations and Risk Measures.” The presentation was given by Mingxin Xu who led the discussion.

Oct 11and Oct 13 2005, “Backward stochastics differential equations and Risk Measure.” The presentation was given by Aytac Ilhan who led the discussion.

Oct 18 2005, “Mutual Fund Portfolio Choice in thew presence of Dynamic Flows.” The Presentation was given by Tao Pang.

Oct 25 2005, “The connection between BSDE and PDE and numerical Methods to solve BSDE.” The presentation was given by Tao Pang.

Nov 8 and 15 2005 “The Stochastic Portfolio Theory.” The presentation was given by Jesus Rodriguez.

Dec 1 2005, “Market Diversity and The Distribution of Capital in Equity Markets.” The discussion was led by Adrian Banner.

Group Members: Jose Enrique Figueroa-Lopez (Purdue University), Jean Pierre Fouque (North Carolina State University), Aytac Ilhan (University of Oxford), Tao Pang (North Carolina State University), Moustapha Pemy (North Carolina State University), Jesus Rodriguez (North Carolina State University), Ronnie Sircar (Princeton University), Mingxin Xu (University of North Carolina at Charlotte)

2.3.13 Courses

1. Special Topics in Financial Mathematics

Course Listing: ECG 790M.001, MA 797M.001, ST 810M.001 at NCSU, MATH 390.55 at UNC, STA 294.02 at Duke

Time and Place: Wednesdays, 4:30-7:00 pm, SAMSI/NISS

Instructors: Jean-Pierre Fouque (NCSU), Paul Fackler (NCSU), and Ronnie Sircar (Princeton)

Prerequisites: A first graduate course in financial mathematics such as MA 547 offered at NCSU (or equivalent).

Average Attendance: 20 graduate students, postdocs and visitors

General Description:

The course started with a review of no-arbitrage pricing in complete and incomplete markets in continuous time in the context of equity, fixed income, and credits markets. Stochastic volatility models and implied volatility smiles an skews were introduced. Derivative pricing and hedging when volatility is uncertain were treated in the context of multiscale modeling with asymptotic methods.

The second part of the course was devoted to real option modeling. Real options arise when decision makers can choose among a set of mutually exclusive activities. The option to switch out of a current activity can be valued by solving an optimal switching problem. Such models arise in numerous situations, including American option pricing problems, entry/exit problems, sequential choice problems, job change problems and many more. This course discusses the general decision problem in which real options arise and computational techniques for solving such problems. Examples from the literature were used to illustrate and motivate the discussion. In addition to covering the onestate models that represent the bulk of current applications, recent innovations in formulating and solving multi-state problems were discussed.

In the last part of the course optimal hedging and risk management problems involving derivatives were studied.

2. Advanced Topics in Financial Econometrics

Course Listing: ECG 790O.002, MA797O.002, ST 810O.002 at NCSU, ECON 388.001 at UNC, STA 294.01 at Duke

Time and Place: Tuesdays, 4:30-7:00 pm, SAMSI/NISS

Instructors: Eric Ghysels (UNC) and Bas Werker, SAMSI University Fellow from Tilburg Univ.

Prerequisites: Graduate Time Series, Econometrics, Empirical Finance and Asset Pricing theory.

Average Attendance: 12-15 attendees including 1 faculty member, 1 SAS employee and graduate students.

General Description:
This course covered a selected list of current empirical research topics in finance and related econometric methods. A list of the topics that were covered includes estimation of continuous time processes, MIDAS regressions, stochastic volatility models and term structure models. The coverage of the different topics depended in part on the interest of the students. The purpose of the course was to overview the current developments and to prepare students for research in the areas.

3. Program on Astrostatistics

3.1 Introduction

The principal purpose of the SAMSI program on Astrostatistics is to identify promising research paths for statistical sciences and applied mathematics in problems of observational astronomy, astrophysics and particle physics, and to initiate research on these problems. Astrostatistics is a small but growing field of collaborative researchers who specialize in identifying and developing methods for astronomical research. The problems in astronomy are unique in some way and require new methodologies to analyze massive data streaming from large and federated sky surveys. The program is organized in collaboration with the Center for Astrostatistics (CASt) at Penn State.

3.2 Program Organization

The program is led by Jogesh Babu (CASt, Penn state) who is in residence at SAMSI during January - May 2006. Four Working Groups have been formed, whose principal functions, as in all SAMSI programs, are to organize the research and ensure communication. In addition, two Intensive Research Sessions are organized. These groups function for the entire Spring Semester of 2006. The majority of participants are from outside the Triangle area. All the groups have representation from at least two different fields. Some of these working groups have some common statistical issues and the groups interact in such cases.

Exoplanets working group led by Bill Jefferys (Universities of Texas and Vermont) and Merlise Clyde (Duke University) include: Jogesh Babu (Pennsylvania State University, Department of Statistics), Susie Bayarri (University of Valencia, Department of Statistics and Operations Research), Jim Berger (Duke University, ISDS), Floyd Bullard (Duke University, ISDS), David Chernoff (Cornell University, Department of Astronomy), Pablo de la Cruz (University of Valencia, Observatori Astronomic), Andrew Cumming (McGill University), Gauri Datta (University of Georgia, Department of Statistics), Peter Driscoll (San Francisco State University), Eric Feigelson (Pennsylvania State University, Department of Astronomy & Astrophysics), Debra Fischer (San Francisco State University), Phil C. Gregory (University of British Columbia, Department of Physics and Astronomy), Eric Ford (University of California, Berkeley, Department of Astronomy), Tom Jefferys (Unaffiliated), Michael Last (NISS), Hyunsook Lee (Pennsylvania State University, Department of Statistics), Jaeyong Lee (Seoul National University, Department of Statistics), Tom Loredo (Cornell University, Department of Astronomy), Barbara McArthur (University of Texas at Austin, Department of Astronomy), Raman Narayan (San Francisco State University), Jeff Scargle (NASA Ames Research Center), Alex Wolszczan (Penn State).

Surveys and Population Studies working group led by Tom Loredo (Cornell University) Currently has 23 members, about 1/3 astronomers and 2/3 statisticians: Jogesh Babu (Penn State University), Ruth Barrera (National University of Colombia), Brendon Brewer (University of Sydney), Alanna Connors (Eureka Scientific), David Chernoff (Cornell University), Pablo de la Cruz (Universitat de Valencia), Gauri S. Datta (University of Georgia), Sam Finn (Penn State University), Matthew Fleenor (University of North

Carolina), Martin Hendry (University of Glasgow), Angela Hugeback (University of Chicago), Woncheol Jang (Duke University), Kristofer Jennings (Purdue University), Chunglee Kim (Northwestern University), Hyunsook Lee (Penn State University), Kuo-Ping Li (University of North Carolina), Ji Meng Loh (Columbia University, Tom Loredo (Cornell University), Vicent Martinez (Universitat de Valencia), Francisco Vera (NISS), Martin Weinberg (University of Massachusetts, Amherst), Haywood Smith (University of Florida), David Wittman (University of California).

Source Detection and Feature Detection working group led by David van Dyk (University of California, Irvine) include: Keith Arnaud (NASA, Goddard Space Flight Center), Jim Chiang (GLAST), Alanna Connors (Eureka Scientific), Peter Freeman (CMU), Jiashun Jin (Purdue University), Vinay Kashyap (Smithsonian Astrophysical Observatory), Taeyoung Park (Havard University), Adam Roy (UC Irvine), Jeff Scargle (NASA, Ames Research Center), Aneta Siemiginowska (Smithsonian Astrophysical Observatory), Alex Young (NASA Goddard Space Flight Center), Yaming Yu (University of California, Irvine), Jogesh Babu (Penn State University), Lingsong Zhang (University of North Carolina), Woncheol Jang (Duke University), Rebecca Willett (Duke University), Eric Feigelson (Penn State), Xiao-Li Meng (Havard University), Thomas Lee (Colorado State University).

Gravitational Lensing working group led by Arlie Petters (Duke) include: Charles Keeton (Rutgers University), Christopher Genevese (CMU), Jogesh Babu (Penn State), Ji Meng Loh (Columbia), Brian Rider (Colorado), Nicholas Robbins (Duke, Grad Student), Francisco Vera (NISS, Postdoc), Liliya Williams (Minnesotta), Yaming Yu (U. C. Irvine), Zhengyuan Zhu (UNC).

Intensive Session on Statistical Issues in Particle Physics led by Louis Lyons (Oxford, UK) met in March, with heavy emphasis during March 6-16. The members include:
Michael Woodroofe (University of Michigan), Kyle Cranmer (Brookhaven Lab), Jim Linnemann (Michgan State), Nancy Reid (University of Toranto), Luc Demortier (Rochefeller University), Joel Heinrich (U. Penn), Giovanni Punzi (Scuola Normale Superiore and INFN), Harrison Prosper (Florida State), Pushpa Bhat (Fermi Lab), Bodi Sen (University of Michigan), Jogesh Babu (Penn State), John Hartigan (Yale University), Hyunsook Lee (Penn State).

Intensive Session on Stellar Evolution led by Bill Jefferys (Universities of Texas and Vermont) met during February 20-23. The members include: Ted von Hippel (University of Texas), Steve DeGennaro (University of Texas), Elizabeth Jeffery (University of Texas), Nathan Stein (University of Texas), David van Dyk (University of California, Irvine), Tom Loredo (Cornell), Theodore Arthur Sande (MIT).

3.3 Achieving Diversity

Refer to Section I.H

3.4 Research Progress

Each Working Group is having regularly scheduled meetings/teleconferences throughout the program period. Each Working Group and intensive sessions have developed a detailed research agenda. It is only few weeks since the goals and outcomes document was prepared.

3.4.1 Exoplanets

The kickoff meeting included presentations by Alex Wolszczan (Penn State) and Barbara McArthur (U Texas) presenting up-to-date information on radial velocity and astrometric data. This was followed by a two-week intensive session including with presentations by Eric Ford and Barbara McArthur on methods for current observations (including the discovery of a planet by gravitational lensing).

Initial weekly working group meetings focused on model specification, prior selection and MCMC methods for fitting models to radial velocity data, with a presentation by Eric Ford. Phil Gregory made a presentation to the group on parallel tempered MCMC approaches for parameter estimation and computation of marginal likelihoods for model selection. As part of the working groups' effort to create a testbed of problems, Ford provided actual data sets from two systems for comparing methods and testing code. A third challenging data set was provided by Gregory, in which period is poorly resolved. These data sets have been analyzed independently by Ford, Bullard and McArthur and the results compared. An important result of that effort was identification of bugs in existing code and the effects of model parameterization on MCMC convergence and mixing. This highlighted the importance of using good parameterizations of the problem to improve performance of MCMC methods, particularly when the eccentricity of the planet is small.

Most of the remaining time (the past month) has been devoted to learning more about and understanding various approaches to model selection and the calculation of marginal likelihoods. Bullard has implemented an importance sampler based on a mixture of Student t distributions with location and scale parameters based on 3 local modes from MCMC output. Jim Berger proposed a higher dimensional t mixture with locations at all draws from MCMC output. Initial comparisons by Bill Jefferys indicates that this works extremely well, and requires that only a small number of MCMC draws be used (say 100) to construct the importance sampler. Clyde made a presentation on John Skilling's nested sampling approach for calculating marginal likelihoods and sampling from posterior distributions, and suggested possible improvements using importance sampling densities rather than the prior for proposing draws. Ongoing work is to determine if these methods can be applied in the marginal likelihood calculations for exoplanet models. To date, the marginal likelihood comparisons have involved zero versus 1 planet models; an important issue is the degree to which the methods scale as the number of planets under consideration increases to two, which almost doubles the number of parameters. The working group will prepare a report comparing the different methods with suggestions for best practice.

The group is now turning to discuss issues in adaptive design.

3.4.2 Surveys and Population Studies

The Surveys and Population Studies (SPS) working group has begun its intensive working session, with many group members gathered at SAMSI. In preparation for this session, the group held six weekly teleconferences, with 10 to 15 group members participating in each meeting. After an initial planning meeting, the subsequent meetings involved substantial interaction around the following five formal presentations:

- "Astronomical size-frequency distributions: The role of measurement error" (Tom Loredo, Dept. of Astronomy, Cornell U.)
- "Lutz-Kelker bias and corrections" (Haywood Smith, Dept. of Astronomy, U. Florida)
- "What do astronomers mean by Malmquist bias?" (Martin Hendry, Dept. of Astronomy, U. Glasgow)
- "Selection biases: Truncation and censoring" (Jogesh Babu andWoncheol Jang, Depts. of Statistics, PSU and Duke)
- "Coincidence assessment with astronomical catalogs" (Tom Loredo, Dept. of Astronomy, Cornell U.)

For each of these topics there is an accompanying report or presentation document posted on the SPS working group web site. The first four presentations established a theme that has become a central topic for the SPS intensive working session: merging model-based and design-based approaches for analyzing survey data in astronomy. A unique aspect of astronomical surveys is the combined presence of truncation (often random) and measurement error (often heteroscedastic). Model-based approaches can readily handle both complications, but astronomers' techniques rely on overly restrictive models. Design-based approaches handle truncation in a more robust way than the model-based methods, but no such methods known to astronomers can account for measurement error. Presentations and collaborations in the March working session will address devising methods that marry the rigorous error handling of modeling with the robustness of product-limit estimators.

The presentation on coincidence assessment introduced a very common astronomical multiple testing problem to statisticians. An intriguing direction for future research suggested by statistician participants involves combining false discover rate (FDR) control techniques with Bayesian modeling. FDR may be used (with a high rate) to generate a subset of data for a subsequent Bayesian analysis. This both reduces the size of the data set, making the subsequent analysis more computationally tractable, and establishes an objective prior for the subset.

The SPS group is also exploring methods for calculating marginal likelihoods for low-dimensional models (this is being pursued jointly with the Exoplanets group). Two avenues the group is pursuing are a generalization of the basic trapezoid rule to n-dimensional simplexes found by Delaunay triangulation of MCMC samples, and methods based on recursive bisection of the parameter space.

The SPS group will be embarking on a new topic area during the March session: spatial statistics for understanding the galaxy distribution. Already astronomers and statisticians have met jointly to set the agenda for the initial meetings on this topic.

3.4.3 Source and Feature Detection

Modeling High-Energy Count Data: A common statistical model for the data generation mechanisms of the RHESSI (Solar data), CHANDRA (X-ray) GLAST (gamma-ray) and EGERT (gamma-ray) telescopes has been formulated. The implementation of statistical software for MCMC and EM fitting of this model is nearing completion. Once this software is available, the group will be able to think more seriously about feature detection and error bars in extended sources.

Modeling Optical and Solar Images: Preliminary methods that automate feature detection in solar images are being explored. These methods aim to identify, track, and monitor the evolution of such solar features as flares, plumes, and sun spot groups. Preliminary results using statistical image processing are very promising in that they appear to be able to automate what has required up until now tedious manual labor.

Statistical Issues with Low-Count Data: Power-based and significance-based definitions of upper limits for the intensities (photon counts per unit time) of weak sources have been formulated. Frequentist and Bayesian methods for computing these limits are currently being discussed, developed and compared with more standard techniques in the literature.

3.4.4 Gravitational Lensing

Collaborations between statisticians, mathematicians, and astronomers have begun in the lensing session. Keeton (astronomy) and Petters (mathematics/physics) are collaborating on magnification probability distributions with applications to dark matter substructures on galactic scales; Williams (astronomy) and Zhu (statistics) are working together on statistical methods in cluster mass reconstruction; Petters (mathematics/physics) and Rider probability) are collaborating on the statistics of image counting in stochastic microlensing; Loh (statistics) and Willaims (astronomy) are exploring applications of spatial statistics to the distribution of dark matter structures.

3.4.5 Intensive Session on Stellar Evolution

The period Feb 20-27 was an intensive session at SAMSI where those in residence (van Dyk, Jefferys) worked with Steve de Gennaro, Elizabeth Jeffery and Nathan Stein on various problems. These included: Improving MCMC sampling, handling field stars, handling binary stars (a major breakthrough here, since the group thought of a way of doing this that avoids reversible jump or other tricks to sample on spaces of variable dimension). The collaboration is continuing via weekly video conferences.

3.4.6 Intensive session on Statistical Issues in Particle Physics

The group has grown out of the PHYSTAT series of Conferences on “Statistical problems in Particle Physics, Astrophysics and Cosmology”, and in particular the participation of professors Babu and Feigelson at the Stanford PHYSTAT meeting. The PHYSTAT05 Organizing Committee also endorsed the need for a series of Workshops focused on specific problems.

The very active intensive session had a series of meetings at SAMSI during the week of March 6th to 10th. The topics focused on include:

- Upper limits in the presence of nuisance parameters. Results and properties of Bayesian and Frequentist approaches to this problem were presented by Heinrich and Punzi respectively. The presentations by statisticians Reid andWoodroofe were very valuable to the physicists in the group.
- Multivariate methods for signal/background separation. Almost every analysis in particle physics involves such a procedure. Prosper discussed very recent results of Bayesian neural networks that showed good behavior with a remarkably small number of training events. He also raised interesting theoretical questions about how to test compatibility between various multi-dimensional distributions, such as those used for training multivariate procedures.
- Goodness of fit with sparse multi-dimensional data; p-values; discovery. With the advent of the new LHC accelerator at CERN in 2007, probably the most crucial question will be assessing the significance of any possible signal for the Higgs boson for new physics beyond the Standard Model (e.g. super-symmetry, quark and/or lepton substructure, extra dimensions). This is usually assessed in terms of significance p-values for the null hypothesis of the Standard Model. As with upper limits, nuisance parameters cause problems; the possibilities were discussed by Cranmer and Demortier. Further studies are in hand to compare the methods they described, and their properties. It was particularly valuable for Particle Physicists to be exposed to Bayesian methods. This topic is so crucial that it is likely to be the topic of a further Workshop at CERN.

Michael Woodroofe, John Hartigan, Hyunsook Lee and Louis Lyons remained at SAMSI for the rest of March. This enabled an ongoing series of less formal interactions, including with members of other Working Groups, and with members of the Duke Statistics Department. In particular, the question of anomaly detection is common over a wide variety of subjects, ranging from Astrophysics to Medical studies; the inter-disciplinary discussions that are readily possible in the SAMSI environment are particularly valuable. Particle Physicists in the past have tended to develop their own methods for dealing with the statistical analysis of their data. It was especially valuable to have contact with statisticians who have an understanding of practical statistical problems. As well as those in the Particle Physics Working Group, particular new links have been forged with Statisticians Tom Banks, Jim Berger, David van Dyk and Robert Wolpert. Discussions with the Astrophysicists Bill Jefferys and Tom Loredo have also been most valuable.

The work done at SAMSI is seen as a stepping-stone towards the Workshop at BIRS in Banff in July 2006. It is planned to continue the studies before and during the Banff Workshop, and to produce status reports after that meeting on the 3 topics listed above.

The presentations at the SAMSI Intensive session on Statistical Issues in Particle Physics include: Nancy Reid “Modifications to Profile Likelihood”; Nancy Reid “p-value Functions”; Luc Demortier “p-values from A to P”; Jim Linnemann “False Discovery Rate”; Jim Linnemann “Statistical Software Repository for Particle Physics”; Kyle Cranmer “Discovery in Presence of Nuisance Parameters”; MichaelWoodroofe “Nuisance Parameters”; Pushpa Bhat “Multivariate Methods”; Harrison Prosper “Signal/Background Discrimination in Particle Physics”; Giovanni Punzi “Ordering Rules for the Neyman Construction with

Nuisance Parameters"; Joel Heinrich "Limits and Nuisance Parameters"; Jim Berger "Bayesian Testing"; John Hartigan "Conditioning"; John Hartigan "Stein's Paradox"; John Hartigan "Bayesian Priors"; Louis Lyons "p-values in Particle Physics".

3.5 Graduate Student Involvement

Refer to Section I.C

3.6 Technical Reports

As the working groups were only constituted toward the end of January, it is too early to produce technical reports. The Exoplanets group is planning on two articles:

- Stochastic Computation of Bayes Factors for Model Selection for Exoplanets (potential authors: Jim Berger, Floyd Bullard, Merlise Clyde, Eric Ford, Phil Gregory, Bill Jefferys and others)
- Nested Importance Sampling (Floyd Bullard and Merlise Clyde)

The *Source and Feature Detection* working group expect to prepare a technical report "Statistical Methods for Computing Upper Limits for the Intensity of Weak Astronomical Sources" in the near future.

Tom Loredo along with some of the Surveys and Population Studies working group members is preparing a paper for SCMA IV conference.

3.7 Workshops

3.7.1 Planning Meeting

In order to begin focusing on the research topics for the Astrostatistics Program, a planning meeting was held at NASA Ames Center during 14-15, 2005. Thursday, July 14, was devoted primarily to scientific discussion, including learning about the wide variety of research interests of astronomers, physicists and statisticians. Each participant had roughly 30 minutes in which to describe his/her interests or applications, although a significant portion of this time was reserved for questions and discussion. Friday, July 15 was devoted mostly to discussion of the SAMSI program itself, especially discussion of potential participants and the planning of workshops and events for the semester long program.

The participants included: Jogesh Babu (Program Leader), James Berger (Director of SAMSI), Peter Bickel (NAC Co-Chair), Floyd Bullard, Merlise Clyde, Alanna Connors, Andrew Connolly, Phil Gregory, Fabrizia Guglielmetti, Bill Jefferys, Tom Loredo, Louis Lyons, Fionn Murtagh, Don Richards, Jeff Scargle, Megan Sosey, David van Dyk, Larry Wasserman.

3.7.2 Opening workshop

The January 23-25, 2006 opening workshop for the program attracted 67 attendees from diverse fields including, statistics, astronomy, physics and applied mathematics, and met the goal of informing the composition and activities of the Working Groups. Details of the program are at http://www.samsi.info/workshops/2005astro-workshop200601.shtml. All the presentations at the opening workshop are available at the SAMSI web site.

3.7.3 Education and Outreach

The Astrostatistics Program began with Tutorials from 1/18/2006-1/22/2006, designed to familiarize statisticians with current trends in astronomy and expose astronomers to modern methodologies in statistics and applied mathematics. These were conducted in collaboration with CASt to prepare astronomers and statisticians for the cross-disciplinary presentations at the opening workshop. The three tutorials were:

Bayesian Astrostatistics (led by Tom Loredo, Cornell University). This three-day session included several lectures and practicum classes, by Tom Loredo, Bill Jefferys (Universities of Texas and Vermont) and Philip Gregory (University of British Columbia), teaching 31 participants the basic theory and practice of Bayesian statistics, using examples from astronomy.

Nonparametric statistics and Machine Learning for Astronomers (Chad Schafer and Larry Wassermann, Carnegie-Mellon University). This two-day tutorial introduced astronomers to modern methods in nonparametric statistics including: kernel regression, local polynomial regression, splines, wavelets, adaptive methods, and density estimation. The tutorial included implementation details in the R language. 24 attendees participated in this.

Astronomy for Statisticians (Bill Jefferys of Universities of Texas and Vermont, and Eric Feigelson of Penn State). In this two-day tutorial modern understanding of our universe was reviewed spanning planetary systems, stars, the Milky Way Galaxy, extragalactic astronomy and cosmology. Statistical issues underlying the astronomical studies were emphasized and discussed. 29 attendees participated in this.

All the presentations at the tutorials are available on-line on the SAMSI website.

In addition to the tutorials, a Seminar Course on Astrostatistics is being held during the Program, led by Jogesh Babu.

4. Education and Outreach Program

The SAMSI Education and Outreach (E&O) Program encompasses a variety of activities which have achieved national stature for both their scientific and pedagogical content. The annual activities include two-day Undergraduate Outreach Days held in November and March, a week-long Undergraduate Workshop (UGS) held in May, and the ten-day Industrial Mathematical and Statistical Modeling (IMSM) Workshop for Graduate Students which is held at the end of July. The activities during the reporting period also included the co-sponsored Professional Enhancement Program (PREP).

Undergraduate Outreach Days

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

Financial Mathematics, Statistics and Econometrics: November 18-19, 2005

The first outreach workshop focused on topics from the SAMSI Program on Financial Mathematics, Statistics and Econometrics (FMSE). The students were provided with an overview of SAMSI and the FMSE Program by Ralph C. Smith (SAMSI-NCSU) and Jean-Pierre Fouque (NCSU). During the morning session, J.P. Fouque and Jeff Scroggs (NCSU) gave presentations and demonstrations on discrete and binomial tree models, and Mike Aguilar (UNC Graduate Student) and Minxin Xu (UNC-Charlotte) discussed financial markets and portfolios in the early afternoon. In the final afternoon sessions, students modified software provided by the presenters so they could predict returns from simulated portfolios. During dinner on Friday, members of the directorate and program interacted with students to discuss graduate and career opportunities in the field. The workshop concluded on Saturday with a session on credit risk led by J.P. Fouque. Details regarding this workshop can be obtained at the website http://www.samsi.info/workshops/2005ug-workshop200511.shtml.

There were 30 student participants which included 13 females, 3 African Americans, and 1 Native American.

National Defense and Homeland Security: March 3-4, 2006

The second workshop focused on topics from the year-long SAMSI Program on National Defense and Homeland Security (NDHS). Following an opening session by Alan Karr (SAMSI-NISS) in which the gave an overview of the program, the workshop was comprised of sessions on Social Networks - David Banks (Duke), Negash Medhin (NCSU), Hoan Nguyen (SAMSI Postdoc), Eric Vance (UNC Graduate Student); Agricultural Systems - Ping Bai (UNC Graduate Student), Sava Dediu (SAMSI Postdoc), Anjela Govan (NCSU Graduate Student); Anomaly Detection - David Dickey (NCSU), Francisco Vera (SAMSI Postdoc), Lisa Denogean (SAMSI Postdoc); and Data Confidentiality - Alan Karr (SAMSI-NISS). As with the FMSE outreach workshop in

November, members of the directorate and SAMSI postdocs met with the students during dinner on Friday to discuss graduate and career opportunities.

Details regarding the workshop, including the presentations, can be obtained at the website http://www.samsi.info/workshops/2005ug-workshop200603.shtml.

There were 23 student attendees, including 10 females and 1 African American.

Undergraduate Workshop: May 30 - June 3, 2005

The one-week SAMSI Workshop for Undergraduates focused on mathematical and statistical topics pertaining to inverse problems. During the initial sessions, students were introduced to physical applications involving structural, acoustic and thermal systems as well as the concepts of forward and inverse problems. Both mathematical and statistical models were derived for a prototypical system comprised of a vibrating beam, and significant attention was focused on the formulation and implementation of least squares relations to estimate material parameters given measured data. The tutorials included substantial exposure to MATLAB and routines for numerical integration and optimization. On the final day of the workshops, each student team presented the results they had obtained during the week. The Undergraduate Workshop encompasses three highly unique components.

- All tutorials and sessions were presented by SAMSI graduate students and postdocs under close supervision of Smith, members of the Education and Outreach Committee, and local faculty. This allowed 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 provided students with an intensive introduction to the synergy between applied mathematics and statistics within the context of timely physical applications.
- During one of the sessions, the students were introduced to a variety of experiments and each team collected their own data from the vibrating beam. This exposure to data collection illustrates both the physical basis for models and various mechanisms yielding uncertainty or noise. Whereas a number of aspects were listed as highly positive in exit evaluations, the laboratory experience was one of the most highly ranked experiences.

Full documentation regarding the presentations, tutorials, software, and student presentations can be found at the website http://www.ncsu.edu/crsc/events/ugw05/.

There were 19 SAMSI participants in the workshop including 6 females, 2 African Americans, and 2 Hispanics. To provide a strong outreach component to the workshop, 8 additional students from the NCSU REU on Modeling and Industrial Mathematics were allowed to attend the workshop as a prelude to the REU, which commenced the following week. The 8 REU students were also selected from a national pool of candidates and hence they will disseminate their workshop experience in a number of programs across the country.

Industrial Mathematical and Statistical Modeling Workshop: July 25 - August 2, 2005

The ten-day Industrial Mathematical and Statistical Modeling Workshop for Graduate Students was the 11th in the series and the 3rd 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.

For the 2005 workshop, 41 SAMSI participants were chosen from a pool of 61 applicants - of the 41, 15 were female, 3 were Hispanic, and 2 were African American. To further augment the outreach component of the workshop, 8 additional students were selected from the NCSU REU which concluded the previous week. These students were selected from a national pool of applicants and hence will take their workshop experience back to a wide range of undergraduate programs.

The 49 attendees were divided into 7 teams to investigate current research problems presented by scientists from Intelligent Automation, Inc., Lord Corporation, MIT Lincoln Lab, Novartis Pharma, Novozymes, Inc., PAR Technologies and Progress Energy. Each team gave a 30 minute oral presentation summarizing their results on the ¯nal day of the workshop and written reports were compiled as the SAMSI Technical Report 2005-11 which can be obtained at http://www.samsi.info/reports/index.shtml.

Additionally, the team investigating the project presented by MIT Lincoln Lab continued their research after the conclusion of the workshop and had the following paper accepted as a refereed proceedings: A. Berger, N. Razouk, G. Angelides, A. Bartlett, A. Langville, Z. Li, C. Lipkin, N. Mavinga, E. Perez, E. Tweedy and E. Wheeler, “Locally constrained shortest paths and an application in mission planning," Proceedings of the 44th ACM Southeast Conference, Melbourne, FL, pp. 766-767, 2006. The corresponding poster presentation won a 2nd place “Best Poster Award."

Details regarding the workshop can be found at http://www.ncsu.edu/crsc/events/imsm05/

Mathematics Meets Biology: Competitive Exclusion, Coexistence and Data Fitting, May 25 - 28, 2005

This Professional Enhancement Program (PREP) was held at the University of Louisiana at Lafayette, Lafayette, LA, and co-sponsored by the Statistical and Applied Mathematical Sciences Institute (SAMSI) and the Mathematical Association of America (MAA). This workshop was designed as part of the MAA-PREP continuing research-oriented education series for faculty at primarily teaching universities and colleges. The SAMSI related component (1.5 days of tutorials/hands-on modeling and computations) was carried by the SAMSI Associate Director for Education and Outreach (H.T. Banks) with the assistance of three SAMSI/CRSC Postdocs and Graduate Fellows (Brian Adams, Sarah Grove and Shuhua Hu). The material, originating in earlier SAMSI research programs, focused on modeling, estimation and control of HIV. Special topics included MCMC vs. Prohorov metric formulations of estimation problems in the presence of intra- and inter-patient variability, Expected Maximization and Censored Data algorithms in

HIV longitudinal population patient data, and control via Structured Treatment Interruption (STI) therapies.

There were 28 participants of whom 13 were female and 2 were Hispanic.

Based on evaluations by participants, Directors of the MAA PREP invited SAMSI to co-sponsor and participate in a short course on the material to be offered at the annual Joint Mathematical Meetings in San Antonio, TX. This was successfully done January 12-14, 2006. Pre-enrollment (at 48) and attendance was among the highest of the 20 short courses offered at the meetings.

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.

5. Planning, and Technology Transfer Workshops

5.1 Random Graphs and Stochastic Computation

This was simultaneously a planning workshop for a potential SAMSI program, and a reprise of the activity following on from the 2002-03 SAMSI program on Stochastic Computation. The workshop was held during June 13-14, 2005, organized by Beatrix Jones (Massey University) and Mike West (Duke University), and had two goals:

- To provide a forum for cross-disciplinary communication on hot-topic research in random graphs, graph theory, related random matrix theory, applied statistical modeling with graphical structures and computational methods development related to graphical model analysis – much of this following on from research done in the 2002-03 SAMSI program on Stochastic Computation; and

- To define specific areas for focus for a future SAMSI research program.

The workshop brought together a small number of leading researchers from statistics, applied probability, applied mathematics and computer science, all of whom work on critical aspects of modeling, analysis and computation involving graphical models, graph theory, associated theories of random graphs and random matrices, and applications of statistical and mathematical models involving such structures in scientific and social scientific applications. Speakers provided initial overview/orientation discussion aimed specifically at communication across the subdisciplinary boundaries, as well as highlights of core problems and research challenges from their own disciplinary perspectives. The invited oral presentations focused on communication across the various flavors of mathematical sciences as opposed to focusing wholly and more traditionally on within-subdiscipline research advances. The oral presentations and discussions were complemented by poster sessions and a culminating "open house" discussion to define next steps towards a follow-on plan for a scientifically diverse but also goal-focused SAMSI research program proposal. The following were the key areas of research identified for such a program.

- Open questions of real theoretical and practical import include those of defining new families of parametric models for dependencies and structure in non-Gaussian multivariate data, and also non-parametric extensions of current parametric models. Extensions of current emerging "core" methodologies to models involving structured mixtures of highly constrained Gaussian or discrete/multinomial graphical models offer some obvious immediate opportunities.
- Hierarchical extensions of graphical models, that aim to combine random-effects/multi-level modeling concepts with highly-structured graphical models is an area likely to emerge as of central relevant to the next-stage developments in applied multivariate analysis in higher dimensions.

- Theoretical questions of definitions and methodological/practical ramifications of key but mathematically imprecise concepts of sparsity, component “size” and “distances”, and multiple aspects of “paths through graphs” in statistical graphical models. Relationships with developments in computer science in data base algorithms should be one focus.
- Key theoretical and methodological questions of how to represent “core structure” in data when using graphical models and in the context of model uncertainty that generates many, many candidate graphs with similar support from the data and prior context; issues and concepts of practical versus theoretical relevance of model uncertainty with many. many candidate models are germane here as in other areas, and problems of statistical management of generalized ”equivalence classes” of models need thought.
- Topologies of families of graphical models generating inter-node distance structure and neighborhood size distributions in given families (e.g., power laws) suggests a need for collaborative theoretical work involving applied probabilists.
- Intersections of Bayesian graphical models, Bayes nets and the broad field of structural equations modeling are really open for theoretical and computational synthesis.
- Computational research in the area of generalized variational methods for graphical models seem to be an opportunity for a near-term “focused research group” program, and should involve stochastic computational experts.
- Multiple application arenas for all of the above across various areas in biology and genomics, IT and information sciences, and many social sciences.
- Stochastic and stochastic-deterministic model search that involves “targeted” interest in local structure around a subset of “special” nodes in a graphical model. MCMC, greedy and variational methods can often be simply useless in cases of focus on a “small” subset of target nodes whose neighborhoods are uncertain.
- Theoretical issues of aggregation in graphs. How to probabilistic models of multivariate structure change under aggregation/disaggregation of variables in neighborhoods, cliques or probabilistically generated subsettings? General questions of definitions of aggregations to underlie potential multi-scale graphical models.
- Stochastic computational methods as yet undeveloped or at least quite immature in graphical modeling include fully practicable reversible jump MCMC and also sequential importance sampling and particle filtering methods.
- Connections between graphical modeling in discrete multivariate distributions, large contingency tables studies and algebraic statistics are known through not well-developed and potentially key in many applications (in biology and genetics, for example, as well as existing/emerging application areas in IT/social sciences).
- Stochastic process modeling on graphs. This is an area that might be usefully piloted by SAMSI via statistics, applied probability and applied mathematics collaborations. Graphs are physical models of a system monitored through time, and data arise on measurements of “flows” of communications between nodes. Connections with “network flow” and “tomography”.
- Dynamic (in time) problems involving stochastically time-evolving distributions on graphical models, and time-evolving graphical model structures.

5.2 Data Mining and Machine Learning Technology Transfer Short Course

The first SAMSI Technology Transfer Short Course was held on July 25-29, 2005. It built on the 2003-04 program on Data Mining and Machine Learning (DMML), consolidating and presenting both background material on DMML and research performed during the SAMSI/DMML program. There were approximately twenty attendees.

David Banks of Duke University was the organizer and principal presenter for the course. The main topics were:

- Background and Overview: Nonparametric Regression, Cross-Validation, the Bootstrap
- Key Ideas and Methods: Smoothing, Bias-Variance Tradeoff
- Search and Variable Selection: Experimental Design, Gray Codes, Fitness
- Nonparametric Regression: Heuristics on Eight Methods
- Comparing Methods: Designing Experiments in Data Mining
- Local Dimension: How to Pick Problems Wisely
- Classification: Boosting, Random Forests, Support Vector Machines
- Cluster Analysis: Hierarchical, k-Means, and Mixture Models; SOM
- Issues with Bases: Hilbert Space, Shrinkage, Overcompleteness
- Wavelets: Introduction, Construction, Examples
- Structure Extraction: Regression and Multidimensional Scaling
- Vapnik-Cervonenkis Classes and PAC Bounds

Other presenters were:

- **Jack Liu** (GlaxoSmithKline): "Visualization and Data Mining for Microarrays"
- **J. S. Marron** (UNC-Chapel Hill): "Issues with High Dimensional, Low Sample Size Data"
- **Feng Liang** (Duke): "Model Complexity and Regularization"
- **Merlise Clyde** (Duke): "Bayesian Model Averaging"

In addition to presentations, there were daily computer labs that gave participants hands-on experience with real data and real DMML software.

5.3 Data Assimilation for Geophysical Systems Transition Workshop

The transition workshop for the Data Assimilation program was held at NISS/SAMSI on October 5, 2005. It featured key speakers discussing the central topics covered in the program. It was a very intense day with four main sessions. Each session started with an overview talk of 45 minutes, followed by a series of 15 minute talks describing specific research efforts. The main talks were given by:

- Dennis McLaughlin (MIT)
- Zoltan Toth (NCEP)
- Achi Brandt (UCLA/Weizmann)
- Mark Berliner (Ohio State)

During lunch, each of the working groups from the program reported on their activities and specific progress that had been made. During the afternoon break there was also a discussion led by Istvan Szunyogh on the THORPEX initiative.

5.4 NISS/SAMSI Workshop on Collaborations in the Mathematical Geosciences (CMG)

5.4.1 Executive Summary

This is an excerpted version of the full report of a workshop organized by the National Institute of Statistical Sciences (NISS) and Statistical and Applied Mathematical Sciences Institute (SAMSI) to convene geoscientists, applied mathematical scientists and statistical scientists to discuss and articulate future research issues lying at the interface(s) of the geosciences, the applied mathematical sciences and the statistical sciences.

The focus of the workshop was essentially identical to that of the Collaboration in Mathematical Geosciences (CMG) program at the National Science Foundation (NSF), a joint activity of the Geosciences and Mathematical and Physical Sciences directorates. An explicit purpose of the workshop was to provide input from the geosciences, applied mathematical sciences and statistical sciences communities to NSF regarding possible future research emphases for the CMG program. Funding for the workshop was provided NISS by the Division of Mathematical Sciences.[2]

The workshop was held on October 6 and 7, 2005 in Research Triangle Park, North Carolina. Of approximately 100 attendees, one-third were geoscientists and the remainder were applied mathematical scientists and statistical scientists.

The “information gathering” component of the workshop consisted of two sets of breakout discussions; each accompanied by reports to the full set of participants and associated discussion. One set of breakout discussions was organized by geosciences topics—atmospheric sciences, earth sciences, ocean sciences, polar sciences and space sciences. Deliberately, each was moderated by an applied mathematical or statistical scientist. The second set of breakout discussions was organized by applied mathematical and statistical sciences topics—data assimilation, massive databases, multiscale modeling and representation of uncertainty. Each of these was moderated by a geoscientist.

Perhaps the most striking “meta-observation” about the workshop is the extent to which breakout discussions with rather different foci articulated essentially the same issues. This is most evident across the geosciences breakout discussions (§I.5.4.5.1), but also occurred across the mathematical/statistical sciences breakout discussions (§I.5.4.5.2).

---

[2] Award DMS–0535212, Workshop on Mathematical Geosciences, Alan F. Karr, principal investigator.

Not surprisingly, these cross-cutting themes are formulated primarily in mathematical and statistical terms, with the "cross-cuttingness" representing multiple geoscience contexts. One interpretation is that future CMG solicitations could[3] be structured to focus on cross-cutting issues, perhaps even to the point that CMG proposals involving multiple geosciences might be seen as particularly attractive.

5.4.2 Scientific Findings

The three themes of the CMG program in 2005 were:

- Mathematical and statistical modeling of large, complex geosystems;
- Representing uncertainty in geosystems;
- Analyzing large geoscience data sets.

Two of these recurred in the workshop, albeit with possibly novel variations:

Multiscale modeling, which is implicit (indeed, almost explicit) in the "Mathematical and statistical modeling of large, complex geosystems" theme. The most common version of this issue raised at the workshop was *mismatches between the model scale and data scale*, surely not a new problem, but one to which, with development of both models and statistical tools, there are promising new approaches.

Representation of uncertainty, although—linking with the previous item—the workshop did identify a distinct and seemingly new cross-cutting theme of *representation of uncertainties on multiple scales*.

The third 2005 theme of "analyzing large geoscience data sets" did not come up as an independent issue, although some of the database issues discussed below can be construed as particular manifestations of it. It seems clear large geoscience data sets have become so commonplace that scale in itself is no longer seen as a pressing concern.

The principal "new" themes that were spelled out at the workshop are described next.[4]

Stochastic modeling, which arose independently in the atmospheric, earth and ocean sciences breakout discussions. A more traditional interpretation of this issue is scale-related interfaces between deterministic (PDE) models and stochastic "noise." However, concerns raised at the workshop deal more with approaches that use (Bayesian) statistical methods to accommodate random variations in the parameters of deterministic models.

Coupling models of different geophysical systems, with the clear principal emphasis

[3] As to some degree have past solicitations.

[4] Placing new within quotation marks recognizes the likelihood that some may have been articulated elsewhere.

being coupled atmosphere –ocean models. However, the talk by William Dewar also drew strong attention to coupling of ocean and biosphere models.

This theme appears to represent a natural and important maturation of CMG: rather than address problems lying at the interface of the mathematical/statistical sciences and "one" geoscience, it is now possible to address problems involving multiple geosciences.

Data issues, most of which depart significantly from previous concerns over scale of the data—although scale remains a contributing factor. Rather, the issues identified at the workshop focus on utility of data. For example, many geoscience data sets are structured as relational databases and stored in relational database management systems, but many computational algorithms (for example, to solve partial differential equations) and statistical software systems cannot (easily if at all) handle such data.

Data quality and provenance received repeated mention. The proliferation of data sets is one factor, as is the inevitable fact that many electronically generated and assembled databases never receive the scrutiny that was once part of all statistical analyses. Some of the concerns could be met by sufficiently detailed metadata, but to some degree this simply renames the problem.

Data integration, which in some statistical settings is known as combining information," was also of broad interest. It seems clear that existing technologies cannot cope with problems such as poorly characterized dependences among databases. Interestingly, this theme ties directly to the data quality and coupled models

Model assessment, evaluation, validation and verification, which at one level is not at all a new issue; indeed, one could claim that this is what geoscience modeling is "all about." Nevertheless, essentially all of the breakout discussions (in both sets) raised this issue in some form. Interpretations of the problem ranged from the absence of suitable abstractions and paradigms to scalability of computational tools.

Sharp gradients in physical systems, for which good mathematical and statistical tools are not yet readily available. To some extent this is a multiscale modeling problem, but classifying it only as that obscures its importance.

Tradeoffs among model complexity, data detail and computational power. For example, the data assimilation breakout discussion (§5.4.5.2.1) raised the question of the relative merits of "low-dimensional stochastic modeling combined with a sophisticated data assimilation scheme" as compared to "highdimensional deterministic modeling combined with a "practical" data assimilation scheme."

More generally, any geoscience approach combines data, models and computational power in varying degrees (of detail). Methods for understanding these tradeoffs—and especially software tools supporting informed modeling decisions—would be a major advance.[5]

5.4.3 Program Findings

Although it was not explicitly part of their charge, several of the breakout discussions considered and reported on their perceptions of the past and potential "value added" of the CMG program to the scientific enterprise as well as to collaboration and work force issues.

There was virtual consensus among the participants that CMG has been very successful at achieving its stated objective of fostering collaborations between geoscientists and mathematical and statistical scientists. There was equal agreement that CMG| has *not yet* succeeded to the point that there is sufficient momentum for ongoing collaborations to continue or new ones to be catalyzed.

5.4.4 Workshop Details

The workshop was held on Thursday and Friday, October 6–7, 2005 at the Radisson Hotel Research Triangle Park in Research Triangle Park, NC. It was hosted jointly by NISS and SAMSI.

5.4.4.1 Workshop Purpose

The purpose of the workshop was to articulate future research issues lying at the interface of the mathematical and statistical sciences and the geosciences. Many of these issues were expected to relate strongly to themes of current CMG program. However, a number of variations and entirely new themes emerged at the workshop.

The workshop was designed to emphasize:

Collaboration and interfaces, using a format (see §5.4.4.3) of paired presentations, one by a geoscientist, in the role of "problem owner," the other by a mathematical or statistical scientist.

Extending CMG into areas of the mathematical and geophysical sciences that are not currently reached, principally by inviting participants such areas. Examples are fuzzy logic, topology, geochemistry, biological oceanography, and mineral physics.

Active participation by as many attendees as possible, through means described in §5.4.4.3.

New researchers and under-represented minorities, to whom support was preferentially directed, and who, in the former case, were given dedicated opportunities to participate.

The goals of the workshop were straightforward:

---

[5] This issue is, of course, not unique to CMG.

1. Identify more sharply emerging research issues and approaches at the interface of the mathematical and statistical sciences and the geosciences;
2. Define the new collaborations to address the issues.

5.4.4.2 Organizing Committee

The organizing committee for the workshop consisted of Kayo Ide (UCLA: data assimilation, transport and mixing), Alan Karr (NISS, chair: statistical sciences; computer science), Douglas R. Macayeal (Chicago: climate dynamics; Antarctica), Cecile Penland (Climate Diagnostics Center, NOAA: climate), Jan Sojka (Utah State: space science), Robert van der Hilst (MIT: geophysics) and Mary Wheeler (Texas: applied mathematics). Christopher Jones (North Carolina: applied mathematics) provided significant assistance.

5.4.4.3 Program

The full program of the workshop is included in Appendix F. Principal features were:

Paired presentations, one by a geoscientist, in the role of "problem owner," and one by a mathematical or statistical scientist, responding with his or her identification of central issues from the mathematical and statistical perspective. Some presenters communicated presentations to each other in advance of the meeting. Table 1 summarizes the paired presentations, electronic copies of which accompany this report.

Breakout discussions by geosciences topic, addressing atmospheric sciences, earth sciences, ocean sciences, polar sciences and space sciences. Table 2 lists the moderators of these discussions.

Breakout discussions by mathematics/statistics topic, Table 2 lists the moderators for this second set of discussions.

The program also offered multiple opportunities for all attendees to present scientific material, including

Poster session, preceded by Poster Sales Talks. A total of 15 posters were presented.

Two-Minute Madness using a highly successful format at NISS and SAMSI events. All participants were allowed up to two minutes to place an idea before the workshop.

New Researcher Session, at which four researchers within six years of the Ph.D. gave presentations on their research. These presenters were Peter Huybers (Woods Hole), Samar Khatiwala (Columbia), Frank Li (Penn State) and Partha Routh (Boise State).

5.4.4.4 Participants

Total attendance at the workshop, including six people from NSF, was 100. Approximately two-thirds of these were from the mathematical and statistical sciences and one-third from the geosciences. A complete list of workshop participants appears in Appendix E.

5.4.4 Breakout Discussion Reports

As noted previously, there were two sets of breakout discussions:

By geosciences topics— atmospheric sciences, earth sciences, ocean sciences, polar sciences and space sciences, each moderated by an applied mathematical or statistical scientist.[6] The reports from these are summarized in §5.4.5.1.

By applied mathematical and statistical sciences topics— data assimilation, massive databases, multiscale modeling and representation of uncertainty, each moderated by a geoscientist. The reports from these are summarized in §5.4.5.2.

5.4.5.1 Geosciences Breakout Discussions

The most striking aspect of these reports is the extent to which the same issues were identified in multiple geoscience contests. While to some degree this may reflect the two–to–one numerical dominance of mathematical and statistical scientists among workshop participants, it is still signal and not noise.

5.4.5.1.1 Atmospheric Sciences

This group reported problems grouped into five principal themes: Model building and assessment; Multiscale processes; Coupling processes; Dimension reduction; and Observations. As an overarching "real issue," the group posed uncertainty and its interaction across all other issues.

5.4.5.1.2 Earth Sciences

This group organized its report into eight themes: Stochastic methods; Uncertainty; Data quality; Scale/resolution; Coupled processes; Time issues; Correctness (of models); Collaboration Model.

5.4.5.1.3 Ocean Sciences

This group identified five "grand challenges:" Predicting climate change; How does the ocean fuel hurricanes?; Can we predict regime transitions?; Detection of (un)natural variability; Use and interpretation of Lagrangian data.

5.4.5.1.4 Polar Sciences

Principal issues raised by this small group were: Multiscale modeling and Biology–geophysics interfaces.

5.4.5.1.5 Space Sciences

Issues presented by the group were: Matching objects across multiple databases; Modeling of non-equilibrium processes; Space-time modeling; Large scale simulations.

[6] With the exception of the polar science group, which was small and elected to operate without a moderator.

5.4.5.2 Mathematical/Statistical Sciences Breakout Discussions

Even though these breakout discussions addressed disparate mathematics/statistics topics, there was substantial similarity among their reports.

5.4.5.2.1 Data Assimilation

Virtually in parallel with the atmospheric sciences group, this group saw Nature—Observation—Model interactions as the central focus of CMG. Two specific classes of issues were articulated: Multiscale modeling; Optimization–performance analysis

5.4.5.2.2 Massive Databases

The group prefaced its report by noting that massively parallel computation is currently "not a problem." What the group *did* identify as problems were: Data reduction; Dimension reduction; Storage standardization; The "provenance" of data; Representation of computational models at different scales; Representation of uncertainty

5.4.5.2.3 Multiscale Modeling

As high-level issues, the group identified: (Current) Deficiencies of homogenization; Using scale separation to go beyond homogenization; Effective use of database (and data structures) ideas from computer science in multiscale computing; Data assimilation in multiscale models.

5.4.5.2.4 Representing Uncertainty

The report began with a sound bite summary: "We are primitive in our understanding of, and approach to, uncertainty." This leads to a strong sense that concepts of uncertainty needs to expand; put differently, existing precise definitions are not necessarily used of useful. Some questions raised in this regard were:

- How is uncertainty in models assessed in the context of "competing"—and not nested!—models.
- How are model inadequacies diagnosed?
- How should model complexity be reduced in the face of uncertainty analysis?

5.4.5.3 The CMG Program

As noted previously, there was near-unanimity among participants that CMG *is working, and has not outlived its usefulness.* More specifically, there was a strong sense that

- CMG science is high-quality science, and that CMG is enabling advances that would not have occurred without it.
- NSF and CMG program directors correctly recognize the efforts necessary to initiate cross-disciplinary collaborations and accept that the initial pace of the research may reflect this.
- CMG is fully compatible with the role of training and development of the scientific work force that is present in all NSF programs, although particular attention needs to be paid to the needs of new researchers.

5.5 Latent Variables in the Social Sciences Transition Workshop

On November 10 and 11, 2005, we held another workshop that grew out of the Latent Variables in the Social Sciences SAMSI program. This provided an update on the projects that were started during the previous year. In addition to a number of papers presented by participants in last year's programs, we had Peter Bentler (Psychology and Statistics, UCLA) and Leo Goodman (Sociology and Statistics, UC, Berkeley) as invited speakers. Over 100 participants from North Carolina and the rest of the country attended the workshop. It was a great success.

5.6 Planning Meeting - Experimental Analysis of Algorithms: Interfaces between Statistics and Computer Science (3/4/06 at DIMACS)

A planning workshop on Experimental Analysis of Algorithms was held at DIMACS (Rutgers University, Piscataway, NJ) on April 4, 2006. It was co-sponsored by DIMACS, NISS and SAMSI. The organizing committee consisted of David Johnson (AT&T Labs Research), Alan Karr (NISS), Regina Liu (Rutgers), Joseph Naus (Rutgers), Catherine McGeoch (Amherst College) and Fred Roberts (DIMACS). Approximately a dozen statisticians and computer scientists from academe, research institutes and industry attended.

The goal of the meeting was to introduce experts in experimental algorithmics and in relevant statistical methodologies to one another, in order to identify research topics of common interest to both communities, and to develop plans for research collaborations and events (such as workshops or speaker series or tutorials) that address these topics.

Most of the meeting was in a brainstorming format to discuss: What types of statistical problems do experimenters face in algorithmic research? What techniques of statistics and data analysis are most suitable for addressing these problems? What techiques need to be invented? What problems if any do statisticians see with currently typical papers in experimental analysis of algorithms?

Among promising paths were:

1. Models to predict running time of algorithms as a function of input parameters, especially those that go beyond the traditional "O(N log N)" predictions. Examples of going beyond include estimation of implied constants in "O" and inputs other than just problem size.
2. Characterization of "typical" algorithm inputs, which relates to operational profiling in the software engineering literature.
3. Software tools that implement prediction methods.
4. Possible use of ideas from experimental design and data editing/quality.

Alan Karr has agreed to lead continuing communication among the group. One possible follow-up activity would be a workshop within the 2006-07 SAMSI program on computer models.

F. Industrial and Government Participation

Government and industry participation in SAMSI programs and activities reflects broad interest in the SAMSI vision. The following summarizes participation during 2005-06.

National Defense and Homeland Security: The following individuals were long-term participants in the working groups.

Name	Affiliation	Working Group(s)
Deepak Agarwal	AT&T Labs Research	AD, SN
Howard Burkom	John Hopkins Applied Physics Laboratory	AD
Lawrence Cox	National Center for Health Statistics	AD, DC
Joe Fred Gonzalez, Jr.	National Center for Health Statistics	AD, DC
Myron Katzoff	National Center for Health Statistics	AD, SN
Henry Rolka	National Center for Public Health	AD
KevinWard Drummey	U.S. Department of Defense	AD, SN
Richard Picard	Los Alamos National Laboratory	
Clifford Wang	Army Research Office	
Abera Wouhib	National Center for Health Statistics	

Mid-program workshops drew participants from the Bureau of Labor Statistics, Census Bureau, Energy Information Administration, National Center for Education Statistics, and the National Center for Health Statistics.

Financial Mathematics, Statistics and Econometrics: The working groups had long term participants from Bank of America and Bank of Canada. The courses had individuals from SAS. Individuals from a large number of financial institutions participated in the workshops, as can be seen from the workshop participant lists in Appendix E.

Astrostatistics: Because of the nature of the program, there was not industrial involvement. However, there was significant participation in the working groups from government agencies and laboratories such as NASA-Ames, NASA-Goddard, Smithsonian Astrophysical Observatory, Brookhaven National Laboratory and Fermi National Laboratory. Individuals from these and other such organizations also participated extensively in the workshops.

Latent Variable Modeling in the Social Sciences: Government participation included David Dunson of NIEHS, who led a working group, and workshop attendees from NIH, Statistics Canada, the Census Bureau and the Bureau of Labor Statistics. Industrial participation included Paul Biemer of RTI International, who led a working group, multiple researchers from MetaMetrics – who participated throughout the program in several working groups – and workshop attendees from MetaMetrics, Pfizer, RTI, and Scientific Software International.

Computational Biology of Infectious Diseases: The program enjoyed the participation of individuals from SAS in the Mathematical Genomics for Vaccine Design working group. The transitional workshop held at the end of the program involved the participation of key representatives of the Burroughs Wellcome Fund, which has a strong and active interest in the diseases of the developing world, and of One World Health, a non-profit organization seeking to facilitate the development of drugs for the developing world.

Data Assimilation in Geophysical Systems: Data assimilation is critical in the area of numerical weather prediction that has great societal importance and impact. A number of Government labs have focused efforts in this area. We have made connections through bringing researchers from Government labs to the workshops and to visit SAMSI for periods of time that have afforded a stronger connection with their work and the beginnings of collaborative efforts with our group. Craig Bishop (NRL), Ron Gelaro (NASA), Jeff Anderson (NCAR), Doug Nychka (NCAR) and Zoltan Toth (NCEP) all delivered lectures at the opening workshop. Jeff Anderson gave a tutorial. Carolyn Reynolds (NRL), Chris Snyder (NCAR) and Joe Tribbia (NCAR) delivered lectures at the second workshop, held at IPAM. In addition, Carolyn Reynolds has returned here for a week to deliver an in-depth lecture and consult further with the assembled group at SAMSI. A summer school on Fusing Geophysical Data and Models was held at NCAR in June as part of the SAMSI DA program.

G. Publications and Technical Reports

I. NATIONAL DEFENSE AND HOMELAND SECURITY

Publication and Technical Reports

- Banks, H.T., Karr, A.F., Nguyen, H.K. and Samuels, J.R., Jr. *"Sensitivity To Noise Variance In A Social Network Dynamics Model"* Quarterly of Applied. Mathematics (2006) To appear.

- Karr, A.F., Fulp, W.J., Lin, x., Reiter, J.P., Vera, F., and Young, S. S. *"Secure, Privacy-Preserving Analysis Of Distributed Databases"* Technometrics (2006) Invited paper, to appear.

- Medhin, N. and Hong, C.C. *"A Nonlinear Programming Approach to the Study of Social Networks"* (2006) To be submitted to Neural Parallel and Scientific Computation.

- Medhin, N. and Hong, C.C. *"Positive And Negative Affinities Model For Social Networks"* (2006)

Reports in Preparation

- Agarwal, D., McGregor, A., Phillips, J.M., Venkatasubramanian, S., and Zhu, Z. *"Spatial Scan Statistics: Improved Approximations and a Performance Study"* (2006) In preparation.

- Airoldi, E., Banks, D. L., and Xing, E. *"Latent Space Mixture Models"* (2006) In preparation.

- Denogean, L., Karr, A.F., and Qaqish. B. *"Doubly Randomized Data Swapping"* (2006) In preparation.

- Karr, A.F., Oganian, A., Reiter, J., and Wu, M-J. *"Combinations of SDC Methods"* (2006) In preparation.

- Karr, A.F., Oganian, A., Reiter, J., and Wu, M-J. *"New Measures of Data Utility"* (2006) In preparation.

- Nettel-Aguirre, A. and Chipman, H. *"Mining Transactional Data using Latent Space Social Network Models"* (2006) In preparation.

II. FINANCIAL MATHEMATICS, STATISTICS AND ECONOMETRICS

Publication and Technical Reports

- Anderson, E., E. Ghysels and J. Juergens *"The Impact Of Risk And Uncertainty On Expected Return"* Discussion paper ASU and UNC.

- Andreou, E. and B. Werker *"An Alternative Asymptotic Analysis of Residual-Based Statistics"* Discussion paper Tilburg University.

- Boes, M.-J., F.C. Drost and Bas *Werker "Nonparametric Risk-Neutral Return And Volatility Distributions"* Discussion paper Tilburg University.

- Boes, M.-J., F.C. Drost and Bas Werker *"The Impact of Overnight Periods on Option Pricing"* Journal of Financial and Quantitative Analysis, to appear.

- Chabi-Yo, F., E. Ghysels and E. Renault *"Disentangling the Effect of Heterogeneous Beliefs and Preferences on Asset Prices"* Discussion paper UNC.

- Chen, X. and E. Ghysels *"Intra-day News Impact Curves and Realized Volatility"* Discussion paper UNC.

- Engle, R., E. Ghysels and B. Sohn *"On the Economic Sources of Stock Market Volatility"* Discussion paper NYU and UNC.

- Fouque, JP and Zhou, X. *"Modeling Correlated Defaults: First Passage Model under Stochastic Volatility"* To be submitted (April 2006).

- Fouque, JP, Sircar, R., and Solna, K. *"Stochastic Volatility Effects on Defaultable Bonds"* Applied Mathematical Finance. To appear 2006.

- Koijen, R.S.J., T. Nijman and B. Werker, ``C`` Labor *"Income and the Demand for Long-Term Bonds"* Discussion paper Tilburg University.

- Pang, T., Pemy, M. and Chang, M-H. *"Optimal Control of Functional Stochastic Differential Equations with Bounded Memory"* International Journal of Probability and Stochastic Processes, submitted.

- Pang, T., Pemy, M. and Chang, M-H. *"Optimal Stopping for Stochastic Functional Differential Equations"* SIAM Journal of Optimization and Control}, submitted.

- Pang, T., Pemy, M. and Chang, M-H. *"Viscosity Solutions of Infinite Dimensional Black-Scholes Equation and Numerical Approximations"* Submitted.

- Pemy, M., Yin, G. and Zhang, Q. *"Liquidation of a Large Block of Stocks"* Journal of Banking and Finance, submitted.

- Renault, E. and B. Werker *"Causality Effects in Return Volatility Measures with Random Times"* Discussion paper Tilburg University.

Reports in Preparation

- Fouque, JP and Zhou, X. *"Perturbed Gaussian Copulas"* In preparation.

- Fouque, JP and Rodriguez, J. *"Singular Perturbations for SPDEÕs"* In preparation.

III. ASTROSTATISTICS

Reports In Preparation

- The Exoplanets working group is planning two articles:

  1. Berger, Jim, Floyd Bullard, Merlise Clyde, Eric Ford, Phil Gregory, Bill Jefferys and others (potential authors) *"Stochastic Computation of Bayes Factors for Model Selection for Exoplanets"*

  2. Bullard, Floyd and Merlise Clyde *"Nested Importance Sampling"*

- The Source and Feature Detection working group expect to prepare a technical report *"Statistical Methods for Computing Upper Limits for the Intensity of Weak Astronomical Sources"* in the near future.

- Loredo, Tom, along with some of the Surveys and Population Studies working group members is preparing a paper for SCMA IV conference.

IV. LATENT VARIABLE MODELS IN THE SOCIAL SCIENCES

Publication and Technical Reports

- Banks, H.T., A.F. Karr, H.K. Nguyen, and J.R. Samuels, Jr. *"Sensitivity to Noise Variance in a Social Network Dynamics Model"* SAMSI 2005-10, November 14, 2005

- Beyerlein, Kraig and John R. Hipp *"From Pews to Participation: The Effect of Congregation Activity and Context on Bridging Civic Engagement."* Social Problems 53(1): 97-117. (2006)

- Beyerlein, Kraig and John R. Hipp *"Social Capital, Too Much of a Good Thing? American Religious Traditions and Community Crime."* Social Forces 84(2): 995-1013. (2005)

- Bollen, K.A., Ray, S. and Zavisca, J. *"A Scaled Unit Information Prior Approximation To The Bayes Factor"* (2006a) Submitted for publication.

- Bollen, Kenneth A., Sharon L. Christ, John R. Hipp *"Growth Curve Models"* In Encyclopedia of Social Science Research Methods, edited by Michael Lewis-Beck, Alan Bryan and Tim Futing Liao. Thousand Oaks, CA: Sage Publications. (2003)

- Cai, B. and Dunson, D. "*Variable Selection In Nonparametric Random Effects Model"* ISDS Discussion Paper 2005-16, submitted.

- Dunson, D.B., Palomo, J. and Bollen, K. "*Bayesian Structural Equation Modeling"* Handbook on Structural Equation Models (ed. S.-Y. Lee). Elsevier SAMSI 2005-5, July 27, 2005

- Hipp, J. *"Neighborhood Networks of Social Distance: Do They Predict Neighborhood Satisfaction?"* Presented at the International Sunbelt Social Network Meeting, Redondo Beach, CA February 2005.

- Hipp, John R., Daniel J. Bauer, and Kenneth A. Bollen *"Conducting Tetrad Tests of Model Fit and Contrasts of Tetrad-Nested Models: A New SAS Macro"* Structural Equation Modeling. 12(1): 76-93. (2005)

- Hipp, John R., Daniel J. Bauer, Patrick J. Curran, and Kenneth A. Bollen *"Crimes of Opportunity or Crimes of Emotion: Testing Two Explanations of Seasonal Change in Crime."* Social Forces 82(4): 1333-1372. (2004)

- Hipp, John R. and Kenneth A. Bollen *"Model Fit in Structural Equation Models with Censored, Ordinal, and Dichotomous Variables: Testing Vanishing Tetrads."* Sociological Methodology 33: 267-305. (2003)

- Guang, Guo, and John R. Hipp. *"Analysis of Linear Longitudinal Data."* Pp. 347-368 in New Handbook on Data Analysis, edited by M.A. Hardy. London: Sage. (2004)

- Kamata, A. and Bauer, D. J. *"A Note On The Relationship Between Factor Analytic And Item Response Theory Models"* (2005) Submitted.

- Kamata, A., Bauer, D. J., & Miyazaki, Y. *"Multilevel Measurement Model"* To appear in A. A. O'Connell & D. B. McCoach (Eds.). *Multilevel Analysis of Educational Data*. Information Age Publishing. (2005)

- Palomo, Jesus, David Dunson, and Kenneth A. Bollen *"Bayesian Structural Equation Modeling"* In Handbook on Structural Equation Models.

- Pastor, Jr. Manuel, Jim Sadd, and John Hipp *"Which Came First? Toxic Facilities, Minority Move-in, and Environmental Justice"* Journal of Urban Affairs 23(1): 1-21. (2001)

- Reiter, J.P., T.E. Raghunathan, and S. Kinney *"The Importance of Modeling the Sampling Design in Multiple Imputation for Missing Data, Survey Methodology"* (2006) In revisions.

- Vance, Eric and David Banks *"Agent-Based Methods for Dynamic Social Networks"* Paper presented at 2005 SAMSI transition workshop.

- Vance, Eric *"Quantifying Elephant Social Structure: Using a Bilinear Mixed Effects Model to Elicit Qualities of Elephant Behavior"* JSM (2005)

- Visser, Ingmar *"Depmix, An R Package For Fitting Mixtures Of (Latent) Markov Models On Multivariate Mixed Timeseries Data"* Package and manual with illustrative examples to be downloaded from www.r-project.org.

- Zavisca, J. *"Does Money Buy Happiness in Unhappy Russia?"* Under Review at the American Journal of Sociology.

Reports in Preparation

- Biemer, P. and Zavisca, J. "*Measurement Error In BLS Unemployment Measures"* (2006) In preparation.

- Bollen, K.A. and Kolenikov *"A Specification Test For Heywood Cases In Latent Variable Models"* (2006) In preparation.

- Bollen, K., Ray, S. and Zavisca, J. *"Bayes Factors In Structural Equation Models (Sems): Schwarz's BIC And Other Approximations*" (2006b) In preparation.

- Bollen, K., A., R. Thomas, L. Wang, and John Hipp *"Limited Dependent Variable Models With Covariate Measurement Error: A Consistent Instrumental Variable Estimator"* (2005) In preparation.

- Dunson, D. B. *"Efficient Bayesian Model Averaging In Factor Analysis"* ISDS Discussion Paper2006-03. In preparation.

- Dunson, D.B., Palomo, J. and Bollen, K. "*Bayesian Structural Equation Modeling"* Handbook on Structural Equation Models (ed. S.-Y. Lee). Elsevier, (2005) In preparation.

- Dunson, D.B., Palomo, J. and Zavisca, J. *"Bayesian Model Selection And Averaging In Structural Equation Models"* Psychometrika, (2005) In preparation.

- Hipp, John R. and Andrew J. Perrin *"Nested Loyalties: Local Networks' Effects on Neighborhood and Community Cohesion."* Urban Studies. In preparation. (2006)

- Hipp, John R. and Daniel J. Bauer *"Local Solutions in the Estimation of Growth Mixture Models"* Psychological Methods. In preparation. (2006)

- Kamata, A. & Bauer, D.J. *"A Note On The Relationship Between Factor Analytic And Item Response Theory Models"* Psychometrika, (2005) In preparation.

- Kelly, C.M., Leibig, P.S., Edwards, L.J. *"Factors Predicting the Extent of Nursing Home Regulatory Activity in the 50 States"* (2005) In preparation.

- Kinney, S. and Dunson, D.B. *"Bayesian Fixed And Random Effects Selection For Binary Response Models"* ISDS Discussion Paper 2006-06. (2006) In preparation.

- Kinney, S. and D.B. Dunson *"Fixed And Random Effects Selection In Linear And Logistic Models"* (2006) In preparation.

- Palomo, J. and D. Dunson *"Bayesian Inference and Computational Issues in Structural Equation Modeling"* In preparation.

- Skinner, C. and K. Chantala *"Use Of Weights In Multilevel Modeling"* In preparation.

- Visser, I., Ray, S., Jang, W., Berger, J., Bayarri, S. and Pericchi, L. *"Generalization of BIC"* (2006) Order of authors to be determined. In preparation.

- Zavisca,J, C. Fischer, M. Hout, A. Raftery *"Extension Of Latent Class Analysis With Application To Cultural Clustering In The United States"* In preparation.

V. From Genomes to Global Health: the Computational Biology of Infectious Disease

Publication and Technical Reports

- Biswas, Atanu, Sujay Datta, Jason Fine and Mark Segal, Wiley *"Statistical Advances in the Biomedical Sciences: Clinical Trials, Epidemiology, Survival Analysis and Bioinformatics"* (edited volume of peer-reviewed papers, eds., in press)

- Guido, N., X. Wang, D. Adalsteinsson, D . McMillen, J. Hasty, C ., Cantor, T. Elston, and J. Collins* *"A Bottom-Up Approach to Gene Regulation"* Nature 439:856-860. * equally contributed (2006)

- He, M, Tomfohr, JK, Devlin, BH, Sarzotti, M, Markert, ML, Kepler T.B. "*SpA: Web-Accessible Sprectratype Analysis: Data Management, Statistical Analysis and Visualization"* Bioinformatics, under review. (2005)

- Ickstadt, Katja, Tina Mueller and Holger Schwender *"Clustering and Disrimination Methods for Single Nucleotide Polymorphism Data"* To appear in a special issue of CHANCE on Genomics.

- Kepler, T.B., H.K. Nguyen, J. Webster-Cyriaque and H.T. Banks *"A Dynamic Model for Induced Reactivation of Latent Virus, CRSC-TR05-44"* December 2005 J. Theoretical Biology, submitted.

- Kepler T.B, He M, Tomfohr JK, Devlin BH, Sarzotti M, Markert ML *"Statistical Analysis of Antigen Receptor Spectratype Data"* Bioinformatics, under review (2005)

- Lloyd, A.L, Valeika, S. & Cintron-Arias, C. *"Epidemic Dynamics on Small World Network*s" In: Modeling the Dynamics of Human Disease: Emerging Paradigms and Challenges" (2006)

- Lu, J., Tomfohr J.K., Kepler, T.B. *"Identifying Differential Expression in Multiple SAGE Libraries: An Overdispersed Log-Linear Model Approach"* BMC Bioinformatics, under review (2005)

- Ray, S. and Mallick, B. *"Model Based Principal Component Analysis : Bayesian Approaches"* Submitted (2004)

- Rodriguez, Abel and Dunson, David and Taylor, Jack *"Analysis of DNA Repair Studies Through Bayesian Hierarchical Models for Mixtures"* ISDS Discussion Paper, Duke University. Submitted to JASA Applications and Methods. (2005)

- Rodriguez, Abel and Schmidler, Scott *"Bayesian Structural Alignment of Proteins"* ISDS Discussion Paper, Duke University (2006)

- Rodriguez, Abel and Schmidler, Scott *"Combining Sequence and Structure Information in Protein Alignments"* ISDS Discussion Paper, Duke University (2006)

- Schwender, Holger, Sya Rabstein and Katja Ickstadt *"Do you Speak Genomish?"* To appear in a special issue of CHANCE on Genomics.

- Shim,E., H.T. Banks, and C. Castillo-Chavez *"Seasonality of Rotavirus Infection with it's Vaccination"* SAMSI 2005-9, November 4, 2005

- Tomfohr, J.K., Lu, J., Kepler, T.B. *"Pathway Level Analysis of Gene Expression Using Singular Value Decomposition"* BMC Bioinformatics, under review (2005)

- Wang, X., N. Hao, H. Dohlman and T. Elston *"Computational and Experimental Analysis of Bistability, Stochasticity and Oscillations in the Mitogen Activated Protein Kinase Cascade"* Biophys. J. 90:1961-1978. (2006)

Reports in Preparation

- Cintron-Arias, Ariel, will be coming in fall 2006 to SAMSI as a Postdoctoral Fellow in the program on *Development, Assessment and Utilization of Complex Computer Models*. He will also work with Banks and Lloyd on parameter estimation and network models, and will then be supported on their grants from fall 2007.

- Cintron-Arias, A., C., Castillo-Chavez, C., Bettencourt, L., and Banks, H.T. *"Analysis of Effective Susceptible Population Sizes for Influenza A H3N2"* In preparation.

VI. Data Assimilation for Geophysical Systems

Publication and Technical Reports

- Abramovich, F., Grinshtein, V., Pensky, M. *"On optimality of Bayesian testimation in the normal means problem"* Submitted.

- Alexander, Francis, Greg Eyink, Juan Restrepo *"Accelerated Monte-Carlo for Optimal Estimation of Time Series"* J. Statistical Physics, Vol 119, pp 1331-1345.

- Anderson, J. *"Exploring the Need for Localization in Ensemble Data Assimilation Using an Hierarchical Ensemble Filter"* Submitted to Special Issue of Physica D

- Berliner, L. M. and Wikle, C. *"Approximate Importance Sampling Monte Carlo for Data Assimilation"* Submitted to Special Issue of Physica D

- Bradshaw, D.J., Pensky, M. *"Decision theory based classification of high-dimensional vectors based on small samples"* Test, accepted.

- Bradshaw, D.J., Pensky, M. *"SVM-like decision theoretical classification of high-dimensional vectors"* Submitted.

- Brandt, A. *"Multiscale Methods of Data Assimilation and Feedback Optimal Control"* Submitted to Special Issue of Physica D

- Broecker, J., and L.A. Smith *"On the Importance of Being Proper: Skill Scores for Probability Forecasts"* Submitted to Weather and Forecasting.

- Budhiraja, A., Lee, C. and Chen, L. *"A Survey of Numerical Methods for Nonlinear Filtering Problems"* Submitted to Special Issue of Physica D

- Clarke, L. and L.A. Smith *"Detecting Transparent Noise"* Submitted to Mechanical Systems and Signal Processing.

- Debreu, L. Blayo, B., Simon, E. and De Visme, Y. *"4D Variational Data Assimilation For Locally Nested Models"* Submitted to Special Issue Physica D

- Foley, K. and Fuentes, M. *"A Statistical Framework To Combine Multivariate Spatial Data And Physical Models For Hurricane Surface Wind Prediction"* Invited paper for special issue of the Journal of Royal Statistical Society-Series A.

- Hansen, J. and Penland, C. *"Data Assimilation; Parameter Estimation; Stochastic Differential Equations"* Submitted to Special Issue of Physica D

- Heard, A., Pensky, M. *"Confidence intervals for reliabilty and quantile function with application to NASA Space Flight data"* Submitted.

- Hunt, B.R. *"Efficient Data Assimilation for Spatiotemporal Chaos: a Local Ensemble Transform Kalman Filter"* Submitted to Special Issue of Physica D

- Ihler, A.T., Kirshner, S., Ghil, M., Robertson, A.W., and Smyth, P. *"Graphical Models for Statistical Inference and Data Assimilation"* Submitted to Special Issue of Physica D

- Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C. and Ballabrera-Poy, J. *"4D-Var: Ensemble Kalman Filter; Variance Inflation; Data Assimilation; Model Errors"* Submitted to Special Issue of Physica D

- Khare, S.P. and J.L. Anderson *"An Examination Of Ensemble Filter Based Adaptive Observation Methodologies"* Tellus A, 58, 179-195.

- Khare, S.P. and J.L Anderson *"A Methodology For Fixed Observational Network Design: Theory And Application To A Simulated Global Prediction System"* Accepted for publication in Tellus A.

- Khare, S.P. *"A Theoretical Framework For Ensemble Based Adaptive Observations"* In review with Nonlinear Processes in Geophysics.

- Khare, S.P. and L.A. Smith *"Nonlinear Ensemble Data Assimilation Using Indistinguishable States"* To be submitted to Journal of Atmospheric Sciences.

- Kleeman, R. *"Statistical Predictability In The Atmosphere And Other Dynamical Systems"* Submitted to Special Issue of Physica D

- Lane, Emily, Synte Peacock, Juan Restrepo *"A Dynamic-Flow Box Model and High Lattitude Sensitivity"* Submitted, Tellus

- Lermusiaux, P.F.J. *"Adaptive Modeling, Data Assimilation and Adaptive Sampling"* Submitted to Special Issue of Physica D

- Liu, L., Ide, K. and Jones, C. *"Assimilation Of Lagrangian Data Into Two Layer Point Vortex Systems"* Preprint.

- Mason, S.J., J.S. Galpin, L. Goddard, N.E. Graham, and B. Rajaratnam, *"Conditional Exceedance Probabilities"* Monthly Weather Review, accepted for publication. (2005)

- Mich, N. and Lozier, S. *"On the Temporally-Varying Northward Penetration of Mediterranean Water"* Preprint

- Miller, R.N. *"Topics In Data Assimilation: Stochastic Processes"* Submitted to Special Issue of Physica D

- Peacock, Synte, Emily Lane, Juan Restrepo *"A Possible Sequence of Events for the Generalized Glacial-Interglacial Cycle"* Accepted, Global Biogeochemical Cycles

- Pensky, M., Sapatinas, T. *"Frequentist optimality of Bayes Factor estimators in wavelet regression models"* Submitted.

- Ravela, S.,Emanuel, K.A. and McLaughlin, D. *"Data Assimilation by Field Alignment"* Submitted to Special Issue of Physica D

- Restrepo, J. *"A Path Integral Formulation For Data Assimilation"* Mon Wea Rev, Submitted.

- Reynolds, C.A. *"Impact Of Nonlinearities And Model Error On Pseudo-Inverse Calculations"* Submitted to Special Issue of Physica D

- Reynolds, C.A., *"The Impact On Nonlinearities And Model Error On Pseudo-Inverse Calculations"* Accepted for publication in the special data assimilation volume of Physica D, Nonlinear Phenomena. (2006)

- Salman, H., Kuznetsov, L., Ide, K. and Jones, C. *"A Method for Assimilating Lagrangian Data into a Shallow-Water Equation Ocean Model"* To appear in Monthly Weather Review

- Smith, L. *"Predictability Past, Predictability Present"* To appear as chapter in book edited by Tim Palmer (ECMWF)

- Stuart, A. Hairer, M., Apte, A. and Voss, J. *"A Bayesian Approach to Data Assimilation"* Submitted to Special Issue of Physica D

- Toth, Z. and Pena, M. *"Data Assimilation And Numerical Forecasting With Imperfect Models: The Mapping Paradigm"* Submitted to Special Issue of Physica D

- Wikle, C.K. and Berliner, L.M. *"A Bayesian Tutorial for Data Assimilation"* Submitted to Special Issue of Physica D

- Wunsch, C. and Heimbach, P. *"Practical Global Oceanic State Estimation"* Submitted to Special Issue of Physica D

- Xie, L., Bao, S., Pietrafesa, L., Foley, K., Fuentes, M. "*A Real-Time Hurricane Surface Wind Forecasting Model: Formulation And Verification"* Monthly Weather Review, in press. (2006)

VII. BOOKS AND MISCELLANOUES PUBLICATIONS

- Buche, Robert T., Mansoor A. Haider, Mette S. Olufsen, Ralph C. Smith, and Hien Tran (2005) "*Eleventh Industrial Mathematical and Statistical Modeling Workshop for Graduate Students*" SAMSI 2005-11, July 25-August 2, 2005

- Dunson, D. (editor) "*Model Uncertainty in Random Effects and Latent Variable Models*" Springer. In preparation. (Will have a number of articles from participants in Model Uncertainty Group.)

- Fouque, JP, Papanicolaou, G., Sircar, R., and Solna, K. The following book was written in part during the FMSE program: Volatility Perturbations in Financial Market, Cambridge University Press. In preparation.

- Kang, W., and A.J. Krener *Control Bifurcations* The notes were developed in part while Krener taught the course MA 797R as detailed in Section 3.13. These notes were also typeset during the program.

- Lloyd is developing a textbook on Infectious Disease Modeling, based on his lecture notes for his modeling course. He is currently approaching publishers with a first draft version.

- Smith, R.C. *Smart Material Systems: Model Development* SIAM, Philadelphia, February 2005 The contents of this research monograph reflect certain concepts investigated during the program.

H. Efforts to achieve Diversity

SAMSI puts considerable emphasis on achieving diversity. We are co-sponsoring, with the Department of Mathematics at UNC, Chapel Hill, the Conference for African-American Mathematical Scientists in June, 2006 (more fully described at the end of the section). In 2008, the Blackwell-Tapia Conference will be held at SAMSI. We played a significant role at the SACNAS (Society for the Advancement of Chicanos and Native Americans in Science) meeting in October 2005. Associate Director Jones delivered a special session tutorial on work that covered some of the work achieved during the Data Assimilation Program and we co-sponsored a reception with other institutes at that meeting. There was an opportunity at that reception to explain the nature of SAMSI and the opportunities for minority scientists at all levels.

Coming out of a discussion at the SACNAS meeting, an institutes' diversity coordination committee has been formed under the leadership of Chris Jones (SAMSI) and Helen Moore (AIM). This committee will coordinate diversity initiatives and activities across the mathematical sciences institutes in the US and Canada.

The diversity efforts and successes within the individual programs are outlined below.

National Defense and Homeland Security: Women and new researchers are well represented throughout the program. One of the program leaders is a woman. Five new researchers and five females presented at the Opening Workshop. At the Opening Workshop there were 25 female, five African-American, four Hispanic, and 33 New Researcher participants in attendance. In the Agricultural Systems working group the leader is a woman. The Agricultural Systems working group also has one regular female participant, one female postdoc and one female student. The Anomaly Detection working group has two active female participants, one female postdoc and one female graduate student. The Data Confidentiality working group has three female postdocs and two female students. The Social Networks working group has one female postdoc and one female student. At the Mid-Year Meeting on Anomaly Detection there were six female, two Hispanic, and ten New Researcher participants in attendance. At the Mid-Year Meeting on Social Networks there were five female, one Hispanic, and eleven New Researcher participants in attendance. At the Mid-Year Meeting on Data Confidentiality there were eight female, three Hispanic, and eight New Researcher participants in attendance

Financial Mathematics, Statistics and Econometrics: Two significant goals of the program were to make it as widely accessible to young people as possible and to recruit a diverse range of participants. Both goals were addressed through aggressive solicitation by the program leaders and committee via personal and research contacts as well as formal symposia and presentations. For example, the majority of participants who attended the Opening Workshop were notified by either the organizers or committee. To illustrate, on the first day of the Opening Workshop there were 78 attendees of whom 20 were women and 2 were African Americans. On the second day, during the inaugural lecture we had approximately 165 attendees with similar proportions of women and minorities.

Similar demographics were observed during the remainder of the workshop as well as at the other workshops. For example, the Credit Risk workshop had 27 attendees which included 5 women.

Astrostatistics: A number of women and two African American researchers are participating in all the working groups. A women is the co-leader of the *Exoplanets* working group and two other women are also participating in this group. Two women are participating via teleconference in the *Surveys and Population studies* and *Source and Feature Detection* working groups and four are participating in the *Source and Feature Detection* working group. Two women participated in the *Statistical Issues in Particle Physics intensive session*, and one in the *Stellar Evolution intensive session*.

The program leadership had one African American and one woman.

Latent Variable Models in the Social Sciences: Because of both the nature of the social sciences and strong efforts by SAMSI, diversity in the program was high. In particular, one of the four faculty fellows was an African American; both the postdoctoral fellow and the postdoctoral associate participating in the program were female; one working group was led by an African American; two of the four graduate fellows were women; one of the two tutorial presenters and four presenters at the opening kickoff workshop were women.

Computational Biology of Infectious Disease: The Leadership of the program consisted of one man and two women. Several of the long-term visitors and targeted experts were women – including a SAMSI Distinguished Lecturer – and one of the long-term visitors is Hispanic. Six of the graduate students of this visitor, of various ethnic backgrounds including Hispanic, spent the semester in the program. The class on computational immunology taught by Drs. Kepler and Cowell was attended by graduate students from NCSU, UNC and Duke, and consisted of approximately one-half women and one-half men. One African-American woman was enrolled.

Data Assimilation in Geophysical Systems: The core group at SAMSI was put together with diversity in mind. Of the group of ten senior researchers (University Fellows, Faculty Fellows, New Researcher and Long Term Visitors), six were male and four were female. Among the students, women outnumbered men by three to one. Diversity was only lacking at the rank of postdoc – both of these were male. The Distinguished Lecturer associated with this program was a female distinguished professor from Maryland.

The New Researcher Fellow was an African-American from Spelman College. She said that she benefited enormously from the program by learning a new area but also in the connections that she made. She started working with Susan Lozier of Duke University as a result of her time at SAMSI.

Work coming out of the SAMSI program was presented by program leader Jones at the SACNAS (Society for the Advancement of Chicanos and Native Americans in Science) Annual Conference in a two hour lecture of a special session. This was held in October, 2005.

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

- PREP (May, 2005): Out of 28 participants, 13 were female and 2 were Hispanic.
- Undergraduate Workshop (May - June, 2005): From the 19 SAMSI participants, 6 were female, 2 were African American, and 2 were Hispanic.
- Industrial Mathematical and Statistical Modeling Workshop (July 2005): Out of 41 SAMSI participants, 15 were female, 3 were Hispanic, and 2 were African American.
- 2-Day Undergraduate Workshop (November, 2005): From the 30 participants, 13 were female, 3 were African American, and 1 was Native American.
- 2-Day Undergraduate Workshop (March, 2006): Out of 23 participants, 10 were female and 1 was African American.

12th Annual Conference for African-American Researchers in the Mathematical Sciences: The 12th annual Conference for African-American Researchers in the Mathematical Sciences will take place June 20-23, 2006. It will be run jointly between the Department of Mathematics at the University of North Carolina at Chapel Hill and SAMSI. The first three days will be held on the campus of UNC and the last day at SAMSI. The workshop will include invited speakers, tutorials, and a graduate student poster session. The conference will spotlight the accomplishments of mathematicians from underrepresented minority backgrounds, and is open to all.

Invited speakers: Farrah Jackson Chandler (University of North Carolina Wilmington - Mathematics); Dominic Clemence (North Carolina A and T State University - Mathematics); Ethelbert N. Chukwu (North Carolina State University - Mathematics); Jeffrey Forbes (Duke University - Computer Science); Charles Hagwood (NIST, A.B. degree from NCA&T); Rudy Horne (FSU-Mathematics, former postdoc at UNC-Chapel Hill); Johnny Houston (Elizabeth City College - Mathematics); Otis Jennings (Duke University - Business School); Arlie Petters (Duke University - Mathematics); Alicia Nicki Washington (Aerospace Corporation, PhD. from NSCU); Kimberly Weems (North Carolina State University - Statistics)

Background: In the early 1990s, William Massey of Bell Laboratories (then AT&T, now Lucent Technologies) had an idea for an organization devoted mainly to addressing critical issues involving African-American researchers and graduate students in the mathematical sciences. It was envisioned that this organization would highlight current research by African-American researchers and graduate students in mathematics, strengthen the mathematical sciences by encouraging increased participation of African-Americans and members of other underrepresented groups, facilitate working relations among them, and provide assistance to them in cultivating their careers.

This organization became known as the Conference for African-American Researchers in the Mathematical Sciences (CAARMS). Massey, together with Raymond Johnson, James Turner and others organized the first meeting of the organization (CAARMS1) which was held at the Mathematical Sciences Research Institute in Berkeley, California, June 1995. CAARMS2, organized by Massey and Nathaniel Dean, was held at DIMACS at Rutgers University in Piscataway, the Institute for Advanced Study in Princeton, and the Bell Laboratories and the AT&T Laboratories in Murray Hill, New Jersey, June 26-28, 1996; CAARMS3 was held at Morgan State University in Baltimore, Maryland, and the National Security Agency in Fort Meade, Maryland, June 1997; and the CAARMS4 at Rice University in Houston, Texas, June 1998; CAARMS5 at the University of Michigan in Ann Arbor, June 1999; CAARMS6 at Morgan State University in June 2000; CAARMS7 at Duke University in June 2001; CAARMS8 at Princeton University in June 2002; CAARMS9 at Purdue University in June 2003; CAARMS10 at The Mathematical Sciences Research Institute in Berkeley in June 2004; and, CAARMS11 at The Institute for Pure and Applied Mathematics at UCLA in June 2005.

I. External Support and Affiliates

1. External Support

Kenan Foundation: provided $50,000 of supplementary support.

Astrostatistics: The planning workshop was partly supported by NASA-AMES and the transition workshop is jointly supported with the Center for Astrostatistics at Penn State University.

Computational Biology of Infectious Disease: The Institute for Genome Sciences and Policy (Duke) provided partial funding for the Collective Computational Biology Workshop. Other external funding was secured for follow-on research as mentioned in the report for the computational biology program.

Data Assimilation in Geophysical Systems: This SAMSI program leveraged other supported efforts around the triangle universities. The ONR supported a project at UNC on data assimilation supported one student and one postdoc, both of whom were very much involved in the working groups and the SAMSI course. One of the SAMSI postdocs, Amit Apte, had split support from SAMSI and this ONR grant. In addition, support of Juan Restrepo, who visited for two months, was shared with this grant. An NSF-CMG grant at UNC supported one student and one postdoc, both of whom were actively involved in the program.

The Summer School at NCAR was jointly funded by NCAR and SAMSI. Furthermore, the second major workshop, held at IPAM, was jointly funded by IPAM and SAMSI.

2. Affiliate Involvement

2.1 Background

The NISS Affiliates and NISS/SAMSI University Affiliates programs both significantly interact with SAMSI activities. Indeed, as a benefit of membership, NISS Affiliates and NISS/SAMSI University Affiliates may receive reimbursement for expenses (from accounts set up based on the affiliate contributions) to attend all SAMSI workshops. Through meetings and other activities, the NISS Affiliates and NISS/SAMSI University Affiliates inform the development of SAMSI programs. To illustrate, the National Defense and Homeland Security program during 2005–06 and the planned computer model program for 2006–07, as 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.

NISS Affiliates and NISS/SAMSI Affiliates are listed below.

Corporations: Avaya Labs, Basking Ridge, NJ Aventis Pharmaceuticals, Bridgewater, NJ; Bell Labs–Lucent Technologies, Murray Hill, NJ; General Motors, Detroit, MI; GlaxoSmithKline, Research Triangle Park, NC and Collegeville, PA; ICAGEN Inc., Durham, NC; Merck Research Laboratories,West Point, PA; Metabolon, Inc., Research Triangle Park, NC; Meta-Metrics, Inc., Durham, NC; Nuevolution, Copenhagen, Denmark; RTI International, Research Triangle Park, NC; SAS Institute, Cary, NC; SPSS, Chicago, IL; Telcordia Technologies, Piscataway, NJ; Wyeth, Collegeville, PA; Xerox Innovation Group, Webster, NY

Government Agencies and National Laboratories: Bureau of the Census,Washington, DC; Bureau of Labor Statistics, Washington, DC; Los Alamos National Laboratory, Los Alamos, NM; National Agricultural Statistics Service, Fairfax, VA; National Cancer Institute, Bethesda, MD; National Center for Education Statistics, Washington, DC; National Center for Health Statistics, Hyattsville, MD; National Institute of Standards and Technology, Gaithersburg, MD; National Security Agency, Ft. George W. Meade, MD

NISS/SAMSI University Affiliates: University of California Berkeley, Department of Statistics; Carnegie Mellon University, Department of Statistics; University of Connecticut, Department of Statistics; Duke University, Institute of Statistics and Decision Sciences and Department of Mathematics; Emory University, Department of Biostatistics; University of Florida, Department of Statistics; Florida State University, Department of Statistics; George Mason University; University of Georgia, Department of Statistics; University of Illinois Urbana–Champaign, Department of Statistics; University of Iowa, Department of Statistics; Iowa State University, Department of Statistics; Johns Hopkins University, Department of Applied Mathematics and Statistics; 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 Statistics and Operations Research; University of North Carolina at Chapel Hill, Department of Biostatistics; University of North Carolina at Chapel Hill, Department of Mathematics; Oakland University, Department of Mathematics and Statistics; Ohio State University, Department of Statistics; Pennsylvania State University, Department of Statistics; Purdue University, Department of Statistics; Rice University, Department of Statistics; Rutgers University, Department of Statistics; University of South Carolina, Department of Statistics; Southern Methodist University, Statistical Science Department; Stanford University, Department of Statistics; Texas A&MUniversity, Department of Statistics; Virginia Commonwealth University, Biostatistics

2.2 Affiliate Participation

Every SAMSI program and event during 2005–06 had strong Affiliate participation, nearing one-half of attendees at some workshops. Expenditures from Affiliates Reimbursement Account expenditures to attend SAMSI events exceeded $25,000.

Participation in the National Defense and Homeland Security program and by government and university affiliates was especially deep. Examples include:

- Lawrence Cox, Joe Fred Gonzalez and Myron Katzoff of the National Center for Health Statistics (NCHS) participated throughout the year in working groups in this program.

- Researchers in computer science and statistics from Carnegie Mellon University (CMU) participated throughout the year in the social networks working group of the program.

- NCHS hosted two of the mid-program workshops for the program, and CMU hosted the third. More than one-half of the attendees at each of these three workshops were from affiliates.

An "Affiliates Workshop on Survey Costs," held in Washington, DC on April 18–19, 2006, was co-hosted by the National Center for Education Statistics (NCES), and drew more than 75 attendees from nearly a dozen affiliates.

2.3 Plans for the Future

There is strong affiliate interest in all SAMSI programs, but interest is especially strong in:

- The 2006 Summer Program on Multiplicity and Reproducibility in Scientific Studies;

- The 2006–07 program on Development, Assessment and Utilization of Complex Computer Models;

- The planned 2007–08 program on Risk Analysis, Extreme Events and Decision Theory.

J. Advisory Committees

The five advisory/oversight committees of SAMSI are as follows:

- The Governing Board (GB), which oversees SAMSI's administration, finances, evaluation and partner organization relationships. The GB meets with the Directorate twice a year. The SAMSI Director also has a conference call with the GB Chair and/or GB every other week.

- The National Advisory Committee (NAC) consists of leading national scholars, and is the primary external input into program choice and development. The NAC met with the Directorate, at SAMSI, on November 11, 2005, to review the progress in the current programs and to consider the pre-proposals and proposals that had been submitted for programs in future years. In addition, there are frequent e-mails to the NAC asking for advice concerning developing or new programs. Finally, a member of the NAC serves as a Liaison with each of the Scientific Committees of the major SAMSI programs.

- The Local Development Committee (LDC) consists of leading local scholars, and has a crucial role to play in the involvement of local individuals in SAMSI programs, including the Faculty Release Fellows, the Graduate Associates, and the University Fellows. The LDC met with the Directorate on August 4, 2005.

- The Chairs Committee consists of the chairs of the following departments at the partner universities:
    - Duke: Biostatistics and Bioinformatics, Institute of Statistics and Decision Sciences, Mathematics
    - NCSU: Mathematics, Statistics
    - UNC: Biostatistics, Mathematics, Statistics and Operations Research

  Note that the Chairs are also ex officio members of the LDC. Meetings with the Chairs were held before (and during) the LDC meeting mentioned above.

- The Education and Outreach Committee provides guidance concerning new initiatives in education and outreach and provides contacts – around the nation – for advertising opportunities in education and outreach at SAMSI.

As discussed in the Executive Summary, a new Scientific Liaison Committee is also being created, to advise the NAC (and SAMSI) as to potential future programs. Individuals are currently being selected for this committee.

The membership of each of the current committees is given in the table on the following page.

Committee	Name	Affiliation	Field
Governing Board	Bruce Carney	UNC, Assoc. Dean	Astronomy
	George Casella	U. Florida	Statistics
	Tom Manteuffel	U. Colorado	Applied Mathematics
	Vijay Nair	U. Michigan	Statistics
	John Simon	Duke, Asst. Provost	Chemistry
	Daniel Solomon	NCSU, Dean	Statistics
National Advisory Committee	Mary Ellen Bock (Co-Chair)	Purdue	Statistics
	Lawrence Brown	Pennsylvania	Statistics
	Raymond Carroll	Texas A&M	Statistics
	Carlos Castillo-Chavez(Co-Chair)	Arizona State	Mathematics & Stat
	Rick Durrett	Cornell U	Probability
	Nancy Kopell	Boston U	Mathematics
	Rod Little	U. Michigan	Biostatistics
	David Mumford	Brown U	Applied Mathematics
	Daryl Pregibon	Google, Inc	CS and Statistics
	G.W. Stewart	Maryland	CS and Mathematics
	Mary Wheeler	U. Texas	Math and Engineering
	Bin Yu	U. California Berkeley	Statistics
Local Development Committee	David Banks	Duke	Statistics
	H. Thomas Banks	NCSU	Mathematics
	Lloyd Edwards	UNC	Biostatistics
	Gregory Forest	UNC	Mathematics
	Montserrat Fuentes	NCSU	Statistics
	John Harer	Duke	Mathematics
	Sharon Lubkin	NCSU	Mathematics
	Sally Morton	Research Triangle Intl	Epidemiology
	Richard Smith	UNC	Statistics
	Butch Tsiatis	NCSU	Statistics
	Mike West	Duke	Bioinfomatics & Stats
Chairs Committee	Jianwen Cai	UNC	Biostatistics
	Patrick Eberlein	UNC	Mathematics
	Richard Hain	Duke	Mathematics
	Loek Helmnick	NCSU	Mathematics
	Thomas Kepler	Duke	Biostatistics
	Vidyadhar Kulkarni	UNC	Statistics
	Sastry Pantula	NCSU	Statistics
	Dalene Stangl	Duke	Statistics
Education & Outreach Committee	Negash Begashaw	Benedict College	Mathematical Science
	Carlos Castillo Chavez	Arizona State U	Mathematics and Stat
	Karen Chiswell	NCSU	Statistics
	Cammey Cole	Meredith College	Mathematics and CS
	Wei Feng	UNC Wilmington	Statistics
	Marian Hukle	U Kansas	Biological Sciences
	Negash Medhin	NCSU	Mathematics
	Masilamani Sambandham	Morehouse College	Mathematics

II. Special Reports: Program Plan

A. Programs for 2006-2007

1. Development, Assessment and Utilization of Complex Computer Models

Background: Peter Bickel had initiated initial discussion of a program in the area in the SAMSI planning stage. It was not pursued, however, because several SAMSI programs were being planned that significantly involved complex computer models of processes. Those programs did prove to be very successful, but there has been a gradually growing feeling that a more formal SAMSI program in the area would be a good idea.

Two workshops in 2004 solidified this idea. A January 2004 meeting on *Opportunities at the Statistics-Operations Research Interface* highlighted the area as one needing a major interdisciplinary effort. (Indeed, the write-up below is largely taken from a section – written by Margaret Wright and Jim Berger – of the report of that workshop.)

A second relevant workshop was *The Design and Analysis of Computer Experiments for Complex Systems* held at BIRS in July, 2004. This was a joint workshop between SAMSI and the Canadian National Program on Complex Data Structures (NPCDS) and was, in part, a planning meeting to gauge interest in the area. Strong enthusiasm was also expressed there for a SAMSI program in the area, with significant co-participation by NPCDS.

1.1 Introduction

Mathematical models intended for computational simulation of complex real-world processes are a crucial ingredient in virtually every field of science, engineering, medicine, and business, and in everyday life as well. Cellular telephones attempt to meet a caller's needs by optimizing a network model that adapts to local data, and people threatened by hurricanes decide whether to stay or flee depending on the predictions of a continuously updated computational model.

Two related but independent phenomena have led to the near-ubiquity of models: the remarkable growth in computing power via Moore's law for raw speed and its analogue for data storage; and the matching gains in algorithmic speed and accuracy. Together, these factors have vastly increased the applicability and reliability of simulation-not only by drastically reducing simulation time, thus permitting solution of larger and larger problems, but also by allowing simulation of previously intractable problems.

The intellectual content of computational modeling comes from a variety of disciplines, including statistics and probability, applied mathematics, operations research, and computer science. However, most likely for historical reasons, models in certain application domains tend to be associated with a relatively small set of specific modeling techniques. For example, models of flow in physics rely on partial differential equations;

statistics is the source for models in medicine (such as of life expectancy and drug side effects) that lack equations but for which data are plentiful; and transportation systems are represented with classical models from operations research. As a consequence of the narrow focus dictated by custom, modelers in some application areas may not be aware of the latest developments (or even of well-known features) in models from other domains.

Despite enormous recent progress in formulating and solving computational models, the issue of modeling in its broadest sense remains a grand challenge because there is so much more to be done. A new, unified paradigm is needed that integrates adaptive mathematical formulation, maximally effective use of observed data, and flexible solution methodologies that can cope with highly nonlinear, very large, specially structured problems. Within the overall context of modeling, the interface between operations research and statistics provides a promising platform for addressing the modeling issues listed next.

1.1.1 Integrating Experimental Data

It is increasingly the case that models must make sense of massive amounts of physical data, often of high dimension. Furthermore, depending on the problem, it may be easy, difficult, or impossible to obtain high-quality experimental data that are well matched to the model being developed.

- In instances such as designing airplanes, drugs, and traffic systems, there is a coordinated interplay between an evolving sequence of models and collection of observed data. When further data are needed, they can be obtained more or less to order and on point.

- In other situations, such as modeling the consequences of individual cancer chemotherapy or the effects of “baby boomer” retirements on the budget of the United States, it might be possible in theory to obtain some experimental data, but in practice this would be impractical, inadequate, or too late to be helpful. Under these circumstances, although new data can be gathered, they will necessarily be inadequate as well as of possibly dubious relevance to the questions of interest.

- In still other situations, such as simulating nuclear weapons without testing them, the only experimental data are incomplete, gathered from systems that no longer exist, and subject to deep uncertainty about their quality.

Allowing models to make the best use of available data in all situations is without question a major challenge.

When comparing the results of a model with observed data, it is crucial to know whether there is a guaranteed band of reliability within which changes in the data do not affect the quality of the solution. It is also important to quantify the potential effects on the model output of perturbations in the data, to provide a concrete bound on how much the solution may change. A further issue is related to experimental data of uneven quality; as noted above with respect to simulating nuclear weapons, the model should not be allowed to degrade its overall fidelity to match suspect data.

Physical data can be used in additional ways in developing models, in particular to identify and ameliorate bias. When, as is often the case, a computational model cannot capture all relevant features of a process that changes rapidly with time (for example, the behavior of the Internet), techniques like "parameterization" are needed to incorporate data from the time-varying elements automatically.

1.1.2 Coping with the Varieties of Uncertainty

Essentially all models of real-world processes contain uncertainties. A common form of uncertainty arises from unknown variations in the inputs or internal processes of the model.

The effects on the solution of small local perturbations in the problem data are understood for a relatively small class of optimization problems—e.g., for linear programming—but analysis is needed of the effects of larger perturbations, structured perturbations, and systematic (non-random) perturbation.

A danger of worst-case perturbation analysis is that it may lead to highly unrealistic conclusions, which may be either overly optimistic because large constants compromise bounds of apparently favorable order, or else overly pessimistic because the scenarios for which the worst case is achieved will never occur in practice. Hence methods are required that can characterize the typical effects of perturbation. The recent development of "smoothed analysis" to explain the complexity of various methods from operations research—notably, the simplex method for linear programming—offers hope for a realistic general approach.

Even at their best, some general techniques of accounting for uncertainty are impractical for nonlinear models with huge input spaces or discrete variables. A less obvious issue is that some models contain numbers that must be treated as exact and unvarying—for example, the coefficients in a geometric formula—even when other quantities are subject to uncertainty.

A further source of uncertainty is associated with model parameters that must be fitted or "tuned" to available data. These parameters are usually determined by optimizing a measure of the quality of the model, and for the most part they do not represent anything physical or real, about which the modeler might have some intuition based on experience. Depending on how the model parameters are chosen, their mathematical properties are not necessarily stable with respect to perturbations in the problem data, nor is there any guarantee that they make sense in the overall framework of the model.

1.1.3 Validation

The fidelity of models to reality (often termed *model validation*) is central to their effectiveness in understanding and predicting real phenomena. However, techniques for validating complex models are limited and are often based on informal heuristics. Validation seems straightforward conceptually—data are collected that represent both the inputs and the outputs of the model, the model is run with those inputs, and the outputs are compared with the observed data corresponding to the outputs. In reality, complications abound: observed data may be expensive, scarce or noisy; the model may be so complex or time-consuming that only a few runs are possible, and uncertainty enters the process at every turn.

1.1.4 Model Improvement

Statistics and optimization can help to improve a flawed model—“flawed” in the sense that it does not match reality well—by identifying and rank-ordering potential causes for the flaws and suggesting corrections. For example, if a model contains several components, statistical and optimization methods may be able to determine that a specific set of components (or perhaps one of those components) is most likely to be the major source of discrepancy from the real process. It may also be possible to determine that more attention is needed to an input variable, or that the model is inadequate when an input variable lies in a certain region.

A major challenge is finding tractable diagnostic methods for the typically limited amount of available data. Further issues are related to the cost of model improvement—if a model is running in real time, it may be impractical to explore even a small part of the full space of possible errors.

1.1.5 Screening Experiments

The purpose of many scientific investigations—and a major use of many computer models— is to identify which factors are most influential on the system response. Depending on the context, the initial set of candidate factors may range in size from a few to several hundred or more. When the computational model is expensive or time-consuming to run, correct determination of the most influential factors requires a careful design of experiments, a process that blends statistics and operations research in defining an “optimal” experiment design, creating algorithms to compute an optimal or near-optimal design, and analyzing the responses.

1.1.6 Model Approximation; Fitting in High Dimensions

For complex simulations, statistical analysis may require thousands of model evaluations (for example, using Markov chain Monte Carlo, or “MCMC”). Thus the computational problem to be solved may be the composition of two highly nonlinear and potentially huge computational subproblems—i.e., an intractable problem. But a solution that is often almost as good, or at least good enough for many purposes, can sometimes be obtained through model approximations or “meta-models”.

The ideas of “model reduction”, “dimension reduction”, “order reduction”, “approximation algorithms”, and “surrogate models” are currently active research topics in several areas of statistics, applied mathematics, computer science, and operations research. One idea is to develop an exact but much less complex (or much smaller) model that retains the most important features of the original; another approach is to compute an approximate solution of the original problem that retains a provable degree of closeness to the exact solution; and obviously a combination of these ideas is possible. In all cases, a mixture of rigor and application-specific knowledge is needed to deal with increasingly high-dimensional spaces and data.

1.1.7 Varying Data Types

Models are associated with input data, parameters, and output data, and it is no longer true in today's models that all of these are well-behaved real numbers. Tomorrow's models will certainly contain data with features that stretch (or exceed) today's capabilities, including:

- Numbers that differ drastically in scale, e.g., by twelve orders of magnitude within a single model;
- Data specified by algebraic functions, logical variables, or logical constraints;
- Data chosen from a discrete set;
- Data that are individual-specific, as is typical for current models of disease progression;
- Spatial or image data.

Dealing with combinations of these data types will require new methodologies in statistics and operations research. To take the most basic example, mixed-integer optimization problems, which involve integer and continuous variables, are extremely common in practice; yet reliable techniques are available for solving them only when the problem is a linear program.

1.1.8 Getting Inside the 'Black Box'

Often the computer model is viewed as a 'black box' from the statistical viewpoint, and a key question is how statistical analysis can be brought 'inside the box.' For instance, assessment of the uncertainty of outputs due to uncertainty in inputs can be approached either by simply simulating inputs and studying the variation in the corresponding outputs (the black box approach) or by formally processing the input distributions through the steps of the code for the computer model (perhaps using approximations such as local linearizations), arriving at a distribution for the output. SAMSI is ideally suited for exploring this issue.

1.2 Subprograms and Leadership

Study of computer models needs to take place in the context of actual computer models. But because of the inherent complexity of computer models, and the very different types of such models, it is proposed to have a SAMSI program with sub-programs, focusing on specific computer modeling scenarios and/or specific issues. This approach allows in-depth exploration of specific types of computer models, while maintaining an overall 'SAMSI umbrella' that allows quick transfer of techniques developed in one sub-program to another. The planned sub-programs and their leadership are as follows.

1.2.1 Environmental/Ecological Models Subprogram

The environmental modeling subprogram will deal with 3 problem and research fields at the interface between statistics and environmental modeling: Problems of model calibration in the presence of structural model deficits and input uncertainty, problems of decision-oriented model application under high uncertainty about model structure and parameter values, and problems of universality or transferability of environmental models. Application areas may be hydrological models, climate models and plankton models. Final decisions about contents and application areas will be taken at or after the opening workshop in September.

*Leader*s *Committee*: Peter Reichert (EAWAG and ETH Zurich, Chair, reichert@eawag.ch), Douglas Nychka (NCAR), Jonathan Rougier (U. Durham), Ken Reckhow (Duke U., Local Scientific Coordinator), Montse Fuentes (North Carolina State U., Local Scientific Coordinator), Nell Sedransk (SAMSI, Directorate Liaison).

*Scientifi*c *Committee*: Jim Clark (Duke U.), Leonard Smith (Oxford U.)

*Planne*d *Activities*: At the opening workshop, there will be talks about problems in all three research fields. This should lead to the formation of working groups which will establish the subprograms in the selected research fields. Establishing the work programs will then need some meetings in the fall of 2006 and will probably lead to intermediate workshops in spring 2007 to support the generation of results for the final program workshop.

1.2.2 Subprogram on Uncertainty in Models of Granular Materials: Sources and Consequences

The goal of the subprogram is to develop a better understanding of the variability that appears in -indeed, often dominates -the observed behavior of granular materials during flow and deformation. Ultimately, science seeks a description of granular flow, mathematical models that can be used in applications -applications ranging from building hoppers to transporting ore to predicting the path of a landslide. Although these applications are familiar and have existed for generations, there is no full scientific explanation of the fundamental physical processes activated during granular flows. New insights arise in statistics, mathematics, physics, and engineering. This program helps build a tapestry of science that, in the end, will provide the desired description of the flow of granular materials. This program brings together scientists whose specializations including: experimental evidence of the significant role of fluctuations in granular deformation; statistical mechanical models of the underlying microscopic physics of grain flow; bridging micro-scale physics and macroscopic scale modeling; analyzing and computing macro-scale mathematical models in the face of uncertainty in those models.

Leaders Committee: Bruce Pitman (U. Buffalo, Chair, pitman@buffalo.edu), Luis Pericchi (U. Puerto Rico); Sorin Mitra (UNC-Chapel Hill, Local Scientific Coordinator); Ralph Smith (SAMSI, Directorate Liaison)

1.2.3 Engineering Models Subprogram

The engineering subprogram will study three frequently occurring problem areas in finite-element and other engineering models. These problems are those of Validation, Calibration, and Combining Data from physical experiments and computer experiments. The emphasis will be on applications where the computer models require substantial running times and the physical models are difficult or expensive, so that, in some cases, physical experiments can be conducted for only subcomponents of the desired system or a physical simulator may only be possible for the desired system. The design of both the physical and computer experiments will be of special interest.

Program Leaders Committee: Tom Santner (Ohio State University, Chair), David Higdon (LANL), Angela Patterson (General Electric Research), Mary Fortier (General Motors); Jim Berger (SAMSI, Directorate Liaison)

Scientific Committee: Laura Swiler (Sandia National Labs), Dave Higdon (Los Alamos National Labs), Scott Mitchell (Sandia National Labs), Shih-Chung Tsai (General Motors)

Planned Activities: A pre-program workshop will be held to sort our issues involving available test-bed problems (including confidentiality issues) and issues of communication with the major engineering subgroups that will not be resident at SAMSI. A later workshop is planned in January to solidify the directions of the working groups and to involve additional people.

1.2.4 Biological Modeling Subprogram

This program will focus on two types of biological models. The first is models of the impact of drug resistance on acute viral infections. These models are based on a multi-scale approach, integrating within-host models (i.e. ones that describe infection within a given individual) with between-host (epidemiological) models that describe the spread of infection at the population level. Numerous questions exist in terms of fitting these models to data, validating the models and using them for assessment of the spread of viral infection. The second focus of the subprogram will be on system biological models, especially calibration and sensitivity analysis of large deterministic and stochastic biochemical network simulators. This latter activity will concentrate in the period of March through May, 2007.

Leaders Committee: Darren Wilkinson (U. of Newcastle, Chair, J.Wilkinson@newcastle.ac.uk),

H. Thomas Banks (North Carolina State U., Local Scientific Coordinator), Ralph Smith (SAMSI, Directorate Liaison).
*Planne*d *Activities*: There will be a year-long working group on models of viral infection. In early March there will be a two-day workshop on systems biology models, followed by intensive research sessions in the area.

1.2.5 Methodology Subprogram

This Subprogram will engage in an in-depth treatment of methodological issues that arise in the design, analysis and utilization of computer models across many fields of application. This Subprogram will be evolved in close collaboration with the three disciplinary subprograms (Ecology/Environmental Models, Engineering Models, and Fluid Flow Models), engaging them in an overall research umbrella.

In trying to predict reality (with uncertainty bounds), some of the key issues that have arisen are: use of model approximations (emulators) as surrogates for expensive simulators, for calibration/prediction tasks and in optimization or decision support; dealing with high dimensional input spaces; validation and utilization of computer models in situations with very little data, and/or functional (possibly multivariate) outputs; non-homogeneity, including jumps and phase changes as we move around the input space; implementation and transference methodology to current practice; efficient MCMC algorithms and prior assessments; optimization and design.

*Leader*s *Committee*: Susie Bayarri (U. Valencia, Chair, bayarri@uv.es), Michael Goldstein (U. Durham), Tony O'Hagan (Sheffield Univ.), Jerry Sacks, Henry Wynn (London School of Economics), Robert Wolpert (Duke U., Local Scientific Coordinator), Jim Berger (SAMSI, Directorate Liaison)

1.2.6 Overall Program Leaders Committee

Susie Bayarri (U. Valencia, Chair), Bruce Pitman (U. Buffalo), Peter Reichert (EAWAG), Tom Santner (OSU), Darren Wilkinson (U. Newcastle); Dave Higdon (LANL Liaison), Scott Mitchell (Sandia Liaison), Derek Bingham (Simon Fraser U., Liaison to the Canadian National Program and Complex Data Structures); James Berger (SAMSI, Directorate Liaison), and Mary Wheeler (National Advisory Committee Liaison)

1.3 Participants

Long-term SAMSI Visitors:

- M.J. Bayarri (U. Valencia): will be at SAMSI during entire year; interests in methodology and engineering models.
- Bruce Pitman (U. Buffalo): Resident in Fall, with visits in Spring; interested in granular flow (e.g. avalanche) models and renal (blood flow) models.

- Peter Reichert (Swiss Federal Institute of Aquatic Science and Technology): will be at SAMSI during entire year; part of leadership in environmental modeling subprogram (Aquatic models).
- Jonathan Rougier (Univ. Durham): February through May; interested in climate change prediction
- Jerry Sacks: September and October, with later visits.
- Tom Santer (Ohio State); September -mid-March, interested in biomedical engineering models.
- Chungsheng Ma (Wichita State Univ.): visiting for entire year; interested in the application of spatial statistics to computer modeling.
- Darren Wilkinson (U. Newcastle): 3 months in Spring
- Xinlei Wang (Southern Methodist Univ.): planning to be resident during the Spring
- Colin Fox (U. Auckland): frequent short visits from Colorado, during visits there
- Michael Goldstein (Univ. Durham): September and Easter week;
- Ken Hansen (LANL): TBD
- Nick Hengartner (LANL): can perhaps visit for 4 weeks; interested in agent-based computer models
- Tony O'Hagan (Sheffield Univ.): October and April, plus workshops; interested particularly in modularization of systems and surrogates for heterogeneous models
- Angela Patterson (General Electric Research) will be a regular visitor throughout the year; interests include design an optimization with surrogates; Engineering models.
- Rui Paulo (U. of Bristol). September 2006; May -?, 2007.
- Luis Pericchi (Univ. Puerto Rico): possible sabbatical; interested in engineering, environment, and biomedical models.
- Steve Pollock (U. Michigan): TBD
- Leonard Smith (London School of Economics): interested in model uncertainty in dynamical models.
- David Steinberg (Tel Aviv Univ.); 3 months, possibly February -April, 2007. Interest in Design, computer experiments, Bayes smoothing and seismology.
- Henry Wynn (London School of Economics): will visit for September 06 and one later month. Interests in experimental design and model surrogates.

Many others have expressed interest in the program, and are being contacted to determine the potential for their involvement. These include Andrew Booker (Boeing), Peter Challoner (Southampton U.), Hugh Chipman (Acadia U.), Dan Cornford (Aston U.), Dennis Cox (Rice), John Dennis (Rice), Gary Foley (EPA), Marc Kennedy (Sheffield U.), Max Morris (Iowa State), William Notz (Ohio State U.), Jeremy Oakley (Sheffield U.), Shane Reese (Bringham Young U.), Randy Sitter (Simon Fraser U.), William Welch (U. British Columbia), Margaret Wright (NYU), and Kenny Ye (SUNY Stony Brook).

Short term visitors: Jeff Wu (Georgia Tech) two weeks; many TBD

New Researcher Fellow: Serge Guillas (Georgia Tech) -functional data analysis

Postdoctoral Fellows and Postdoctoral Associates: Curtis Storlie, Guillaume Vernieres, James Crooks, Elaine Spiller, Cari Kaufman, Ariel Cintron-Arias, Gentry White

Faculty Releases from Triangle Universities: Sorin Mitran and Jingfan Huang (Math, UNC); Robert Wolpert (Stat, Duke); Montse Fuentes (Stat, NCSU)

Affiliates and National Lab involvement:

- General Motors has significant interests in this area, and could be a good source of people, models and data sets. Interested individuals include Mary Fortier and Shih-Chung Tsai.
- LANL will be heavily involved in the program; indeed David Higdon (Head of Statistical Sciences) will be a program leader. Other possibly involved individuals include Mark Anderson, Scott Doebling, Michael Hamada, Ken Hanson, Francois Hemez, Nick Hengartner, Randy Michelsen, Lisa Moore, Charlie Nakhleh, Amanda Rutherford, and Jim Smith.
- Many university affiliates are heavy players in the area.
- Sandia has a large and active group in the area, including John Helton, Scott Mitchell (who is a program leader), William Oberkampf, Martin Pilch, Laura Painton Swiler, and Tim Trocano.
- NCAR has significant interests in the area and Doug Nychka is a program leader.

International involvement: Eurandom, the European equivalent of SAMSI, has expressed interest in the area, and this will be actualized through Henry Wynn.

1.4 Activities and Timing

- A SANDIA workshop, tentatively planned for May 22-23, 2006 in Albuquerque, on Validation Challenge Problems, will be an excellent way to jump-start interactions and assess the current state of validation.
- There will be a short course, for students, postdocs and others August 11-16, 2006 at Simon Fraser University.
- The opening workshop will be September 10-14, 2006.
- A planning workshop or week will be held during Spring 2007, for the possible future SAMSI program *Agent-Based Modeling in the Social Sciences*. (Potential workshop leaders include Nick Hengartner, LANL; Lisa Moore, LANL; Judith Lessler, RTI; Igor Mezic, UCSB).
- There will likely be an intensive mini-program on *Algorithms and OR Models*, run jointly with DIMACS and NISS.
- The subprograms will operate a semester, all year, or for a smaller interval of time. Each will have its own workshops/activities.
- Working groups will pursue research in the area throughout the year.

2. High Dimensional Inference and Random Matrices

Random matrix theory lies at the confluence of several areas of mathematics, especially number theory, combinatorics, dynamical systems, diffusion processes, probability and statistics. At the same time random matrix theory may hold the key to solving critical problems for a broad range of complex systems including: biophysics, quantum chaos, signals and communication theory, machine learning, finance and geophysical modeling. These applications will raise the issue of statistical inference in high dimensional systems and a central theme of the program will be to see how recent advances on random matrices may inform such inference. In this program, the researchers involved will explore the interplay of stochastic and mathematical aspects in random matrix theory and application.

2.1 Research Foci

The aim of the program will be to bring together researchers interested in the theory and applications of random matrices to share their results, discuss new research directions and develop collaborations. The focus of the program will be on large-dimensional random matrices and the problems which make use of them.

At least three broad challenges stand out: (1) Furthering our understanding of the spectral properties of random matrices under various models and assumptions (this is a "direct" problem); (2) Recovering from an observed random matrix information about the process from which it was generated (this is an "inverse" problem); and (3) Making an efficient use of newly gained understanding of random matrices to advance research in a wide array of scientific disciplines, including statistics, dynamical systems, climatology, machine learning, signal processing, and finance.

Although specific research foci will be determined by the participants following the Kick-off Workshop, potential research topics might draw from the following:

2.1.1 Direct Problems:

(1) <u>Extreme eigenvalues of random covariance matrices</u>: Asymptotic and non-asymptotic distributions, in the Gaussian setting. Robustness of results to Gaussian assumption. Focus on non-diagonal covariance matrices. Difficulties arising with real-valued random variables.

(2) <u>Dynamic behavior of eigenvalues of matrix processes</u>: Matrices whose elements undergo diffusion (Dyson Processes); stochastic differential equations for their eigenvalues. Scaling limits (Airy processes) and their descriptions via Partial Differential Equations. Connections to growth processes.

(3) <u>Eigenvector problems: Limiting theory for eigenvectors</u>. Free probability techniques.

(4) Spectral properties of other random matrices arising in multivariate Statistics: Techniques such as Canonical Correlation Analysis and related random matrix problems. Covariance matrices with covariance in space and time.

2.1.2 Inverse Problems:

(1) Estimation of large covariance matrices: Regularization techniques by banding, filtering and using L1 penalties; their theoretical and practical properties. Problems of space-time models. Study of methods used in scientific disciplines such as machine learning and "econophysics".

(2) Consistency and estimability problems: Consistency of sample eigenvectors in covariance estimation problems. Estimability issues for eigenvectors and eigenvalues. Implications for Principal Component Analysis (PCA).

2.1.3 Applications:

(1) Climatology: Empirical Orthogonal Functions and related techniques: summarization of evolving geophysical fields via spectral techniques. Approximation via truncated expansions, properties of these truncated expansions (feature significance). Regularization techniques including tapering and inflation methods. Application to detection and estimation of climate change signal.

(2) Dynamical systems: Design of snapshots, i.e., finding a minimal number of functions, tailored to a specific problem, that accurately represent the problem's dynamics.

(3) Data Assimilation: Combination of numerical models and observations. Ensemble Kalman filtering: impact of sample information on propagation of covariance, effect of ensemble size, tapering and inflation methods.

(4) Graphical Gaussian models: Consistent estimation of the model structure,study of the rates of convergence. Consistent estimation of the covariance matrix. Other problems arising in graphical models with large data sets.

(5) Computation of moments: Large random Wishart matrices, connection to graph theory and connection to methods in free probability.

(6) General statistical inference: Consistency of regression functions dependent upon the behavior of certain large random matrices. Problems arising in model selection, estimation and testing. Principal components analysis and choice of the best projections.

2.2 Organization and Program Leadership

The program leaders are Iain Johnstone (Stanford University, Chair), Peter Bickel (UC Berkeley), Helene Massam (York University), Douglas Nychka (NCAR), Craig Tracy (UC Davis); G W. Stewart (Univ. of Maryland, National Advisory Committee Liaison), Chris Jones (SAMSI, Directorate Liaison), and Nell Sedransk (SAMSI Directorate)

The following Scientific Committee provides advice as needed on specific program components:

Myles Allen (Oxford), Estelle Basor (California Polytechnic, San Luis Obispo), David Donoho (Statistics, Stanford), Persi Diaconis (Statistics, Stanford), Jianqing Fan (Princeton), Ken McLaughlin (Mathematics, Univ. of Arizona), Neil O'Connell (Univ. of Warwick, UK), Ben Santner (Lawrence Livermore), Jack Silverstein (Mathematics, N.C. State), Ofer Zeitouni (Univ. of Minnesota).

2.3 Program Development

The organization of a SAMSI program requires considerable pre-program planning. For the planning, organization and design of this program, the program leaders have held a conference call once per week with the SAMSI liaisons.

2.4 Major Participants

Various senior visitors will play a critical role during the program through their extended presence at SAMSI, including Mikhail Belkin (Ohio State University), Helene Massam (York University), Chrsitian Houdré (Georgia Tech), Ofer Zeitouni (Minnesota), Greg Rempala (Louisville), and Nanny Wermuth (Chalmers).

University Fellows: Eitan Greenshtein (Purdue), Peter Miller (Michigan)

New Researchers: Makram Talih (Hunter College), Ming Yuan (Georgia Tech), Debashis Paul (UC Davis), Nourredine El Karoui (UC Berkeley).

Postdoctoral Fellows: Manjudar Krishnapur (Statistics, UC Berkeley-tenure track at Toronto from 1/07), Jayanta Pal (U. Michigan), Bala Rajaratnam (Statistics, Cornell-second year offer at Stanford with Iain Johnstone).

Graduate Students: There will be various graduate students involved in the program from the triangle universities. In addition, we are anticipating four graduate student fellows from outside the triangle area. Two from Georgia Tech will accompany Christian Houdré and two will come from Princeton University (students of Jianqing Fan).

Local Participants: Faculty from the three affiliated universities in the Research Triangle who will be active in the program and will have faculty release time are Sayan

Mukherjee (ISDS, Duke), Stefanos Venakides (Math, Duke), Jack Silverstein (Math, NCSU), Ilse Ipsen (Math, NCSU), Len Stefanski (Statistics, UNC), Yufeng Liu (Statistics, UNC), and Young Truong (Biostatistics, UNC).

2.5 Description of Activities

Opening Workshop: The Kickoff Workshop will be September 17, 2006 - September 20, 2006 will include a one-day Opening Tutorial to present background on random matrices viewed from each key mathematical discipline. The principal goal of the workshop itself will be to engage a broad mathematical base to focus on open questions that are amenable to solution by combining probabilistic and applied mathematical approaches. Invitations have been issued to Estelle Basor (Cal Poly SLO), Persi Diaconis (Stanford), Alexander Its (IUPUI), Gabriele Hegerl (Duke), Thomas Marzetta (Bell Labs), Sara Van de Geer (Leiden), John Lafferty (CMU), Steven Smith (MIT), Ioanna Dumitriu (UC Berkeley), Alan Edleman (MIT), Dave Donoho (Stanford), and Liza Levina (Michigan).

Course: A team-taught course will be held at the NISS/SAMSI building during the fall semester. The course is currently being planned and will be started by Ofer Zeitouni, who will be visiting during the first part of the program. It is anticipated that Jack Silverstein (NCSU) and some of the senior visitors will take part in teaching the course. It will cover mathematical, statistical and computational issues of random matrices.

Summer Pre-Program: A few researchers will gather during July and August to consider problems in Random matrices that pertain to wireless communications. This effort will be led by Brian Rider (Mathematics, Colorado) and Jack Silverstein (NCSU). It will also involve Manjudar Krishnapur who will be starting his postdoctoral fellowship in July at SAMSI. It is anticipated that a small visitor program will be held in concert with a working group activity that will anticipate a continued effort during the program itself.

Mid-Program Workshops: Various mid-program workshops are anticipated Some mini-workshops will be planned later as needs of working groups become more apparent. Two workshops are currently in the planning stage:

- Bayesian methods workshop to be held in lat October
- Workshop on Random Matrices and Climate, to be held in late January or early February at NCAR.

Transition Workshop: will be held at AIM in Palo Alto, California, April 10-13, 2007. This will provide a forum for the presentation of results and applications as well as planning for future collaborations and research efforts stemming from the SAMSI program.

Working Groups: The working groups will meet regularly throughout the program to pursue particular research topics identified in the kickoff workshop (or subsequently chosen by the working group participants). The working groups consist of SAMSI visitors, postdoctoral fellows, graduate students, and local faculty and scientists. It is not necessary to be continually resident at SAMSI to maintain connection to the working groups as effective methods for remote participation have been developed.

Multimodal Working Group: One working group will be set up with two planned nodes connected by full video conferencing. One base will be at SAMSI and the other will be set up at UC Berkeley. There is a considerable group of high level researchers interested in these topics based in the Bay Area. This will allow the program to engage these researchers with full participation in the working group and even co-lead it with people based at SAMSI. The West Coast group will stay at SAMSI for a period of one week after the opening workshop to plan and lay the foundations for this working group. There will also most likely be researchers at NCAR involved in this group, creating possibly a third node.

Spring Program: There will be a dedicated effort during the spring semester, as a continuation of the originally planned fall program, on geometry, random matrices and statistical inference. This will be led by Mikhail Belkin (Ohio State) who will be in residence, together with Sayan Mukherjee (Duke). This part of the program will focus on aspects that bring up computer science issues.

3. Multiplicity and Reproducibility in Scientific Studies

3.1 Introduction

Concerns over multiplicities in statistical analysis and reproducibility of scientific experiments are becoming increasingly prominent in almost every scientific discipline, as experimental and computational capabilities have vastly increased in recent years. This 2006 SAMSI summer program will look at the following key issues that arise.

3.1.1 Reproducibility

A scientist plans and executes an experiment. A clinical trials physician runs a clinical trial assigning patients to treatments at random and blinding who has what treatment. A survey sampling person collects a survey. Scientists use statistical methods to help them judge if something has happened beyond chance. They expect that if others replicate their work, that a similar finding will happen. To clear a drug the FDA requires two studies, each significant at 0.05.

A recent paper by Ioannidis (JAMA 2005; 294:218-228) showed startling and disconcerting lack of reproducibility of influential statistical studies published in major medical journals. It found that about 30% of randomized, double blinded medical trials failed to replicate and that 5 out of 6 non-randomized studies failed to replicate -about an 80% failure rate. We aim to explore and clarify the causes of failures to reproduce in more detail, not only in the Ioannidis paper, but also more broadly, identifying commonalities that lead to these problems, and attempting to estimate its prevalence. Multiplicities (both obvious and hidden) will be considered in particular, along with selection biases and regression to the mean. At the conclusion of the program, recommendations for scientific reporting and publication will be made.

3.1.2 Subgroup Analysis

Large, complex data sets are becoming more commonplace and people want to know which subgroups are responding differently to one another and why. The overall sample is often quite large, but subgroups may be very small and there are often many questions. Genetic data is being collected on clinical trials. Which patients will respond better to a drug and which will have more severe side effects? Disease, drug, or side effects can result from different mechanisms. Identification of subgroups of people where there is a common mechanism is useful for diagnosis and prescribing of treatment. Large educational surveys involve groups with different demographics, different educational resources and subject to different educational practices. What groups are different; how are differences related to resources and practices? What really works and why? Is the finding the result of chance? There is a need for effective statistical methods for finding subgroups that are responding differently. There is a need to be able to identify complex patterns of response and not be fooled by false positive results that come about from multiple testing. Our idea is to bring together statisticians and subject experts to develop and explore statistical strategies to address the subgroup problem. The benefit will be creditable statistical methods that are likely to produce results that will replicate in future studies.

3.1.3 Massive Multiple Testing

The routine use of massively multiple comparisons in inference for large scale genomic data has generated a controversy and discussion about appropriate ways to adjust for multiplicities. We will study different approaches to formally describe and address the multiplicity problem, including the control of various error rates, decision theoretic approaches, hierarchical modeling, and probability models on the space of multiplicities, and model selection techniques. Besides applications in inference for genomic data we will consider similar problems arising in clinical trial design and analysis, record matching problems, classification in spatial inference, anomaly discovery and syndrome surveillance. The goal of this program is to identify the relative merits and limitations of the competing approaches for diverse applications, and to understand which features of reproducibility are addressed.

3.1.4 Political Issues

There appears to be disagreement in some scientific disciplines as to the need to correct for multiplicities. The program will explore the motivations behind these disagreements, and seek ways to achieve a scientific consensus on the issues.

3.2 Participants

3.2.1 Program Leaders Committee

Peter Muller (M.D. Anderson Cancer Center), Juliet Shaffer (U. Calif. Berkeley), Peter Westfall (Texas Tech. Univ., Chair); Stan Young (NISS, Scientific Coordinator); James Berger (SAMSI, Directorate Liaison), and Ray Carroll (Texas A&M, National Advisory Committee Liaison)

3.2.2 Full Program Participants; Most Confirmed, Some Tentative

Code: MT=Multiple Testing; SUB=Subgroup Analysis; REP=Replication

- Susie Bayarri (MT), Univ. Valencia
- Yoav Benjamini (MT), Tel Aviv U.
- Anirban Das Gupta (MT), Purdue U.
- Alex Dmitrienko, Lilly Corporation
- Sandrine Dudoit, U. Calif. Berkeley
- Helmet Finner, U. Dusseldorf
- Chris Genovese, Carnegie Mellon U.

- Peter Hoff (MT), U. Washington
- Peter Imrey (REP), Cleveland Clinic
- Sonia Jain, UCSD
- Valen Johnson, MD Anderson
- Robert Obenchain (SUB), Lilly Corp.
- Keith O'Rourke (REP), Ottawa Health Research Institute
- Luis Pericchi (MT), U. Puerto Rico
- Joseph Romano (SUB), Stanford U.
- Ken Rice (MT), U. Washington
- Sanat Sarkar, Temple U.
- Azeem Shaikh (SUB), Stanford U.
- Frank Shen (SUB), Bristol-Myers Squib
- S. Sivaganesan (SUB), U. Cincinatti
- Rochelle Tractenberg (REP), Georgetown University Hospital
- Dani Yekutieli, Tel Aviv U.

3.2.3 Potential local participants

A large number of faculty and students in the local statistics and biostatistics departments, as well as at local companies (such as Glaxo-Smith-Kline) have expressed great interest in the program and will be participating.

3.3 Activities and timing

- July 10-12, 2006: Opening workshop, laying out the challenges in each area, and setting the stage for the research.

- July 12-26, 2006: Working groups in the four areas meet daily.

- July 27-28, 2006: Transition workshop, summarizing progress and remaining challenges.

- A selection of papers arising from the workshop will be published in a special section of *The American Statistician*.

4. Brainstorming Meetings

As new science research areas emerge, the incorporation of mathematical sciences as research foci often comes well after the other sciences have well-developed agendas that may have involved the application of extant mathematical tools and computational methods. It is not necessarily the case that the tools and methodologies applied are inherently suited to the particular application; moreover these may be utilized in the absence of mathematical formulation of the essential issues.

In consequence, SAMSI is inaugurating a series of small workshops, each with the purpose of examining a new area of interdisciplinary research to determine what, if any, critical role the statistical and mathematical sciences might play in the research and also to determine what kinds of new development in the mathematical sciences might be required.

As a first example, in his address to the nation earlier this year, President Bush proposed significant national investment of research resources in Nanoscience and Nanotechnology. The National Science Foundation Committee for Review of the National Nanotechnology Program (as of January 2005) was comprised 24 members with expertise in physics, chemistry, engineering sub disciplines, medicine, law, health policy and business. ***No*** mathematician. Similarly, of awards made under the several Nanoscience programs at NSF, a handful involve modeling (although not by mathematicians); and some awards have been made in computer science. Yet at the *NNI* (National Nanotechnology Initiative) *Grand Challenge Workshop on Instrumentation and Metrology* held in 2004 at NIST (National Institute of Standards and Technology) the several working groups cited challenges and needs for new mathematical research and implementation as critical to progress.

The first challenge is to ascertain what kinds of mathematical problems need to be addressed in terms of fundamental research, in terms of adaptation to the interdisciplinary research, in terms of formulation of the scientific research itself.

A new kind of workshop for SAMSI will examine these questions in small groups of statistical and applied mathematical researchers plus scientists working in Nanotechnology. The goal is to outline an agenda for interaction and for initial exploration of fundamental as well applied research. There are 29 National Nanotechnology Initiative Centers including one at Purdue, a NISS-SAMSI affiliate. Therefore, Purdue was selected to host a “Brainstorming Workshop on Statistical and Mathematical Sciences for Nanoscience and Nanotechnology” in the fall of 2006. The local Chair for the Workshop will be Mary Ellen Bock, Chair of Statistics at Purdue University and Chair of the National Advisory Council for SAMSI.

This Workshop will consist of one day of open presentations by nanoscientists and nanoengineers followed by one day of closed sessions. Invited participants at this “Math & Nano Workshop” workshop will consist of a team from each of 10-15 institutions represented: One statistician or mathematician plus one nanoscientist or nanoengineer.

Opening presentations of the science (by the scientists) will provide a basis for discussion during the closed working sessions to delineate possible roles for Statistics and Applied Mathematics. The output from the workshop will be an outline proposing mathematical work from fundamental to methodological to formulational. It is hoped that the team from each institution will continue to function in interdisciplinary fashion upon return home. If these efforts are successful, a follow-on workshop will be held with the next 12 months to reappraise the outlined research possibilities and to consider planning a full-blown SAMSI Program within the next few years.
While there are few statisticians and applied mathematicians working on mathematical research to underpin nanoscience, the work of those few may prove to be groundbreaking. Nanoscale problems present unique challenges because at the nanoscale things operate as individual units not aggregates: for example, tiny forces that are either ignorable or averageable at the microscale are NOT ignorable at the nanoscale. Computational demands are unprecedented, so that pure computing power is not a solution; however, Bayesian computation and other self-informing methods may reduce the reconstruction time for a single nano-image from over 24 hours by at least an order of magnitude. Understanding these fundamental differences from aggregate behaviors may both require and inspire new mathematics.

A second workshop is also in the planning stages to examine Quantum Computation and Quantum Communication. While mathematicians as well as physicists and computer scientists have been involved for some time, statisticians have not; and the problems in Quantum Communication in particular cannot be solved without accounting for the random processes inherent in the photon transmission and reception. It is anticipated that this second Brainstorming Workshop in Spring, 2006 will be located off-site (currently under negotiation with a NISS-SAMSI Affiliate) and will follow the paradigm of the first, with modifications based on that experience.

B. Scientific Themes for Later Years

The programs listed below have not yet been formally approved, but all are well along in the development cycle and we are confident that they will be approved and implemented.

1. Challenges in Dynamic Treatment Regimes and Multistage Decision-Making (Tentative for June 18-29, 2007)

1.1 Scientific Overview

The management of chronic disorders, such as mental illness, substance dependence, cancer, and HIV infection, presents considerable challenges. In particular the heterogeneity in response, the potential for relapse, burdensome treatments, and problems with adherence demand that treatment of these disorders involve a series of clinical decisions made over time. Decisions need to be made about when to change treatment dose or type and regarding which treatment should be used next. Indeed, clinicians routinely and freely tailor treatment to the characteristics of the individual patient with a goal of maximizing favorable outcomes for that patient. To a large extent the tailoring of sequences of treatments is based on clinical judgment and instinct rather than a formal, evidence-based process.

These realities have led to great interest in the development of so-called “dynamic treatment regimes” or “adaptive treatment strategies.” A dynamic treatment regime is an explicit, operationalized series of decision rules specifying how treatment level and type should vary over time. The rule at each stage uses time-varying measurements of response, adherence, and other patient characteristics up to that point to determine the next treatment level and type to be administered, thereby tailoring treatment decisions to the patient. The objective in developing such multistage decision making strategies is to improve patient outcomes over time.

Methodology for designing dynamic treatment regimes is an emerging area that presents challenges in two areas. First, experimental designs for collecting suitable data that can be used efficiently to develop dynamic regimes are required. Second, techniques for using these and other data to deduce the decision-making rules involved in a dynamic regime must be developed. In both areas, input from researchers in a variety of disciplines and collaborations among them will be critical.

Currently, decision rules that make up a dynamic regime are formulated based on clinical experience, disease-based theories, and expert opinion and then evaluated in randomized two-group trials against a usual or standard treatment. Such an experimental approach assesses a dynamic regime as a “package” and hence does not provide evidence on usefulness of components of the strategy (e.g., timing of treatment alterations and the choices of treatment to which to switch). Accordingly, trials in which patients are randomized to different treatment options at each decision point have been proposed; however, little is known about when such trials should be conducted in lieu of the current approach of melding clinical judgment and expert opinion to formulate decision rules and using the standard two-group paradigm. An alternative approach is to conduct a series of randomized trials, as in agriculture and engineering; again, there is little guidance on how to implement this approach when the goal is to develop a dynamic regime.

Methods to make use of data in developing dynamic regimes involve complex considerations. The construction of optimized decision rules requires incorporating the effects of future decisions when evaluating present decisions, as is well-known to scientists working on improving multistage decision-making. Treatment given at any time may set a patient up for improved response to subsequent treatments or have delayed effects that either enhance or reduce effectiveness of subsequent treatments. The development of a dynamic regime hinges on how one operationalizes the relative importance of patient outcomes over time. Researchers who work on multistage decision problems in other contexts (robotics, artificial intelligence, control theory) readily recognize these types of issues. A key challenge is to determine how to collect sufficient information to ascertain the "state" of an individual insofar as making treatment decisions goes. Computer scientists working the field of reinforcement learning and statisticians working in medical decision-making have quantified the properties that the state must possess; in addition, practical considerations associated with feasibility of collection, cost, and patient burden must be taken into account. Typically, a great deal of information is available at each decision point, and methods for best summarizing this information yet maintaining the summary's usefulness for deciding on treatment alteration or type are required. Methods for feature extraction developed by statisticians and computer scientists are well-suited to this problem, but the focus on multistage decision-making rather than prediction requires evaluation of these methods from a different perspective. Computational and inferential challenges arise in all of these endeavors; e.g., complexities of optimizing dynamic regimes can invalidate standard statistical inferential techniques, scientific considerations entail thinking beyond the standard loss functions familiar to statisticians, and the abundance of information at each decision point quickly leads to a "small n, large p" problem and the attendant computational issues. For some disorders, e.g., HIV infection, knowledge of the underlying within-subject biological has led to development of sophisticated mechanistic models for the processes governing disease progression and effect of treatment, which offer a scientific basis (via closed loop control methods) for designing dynamic regimes; however, this approach has not been widely explored or tested in samples of patients in this context.

In summary, the critical need for development of methodology for dynamic treatment regimes, the considerable theoretical and practical challenges, and the relevance of work across a range of disciplines requires a forum in which the essential exchange of ideas and collaboration necessary to propel advances is fostered. Moreover, many in the statistical and applied mathematical research communities are unfamiliar with this area and the research opportunities it presents. The proposed SAMSI summer program will bring this area to the attention of statistical and applied mathematical scientists, whose expertise is critical; jump-start the necessary methodological development; and nurture the necessary interdisciplinary collaboration and communication between statisticians/applied mathematicians and computer scientists and health and behavioral science researchers. We especially hope to intrigue and engage more junior researchers with relevant expertise.

1.2 Program Scope, Timing, and Activities

The program will take place in the summer of 2007 over a two-week period (two successive Monday–Friday periods straddling one weekend). The following activities are planned:

Tutorials (Monday-Wednesday, Week 1). Six tutorials will be held, two per day, to provide participants unfamiliar with the foundation necessary to assimilate more advanced development. All tutorials will be taught by experts in these areas. Suggested names of instructors are given; instructors will be recruited and finalized following approval of the program:

- Day 1: *Introduction to Causal Inference* (Butch Tsiatis or Dan Scharfstein) and *Introduction to Dynamic Treatment Regimes* (Susan Murphy or Marie Davidian or Dan Scharfstein)

- Day 2: *Computational Challenges with High Dimensional Data* (Joelle Pineau) and *Introduction to Mechanistic Models and Control Theory* (Tom Banks or Daniel Rivera or Joelle Pineau)

- Day 3: *Classification and Reinforcement Learning* (Peter Bartlett or Doron Blatt or John Langford or Michail Lagoudakis or Ronald Parr) and *Problems in Nonstandard Statistical Inference* (James Robins or Susan Murphy)

Workshop (Thursday-Friday, Week 1). The workshop will feature one or two overview talks presenting the "big picture" of the methodological challenges followed by more advanced and targeted talks on research relevant to development of dynamic treatment regimes that build on the foundation provided by the tutorials. There will be talks the first full day and second morning, after which participants will be divided into discussion groups centered around four key areas that will form the basis for "brainstorming" by working groups during the next week of the program (below). The discussion groups will develop lists of important challenges and questions centered around their theme.

We believe that the tutorials and workshop could attract a significant audience. A conference on causal inference in January 2006 at Johns Hopkins, which features sessions on dynamic treatment regimes, was attended by 200 people (registration was closed when the audience became this large).

Working Groups (Monday-Wednesday, Week 2). Four Working Groups will convene to discuss and prioritize challenges in their respective areas. Participants will identify the most pressing problems and outline modes of attack and specific research directions to be pursued. The proposed working group foci and potential lead participants are:

- *Difficulties In Statistical Inference* (Peter Bartlett, Susan Murphy, Sasha Rakhlin, Jamie Robins).
- *Bayesian Approaches* (Giovanni Parmigiani, Dan Scharfstein, Peter Thall).

- *The Role of Mechanistic Models*. (Tom Banks, Victoria Chen, Marie Davidian, Daniel Riviera).
- *Practical Challenges and Applications* (Erica Moodie, Joelle Pineau, Butch Tsiatis).

Working Groups will meet daily according to a schedule that will allow participants to be involved with more than one group if desired.

Dr. Murphy has led two successful, small (~ 20 individuals) meetings of a "network" of diverse researchers from statistics, computer science, engineering, behavioral science, and medicine devoted to this topic, and the willingness of these individuals to participate in several such meetings suggests to us that the Working Groups will be enthusiastically attended by our envisioned 5–6 participants per group. We hope to attract at least 2–3 junior researchers to each group.

Summary and Transitional Workshop (Thursday-Friday, Week 2). Each Working Group will present their results, findings, and recommendations for the future to the all Working Group members along with additional participants who may return for this final activity. Discussion will follow each group's presentation. These presentations and discussions will form the basis for a white paper outlining methodological challenges in the area of dynamic treatment regimes to be written by the Program Leaders for submission to a leading statistical or mathematical science journal, with input from participants.

Because this area is so new and does not have a long-standing and established groundwork of methodological research, we believe that a two week period is most appropriate and sufficient to achieve the objective of raising awareness and spearheading the necessary activities and connections.

1.3 Program Leadership

The Program Leaders Committee is currently Susan Murphy (University of Michigan), Daniel Scharfstein (Johns Hopkins Bloomberg School of Public Health), Joelle Pineau (McGill University); Local Scientific Coordinators: Marie Davidian and Butch Tsiatis (North Carolina State University).

1.4 Program Participants

We envision that the Tutorials and Workshop will attract a diverse group of about 50 participants from statistics, applied mathematics, computer science, and the health and behavioral sciences. We expect 20–25 participants to remain in residence during the second week to participate in Working Groups. Strong efforts will be made to recruit junior investigators for participation in the Working Groups.

1.5 Program Outcome

The goal of the Program is to stimulate and delineate the needed statistical/mathematical/algorithmic research in this important area. The desired outcome for the program will be a well-defined, concrete list of specific research directions that should be pursued to advance the needed methodological development. These will be brought to the attention to the research community via the planned white paper summarizing these directions.

2. Random Media (Tentative for 2007-08)

2.1 Summary

The proposed program will address a number of fundamental issues pertaining to random media including scattering theory in highly discontinuous and random media, time reversal, model development, analysis, and numerical approximation for interface methods, and imaging in random media. The inherent synergy between deterministic, statistical, and physical analysis necessitates a concerted collaboration between applied mathematicians, statisticians, engineers, geologists, and material scientists which has often been absent in the past and is necessary to provide fundamental advances to the field.

2.2 Background

The field of random media is a classical one which is presently receiving widespread attention as new theory, approximation techniques, and computational capabilities are applied to emerging applications. Due to the breadth of the field, the inherent deterministic, stochastic and applied components have typically been investigated in isolation. However, it is increasingly recognized that these components are inexorably coupled and that synergistic investigations are necessary to provide significant fundamental and technological advances in the field. The proposed SAMSI Program on Random Media will provide a forum to investigate statistical and deterministic components of random media for applications including, but not limited to, time reversal, interface problems, imaging in random media, and scattering theory for discontinuous media.

The component on time reversal will build upon recent analysis and experimental observations that time reversal of waves propagating in disordered media permit refocusing. This somewhat unexpected property has profound ramifications in domains such as wireless communications, medical imaging, nondestructive evaluation, and underwater acoustics. Whereas the behavior of one-dimensional acoustic waves is mathematically and statistically understood, questions regarding multidimensional media remain widely open with the exception of the baraxial wave equation.

Interface problems arise in a diverse range of applications including multiphase flows and phase transitions in fluid mechanics, thin film and crystal growth simulations in material science, and mathematical biology problems modeled by partial differential equations involving moving fronts. In computational fluid dynamics, electromagnetic scattering and ground water flows, efficient numerical approximation are essential for quantifying the effective property of the medium due to fluctuating inhomogeneous and random medium. The level set method has proven to be an extremely versatile tool for tracking deformations in shape geometries, moving interfaces, and free boundaries in a number of related applications, and one facet of the program will focus on extensions of this approach to include the effects of random media and stochastic processes. Other aspects of the interface component will focus on modeling and analysis of random interface growth processes including crystal growth and solidification, Monte-Carlo

Wiener-Chaos expansion and homogenization methods for stochastic partial differential equations, and level set methods and Lagrangian formulations (particle approaches) for random media simulations.

Imaging problems in random media arise in a number of applications including biomedical imaging and seismic analysis. In the latter category, a detailed knowledge of earth medium heterogeneities is necessary for oil and gas recovery, earthquake and volcanic predictions, and environmental analysis. One fundamental issue involves the multiscale relation between large scale structures, which are considered as deterministic, and small scale heterogeneities which are considered to be random fluctuations form the deterministic structures. A related issue concerns the analysis of coupled processes.

Whereas mathematical scattering theory for one-dimensional regimes is fairly mature, little of the analysis extends to multidimensional media with the exception of the baraxial wave equation. Hence this facet will focus primarily on the development of theory, numerical methods and validation techniques pertaining to scattering theory for multidimensional media.

2.3 Program Timing and Previous Related Workshops

Within the last ten years, the field of random media has experienced a number of advances which delineate both the potential for the field and fundamental and technological limitations which must be addressed. Whereas there have been workshops and conference on related subtopics within the last five years, there have been no concerted forums devoted to the combined statistical, mathematical and physical analysis of the full topic. Hence the proposed SAMSI program for 2007-08 is very timely.

Recent related workshops and conferences include the following.

Joint Summer Research Conference in the Mathematical Science: Waves in Periodic and Random
Media (Mount Holyoke College, June 2002) – This conference focused primarily on analytic, numerical and physical perspectives of topics including photonic crystals, wave propagation in linear and nonlinear periodic media, waves in mesoscopic media, and surface waves.

Workshop on Probing Earth Media having Small-Scale Heterogeneities (Tohoku University, Japan,
November 2004) – This conference focused on advances in the field of seismic wave propagation in earth media containing small-scale heterogeneities.

IMA Annual Program Year Workshop: Imaging from Wave Propagation (Institute for Mathematics
and Its Applications, October 2005) – This program combines aspects of seismic imaging and wave propagation.

2.4 Participants and Personnel

Program Leaders Committee: Maarten De Hoop (Purdue University), Rick Durrett (Cornell University – NAC), Josselin Garnier (Universite Paris VII), George Papanicolaou (Stanford University), Lenya Ryzhik (University of Chicago), Ralph Smith (SAMSI, Directorate Liaison), Chrysoula Tsogka (University of Chicago), Jack Xin (UC Irvine), Wojbor Woyczynski (Case Western Reserve University), Hong-Kai Zhou (UC Irvine)

Long-Term SAMSI Visitors: Hong-Kai Zhou

Potential Participants and Visitors: Boris Rozovski (USC), Russel Caflish (UCLA), Randy LeVeque (Washington), J.P. Fouque (UC Santa Barbara), Kazi Ito (NCSU), Zhilin Li (NCSU), John Trangenstein (Duke), Weinan E (Princeton), Chrysoula Tsogka (Chicago), Lenya Ryzhik (Chicago), Knut Solna (UC Irvine), David Donoho (Stanford), Robert Burridge (MIT), John Cushman (Purdue University)

Postdocs: This is a field in which there will be numerous excellent postdoctoral candidates.

Faculty Releases from Partner Universities: There are a number of faculty in the partner universities with considerable interest in mathematical, statistical and physical aspects of random media.

2.5 Working Groups

Each of the previously noted areas poses a potential working group topic. Additional working groups will be identified both as the program is developed and as future directions are identified in the Opening Workshop. Potential working groups can be summarized as follows.

Time Reversal: This group will focus on mathematical and statistical aspects of time reversal for random media.

Interface Problems: The working group on interface problems will focus on model development, deterministic and statistical analysis, and numerical approximation for a range of applications. One goal of this group is to identify and quantify the effects of random media and stochastic processes on level set methods.

Imaging Problems in Random Media: This group will investigate statistical aspects such as the stability of imaging operators in background media with random fluctuations. The second focus will be on the characterization of interferometric methods such as virtual source imaging using limited aperture arrays.

Scattering Theory: This group will focus on the analysis and characterization of waves in multidimensional media.

Porous Media: This working group will focus on topics pertaining to stochastic transport processes and physics associated with porous media.

2.6 Leveraging

Aspects of this program will build on previous SAMSI programs on environmental models (200203), multiscale modeling (2003-04), and computer models (2006-07). It is anticipated that issues pertaining to data collection and analysis in the concurrent program on Wireless Environmental Sensor Networks will also be germane to this program.

3. Wireless Environmental Sensor Networks (Tentative for Spring 08)

3.1 Summary

Data gathered by wireless sensor networks, either fixed or mobile, pose unique challenges for environmental modeling: a complex system is being observed by a dynamical network. On the one hand an ideal information-sampling process across a sensor network would be governed by the complex system model that must simultaneously be informed by these data; on the other hand, the network itself is a dynamic system of self-organizing nodes which exhibit both independent and dependent behaviors. This presents a unique opportunity to organize the sensor system so that a local or micro event can trigger a broad or macro observation – or conversely, a macro observation can trigger highly detailed local data gathering.

A collaborative effort of statistician, mathematicians, computational scientists and environmental scientists is required to formulate and to address the canonical modeling questions of what data to gather, how to fit the model, how to assess the uncertainty of inferences and how to reconfigure or relocate the network nodes. Most importantly this interdisciplinary effort will consider the problems and the consequences of utilizing or directly influencing the network dynamics to optimize data gathering, model assessment and prediction. The principal application will be forest ecosystems, with the objective of understanding how climate change and CO2 impact ecosystems from the cellular level within a leaf to the macroscale of the forest stand.

3.2 Background

Environmental sensor networks have the capability of capturing local and broadly information simultaneously; they also have the capacity to respond to sudden change in once location by triggering simultaneous observation across the network and/or by changing the observation frequency or otherwise altering the sampling plan. When different kinds or sensors are mounted together on each platform, this dynamic response capability can operate in a multimodal fashion. The open challenge is to design the observation process (algorithms), to model the environmental system in evolution using the multimodal data (“data fusion”) and to predict environmental response. Simulation is one chief tool for examining the complex environmental system – dynamic sensor network interaction; exploration of real data from one or more environmental systems is another.

This program will bring together an interdisciplinary group of ecologists, mathematicians, statisticians, and computer scientists with the objective of formulating a methodology for wireless sensor networks used for environmental modeling: 1) Network ➔ Environmental Complex System Model: to define the modeling and technical challenges related to processes that will be informed by wireless networks, including the nature of spatio-temporal designs, 2) Environmental Complex System Model ➔

Network: to formulate probability-based adaptive sampling processes and to develop the algorithms needed for implementation in the context of environmental research, 3) Environmental Complex System Model ➔ Sensor motion control: to reposition mobile sensors to optimally co-locate active sampling, 4) Data Fusion, and 5) Network Data ➔ Model-based Prediction: to develop scalable process models that can be implemented in an inferential framework and used for prediction. Part of this effort will be general, and we plan to assemble a broad group of scientists who work in different systems to help define the intensity of sampling in space and time that is needed for understanding and prediction. Part will be focused on individual case studies, including a study of biosphere-atmosphere interactions in forest canopies, entailing a range of issues as broad as exchange of CO2 and H2O, C storage, microclimate, and biodiversity, through competition for light and soil moisture.

The motivation for this research and its mathematical, statistical and computational complexity are clearly illustrated in the following example. Understanding how changing climate and CO2 will impact ecosystems, including storage of carbon and biodiversity, depends on the capacity to link processes that affect carbon uptake by individual leaves to the growth of whole trees in competitive forest stands. This ‘scaling problem’ has emerged as one of the most important, challenging, and sustaining issues for forecasting consequences of global change. The linkage from leaf to forest is necessary, because we can only measure the direct effects of temperature, CO2, and moisture at the scale of leaves, but the consequences of those effects are most relevant and profound at the level of a forest stand and a landscape.

The process models of ecosystem change that will ingest real time environmental data build in assumptions about how data are collected, the traditional realm of statistics. Even standard data types present serious problems for modelers, often met with ad hoc 'inversion' techniques that are not based on probabilistic models and, thus, do not yield probability statements, in the form of confidence and prediction intervals. Bayesian methods are beginning to meet the demands of high-dimensional processes informed by heterogeneous data, building on graphical approaches. They open the potential to assimilate information obtained from entirely novel scales.

Complementary to advances in modeling are wireless sensor networks, which promise to bridge key gaps in our capacity to observe scales of variation that affect forests (http://www.neoninc.org/documents/sensor_meet1_report.pdf). Given infinite power supply, there is now a capacity to collect infinitely dense data in space and time. This capacity brings with it new challenges of developing adaptive sampling algorithms that will be controlled by software and based on the ecosystem models that must ingest these data. Despite a proliferation of creative techniques for modeling spatiotemporal data, 'smart' wireless networks extend many of the challenges that are already well-recognized in the statistical community (e.g., non-stationary covariances, non-separable spatio-temporal covariances, parallel models needing state variables at different scales,…) and present new ones. Wireless networks lack infrastructure, consisting of nodes that self-organize for purposes of data collection and transmission. We can begin to think about scalable process models directly informed by highly unbalanced data, as opposed to, say, nested (but rigid) lattices. Sampling efficiency is paramount, because power is supplied by batteries, and data transmission is energy demanding. The need to minimize the transfer of redundant information (e.g., oversampling in space and time) must be weighed against the risk of missing rapid changes and potentially short-term pulses that are missed by coarse sampling designs. Clearly, sampling must be adaptive and, to the extent possible, regulated within the network. Efficient exploitation of this technology will depend on the capacity to obtain observations that have ‘value’ for (provide learning) in ecosystem models that simultaneously ingest alternative sources of information.

Combining unusually extensive field data, including wireless networks, within a powerful modeling framework, makes it possible to address the properties, or ‘scaling relationships’, that describe how the rapidly changing atmosphere translates from leaf level processes to whole-tree and forest response. A start in this direction would be to develop a graphical modeling structure that allows for decomposition into tractable subunits the complex set of processes that link atmospheric change, leaf physiology, carbon allocation and growth, and population demography. We anticipate novel implementation of a hierarchical framework to assimilate data from many sources and represent processes at different scales, from ½ hourly gas exchange by individual leaves to annual changes in tree size and population density. The graphical framework allows knowledge and uncertainty to propagate across scales and represents a full synthesis of what existing data can tell us. The scaling relationships needed to understand and forecast consequences of global change enter as hypotheses that can be weighed by the data themselves.

A second cogent illustration is the observation of dynamic oceanographic processes. As our understanding of the oceans has advanced, it has become apparent that many critical processes occur at temporal and spatial scales that cannot be effectively sampled with traditional ship-, moored buoy-, or satellite-based approaches. The present generation of oceanographic field programs is fundamentally limited by too few measurements, taken too slowly, and at too great a cost. Traditional studies are particularly limited in their ability to investigate the onset and immediate aftermath of episodic events such as earthquakes, tsunamis, submarine volcanic eruptions, or hurricanes. While satellite observations have provided oceanographers with a unique global perspective of the ocean, they remain primarily limited to measuring properties at the air-sea interface. This urgent need to advance the understanding of the ocean has resulted in a new paradigm for doing ocean research in a more adaptive way.

The NSF Ocean Observatories Initiative Science Plan (see http://www.orionprogram.org) calls for the exploitation of small, low-cost autonomous underwater vehicles that function as mobile ocean observing systems and provide broad spatial coverage. Reacting to short-lived events and changes observed in real time will require sophisticated mobile sensing platforms integrated into the ocean observatory infrastructure. Self-propelled mini-buoys, aerial and underwater vehicles can be remotely commanded to investigate episodic events, thus enhancing opportunities to capture key processes that do not occur in the vicinity of an observatory node. Mobile platforms can feed real-time data into mathematical models, provide in-situ predictions and adapt their behavior accordingly.

Although technology provides the physical components of such sensor networks, their potential benefits are not yet being realized. As of today, one fundamental limitation is a lack of understanding on how to jointly address the distributed data fusion and motion coordination problems. Both problems are intimately intertwined. To coordinate their motion, agents need to collaboratively construct in-situ representations of the observed phenomena based on the collected data. Moreover, satisfactory solutions to the distributed data fusion problem rely upon motion coordination strategies that optimize information gathering of the physical processes. Most currently-available algorithms for mobile networks send all data back to a central station, where aggregation is performed. While such strategies are sometimes appropriate, they do not enable the network to rapidly adapt to the evolution of physical phenomena. Instead, data aggregation should be performed, at least partially, in-network to allow for decision making on the fly. Rapid adaptation must also rely on intelligent reduction of the computational complexity. Utilizing computational geometry and topology to characterize network configurations, even if continually reorganizing, may allow reduction of computational complexity by minimizing the amount of sensing in addition to minimizing the computation per se. For example, computational complexity, time and cost are reduced when precise localization of an individual mobile sensor can be bypassed in favor of its position within the organization of the network.

3.3 Scientific Description

The five problems identified above interlace environmental science and ecology with computational mathematics and computer science, applied mathematics, statistics, and engineering. Therefore a successful approach to any one of these will require an interdisciplinary working group of ecologists, engineers, mathematicians, statisticians, and computer scientists. Therefore the first goal of this program is, from an interdisciplinary point of view, to define the modeling and technical challenges related to processes that will be informed by wireless networks and to identify among the combined disciplines the existing expertise and to determine the integrated resources to be brought to bear.

One class of challenges implicit in pervasive sensor networks is based on the local-to-global transition; and as sensors become more miniaturized they return data which is increasingly local. Computational geometry and topology provide a number of novel techniques for converting local combinatorial data into global descriptors.

A second class of challenges is in addressing the computational requirements efficiently when the model systems are complex, the networks themselves (either fixed and pervasive or mobile) are complex and dynamic or, under adaptive sampling schemes, quasi-dynamic, and data aggregation/analysis cannot be efficient if it resides solely in a centralized station. The initial step in finding possible approaches to the apparent [computational] complexity problems is distinguishing the elements of these that appear as data fusion problems, model complexity problems, “sensing complexity” problems, adaptive sampling problems and computational complexity. Then suitable integration of formulations and approaches by the varied disciplines can follow.

Once research approaches have been defined, more specific goals in the context of forest ecology will be to develop algorithms needed for adaptive sampling of environmental variables and to develop scalable process models that can be implemented in an inferential framework and used for prediction. This example will provide tangible results and direct implementation to the study of biosphere-atmosphere interactions in forest canopies, entailing a range of issues as broad as exchange of CO_2 and H_2O, C storage, microclimate, and biodiversity, through competition for light, nutrients, and soil moisture.

3.4 Research Foci

Research foci will be built from themes involving 1) sampling: adaptive sampling and network-triggered observation, 2) computational issues: complexity, data fusion, mapping of algorithms to sensors, and 3) modeling: prediction and uncertainty, especially multiscale modeling - all with primary application to modeling forest response to global environmental change. Although the specific agendas for the research foci will be determined by the participants at the Opening Workshop, potential research topics might be drawn from the following questions.

3.4.1 Sampling from wireless networks:

Cost of spatio-temporal data in terms of both energy and delay: Each sample has a footprint in power in space and time, some value to one or more process models (e.g., importance of parameters in space and time, sensitivity of estimates to the observation), and some cost (e.g., data transmission). Is it possible to derive frameworks such that the utility of each sample exceeds its cost?

Frameworks for adaptive sampling: game-theoretic, reinforced learning, dynamic experimental design, topology-based optimization. How can a sampling scheme to respond to highly non-stationary dynamics in energy-constrained sampling networks. Are modes of operation controlled by adaptive state machines enough?

3.4.2 Computational issues for wireless network data:

Complexity of process models: What is the trade-off between accuracy (in terms of predictive power) and cost for various implementations, ranging from completely in-network computation to the case where the network just reports data based on commands from grid-based models?

Mapping of algorithms to wireless sensor networks: What is efficient integration of local and centralized computation? How should multi-scale computations be assigned to nodes, subnets and a centralized station? To what extent should incoming data determine computation allocation?

Data fusion for networks: How can model-updating or Bayesian approaches to estimation and signal processing decrease the rigidity distributed data fusion for mobile networks? What are the current computational limits for detection of small-scale features and/or speed of feature evolution?

3.4.3 Environment modeling from sensor networks:

Model complexity and adequacy: In the trade-off of dimensionality and predictive accuracy, what are the diagnostics for excessive vs. insufficient parametrization? Can models be developed so that reduced forms (e.g., deleting submodels, subsets of parameters, or reducing resolution of observations and/or parameter specification) still function simultaneously with near-optimality at several scales?

Prediction Uncertainty: Appropriate modeling of sources of uncertainty due to sampling (e.g., “lost” samples, outliers, bad sensors, measurement noise, and unreliable communication), and integration of data models with process models.

Model adequacy: Is there coherent noise in bio-micrometeorological systems that should drive exploration of new regions of the state space?

Broader goals of the proposed project include the capacity to integrate data and process across the range of scales that define how forested ecosystems respond to global change. This process level understanding is critical for anticipating consequences of human impacts on landscapes. The approach entails integrated models of a complex system and involves heterogeneous data, process understanding, uncertainty, and model selection. It is interdisciplinary, involving global change, ecophysiology, forest dynamics, statistics, and computer sciences.

Broader applications will include increased capacity to forecast consequences of global change and the integration of mathematics, statistics, and environmental sciences. The products of our analysis include mathematical and statistical models that can be directly assimilated with global models of climate and CO2 change. Inference on scaling relationships will be implemented in simulation of whole forests to examine potential consequences of changing climate and atmospheric CO2 for forest diversity and carbon sequestration. These results will have immediate application to the problem of forecasting biosphere responses to atmospheric change.

3.5 Organization and Program Leadership

Program development of a SAMSI Program requires considerable pre-program planning involving the Program’s Scientific Advisory Committee, the primary Program Leaders, local Scientific Liaisons and SAMSI support staff. The Scientific Advisory Committee serves as a scientific resource to the Program Leaders during the formulation of the proposal, and a sounding board for the proposed research agenda, offering advice on researchers to include as speakers at the various Program workshops and other key visitors, and as a network for recruiting suitable Post-Docs.

Program Leaders Committee: Paul Flikkema, NAU Chair; Zoe Cardon, UCONN; Jim Clark, Duke; Don Estep, CSU; Deborah Estrin, UCLA; Mark Hansen, UCLA; Nell Sedransk (SAMSI Directorate Liaison); Chris Jones (SAMSI Directorate Liaison); Alan Gelfand, Duke (Local Science Coordinator)

Scientific Advisors: Jorge Cortes, UC Santa Cruz; Robert Ghrist, UIUC

Potential Participants: Pankaj Agarwal, Duke; Anstassia Ailamaki, Carnegie Mellon; Peter Arzberger, UCSD; Dennis Baldocchi, UC Berkeley; Bruce Beck, Georgia; Kate Calder, Ohio State; Benoit Courband, CEMAGREF; Jerry Davis, NCSU; Todd Dawson, UC Berkeley; Julia Downes, North Florida; Martin Haenggi, Notre Dame; Richard Han, Colorado; Leana Golubchik, USC; Robert Gray, Stanford; Hoshin Gupta, Arizona; Tom Harmon, UCLA; William Kaiser, UCLA; Gaby Katul, Duke; George Koch, NAU; Tim Kratz, Wisconsin; Martin Lechowicz, McGill; Mark Lewis, Alberta; Kirk Martinez, U Southampton; Margaret Martonosi, Princeton; Jerry Melillo, Woods Hole; Russell Monson, CSU; Gregory Pottie, UCLA; David Neuhoff, Michigan; Rob Nowak, Wisconsin; Doug Nychka, NCAR; Ram Oren, Duke; Chris Paciorek, Harvard; John Porter, UVA; Phil Rundel, UCLA; Lindsey Seders, Notre Dame; Chris Shoemaker, Cornell; Kieth Smettem, Western Australia; Mani Srivastava, UCLA; Kannan Ramchandran, UC Berkeley; Lang Tong, Cornell; Chris Wikle, Missouri; Steve Wofsy, Harvard; Jun Yang, Duke; Don Zak, Michigan.

3.6 Description of Activities

3.6.1 Workshops

The Opening Workshop for the Wireless Environmental Sensor Networks Program will engage a broad scientific and mathematical community in the exploration of the varied aspects of interdisciplinary work in this area. Tutorials and presentations will provide an initial orientation by researchers in each of these aspects to those from other disciplines. During the Program, shorter workshops involving speakers with specialized expertise will serve to focus on specific aspects, technologies or substantive issues. A Closing Workshop will follow the Program, most probably deferred for up to a year to allow continuation of the research to bring specific projects to fruition.

3.6.2 Working Groups

The Program will revolve around four to six Working Groups, formed at the end of the Opening Workshop, based on the affinities and expertise of the participants. Every Working Group will have multi-disciplinary membership. The Working Groups will include senior researchers, one of whom will be a University Fellow, plus junior visitors, several SAMSI Post-Docs and graduate students (with senior visitors particularly encouraged to bring their own graduate students to participate when suitable). All students and Post-Docs will have personal mentors in residence at SAMSI, either drawn from the Program Visitors or from local university faculty released from duty to participate in the Program. Working Groups will meet at least weekly and will set their own research agendas.

3.6.3 University Courses

The Program will also include course(s) taught at SAMSI that may be taken for graduate credit at any of the local universities.

3.6.4 Leveraging Past SAMSI Programs

Aspects of this program will build on previous SAMSI programs on Large-Scale Computer Models (environmental models) in 2002-3 and Multi-Scale Modeling in 2003-4. Technical work on data fusion and on the network effects on complex system models necessarily encounters random matrices, so this program will interleave naturally and beneficially with the SAMSI program in High-dimensional Inference and Random Matrices in 2006-7.

4. Sequential Monte Carlo Methods for Scientific Computing (Tentative for 2008-09)

4.1 Introduction

Many physical phenomena and much data can be accurately modeled using probabilistic or statistical models. Typical instances of these phenomena/data are the movement of a particle in a random medium, the volatility of the stock market, phylogenetic trees, the world wide web, auctions at ebay or images. However, even if it is possible to obtain realistic physical models or satisfactory statistical models, only the simplest can be solved without the use of numerical methods. Examples of the need for such numerical methods include the popular Ising model in physics and statistical image processing, non-linear non-Gaussian time series models, semi-parametric Bayesian models, and so on. Fortunately, the advent of enormous, cheap computational power and the development of a plethora of complex inference mechanisms has allowed scientists to develop powerful numerical methods.

Amongst numerical techniques simulation-based methods, especially Monte Carlo (MC) methods, now play a central rôle. This approach, developed initially in physics during the early days of electronic computing, has been adopted by researchers in numerous scientific fields including computer science, statistics, computational biology, chemistry and physics. Broadly speaking MC methods are computer-based statistical sampling approaches for optimization, integration, and simulation of complex systems. MC techniques are an essential tool to scientists and, with the increasing complexity of problems to be addressed, these techniques are going to become even more important in the 21st century. In particular, two classes of MC methods have emerged as the tools of choice for analysis and exploration of complex stochastic models and systems that are otherwise effectively impossible to analyze: Markov chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC).

MCMC methods are iterative algorithms for sampling from high-dimensional probability distributions. This problem appears in so many scientific applications that the Metropolis algorithm (the most basic MCMC algorithm) was ranked first in the "Top Ten Algorithms of the 20th Century" selected by computer scientists, applied mathematicians and physicists (http://amath.colorado.edu/resources/archive/topten.pdf). MCMC algorithms have provided enormous scope for realistic statistical modeling and have attracted much attention from statisticians. Indeed, both Bayesians and frequentists need to integrate over possibly high-dimensional probability distributions in situations such as dealing with missing data or marginalising nuisance parameters in order to make inference about the parameters of interest or to make predictions. MCMC methods make this integration straightforward conceptually. The past few years have witnessed an explosive growth of interest in MCMC methodology from researchers in almost all areas of statistics and this has led to a real revolution, especially in Bayesian statistics; e.g. [4].

However, many scientific problems cannot be addressed using MCMC. Indeed, when massive datasets are available, MCMC algorithms are generally far too cumbersome. Also, when data have to be processed "on-the-fly" because of real-time constraints, MCMC methods are not appropriate because of their batch nature. More recently a set of alternative MC techniques known as Sequential Monte Carlo (SMC) methods has emerged to solve such problems. They have become very popular for solving optimal filtering problems; i.e. estimating the sequence of posterior distributions for non-linear non-Gaussian state-space models. In this context they are known as particle filtering methods. This class of methods has already had a massive impact in time series analysis, signal processing, robotics, computer vision or target tracking where they have become routine [2].

4.2 Challenges

In statistics and related fields, SMC methods are still closely associated with optimal filtering. This association has somehow hindered their development and hidden the main principles behind this elegant and principled methodology. It has now been realized that the potential of these methods is much greater than was first imagined [1], [3].

SMC methods are by no means limited to time-varying parameter estimation. Broadly speaking, SMC methods rely on a divide-and-conquer strategy for solving computational problems. Specifically, when a problem (which may be static) has a complex structure, it is often useful to decompose the target structure into a sequence of simpler but dynamically evolving substructures. One can then use the information obtained by solving a sequence of easier and smaller problems to help solve the ultimate target problem. The key to the success of this method is often the ability to gradually update the system from the simplest structure to the target structure.

SMC methods can be applied to sample approximately from any sequence of probability distributions of interest and also to estimate their unknown normalizing constants. It is crucial to realize that many important problems can be formulated as special instances of this general framework: sequential and batch Bayesian inference, computation of Bayes factors, computation of p-values, inference in contingency tables, rare event probabilities, optimization, counting the number of objects with a certain property for combinatorial structures, computation of eigenvalues and eigenmeasures of positive operators, PDE's admitting a Feynman-Kac representation and so on.

Despite its success, this young research area is still in its infancy and there are many challenges to address. Like MCMC, SMC are not black box methods and a brute force application of standard algorithms will inevitably fail for complex stochastic models. For example, standard SMC algorithms perform poorly when applied to the high-dimensional models such as space-time models typically used in environmental and ecological statistics or atmospheric sciences.

The main objectives of this program are to

- Address methodological and theoretical problems of SMC methods
- Identify and address important applied problems (e.g. data assimilation)

We anticipate that this program would create an exceptional opportunity for exchanging ideas between the communities; and help to shape the future of stochastic computation. Input from statisticians, computer scientists, engineers or physicists would be absolutely crucial.

4.3 Program Leaders

Anthony E. Brockwell (Statistics, Carnegie Mellon University), Arnaud Doucet (Statistics & Computer Science, University of British Columbia), Simon J. Godsill (Engineering, Cambridge University), Raquel Prado (Applied Mathematics and Statistics, University of California Santa Cruz); National Advisory Council Liaison - David Mumford (Brown University); Directorate Liaison - Jim Berger (SAMSI).

4.4 Suggested Participants

Christophe Andrieu, Bristol, UK; Bhavik R. Bakshi, Ohio, USA; Andrew Blake, Microsoft Cambridge, UK; Mark Briers, Cambridge, UK; Wolfram Burgard, Freiburg, Germany; Mark Coates, McGill, Canada; Dan Crisan, Imperial College, UK; Rong Chen, Illinois, USA; Yuguo Chen, Illinois, USA; Nicolas Chopin, Bristol, UK; Manuel Davy, Lille, France; Nando De Freitas, British Columbia, Canada; Maria De Iorio, Imperial College, UK; Pierre Del Moral, Nice, France; Frank Dellaert, GeorgiaTech, USA; Persi Diaconis, Stanford, USA; Petar Djuric, Stony Brook, USA; Michael Dowd, Dalhousie, Canada; Geir Evensen, Bergen, Norway; Paul Fearnhead, Lancaster, UK; Dieter Fox, Washington, USA; Neil Gordon, DSTO, Australia; Peter Grassberger, Jülich, Germany; Fredrik Gustafsson, Linkoping, Sweden; Simon Haykin, McMaster, Canada; Alfred Hero, Michigan, USA; Michael Isard, Microsoft, USA; Ajay Jasra, Imperial College, UK; Jun Liu, Harvard, USA; Tom Higuchi, ISM Tokyo, Japan; T. Kirubarajan, McMaster, Canada; Genshiro Kitagawa, ISM Tokyo, Japan; Chris Kreucher, Michigan, USA; Hans Kuensch, Zurich, Switzerland; Francois LeGland, IRISA, France; Simon Maskell, QinetiQ, UK; Peter Mueller, Texas, USA; Anastasia Papavasiliou, Warwick, UK; Mike Pitt, Warwick, UK; Avi Pfeffer, Harvard, USA; Nick Polson, Chicago, USA; Raquel Prado, Santa Cruz, USA; James Rawling, Winsconsin, USA; Tobias Ryden, Lund, Sweden; Christian Robert, Paris, France; Gareth Roberts, Lancaster, UK; Chris Rogers, Cambridge, UK; Neil Shephard, Oxford, UK; Sumeetpal Singh, Cambridge, UK; Vladislav B. Tadic, Sheffield, UK; Sebastian Thrun, Stanford, USA; Peter Jan van Leeuwen, Utrecht, Netherlands; Namrata Vaswani, Iowa, USA; Ba-Ngu Vo, Melbourne, Australia; Darren Wilkinson, Newcastle, UK; Patrick Wolfe, Harvard, USA; Junni Zhang, Beijing, China.

References

[1] Del Moral P. (2004) *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications.*New York: Springer-Verlag.
[2] Doucet A., de Freitas J.F.G. & Gordon N.J. (eds.) (2001) *Sequential Monte Carlo Methods in Practice.* New York: Springer-Verlag.
[3] Liu, J.S. (2001) *Monte Carlo Strategies in Scientific Computing.* New York: Springer-Verlag.
[4] Robert C.P. & Casella G. (2004) *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 2nd edition.

5 Risk Analysis, Extreme Events and Decision Theory (Tentative for 2007-08)

5.1 Summary

This is a proposal for full-year SAMSI program that will address fundamental issues in risk analysis, as well as associated problems associated with extreme events and decision theory. The program will engage researchers from the statistical sciences, applied mathematics and the decision sciences, including operations research, and which can yield genuine impact on the practice of risk analysis and assessment,[1] and on theory and methodology for extreme events and decision theory.

5.2 Background

Over the past several years, there has been a wealth of scientific progress on risk analysis. But, the same forces of diversity that have stimulated the progress have to a significant degree engendered fragmentation. To illustrate, the set of underlying problems has become increasingly diverse, drawing from areas ranging from national defense and homeland security to genetically modified organisms to critical infrastructure. This diversity has led to a wealth of effective but often narrowly focused research.

The diversity of problem contexts also has a significant public policy component, which is driven in part by the increasing stakes and the multiplicity of stakeholders. In particular, policy concerns direct attention not only to the dramatic risks for huge numbers of people associated, for example, with events of the magnitude of Hurricane Katrina or bioterrorism, but also to "small-scale" risks such as drug interactions driven by genetic factors that may be relatively rare.

Notwithstanding this diversity, is clear that overarching needs exist. It is equally clear that specialization of research on risk analysis has to some extent inhibited the communication and collaboration needed to address the overarching needs. For example, early detection of extremely rare events is needed in contexts ranging from bioterrorism to such "traditional" areas as automobile safety, the latter exemplified by the Ford Explorer–Firestone tire problem.

On October 27–29, the National Institute of Statistical Sciences (NISS) and Iowa State University (ISU) co-sponsored a *Workshop on Overarching Issues in Risk Analysis*, held in Ames, IA.[2] The workshop was meant to be a "stock-taking" of the proliferation of exciting new research on risk analysis over the past several years, seeking answers to such questions as:

- What are the high-leverage gaps?
- What issues span multiple problem contexts?
- What kinds of collaborations among researchers in the statistical, applied mathematical and decision sciences and domain scientists are needed to carry out the research?

Underlying risk analysis are theoretical and methodological issues associated with extreme events (such as major hurricanes) and decision theory.

Some of the material in §5.4 is drawn from this workshop. Other sources of input are the opening workshop of the 2005–06 SAMSI program on National Defense and Homeland Security (NDHS), a recent

[1] For simplicity, the term "risk analysis" is used below.

[2] Details of the program are available at www.niss.org/affiliates/riskanalysis200510/riskanalysis200510-home.html.

Name	Institution	Discipline	Status
Vicki Bier	Wisconsin	Industrial Engineering	Committed in principle as program leader
Alicia Carriquiry	Iowa State	Statistics	Committed in principle as program leader
Dipak Dey	University of Connecticut	Statistics	Committed in principle as program leader
Wolfgang Kliemann	Iowa State	Applied Mathematics	Committed in principle as program leader
Stephen Pollock	Michigan	Operations Research	Committed in principle as program leader
David Banks	Duke	Statistics	Committed to act as local scientific coordinator
Lawrence Brown	University of Pennsylvania	Statistics	Committed to act as NAC liaison
Alan Karr	NISS	Statistics	Committed to serve as directorate liaison

Table 1: Committed and potential program leaders.

Name	Institution	Discipline	Status
Susan Ellenberg	Pennsylvania	Biostatistics	Attended ISU workshop
Roger Hoerl	GE Research	Statistics	Particular interest in financial risk
Robert Obenchain	Eli Lilly	Statistics	Interest in pharmaceutical risk
Paul Slovic	Oregon	Psychology	Attended ISU workshop; interest in risk perception and communication
Detlof von Winterfeldt	USC	Public Policy	Speaker at NDHS workshop; agreed in principle to serve on scientific committee
Stanley Young	NISS	Statistics	Committed to participate

Table 2: Other potential participants.

Board on Mathematical Sciences and their Applications (BMSA) workshop on statistics of networks[3] and discussions with potential program leaders.

5.3 Personnel

5.3.1 Program Leadership

Table 1 lists leaders of the program, their disciplines and the status of SAMSI contact with them.

Others who have indicated major interest in the program are listed in Table 2.

5.3.2 Faculty Releases

Numerous faculty at Duke, North Carolina State University (NCSU) and the University of North Carolina at Chapel Hill (UNC), in applied mathematics, statistics, operations research, biomedical and medical sciences, environmental sciences, engineering and cognitive sciences can participate in the program as faculty releases. A small set of examples is listed in Table 3.

[3] See www7.nationalacademies.org/bms/statistics_of_networks.html.

Name	Institution	Discipline	Scientific Interest
David Banks	Duke	Statistics	Risk assessment
Barrett Slenning	NCSU	Veterinary Medicine	Diseases of animals
Richard Smith	UNC	Statistics	Extreme events

Table 3: Potential Faculty Releases.

5.3.3 Postdoctoral Fellows

Postdoctorals for the program can be drawn from the statistical, applied mathematical, decision and domain sciences.

5.4 Potential Working Groups

In this section are listed a number of potential working groups in a SAMSI risk program, broken into two categories. Those in the "more certain" category have been discussed to some extent with program leaders and others, and it appears that there is some constituency for them. Those in the "less certain" appear promising, but are not likely to materialize unless leaders can be identified and are willing to commit to significant engagement in the program.

The "more certain" potential working groups are:

Cost prediction. To a significant extent, costs are not treated as random in risk analyses, and are instead "estimated" without attention to associated uncertainties.[4] A number of researchers feel that serious attention to predicting costs is central to improving risk analyses.[5]

Data inadequacies, which were perhaps the most clearly identified—but also in some sense too general—theme at the ISU workshop. Data quality is a major concern, especially in situations (for instance, in post-marketing surveillance for adverse affects of pharmaceuticals and even in clinical trials) that require reliance on self-reported data. Risk surveys (see **Exposure assessment** below) pose similar problems.

One possible and wide-reaching goal is to develop strategies for informed tradeoffs between the value of data and the cost of data. Clearly this research requires appropriate decision-theoretic underpinnings.

Data–model–computation tradeoffs. In risk analyses as in many other settings, there are tradeoffs among data (detail and precision, as well as quality), model complexity and computation power, which are shown schematically in Figure 1.[6] Methods for understanding these tradeoffs—and especially software tools supporting informed modeling decisions—would be a major advance.

Here again, the decision-theoretic infrastructure is essential. Moreover, models of extreme events, such as hurricanes, present additional challenges. For example, prediction of hurricane tracks poses

[4] In an entirely different setting, but confirming attention to the issues, an April 2006 workshop on survey costs being organized by NISS will have cost prediction as one of its principal themes, and can inform this working group.

[5] This working group would have significant interaction with other initiatives at NISS and elsewhere, including efforts to predict survey costs.

[6] This issue is widespread. For example, it also arose at a recent NISS/SAMSI Workshop on Collaborations in the Mathematical Geosciences; see http://www.niss.org/cmg200510/cmg200510-index.html.

clear data, model and computation stresses. Within this setting, we cannot currently answer question such as: "What would yield the most improved and timely predictions–more or better data, better models, or more computation (e.g., to explore multiple scenarios)?" Relating improved predictions to the costs (not only monetary, but also in terms of time) to the associated benefits would introduce risk considerations in settings in which they have hitherto been absent.

Exposure assessment has long been a weakness of many risk analyses, because of the large unobserved variability. In some settings, such as environmental risk—a natural point of intersection with other SAMSI themes, the problems are well-recognized. However, they occur in other settings as well; for example, excessive exposure to nutrients can be dangerous, but existing means of estimating exposure are simply not effective at identifying at-risk individuals with extreme exposure.

Even estimation of levels of extreme exposure raises non-standard issues.

Model validation. This ubiquitous issue has special significance in risk analysis, since for some rare events whose risk must be assessed, there are no data. An example drawn from the SAMSI 2005–06 program on NDHS is an outbreak of foot and mouth disease among cattle in the US. The opportunity to link this working group with the 2006–07 SAMSI program on "Development, Assessment and Utilization of Complex Computer Models" is also attractive.

Risk in the pharmaceutical industry. Although some may consider this topic too "specialized" for SAMSI, it is of broad interest. Issues raised at the ISU workshop included the increasing importance of non-randomized clinical trials,[7] multiplicity[8] and asymmetries between null and alternative hypotheses—at the Food and Drug Administration (FDA) and elsewhere.

Elicitation of expert opinion. It is well-recognized that the "data" underlying many risk analyses constitute expert opinions rather than conventional measurements. Despite ongoing attention to eliciting and incorporating expert opinion, there do not seem to be effective, available tools.

The "less certain" potential working groups, in part because leaders are less apparent, are: are:

Modeling resource allocation decisions by multiple stakeholders. In many contexts (Hurricane Katrina provides one), decisions to mitigate future risks are made by multiple stakeholders (e.g., federal, state and local governments), but methods do not exist to integrate these decisions. Existing decision-theoretic abstractions and tools may not be adequate to deal with these problems.

Multiscale risk analysis. Risks exist on multiple time and space scales, but no good methods exist for addressing multiscale risks. Previous SAMSI programs on multiscale modeling and data assimilation may have initiated threads of research that are relevant to this problem.

Risk perception. An especially challenging issue is the increasing divergence between risk perception and risk reality. (For example, based on numbers of deaths, in most years it is more dangerous to drive to the airport than to fly.) Tools to communicate and visualize risk represent a research as well as a policy need.

A particular intriguing problem is interactions among the risk receptions of related individuals (e.g., members of the same household, or employees of the same company). Relevant ideas come from existing research on risk perceptions of individuals, group decision and social networks.[9]

---

[7] Patients routinely thwart randomized trials by having the medications they receive analyzed externally.

[8] The focus the SAMSI Summer 2006 program on Multiplicity and Reproducibility in Scientific Studies.

[9] A working group in the 2005–06 SAMSI NDHS program.

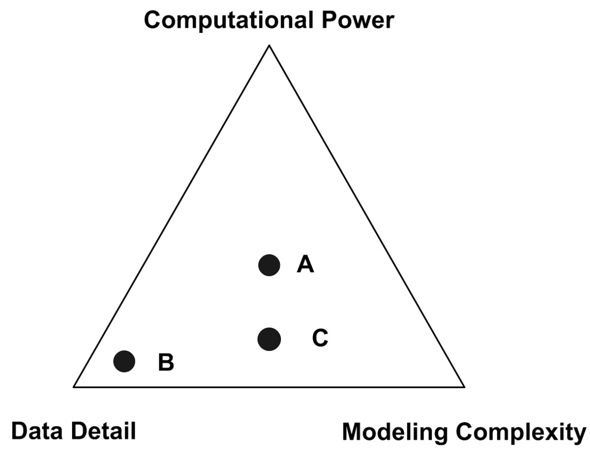


Figure 1: Conceptual representation of tradeoffs among data detail, model complexity and computational power in geoscience (or other) modeling. Three approaches are illustrated. Approach A employs data, models and computation in equal proportions. Approach B depends almost entirely on data. Approach C uses data and models equally, but with less dependence on computation.

Given the strong interest in risk analysis at ISU, not only among statisticians and applied mathematics who might commit to spend significant amounts of time at Statistical and Applied Mathematical Sciences Institute (SAMSI), but also among domain scientists who are much less likely to do so, and if the Fall, 2006, experiment of a bi-nodal working group[10] is successful, is is possible that ISU might become the "base location" for one or more working groups.

5.5 Leveraging

The SAMSI risk analysis program will inherit momentum from other SAMSI programs, NISS and affiliates. Here is a brief summary.

From other SAMSI programs. As noted in §5.2 and §5.4, this program would build on previous SAMSI programs on environmental models (2002–03), internet modeling (2003–04), multiscale modeling (2003–04), latent variable methods in the social sciences (2004–05), national defense and homeland security (2005–06)[11] and computer models (2006–07).

From NISS. NISS research projects on data confidentiality,[12] data quality, software engineering, social networks and survey costs all intersect this program.

From affiliates. Virtually every federal agency, industrial and national laboratory affiliate has substantial interest in risk analysis. The **Risk in the pharmaceutical industry** working group would be a major draw for affiliates from the pharmaceutical industry.

[10] In the program on High Dimensional Inference and Random Matrices.
[11] In which three of the four working groups involved major attention to risk, extreme events or decision theory.
[12] In which *disclosure risk* is a major focus.

APPENDIX E – Workshop Participants

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 (to be added later) were compiled from this data.

The key top Status entry is as follows:

NRG -- New Researcher or Graduate Student
FP – Faculty/Professional
S – Students (Education & Outreach)
A – Faculty (Education & Outreach)

Latent Variables in the Social Sciences
GLAMM Seminars
NISS-SAMSI Building
April 13-15, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	1	1	0	2	0	1	0	1	2	2
Unsuppted	11	3	0	7	6	5	1	7	7	3
SAMSI	2	1	0	0	2	1	0	0	NA	NA

Latent Variables in the Social Sciences
GLAMM Seminars
NISS-SAMSI Building
Support Workshop Participants
April 13-15, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Rabe-Hesketh	Sophie	F	U of California, Berkeley	Education	FP
Skrondal	Anders	M	London Sch of Economics	Statistics	FP

Latent Variables in the Social Sciences
GLAMM Seminars
NISS-SAMSI Building
Workshop Participants
April 13-15, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Berger	James	M	SAMSI & Duke U	Statistics	FP
Bollen	Ken	M	U of North Carolina	Sociology	FP
Edwards	Lloyd	M	U of North Carolina	Biostatistics	FP
Ghosal	Subhashis	M	North Carolina State U	Statistics	FP
Gu	Jiezhun	F	North Carolina State U	Statistics	NRG
Hipp	John	M	U of North Carolina	Sociology	NRG
Kamata	Aki	M	Florida State U	Statistics	FP
Karr	Alan	M	NISS		FP
Kinney	Saki	F	Duke U	Statistics	NRG
Nguyen	Hoan	F	SAMSI		NRG
Palomo	Jesus	M	SAMSI		NRG
Rabe-Hesketh	Sophie	F	U of California, Berkeley	Education	FP
Samuels	John	M	North Carolina State U	Mathematics	NRG
Skrondal	Anders	M	London Sch of Economics	Statistics	FP
Visser	Ingmar	M	U of Amsterdam	Psychology	FP

Latent Variables in the Social Sciences
Year-End Working Group Summaries
Radisson Hotel Research Triangle Park
May 19, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	0	0	0	0	0	0	0	0	0	0
Unsuppted	24	5	0	17	12	6	3	20	12	3
SAMSI	6	1	0	3	4	0	2	2	NA	NA

Latent Variables in the Social Sciences
Year-End Working Group Summaries
Radisson Hotel Research Triangle Park
Workshop Participants
May 19, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Banks	David	M	Duke U	Statistics & Decision Sciences	FP
Banks	H.T.	M	SAMSI & North Carolina State U		FP
Bauer	Dan	M	U of North Carolina	Psychology	FP
Bollen	Ken	M	SAMSI & U of North Carolina	Sociology & Odum Institute	FP
Cai	Bo	M	NIEHS		FP
Dunson	David	M	NIEHS	Biostatistics Branch	FP
Edwards	Lloyd	M	U of North Carolina	Biostatistics	FP
Ghosal	Subhashis	M	North Carolina State U	Statistics	FP
Gu	Jiezhun	F	North Carolina State U	Statistics	NRG
Hipp	John	M	U of North Carolina	Sociology	NRG
Hong	Chung-Chien	M	North Carolina State U	mathematics	NRG
Ickstadt	Katja	F	U of Dortmund	Statistics	FP

Kamata	Aki	M	Florida State U		FP
Karr	Alan	M	NISS		FP
Kelly	Christopher	M	U of North Carolina	Institute on Aging	NRG
Kenney	Melissa	F	Duke U	Environmental & Earth Sciences	NRG
Kinney	Saki	F	Duke U	ISDS	NRG
Medhin	Negash	M	North Carolina State U	Mathematics	FP
Miyazaki	Yasuo	M	Virginia Tech	Educational Leadership & Policy Studies	FP
Munk	Tom	M	U of North Carolina	Education	FP
Palomo	Jesus	M	SAMSI		NRG
Ray	Surajit	M	SAMSI		NRG
Samuels	John	M	SAMSI & North Carolina State U	Mathematics	NRG
Stenner	Jack	M	MetaMetrics, Inc		FP
Thomas	Roland	M	Carleton U	Sprott School of Business	FP
Truong	Young	M	SAMSI & U of North Carolina	Biostatistics	FP
Vance	Eric	M	Duke U	Statistics & Decision Sciences	NRG
Visser	Ingmar	M	U of Amsterdam	Psychology	NRG
Zavisca	Jane	F	SAMSI		NRG

Genomes to Global Health: Computational Biology of Infectious Diseases
Transition Workshop
Radisson Hotel Research Triangle Park
May 22-24, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	5	1	0	6	0	0	0	6	5	3
Unsuppted	14	5	0	11	8	0	1	6	6	2
SAMSI	3	0	0	2	1	1	0	2	NA	NA

Genomes to Global Health: Computational Biology of Infectious Diseases
Transition Workshop
Radisson Hotel Research Triangle Park
Support Workshop Participants
May 22-24, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Haak	Ron	M	Institute for OneWorld Health		FP
Isokpehi	Raphael	M	Jackson State U		FP
Kalyanaraman	Chakrapani	M	U of California San Francisco	Pharmaceutical Chemistry	FP
Kissinger	Jessica	F	U of Georgia	Genetics	FP
Marti-Renom	Marc A.	M	U of California San Francisco	Biopharmaceutical Sciences	FP
Maurer	Stephen	M	U of California Berkeley	Info Tech & Homeland Security Project	FP

Genomes to Global Health: Computational Biology of Infectious Diseases
Transition Workshop
Radisson Hotel Research Triangle Park
Workshop Participants
May 22-24, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Awadalla	Philip	M	North Carolina State U	Genetics	FP
Berger	Jim	M	SAMSI		FP

Boyle	James	M	Duke U	Law School	FP
Burk	Dan	M	U of Minnesota	Law School	FP
Chan	Cliburn	M	Duke U	Bioinformatics & Computational Biology	NRG
Cooke	Ben	M	Duke U	Mathematics	NRG
Cowell	Lindsay	F	Duke U	Bioinformatics & Computational Biology	NRG
Crossman	Colin	M	Duke U	School of Law	NRG
Desowitz	Bob	M	U of North Carolina		FP
Haak	Ron	M	Institute for OneWorld Health		FP
Haystead	Tim	M	Duke U	Pharmacology	FP
He	Min	M	Duke U	Bioinformatics & Computational Biology	NRG
Isokpehi	Raphael	M	Jackson State U		FP
Kalyanaraman	Chakrapani	M	U of California San Francisco	Pharmaceutical Chemistry	FP
Kepler	Tom	M	Duke U	Bioinformatics & Computational Biology	FP
Kissinger	Jessica	F	U of Georgia	Genetics	FP
Marti-Renom	Marc A.	M	U of California San Francisco	Biopharmaceutical Sciences	FP
Maurer	Stephen	M	U of California Berkeley	Info Tech & Homeland Security Project	FP
McGovern	Victoria	F	Burroughs Wellcome Fund		FP
Munshaw	Supriya	F	Duke U	Bioinformatics & Genome Technology	NRG
Rai	Arti	F	Duke U	School of Law	FP
Ray	Surajit	M	SAMSI		NRG
Sung	Nancy	F	Burroughs Wellcome Fund		FP

Truong	Young	M	SAMSI		FP
Volpe	Joe	M	Duke U	Bioinformatics & Genome Technology	NRG

Education and Outreach
MAA Prep Workshop: Mathematics Meets Biology
University of Louisiana-Lafayette
May 25-28, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	0	0	0	0	0	0	0	0	0	0
Unsuppted	14	13	2	17	12	0	20	9	15	10
SAMSI	0	0	0	0	0	0	0	0	NA	NA

Education and Outreach
MAA Prep Workshop: Mathematics Meets Biology
University of Louisiana-Lafayette
Workshop Participants
May 25-28, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Ackleh	Azmy	M	U of Louisiana at Lafayette	Mathematics	FP
Adams	Brian	M	North Carolina State U	CRSC	NRG
Aggarwal	Shiv		Embry-Riddle Aero U	Physical Sciences	FP
Atkinson	Graham	M	Self-employed		FP
Banks	H.T.	M	North Carolina State U	CRSC	FP
Becerra	Linda	F	U of Houston-Downtown	Computer & Mathematical Sciences	FP
Boudreaux	Seth	M	U of Louisiana at Lafayette	Mathematics	NRG
Comar	Timothy	M	Benedictine U	Mathematics	FP

Deng	Keng	M	U of Louisiana at Lafayette	Mathematics	FP
Dib	Youssef	M	U of Louisiana at Lafayette	Mathematics	NRG
Frantz	Michael	M	U of LaVerne	Math, Physics & Computer Sciences	FP
Friedman	Jane	F	U of San Diego	Mathematics	FP
Grove	Sarah	F	North Carolina State U	CRSC	NRG
Hu	Shuhua	F	North Carolina State U	CRSC	NRG
Jang	Sophia	F	U of Louisiana at Lafayette	Mathematics	FP
Karhbet	Sam	M	U of Louisiana at Lafayette	Mathematics	NRG
Lanminh	Pham	F	U of Louisiana at Lafayette	Biology	NRG
Lee	Namyong	M	Minnesota State U, Mankato	Mathematics & Statistics	FP
Myers	Leigh	F	Northwestern State U	Mathematics	FP
Nguyen	Hanh	F	U of Louisiana at Lafayette	Biology	NRG
Nunez	Joanna	F	U of Louisiana at Lafayette	Biology	NRG
Oppenheimer	Bonnie	F	Mississippi U for Women	Science & Mathematics	FP
Oppenheimer	Seth	M	Mississippi State U	Mathematics & Statistics	FP
Patterson	Sam	M	Carleton College	Mathematics & Computer Science	FP
Pham	Lanminh	F	U of Louisiana at Lafayette	Biology	NRG
Saydam	Azime	F	U of Louisiana at Munroe	Mathematics & Physics	FP
Thibodeaux	Jeremy	M	U of Louisiana at Lafayette	Mathematics	NRG
Wang	Xubo	F	U of Louisiana at Lafayette		NRG
Zeleke	Aklilu		Alma College	Mathematics & Computer Science	FP

Education & Outreach
SAMSI-CRSC Interdisciplinary Workshop for Undergraduates
North Carolina State University
May 30-June 3, 2005

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Colleges/Univ	Number of Home State
Supported	16	9	0	0	24	5	19	22	14
Unsuppted	7	5	0	2	10	9	3	5	2
SAMSI	5	4	0	5	4	7	2	NA	NA

Education & Outreach
SAMSI-CRSC Interdisciplinary Workshop for Undergraduates
North Carolina State University
Supported Workshop Participants
May 30-June 3, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Benim	Robert	M	U of Portland	Mathematics	S
Deems	Thomas	M	Slippery Rock U	Mathematics	S
Geis	Jennifer	F	Augsburg C	Mathematics	S
Gerdes	Adam	M	Bemidji State U		S
Groark	Erin	F	Indiana U – Bloomington		S
Hammerbacher	Jeff	M	Harvard U		S
Ibarra	Leonel	M	U of San Diego		S
Isaacs	Abby	F	Antioch College		S
Kofron	David	M	U of Miami		S
Murray	Kelly Rae	F	Indiana U		S
Owais	Aatekah	F	Mount Holyoke Co		S
Patisteas	Dan	M	Drew U		S

Porter	Jacob	M	UC Berkeley		S
Rodriguez	Nancy	F	U of San Diego		S
Slater	Carl	M	Morehouse College		S
Small	Stephan	M	Norfolk State U		S
Tingey	Troy	M	Arizona State U	Mathematics	S
Tolosa	Andres	M	Penn State U		S
Toutain	Genevive	F	Simon's Rock C of Bard		S
Varela	Dario	M	Arizona State U	Mathematics	S
Vogl	Chris	M	Illinois Wesleyan U		S
Wheeler	Micah	M	Christian Brothers U		S
Wilkoff	Sean	M	U of C at Berkeley		S
Wong	Anastasia	F	Mills C	Statistics	S

Education & Outreach
SAMSI-CRSC Interdisciplinary Workshop for Undergraduates
North Carolina State University
Workshop Participants
May 30-June 3, 2005

Last Name	**First Name**	**Gender**	**Affiliation**	**Department**	**Status**
Banks	H.T.	M	SAMSI & North Carolina State U	CRSC	F
Benim	Robert	M	U of Portland	Mathematics	S
Deems	Thomas	M	Slippery Rock U	Mathematics	S
Geis	Jennifer	F	Augsburg C		S

Gerdes	Adam	M	Bemidji State U		S
Groark	Erin	F	Indiana U – Bloomington		S
Gu	Sherry	F	SAMSI & North Carolina State U	Statistics	NRG
Hammerbacher	Jeff	M	Harvard U		S
Ibarra	Leonel	M	U of San Diego		S
Isaacs	Abby	F	Antioch College		S
Joyner	Sarah Lynn	F	Meredith College	Mathematics	S
Jung	Minjung	F	SAMSI & North Carolina State U	Statistics	NRG
Kofron	David	M	U of Miami		S
Latterman	Russell	M	Arizona State U	Mathematics	S
Murray	Kelly Rae	F	Indiana U		S
Nguyen	Hoan	F	SAMSI & North Carolina State U	Mathematics	NRG
Owais	Aatekah	F	Mount Holyoke Co		S
Patisteas	Dan	M	Drew U		S
Porter	Jacob	M	UC Berkeley		S
Ray	Surajit	M	SAMSI & U of North Carolina	Biostatistics	NRG
Rodriguez	Nancy	F	U of San Diego		S
Root	Morgan	M	SAMSI & North Carolina State U	Mathematics	NRG
Samuels	John	M	SAMSI & North Carolina State U	Mathematics	NRG
Slater	Carl	M	Morehouse College		S
Small	Stephan	M	Norfolk State U		S

Smith	Ralph	M	SAMSI & North Carolina State U	CRSC	F
Tingey	Troy	M	Arizona State U	Mathematics	S
Tolosa	Andres	M	Penn State U		S
Toutain	Genevive	F	Simon's Rock C of Bard		S
Tweedy	Eamonn	M	NC State U	Physics	S
Varela	Dario	M	Arizona State U	Mathematics	S
Vogl	Chris	M	Illinois Wesleyan U		S
Wheeler	Micah	M	Christian Brothers U		S
Wilkoff	Sean	M	U of C at Berkeley		S
Wong	Anastasia	F	Mills C	Statistics	S
Zavisca	Jane	F	SAMSI		NRG

Data Assimilation for Geophysical Systems
Summer School on Fusing Geophysical Models with Data: From Theory to Practice to Theory
NCAR – Boulder, CO
June 13-17, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	15	9	6	6	24	12	8	3	20	12
Unsuppted	7	3	0	8	2	0	1	9	5	4
SAMSI	2	0	0	0	2	1	1	0	NA	NA

Data Assimilation for Geophysical Systems
Summer School on Fusing Geophysical Models with Data: From Theory to Practice to Theory
NCAR – Boulder, CO
Supported Workshop Participants
June 13-17, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Apte	Amit	M	SAMSI & UNC	Mathematics	NRG
Arab	Ali	M	U of Missouri, Columbia	Statistics	NRG
Barbu	Alina	F	Delft U of Technology	Applied Mathematics	NRG
Bengtsson	Thomas	M	U of Califonia Berkeley	Statistics	FP
Broeker	Jochen	M	London School of Economics		NRG
Chen	Li		U of Chicago	CISES	NRG
El Moghraby	Amal	F	U of North Carolina & Brown U	Mathematics	NRG
Fitzmaurice	Jean	F	Massachusetts Institute of Technology	Hydrology	NRG
Foley	Kristen	F	North Carolina State U	Mathematics	NRG
Herbei	Radu	M	Florida State U	Statistics	NRG
Judd	Kevin	M	U of Western Australia	Mathematics	FP
Jun	Mikyoung	F	U of Chicago	Statistics	NRG
Khare	Shree	M	SAMSI		NRG
Kim	Yongku		Ohio State U	Statistics	NRG
Lam	Eric	M	Ohio State U	Statistics	NRG
Liu	Liyan	F	U of North Carolina at Chapel Hill	Mathematics	NRG
Liu	Yuqiong		U of Arizona	Hydrology	NRG

Lokupitiya	Ravindra		Colorado State U	Statistics	NRG
Pendleton	Dan	M	Cornell U	Earth & Atmospheric Sci.	NRG
Reynolds	Carolyn	F	Naval Research Laboratory	Marine Meteorology Division	FP
Shaby	Ben	M	Cornell U	Statistics	NRG
Shao	Xiaofeng		U of Chicago	Statistics	NRG
Smith	Leonard	M	U of Oxford		FP
Stroud	Jonathan	M	U of Pennslyvania	Statistics	FP
Sun	Zhibin		U of Maryland	Mathematics and Statistics	NRG
Wikle	Chris	M	U of Missouri, Columbia	Statistics	FP
Yang	Wenzheng		U of Southern California	Earth Science	NRG
Yao	Yonggang		Ohio State U	Statistics	NRG
Zhang	Zepu		U of Chicago	Geo. & Environmental Sci and Statistics	NRG
Zhou	Jie	F	U of North Carolina at Chapel Hill	Mathematics	NRG

Data Assimilation for Geophysical Systems
Summer School on Fusing Geophysical Models with Data: From Theory to Practice to Theory
NCAR – Boulder, CO
Workshop Participants
June 13-17, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Anderson	Jeff	M	NCAR		FP
Apte	Amit	M	SAMSI & UNC	Mathematics	NRG
Arab	Ali	M	U of Missouri, Columbia	Statistics	NRG

Barbu	Alina	F	Delft U of Technology	Applied Mathematics	NRG
Bengtsson	Thomas	M	U of Califonia Berkeley	Statistics	FP
Broeker	Jochen	M	London School of Economics		NRG
Chen	Li	F	U of Chicago	CISES	NRG
El Moghraby	Amal	F	U of North Carolina & Brown U	Mathematics	NRG
Fitzmaurice	Jean	F	Massachusetts Institute of Technology	Hydrology	NRG
Foley	Kristen	F	North Carolina State U	Mathematics	NRG
Gettleman	Andrew	M	NCAR		FP
Herbei	Radu	M	Florida State U	Statistics	NRG
Huang	Xiang-Yu	M	NCAR		FP
Ide	Kayo	F	U of California Los Angeles	Atmospheric Sciences	FP
Jones	Christopher	M	U of North Carolina at Chapel Hill	Mathematics	FP
Judd	Kevin	M	U of Western Australia	Mathematics	FP
Jun	Mikyoung	F	U of Chicago	Statistics	NRG
Khare	Shree	M	SAMSI		NRG
Kim	Yongku		Ohio State U	Statistics	NRG
Lam	Eric	M	Ohio State U	Statistics	NRG
Liu	Liyan	F	U of North Carolina at Chapel Hill	Mathematics	NRG
Liu	Yuqiong		U of Arizona	Hydrology	NRG
Lokupitiya	Ravindra	M	Colorado State U	Statistics	NRG
Nychka	Doug	M	NCAR		FP

Pendleton	Dan	M	Cornell U	Earth & Atmospheric Sci	NRG
Reynolds	Carolyn	F	Naval Research Laboratory	Marine Meterology Division	FP
Ross	Natalie	F	U of Colorado at Boulder	Computer Sciences	NRG
Shaby	Ben	M	Cornell U	Statistics	NRG
Shao	Xiaofeng	M	U of Chicago	Statistics	NRG
Smith	Leonard	M	U of Oxford		FP
Snyder	Chris	M	NCAR		FP
Stroud	Jonathan	M	U of Pennslyvania	Statistics	FP
Sullivan	Peter	M	NCAR		FP
Sun	Zhibin		U of Maryland	Mathematics and Statistics	NRG
Wikle	Chris	M	U of Missouri, Columbia	Statistics	FP
Yang	Wenzheng		U of Southern California	Earth Science	NRG
Yao	Yonggang		Ohio State U	Statistics	NRG
Zhang	Zepu		U of Chicago	Geo & Env Sciences and Statistics	NRG
Zhou	Jie	F	U of North Carolina at Chapel Hill	Mathematics	NRG
Zubrow	Alexis	F	U of Chicago	CISES	NRG

Hot Topics Workshop
Random Graphs and Stochastic Computation
NISS-SAMSI Building
June 13-14, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	12	8	0	11	9	6	6	8	19	10
Unsuppted	15	1	0	11	5	7	2	28	6	3
SAMSI	1	0	0	0	1	1	0	0	NA	NA

Hot Topics Workshop
Random Graphs and Stochastic Computation
NISS-SAMSI Building
Supported Workshop Participants
June 13-14, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Bayarri	M.J.	F	U of Valencia	Statistics	FP
Belkin	Mikhail	M	U of Chicago	Computer Science	FP
Everitt	Richard	M	U of Bristol	Mathematics	NRG
Handcock	Mark S.	M	U of Washington	Statistics	FP
Jones	Beatrix	F	Massey U	Information & Mathematical Sciense	NRG
Kang	Yun	F	Arizona State U	Mathematics	NRG
Massam	Helene	F	York U	Mathematics & Statistics	FP
Rodriguez	Carlos	M	State U of New York at Albany	Mathematics	FP
Sainudiin	Raazesh	M	Cornell U	Statistics	FP
Servatius	Brigitte	F	WPI	Mathematics	FP
Talih	Makram	M	CUNY-Hunter	Mathematics & Statistics	NRG
Tarantola	Claudia	F	U of Pavia	Economics	FP

Thomas	Alun	M	U of Utah		FP
Thompson	Steven	M	Penn State U	Statistics	FP
Vitek	Olga	F	Purdue U	Statistics	FP
Xing	Eric	M	Carnegie Mellon U	Computer Science	NRG
Xu	Hong	M	Penn State U	Statistics	NRG
Yuan	Ming	M	GA Tech		NRG
Zhou	Jasmine	F	U of Southern California	Computational and Molecular Biology	NRG
Znamenskiy	Dmitri	M	Eindhoven Technical U	EURANDOM	NRG

Hot Topics Workshop
Random Graphs and Stochastic Computation
NISS-SAMSI Building
Workshop Participants
June 13-14, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Bayarri	M.J.	F	U of Valencia	Statistics	FP
Belkin	Mikhail	M	U of Chicago	Computer Science	FP
Cai	Bo	M	NIEHS		FP
Carvalho	Carlos	M	Duke U	Statistics	NRG
Dewal	Rahul	M	Duke U	Statistics	NRG
Dinwoodie	Ian	M	Duke U	Statistics	FP
Dobra	Adrian	M	Duke U	Molecular Genetics & Microbiology	FP
Dunson	David	M	NIEHS & Duke U	Statistics	FP

Everitt	Richard	M	U of Bristol	Mathematics	NRG
Handcock	Mark S.	M	U of Washington	Statistics	FP
Hartemink	Alex	M	Duke U	Computer Science	FP
Hengartner	Nick	M	Los Alamos National Laboratory		FP
Hobolth	Asger	M	North Carolina State U		FP
Huber	Mark	M	Duke U	Mathematics & Statistics	FP
Jones	Beatrix	F	Massey U	Information & Mathematical Sciense	NRG
Kang	Yun	F	Arizona State U	Mathematics	NRG
Lim	Johan	M	Texas A&M U	Statistics	FP
Lucas	Joseph	M	Duke U	Statistics	NRG
Massam	Helene	F	York U	Mathematics & Statistics	FP
Mukherjee	Shayan	M	Duke U	Biostatistics & Bioinformatics	FP
Palomo	Jesus	M	SAMSI		NRG
Rodriguez	Carlos	M	State U of New York at Albany	Mathematics	FP
Sainudiin	Raazesh	M	Cornell U	Statistics	FP
Servatius	Brigitte	F	WPI	Mathematics	FP
Talih	Makram	M	CUNY-Hunter	Mathematics & Statistics	NRG
Tarantola	Claudia	F	U of Pavia	Economics	FP
Thomas	Alun	M	U of Utah		FP
Thompson	Steven	M	Penn State U	Statistics	FP
Vitek	Olga	F	Purdue U	Statistics	FP

West	Mike	M	Duke U	Statistics	FP
Xing	Eric	M	Carnegie Mellon U	Computer Science	NRG
Xu	Hong	M	Penn State U	Statistics	NRG
Yoshida	Ruriko	F	Duke U	Mathematics	NRG
Yuan	Ming	M	GA Tech		NRG
Zhou	Jasmine	F	U of Southern California	Computational and Molecular Biology	NRG
Znamenskiy	Dmitri	M	Eindhoven Technical U	EURANDOM	NRG

Data Assimilation for Geophysical Systems
Bridging Statistical Approaches and Sequential Data Assimilation
NISS-SAMSI Building
June 27, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	5	0	0	3	2	3	2	0	4	3
Unsuppted	11	11	0	6	16	15	4	3	6	2
SAMSI	4	0	0	2	2	2	1	1	NA	NA

Data Assimilation for Geophysical Systems
Bridging Statistical Approaches and Sequential Data Assimilation
NISS-SAMSI Building
Supported Workshop Participants
June 27, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Berliner	Mark	M	Ohio State U	Statistics	FP
Cressie	Noel	M	Ohio State U	Statistics	FP
Kim	Sangil	M	U of Arizona		NRG

Kouritzin	Michael	M	U of Alberta	Mathematical and Statistical Sciences	FP
Stroud	Jonathan	M	The Wharton School, U of Pennsylvania	Statistics	NRG

Data Assimilation for Geophysical Systems
Bridging Statistical Approaches and Sequential Data Assimilation
NISS-SAMSI Building
Workshop Participants
June 27, 2005

Last Name	**First Name**	**Gender**	**Affiliation**	**Department**	**Status**
Apte	Amit	M	SAMSI & U of North Carolina	Mathematics	NRG
Berger	James	M	SAMSI		FP
Berliner	Mark	M	Ohio State U	Statistics	FP
Bhattacharya	Sourabh	M	Duke U	Institute of Statistics and Decison Science	NRG
Budhiraja	Amarjit	M	U of North Carolina	Statistics & Operations Research	FP
Chiswell	Karen	F	North Carolina State U	Statistics	NRG
Choi	Jungsoon	F	North Carolina State U	Statistics	NRG
Cressie	Noel	M	Ohio State U	Statistics	FP
El Moghraby	Amal	F	Brown U / U of North Carolina-Chapel Hill	Applied Mathematics	NRG
Foley	Kristen	F	North Carolina State U	Statistics	NRG
Fuentes	Montserrat	F	North Carolina State U	Statistics	FP
Hughes-Oliver	Jacqueline	F	North Carolina State U	Statistics	FP
Khare	Shree	M	SAMSI		NRG
Kim	Sangil	M	U of Arizona	Mathematics & Statistics	NRG

Kouritzin	Michael	M	U of Alberta	Mathematical and Statistical Sciences	FP
Liu	Liyan	F	U of North Carolina	Mathematics	NRG
Liu	Fei	F	Duke U	Statistics	NRG
Nail	Amy	F	North Carolina State U	Statistics	NRG
Ouyang	Zhi	M	Duke U	Institute of Statistics and Decison Science	NRG
Park	Man Sik	M	North Carolina State U	Statistics	NRG
Sang	Huiyan	M	Duke U	Institute of Statistics and Decison Science	NRG
Smith	Leonard	M	SAMSI/LSE/Oxford	Statistics and Physics	FP
Song	Hae Ryoung	F	North Carolina State U	Statistics	NRG
Stroud	Jonathan	M	The Wharton School, U of Pennsylvania	Statistics	NRG
Wan	Xiaohai	M	North Carolina State U	Mathematics	NRG
Wolpert	Robert	M	Duke U	Institute of Statistics and Decison Science	FP
Zhang	Kaishan	F	North Carolina State U	Civil Engineering	NRG

Astrostatistics
Planning Workshop
NASA Ames Research Center, California
July 14-16, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	12	3	0	14	1	6	0	9	12	6
Unsuppted	2	1	0	3	0	1	0	2	3	2
SAMSI	1	0	0	1	0	1	0	0	NA	NA

Astrostatistics
Planning Workshop
NASA Ames Research Center, California
Supported Workshop Participants
July 14-16, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Babu	Jogesh	M	Penn State U	Statistics	FP
Bickel	Peter	M	U of California, Berkeley	Statistics	FP
Bullard	Floyd	M	Duke U	Institute of Statistics and Decision Sciences	NRG
Clyde	Merlise	F	Duke U	Institute of Statistics and Decision Sciences	FP
Connolly	Andrew	M	Carnegie Mellon U	Astronomy	FP
Gregory	Phil	M	U of British Columbia	Astronomy	FP
Guglielmetti	Fabrizia	F	Max-Planck Institute		FP
Jefferys	Bill	M	U of New Hampshire	Astronomy	FP
Loredo	Tom	M	Cornell U	Astronomy	FP
Lyons	Louis	M	Oxford U	Physics	FP
Murtagh	Fionn	M	U of London	Computer Sciences	FP
Richards	Don	M	Penn State U	Statistics	FP
Sosey	Megan	F	Space Telescope Science Institute		FP
van Dyk	David	M	U of California, Irvine		FP
Wasserman	Larry	M	Carnegie Mellon U	Statistics	FP

Astrostatistics
Planning Workshop
NASA Ames Research Center, California
Workshop Participants
July 14-16, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Babu	Jogesh	M	Penn State U	Statistics	FP
Berger	James	M	SAMSI		FP
Bickel	Peter	M	U of California, Berkeley	Statistics	FP
Bullard	Floyd	M	Duke U	Institute of Statistics and Decision Sciences	NRG
Clyde	Merlise	F	Duke U	Institute of Statistics and Decision Sciences	FP
Connors	Alanna	F	Eureka Scientific		FP
Connolly	Andrew	M	Carnegie Mellon U	Astronomy	FP
Gregory	Phil	M	U of British Columbia	Astronomy	FP
Guglielmetti	Fabrizia	F	Max-Planck Institute		NRG
Jefferys	Bill	M	U of New Hampshire	Astronomy	FP
Loredo	Tom	M	Cornell U	Astronomy	FP
Lyons	Louis	M	Oxford U	Physics	FP
Murtagh	Fionn	M	U of London	Computer Sciences	FP
Richards	Don	M	Penn State U	Statistics	FP
Scargle	Jeff	M	NASA-AMES	Astronomy	FP
Sosey	Megan	F	Space Telescope Science Institute		FP
van Dyk	David	M	U of California, Irvine		FP

Wasserman	Larry	M	Carnegie Mellon U	Statistics	FP

Data Mining and Machine Learning
Technology Transfer Short Course
NISS-SAMSI Building
July 25-29, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	0	2	0	1	1	1	0	1	2	2
Unsuppted	11	3	0	9	5	6	0	10	10	2
SAMSI	0	0	0	0	0	0	0	0	NA	NA

Data Mining and Machine Learning
Technology Transfer Short Course
NISS-SAMSI Building
Supported Workshop Participants
July 25-29, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Denby	Lorraine	F	Avaya		FP
Wu	Yuping	F	U of South Carolina	Statistics	NRG

Data Mining and Machine Learning
Technology Transfer Short Course
NISS-SAMSI Building
Workshop Participants
July 25-29, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Banks	David	M	Duke U	Statistics	FP
Bing	Nan	M	GlaxoSmithKline		FP
Choi	Jungsoon	F	North Carolina State U	Statistics	NRG
Clarke	Bertrand	M	U of British Columbia	Statistics	FP

Denby	Lorraine	F	Avaya		FP
Drummey	Kevin Ward	M	Department of Defense		FP
Ghosh	Joyee	F	Duke U	Statistics	NRG
Gong	Xiaohua	M	North Carolina State U	Bioinformatics Research Cen and Statistics	NRG
Karr	Alan	M	NISS		FP
Liu	Delong	M	CIIT Centers for Health Research		FP
Maynard	Andrew	M	ICAgen	Chemistry	FP
Pantula	Sastry	M	North Carolina State U	Statistics	FP
Song	Hae Ryoung	F	North Carolina State U	Statistics	NRG
Vera	Francisco	M	National Institute of Statistical Sciences		NRG
Wang	Cliff	M	ARO		FP
Wu	Yuping	F	U of South Carolina	Statistics	NRG

Education and Outreach
SAMSI-CRSC Industrial Mathematical and Statistical Modeling Workshop for Graduates

North Carolina State University
July 25-August 2, 2005

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Colleges/Univ	Number of Home State
Supported	18	12	0	4	26	23	7	17	16
Unsuppted	17	13	0	5	25	29	1	18	12
SAMSI	1	0	0	1	0	1	0	NA	NA

Education and Outreach
SAMSI-CRSC Industrial Mathematical and Statistical Modeling Workshop for Graduates
North Carolina State University
Supported Workshop Participants
July 25-August 2, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Affane Aji	Chandia	F	Auburn U	Mathematics	S
Bandulasiri	Ananda	M	Texas Tech U	Mathematics	S
Bartlett	Andy	M	College of Charleston	Mathematics	S
Berger	Andre	M	Emory U	Mathematics & Computer Science	S
Damu	Seshadri	M	U of Tennessee	Engineering	S
Froehlich	Mihaela	F	Duke U	Mathematics	S
Georgieva	Anna	F	New Jersey Institute of Technology	Mathematics	A
Gonzalez	John	M	Northeastern U	Mathematics	S
Gruber	Mary	F	Michigan State U	Mathematics	S
Han	Xiaoying	F	U of Buffalo	Mathematics	S
Hanlon	Bret	M	Texas Tech U	Statistics	S
Hazaveh	Kamyar	M	Georgia Tech	Engineering	S
Howard	Anna	F	Self-Employed		A
Jian	Bing	M	U of Florida	Computer Science	S
Kose	Emek	F	Drexel U	Mathematics	S
Li	Zheng	M	Brown U	Mathematics	S
Mavinga	Nsoki	F	U of Alabama	Mathematics	S

Modlin	Danny	M	U of North Carolina, Wilmington	Statistics	S
Navaratna	Menaka	M	Texas Tech U	Mathematics	S
Osorio	Eduardo	M	Rutgers U	Mathematics	S
Ott	William	M	U of Maryland	Mathematics	A
Park	Min-Hyong	F	Cornell U	Statistics	S
Perez	Elizabeth	F	U of South Carolina	Mathematics	S
Ranjan	Priya	M	U of California-Irvine	Economics	A
Rupani	Sidharth	M	Worcester Polytech	Engineering	S
Saglanmak	Nesli	F	Roskilde U	Mathematics	S
Thukani	Thobile	M	Southern U A&M College	Mathematics	S
Vijayat	Amith	M	Clarkson U	Engineering	S
Wheeler	Erik	M	Montana State U	Mathematics	S
Xia	Xiaofeng	M	Emory U	Mathematics & Computer Science	S
Zhou	Guohua	F	Clarkson U	Mathematics	S

Education and Outreach
SAMSI-CRSC Industrial Mathematical and Statistical Modeling Workshop for Graduates
North Carolina State University
Workshop Participants
July 25-August 2, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Affane Aji	Chandia	F	Auburn U	Mathematics	S
Ball	Brian	M	North Carolina State U	Mathematics	S

Bandulasiri	Ananda	M	Texas Tech U	Mathematics	S
Bartlett	Andy	M	College of Charleston	Mathematics	S
Benim	Robert	M	U of Portland	Mathematics	S
Berger	Andre	M	Emory U	Mathematics & Computer Science	S
Bowman	Mike	M	Montana State U	Mathematics	S
Braun	Thomas	M	North Carolina State U	Mathematics	S
Chertock	Alina	F	North Carolina State U	Mathematics	F
Currin	Brenda	F	North Carolina State U	Mathematics	F
Damu	Seshadri	M	U of Tennessee	Engineering	S
Deems	Thomas	M	Slippery Rock U	Mathematics	S
Diawara	Norou	M	Auburn U	Mathematics	S
Froehlich	Mihaela	F	Duke U	Mathematics	S
Geis	Jennifer	F	Augsburg College	Mathematics	S
Georgieva	Anna	F	New Jersey Institute of Technology	Mathematics	A
Gonzales	Chad	M	Arizona State U	Mathematics	S
Gonzalez	John	M	Northeastern U	Mathematics	S
Gruber	Mary	F	Michigan State U	Mathematics	S
Gunnarsdottir	Ingunn	F	Roskilde U	Mathematics	S
Haider	Mansoor	M	North Carolina State U	Mathematics	F
Han	Xiaoying	F	U of Buffalo	Mathematics	S
Hanlon	Bret	M	Texas Tech U	Statistics	S

Haugh	Janine	F	U of New Hampshire	Mathematics	S
Hazaveh	Kamyar	M	Georgia Tech	Engineering	S
Howard	Anna	F	Self-Employed		A
Jian	Bing	M	U of Florida	Computer Science	S
Joyner	Sarah Lynn	F	Meredith College	Mathematics	S
Kose	Emek	F	Drexel U	Mathematics	S
Langville	Amy	F	North Carolina State U	Mathematics	S
Li	Zheng	M	Brown U	Mathematics	S
Lin	Chuan	M	North Carolina State U	Mathematics	S
Lipkin	Craig	M	Drexel U	Statistics	S
Lubkin	Sharon	F	North Carolina State U	Mathematics	F
Mahyari	Morteza	M	Cal Tech	Mathematics	S
Mavinga	Nsoki	F	U of Alabama	Mathematics	S
Modlin	Danny	M	U of North Carolina, Wilmington	Statistics	S
Navaratna	Menaka	M	Texas Tech U	Mathematics	S
Osorio	Eduardo	M	Rutgers U	Mathematics	S
Ott	William	M	U of Maryland	Mathematics	A
Park	Min-Hyong	F	Cornell U	Statistics	S
Perez	Elizabeth	F	U of South Carolina	Mathematics	S
Prabhakaran	Jyothsna	F	U of Alabma	Mathematics	S
Ranjan	Priya	M	U of California-Irvine	Economics	A

Rupani	Sidharth	M	Worcester Polytech	Engineering	S
Saglanmak	Nesli	F	Roskilde U	Mathematics	S
Smith	Ralph	M	SAMSI & North Carolina State U	CRSC	F
Stepp	April	F	West Carolina U	Mathematics	S
Thukani	Thobile	M	Southern U A&M College	Mathematics	S
Tingey	Troy	M	Arizona State U	Mathematics	S
Tuncer	Necibe	F	Auburn U	Mathematics	S
Tweedy	Eamonn	M	North Carolina State U	Physics	S
Vijayat	Amith	M	Clarkson U	Engineering	S
Vogl	Christopher	M	Illinois Wesleyan U	Mathematics	S
Wan	Xiaohai	M	North Carolina State U	Mathematics	S
Wheeler	Erik	M	Montana State U	Mathematics	S
Wong	Anastasia	F	Mills College	Statistics	S
Xia	Xiaofeng	M	Emory U	Mathematics & Computer Science	S
Zhang	Wei	F	Tufts U	Mathematics	S
Zhou	Guohua	F	Clarkson U	Mathematics	S

National Defense and Homeland Security
Opening Workshop
Radisson Hotel Research Triangle Park
September 11-14, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	21	3	0	16	8	8	3	13	19	17
Unsuppted	47	22	0	44	25	15	12	42	34	11
SAMSI	6	2	0	1	7	3	5	0	NA	NA

National Defense and Homeland Security
Opening Workshop
Radisson Hotel Research Triangle Park
Supported Workshop Participants
September 11-14, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Agarwal	Deepak	M	AT&T Labs	Statistics	FP
Al-Shameri	Faleh	M	George Mason U	Information System	NRG
Bier	Vicki	F	U of Wisconsin-Madison	Industrial and Systems Engineering	FP
Czajkowski	Michal	M	U of Louisville	Math	NRG
Fan	Jianping	M	U of North Carolina at Charlotte	Dept. of Computer Science	NRG
Jun	Sung Chan	M	Los Alamos National Laboratory	Biological & Quantum Physics	FP
Kettenring	Jon	M	Drew U	RISE	FP
Larson	Richard	M	Massachusetts Inst of Technology	Civil and Environmental Engineering	FP
Li	Tao	M	Florida International U	School of Computer Science	FP
Lin	Xiaodong	M	U of Cincinnati	Mathematical Sciences	NRG
Mallick	Bani	M	Texas A&M U	Statistics	FP
Maxion	Roy	M	Carnegie Mellon U	Computer Science	FP

Mezic	Igor	M	U of California, Santa Barbara	Mechanical Engineering	NRG
Michailidis	George	M	U of Michigan	Statistics	FP
Nair	Vijay	M	U of Michigan	Statistics/Industrial & Operations Engineering	FP
Peng	Mei	F	Florida International U	School of Computer Science	NRG
Porter	Michael	M	U of Virginia	Systems and Information Engineering	NRG
Rempala	Greg	M	U of Louisville	Mathematics	FP
Schafer	Chad	M	Carnegie Mellon U	Statistics	FP

National Defense and Homeland Security
Opening Workshop
Radisson Hotel Research Triangle Park
Workshop Participants
September 11-14, 2005

Last Name	**First Name**	**Gender**	**Affiliation**	**Department**	**Status**
Agarwal	Deepak	M	AT&T Labs	Statistics	FP
Al-Shameri	Faleh	M	George Mason U	Information System	NRG
Anderson	Dale	M	Pacific Northwest National Laboratory	National Security	FP
Apte	Amit	M	SAMSI and U of North Carolina at Chapel Hill	Mathematics	NRG
Bai	Ping	F	U of North Carolina	Statistics	NRG
Banks	David	M	Duke U	Statistics & Decision Sciences	FP
Banks	H.T.	M	North Carolina State U	Mathematics	FP
Begashaw	Negash	M	Benedict College	Mathematics and Computer Science	FP
Bier	Vicki	F	U of Wisconsin-Madison	Industrial and Systems Engineering	FP

Chipman	Hugh	M	Acadia U	Mathematics and Statistics	FP
Cox	Lawrence	M	National Center for Health Statistics	Office of Research and Methodology	FP
Czajkowski	Michal	M	U of Louisville	Math	NRG
Datta	Gauri	M	SAMSI/Duke U/U of Georgia	Statistics	FP
Dediu	Sava	M	SAMSI & North Carolina State U	Center for Research & Scientific Computation	NRG
Denogean	Lisa	F	SAMSI and North Carolina State U		NRG
Dickey	David	M	North Carolina State U	Statistics	FP
Drummey	Kevin Ward	M	U. S. Federal Government	Department of Defense	FP
Ernstberger	Stacey	F	North Carolina State U	Mathematics	NRG
Fan	Jianping	M	U of North Carolina at Charlotte	Dept. of Computer Science	NRG
Fienberg	Stephen	M	Carnegie Mellon U	Statistics	FP
Filliben	James	M	National Inst of Standards & Technology		FP
Fricker	Ron	M	Naval Postgraduate School	Operations Research	FP
Gallaher	Mike	M	RTI International	Technology Economics & Policy	FP
Ghosh	Joyee	F	Duke U	statistics	NRG
Gill	Ryan	M	U of Louisville	Mathematics	NRG
Gonzalez, Jr.	Joe Fred	M	CDC, NCHS, Office of Research & Methodology		FP
Govan	Anjela	F	North Carolina State U	Mathematics	NRG
Heyward	Shenek	F	North Carolina State U	Statistics	NRG
Holte	Sarah	F	Fred Hutchinson Cancer Research Center	Modeling and Methods	FP

Hubal	Elaine Cohen	F	National Cen for Computation Toxicology & USEPA		FP
Ju	Wen-Hua	M	Avaya Labs Research		FP
Jun	Sung Chan	M	Los Alamos National Laboratory	Biological & Quantum Physics	FP
Karr	Alan	M	National Inst of Statistical Sciences		FP
Katzoff	Myron	M	Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Kelly	Douglas	M	U of North Carolina at Chapel Hill	Statistics and Operations Research	FP
Kettenring	Jon	M	Drew U	RISE	FP
Kim	Se Hee	F	U of North Carolina at Chapel Hill	Biostatistics	NRG
Lacey	Michelle	F	SAMSI		NRG
Larson	Richard	M	Massachusetts Inst of Technology	Civil and Environmental Engineering	FP
Last	Michael	M	National Inst of Statistical Sciences		NRG
Lee	Chihoon	M	U of North Carolina at Chapel Hill	Statistics and Operations research	NRG
Li	Tao	M	Florida International U	School of Computer Science	FP
Lin	Xiaodong	M	U of Cincinnati	Mathematical Sciences	NRG
Lynch	James	M	U of South Carolina	Statistics	FP
Madigan	David	M	Rutgers U		FP
Mallick	Bani	M	Texas A&M U	Statistics	FP
Martinez	Wendy	F	Office of Naval Research		FP
Mattingly	Jonathan	M	Duke U	Mathematics	FP
Maxion	Roy	M	Carnegie Mellon U	Computer Science	FP
Medhin	Negash	M	North Carolina State U	Mathematics	FP

Mei	Yajun	M	FHCRC		NRG
Mezic	Igor	M	U of California, Santa Barbara	Mechanical Engineering	NRG
Michailidis	George	M	U of Michigan	Statistics	FP
Mills	Kathryn	F	Government of Canada	Dept. of Defence	NRG
Mills	Shirley	F	Carleton Univeristy	School of Mathematics and Statistics	FP
Nair	Vijay	M	U of Michigan	Statistics/Industrial & Operations Engineering	FP
Nguyen	Hoan	F	North Carolina State U	Mathematics	NRG
Norminton	Ted	M	Carleton U	School of Mathematics and Statistics	FP
Oganyan	Anna	F	National Inst of Statistical Sciences		NRG
Olufsen	Mette	F	North Carolina State U	Mathematics	FP
Park	Cheolwoo	M	U of Georgia	Statistics	NRG
Pemy	Moustapha	M	SAMSI and North Carolina State U		NRG
Peng	Wei	F	Florida International U	School of Computer Science	NRG
Picard	Richard	M	Los Alamos National Laboratory		FP
Porter	Michael	M	U of Virginia	Systems and Information Engineering	NRG
Qaqish	Bahjat	M	U of North Carolina at Chapel Hill	Biostatistics	FP
Reiter	Jerome	M	Duke U	ISDS	FP
Rempala	Greg	M	U of Louisville	Mathematics	FP
Rigsby	John	M	Naval Surface Warfare Center	B 10	FP
Rodriguez	Jesus	M	SAMSI and North Carolina State U		NRG
Rowe	Brent	M	RTI International		FP

Said	Yasmin	F	George Mason U	Center for Computational Statistics	FP
Samuels	John	M	North Carolina State U	Mathematics	NRG
Schafer	Chad	M	Carnegie Mellon U	Statistics	FP
Schmidt	Lara	F	RAND Corporation	ES	FP
Scott	Sherry	F	U of North Carolina at Chapel Hill	Mathematics	NRG
Sedransk	Nell	F	National Inst of Statistical Sciences		FP
Shmueli	Galit	F	U of Maryland	Decision & Information Technologies, Smith School of Business	FP
Slenning	Barrett	M	North Carolina State U, College of Veterinary Medicine	Population Health & Pathobiology	FP
Spruill	Nancy	F	Office of the Under Secretary of Defense for Acquisition, Technology & Logistics		FP
Stamey	James	M	Baylor U	Statistical Science	FP
Storlie	Curtis	M	SAMSI and North Carolina State U	Statistics	NRG
Tubbs	Jack	M	Baylor U	Statistical Science	FP
Vardeman	Stephen	M	Iowa State U	Statistics and IMSE Departments	FP
Vera	Francisco	M	NISS & SAMSI		NRG
Verkhovsky	Boris	M	New Jersey Inst of Technology	Computer Science	FP
von Winterfeldt	Detlof	M	U of Southern California	Public Policy and Management	FP
Wang	Cliff	M	Army Research Office		FP
Wegman	Edward	M	George Mason U	Center for Computational Statistics	FP
Wilson	Alyson	F	Los Alamos National Laboratory	Statistical Sciences Group	FP
Wouhib	Abera	M	Centers for Disease Control and Prevention	National Center of Health Statistics	FP

Wu	Yichao	M	U of North Carolina at Chapel Hill	Statistics	NRG
Ziya	Serhan	M	U of North Carolina at Chapel Hill	Statistics	NRG

Financial Mathematics, Statistics and Econometrics
Opening Workshop
Radisson Hotel Research Triangle Park
September 18-21, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	41	10	0	22	29	12	21	18	42	19
Unsuppted	79	21	6	50	54	23	17	64	32	10
SAMSI	1	0	0	0	1	0	1	0	NA	NA

Financial Mathematics, Statistics and Econometrics
Opening Workshop
Radisson Hotel Research Triangle Park
Supported Workshop Participants
September 18-21, 2005

Last Name	**First Name**	**Gender**	**Affiliation**	**Department**	**Status**
Barndorff-Nielsen	Ole	M	Aarhus U		FP
Bayraktar	Erhan	M	U of Michigan	Mathematics	NRG
Berndt	Antje	F	Carnegie Mellon U	Tepper School of Business (Finance)	FP
Cai	Zongwu	M	U of North Carolina at Charlotte	Mathematics and Statistics	NRG
Chen	Lisha	F	U of Pennsylvalnia	Statistics	
Chen	Yuguo	M	U of Illinois at Urbana-Champaign	Statistics	FP
Cheng	Ai-re "Meg"	F	U of California, Santa Cruz	Economics	NRG
Cline	Daren B.H.	M	Texas A&M U	Statistics	
Deng	Shijie	M	Georgia Institute of Technology	Industrial and Systems Engineering	NRG

Engle	Robert	M	New York U, Stern School of Business	Salomon Center	FP
Figueroa-Lopez	Jose	M	Purdue U	Mathematics	
Galloway	Mack	M	Florida State U	Mathematics	NRG
Goldman	Elena	F	Pace U, Lubin School of Business	Finance and economics	NRG
Goldman	Irina	F	Stevens Institute of Technology	Financial Engineering	NRG
Hansen	Lars Peter	M	U of Chicago	Economics	FP
Hillebrand	Eric	M	Louisiana State U	Economics	FP
Jabri	Hannah	F	Rice U	Department of Statistics	NRG
Jeanblanc	Monique	F	U of Evry	Mathematics	FP
Kercheval	Alec	M	Florida State U	Mathematics	NRG
Khaliq	Abdul	M	Middle Tennessee State U	Mathematical Sciences	FP
Kodippili	Asitha	M	Fayetteville State U	Mathematics and Computer Science	FP
Lai	Tze Leung	M	Stanford U	Statistics	NRG
LeBaron	Blake	M	Brandeis U	international Business School	
Lee	Beom	M	U of Alabama	Mathematics	NRG
Lee	Kiseop	M	U of Louisville	Mathematics	NRG
Levine	Michael	M	Purdue U	Statistics	NRG
Li	Hongcheng	M	Kent State U	Mathematics	NRG
Li	Min	M	California State U Sacramento	Management Information Science	NRG
Linetsky	Vadim	M	Northwestern U	Industrial Eng and Management Science	FP
Ma	Huaqiang	M	Univeristy of Maryland	Mathematics	NRG

Ma	Xiaofang	M	U of Toronto	Computer Science	NRG
Miao	Linyan	F	Stevens Institute of Technology	Financial Engineering	NRG
Nagaraja	Chaitra	F	U of Pennsylvania, Wharton School	Statistics	FP
Okten	Giray	M	Florida State U	Mathematics	FP
Ostrovsky	Dmitry	M	Lehigh U	Mathematics	FP
Qiao	Zhihua	M	U of Pennsylvania, Wharton School		NRG
Qingkai	Mo	M	U of Toronto	Computer Science	NRG
Sayit	Hasanjan	M	Cornell U	Mathematics Department	NRG
Shepherd	Neil	M	Nuffield College, U of Oxford		FP
Shyamalkumar	Nariankadu	M	U of Iowa	Statistics and Actuarial Science	NRG
Solna	Knut	M	U of California, Irvine	mathematics	FP
Stoikov	Sasha	M	New York U		NRG
Taqqu	Murad	M	Boston U	Mathematics and Statisitics	FP
Tompaidis	Stathis	M	U of Texas, McCombs School of Business	MSIS	FP
Trindade	Alex	M	U of Florida	Statistics	NRG
Tsay	Ruey	M	U of Chicago	Graduate School of Business	FP
Wang	Li	F	Michigan State U	Statistics and Probability	NRG
Yan	Jun	M	U of Iowa	Statistics and Actuarial Science	NRG
Zhang	Bing	M	U of Maryland, College Park	Mathematics	NRG
Zhang	Peng	M	U of Pennsylvania, Wharton School	Statistics	NRG

Financial Mathematics, Statistics and Econometrics
Opening Workshop
Radisson Hotel Research Triangle Park
Workshop Participants
September 18-21, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Aguilar	Michael	M	U of North Carolina at Chapel Hill	Economics	NRG
Allen	Jackie	F	SAS Institute, Inc.		FP
Antoine	Bertille	F	U de Montreal, U of North Carolina	Statistics and OR	NRG
Banner	Adrian	M	INTECH		FP
Barndorff-Nielsen	Ole	M	Aarhus U		FP
Bayraktar	Erhan	M	U of Michigan	Mathematics	NRG
Berndt	Antje	F	Carnegie Mellon U	Tepper School of Business (Finance)	FP
Bester	C. Alan	M	U of Chicago	Graduate School of Business	NRG
Bishwal	Jaya	F	U of North Carolina at Charlotte	Mathematics	NRG
Bloomfield	Peter	M	North Carolina State U	Statistics	FP
Bollerslev	Tim	M	Duke U	Economics	FP
Boone	Bryan	M	SAS Institute, Inc.		FP
Budhiraja	Amarjit	M	U of North Carolina at Chapel Hill	Statistics & Operations Research	FP
Busch	Thomas	M	U of Aarhus & Duke U	Economics	NRG
Cai	Zongwu	M	U of North Carolina at Charlotte	Mathematics and Statistics	NRG
Carmona	Rene	M	Princeton U	ORFE & Bendheim Center for Finance	FP
Chamu	Francisco	M	U of North Carolina at Chapel Hill	Statistics & Operations Research	FP

Chen	Lisha	F	U of Pennsylvalnia	Statistics	NRG
Chen	Yuguo	M	U of Illinois at Urbana-Champaign	Statistics	FP
Chen	Wei	M	SAS Institute, Inc.		FP
Cheng	Ai-re "Meg"	F	U of California, Santa Cruz	Economics	NRG
Cheng	Yungchen	M	Southwest Missouri State U	Mathematics	NRG
Choi	Yangho		U of Iowa	Applied Mathematics and Computational Sci	NRG
Chvosta	Jan	M	SAS Institute, Inc.		FP
Cline	Daren B.H.	M	Texas A&M U	Statistics	FP
Cont	Rama	M	Ecole Polytechnique	Mathematics	FP
Cotton	Peter	M	Morgan Stanley	Fixed Income	FP
Covas	Eurico	M	HSBC Bank	Derivative Models Review Group	FP
Crotty	Michael	M	North Carolina State U	Statistics	NRG
Dedov	Maxym	M	Duke U	Economics	NRG
Deng	Shijie	M	Georgia Institute of Technology	Industrial and Systems Engineering	NRG
Dowla	Arif	M	ACI LTD.		FP
Eccles	Brad	M	Progress Energy Ventures, Inc.		FP
Engle	Robert	M	New York U, Stern School of Business	Salomon Center	FP
Eraker	Bjorn	M	Duke U	Economics	FP
Erdman	Don	M	SAS Institute, Inc.		FP
Fackler	Paul	M	North Carolina State U	Agricultural and Resource Economics	FP
Figueroa-Lopez	Jose	M	Purdue U	Mathematics	NRG

Fouque	Jean-Pierre	M	North Carolina State U	Mathematics	FP
Gallant	Ronald	M	Duke U	Fuqua School of Business	FP
Galloway	Mack	M	Florida State U	Mathematics	NRG
Gao	Yan	M	Progress Energy Ventures, Inc.		FP
Ghosh	Sujit	M	North Carolina State U	Statistics	FP
Ghysels	Eric	M	U of North Carolina at Chapel Hill	Economics	FP
Goldman	Elena	F	Pace U, Lubin School of Business	Finance and economics	NRG
Goldman	Irina	F	Stevens Institute of Technology	Financial Engineering	NRG
Grier	Kevin	M	U of Oklahoma & Duke U	Economics (OK) / Political Science (Duke)	FP
Groth	Martin	M	U of Oslo	Centre of Mathematics for Applications	NRG
Hall	Alastair	M	North Carolina State U	Economics	NRG
Han	Sanggohn		North Carolina State U	Economics	NRG
Hansen	Lars Peter	M	U of Chicago	Economics	FP
Hillebrand	Eric	M	Louisiana State U	Economics	FP
Hochberg	Kenneth	M	Bar-Ilan U	Mathematics	FP
Hoe	SingRu	F	U of Texas at Arlington	Finance and Real Estate	NRG
Holmfeldt	Mia	F	Lund U and Duke U	Economics	NRG
Houdre	Christian	M	Georgia Institute of Technology	Mathematics	FP
Hu	Ping	M	SAS Institute, Inc.		FP
Huang	Xin	F	Duke U	Economics	NRG
Hyde	John	M	Duke U	Mathematics	NRG

Ilhan	Aytac	F	U of Oxford	Mathematical Institute	NRG
Jabri	Hannah	F	Rice U	Statistics	NRG
Jackson	Laura	F	SAS Institute, Inc.		FP
Jahan-Parvar	Mohammad	M	U of North Carolina at Chapel Hill	Economics	NRG
Jarrow	Robert	M	Cornell U	Business School	FP
Jeanblanc	Monique	F	U of Evry	Mathematics	FP
Ji	Chuanshu	M	U of North Carolina at Chapel Hill	Statistics & Operations Research	FP
Juergens	Jennifer	F	Arizona State U	Finance	NRG
Kaniel	Ron	M	Duke U	Fuqua School of Business	FP
Kanter	Yakov	M	Morgan Stanley	Interest Rate Analytical Modelling	FP
Karr	Alan	M	National Institute of Statistical Sciences		FP
Kercheval	Alec	M	Florida State U	Mathematics	NRG
Khaliq	Abdul	M	Middle Tennessee State U	Mathematical Sciences	FP
Kodippili	Asitha	M	Fayetteville State U	Mathematics and Computer Science	FP
Kretschmer	Uta	F	U of Bonn		NRG
Kumar	Nitesh	M	North Carolina State U	Mathematics	NRG
Kyriakoulis	Kostas	M	North Carolina State U	Economics	NRG
Lai	Tze Leung	M	Stanford U	Statistics	FP
LeBaron	Blake	M	Brandeis U	international Business School	FP
Lee	Beom	M	U of Alabama	Mathematics	NRG
Lee	Kiseop	M	U of Louisville	Mathematics	NRG

Levine	Michael	M	Purdue U	Statistics	NRG
Li	Hongcheng	M	Kent State U	Mathematics	NRG
Li	Min	M	California State U Sacramento	Management Information Science	NRG
Linetsky	Vadim	M	Northwestern U	Industrial Engineering and Management Science	FP
Liu	Feng	F	North Carolina State U	Statistics	NRG
Lumba	Varun	M	Progress Energy Ventures, Inc.		FP
Lyadina-Saville	Elena	F	Saville Solver Ltd	Management	FP
Ma	Huaqiang	M	Univeristy of Maryland	Mathematics	NRG
Ma	Xiaofang	M	U of Toronto	Computer Science	NRG
Miao	Linyan	F	Stevens Institute of Technology	Financial Engineering	NRG
Miller	Randy	M	Bank of America	Global Portfolio Strategies	FP
Mykland	Per	M	U of Chicago	Statistics	FP
Na	Sungsoo	M	North Carolina State U	Industrial Engineering	NRG
Nagaraja	Chaitra	F	U of Pennsylvania, Wharton School	Statistics	FP
Nyamadi	Tsatsu	M	North Carolina State U	Financial Mathematics	NRG
Okten	Giray	M	Florida State U	Mathematics	FP
Orhan	Mehmet Aras	M	U of North Carolina at Chapel Hill	Finance	NRG
Ostrovsky	Dmitry	M	Lehigh U	Mathematics	FP
Palandri	Alessandro	M	Duke U	Economics	NRG
Pang	Tao	M	North Carolina State U	Mathematics	FP

Pantula	Sastry	M	North Carolina State U	Statistics	FP
Pelletier	Denis	M	North Carolina State U	Economics	NRG
Pemy	Moustapha	M	North Carolina State U	Mathematics	NRG
Pigorsch	Christian	M	U of Munich		FP
Pong	Der-Chuang		North Carolina State U	Financial Mathematics	NRG
Puggioni	Gavino	M	Duke U	Statistics	NRG
Qiao	Zhihua	M	U of Pennsylvania, Wharton School		NRG
Qingkai	Mo	M	U of Toronto	Computer Science	NRG
Rambharat	Ricky	M	Duke U	Statistics	NRG
Renault	Eric	M	U of North Carolina at Chapel Hill	Statistics & Operations Research	FP
Rodriguez	Jesus	M	SAMSI		NRG
Rodriguez de Almeida	Caio	M	Ibmec-RJ	Business School	NRG
Ross	Kevin	M	U of North Carolina at Chapel Hill	Statistics & Operations Research	NRG
Sanchez	Emmanuel	M	Progress Energy		FP
Sayit	Hasanjan	M	Cornell U	Mathematics	NRG
Shaliastovich	Ivan	M	Duke U	Economics	NRG
Shamseldin	Elizabeth	F	U of North Carolina at Chapel Hill	Statistics & Operations Research	NRG
Shao	Haimei	F	U of North Carolina at Charlotte	Mathematics	NRG
Shen	Haipeng	M	U of North Carolina at Chapel Hill	Statistics & Operations Research	NRG
Shepherd	Neil	M	Nuffield College, U of Oxford		FP
Shyamalkumar	Nariankadu	M	U of Iowa	Statistics and Actuarial Science	NRG

Sinko	Arthur	M	U of North Carolina at Chapel Hill	Economics	NRG
Sircar	Ronnie	M	Princeton U & SAMSI	Operations Research & Financial Engineering	FP
Sloan	Jennifer	F	North Carolina State U	Statistics	NRG
Smith	Ralph	M	North Carolina State U	Mathematics	FP
Smith	Richard	M	U of North Carolina at Chapel Hill	Statistics & Operations Research	FP
Snider	Glen	M	Progress Energy Ventures, Inc.		FP
Sokolov	Denis	M	Duke U	Economics	NRG
Solna	Knut	M	U of California, Irvine	mathematics	FP
Sterijevski	Greg	M	SAS Institute, Inc.		FP
Stoikov	Sasha	M	New York U		NRG
Sun	Linman	F	U of North Carolina at Charlotte	Mathematics	NRG
Taqqu	Murad	M	Boston U	Mathematics and Statisitics	FP
Tauchen	George	M	Duke U	Economics	FP
Todorov	Viktor	M	Duke U	Economics	FP
Tompaidis	Stathis	M	U of Texas, McCombs School of Business	MSIS	FP
Tran	Duy Tien	M	Duke U	Economics	NRG
Trindade	Alex	M	U of Florida	Statistics	NRG
Trovero	Michele	F	SAS Institute and U of North Carolina at Chapel Hill	Statistics	FP
Tsay	Ruey	M	U of Chicago	Graduate School of Business	FP
Tu	Chong	M	Duke U	Statistics	NRG
van Binsbergen	Jules	M	Duke U	Fuqua School of Business	FP

Vestal	Douglas	M	North Carolina State U	Mathematics	NRG
Wang	Li	F	Michigan State U	Statistics and Probability	NRG
Wang	Xian	F	U of North Carolina at Charlotte	Mathematics	NRG
Wang	Yunfei	F	U of North Carolina at Charlotte	Mathematics	NRG
Werker	Bas	M	Tilburg U	Finance and Econometrics Groups	FP
Wu	Yichao	M	U of North Carolina at Chapel Hill	Statistics & Operations Research	NRG
Xu	Mingxin	F	U of North Carolina at Charlotte	Mathematics and Statistics	NRG
Yan	Jun	M	U of Iowa	Statistics and Actuarial Science	NRG
Yang	Jack	M	SAS Institute, Inc.		FP
Yoon	Jungyeon	F	U of North Carolina at Chapel Hill	Statistics & Operations Research	NRG
Zhang	Bing	M	U of Maryland, College Park	Mathematics	NRG
Zhang	Peng	M	U of Pennsylvania, Wharton School	Statistics	NRG
Zhang	Qianyi		North Carolina State U	Statistics	NRG
Zhou	Jun	M	U of North Carolina at Charlotte	Mathematics & Statistics	NRG
Zhou	Xianwen	M	North Carolina State U	Mathematics	NRG

Data Assimilation for Geophysical Systems
Transition Workshop
NISS-SAMSI Building
October 5, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	7	1	0	4	4	0	1	7	8	5
Unsuppted	22	9	0	18	13	6	7	18	21	13
SAMSI	3	0	0	2	1	1	2	0	NA	NA

Data Assimilation for Geophysical Systems
Transition Workshop
NISS-SAMSI Building
Supported Workshop Participants
October 5, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Brandt	Achi	M	U of California Los Angeles and Weizmann Institute	Mathematics	FP
Etherton	Brian	M	U of North Carolina at Charlotte	Geography and Earth Sciences	NRG
Hunt	Brian	M	U of Maryland	Institute for Physical Science and Technology	NRG
Khare	Shree	M	National Center for Atmospheric Research	Institute for Mathematics Applied to Geosciences	NRG
Li	Zhijin	M	Raytheon & Jet Propulsion Laboratory		FP
Majumdar	Sharanya	M	U of Miami	RSMAS, Meteorology and Physical Oceanography	FP
Szunyogh	Istvan	M	U of Maryland	Atmospheric and Ocean Sciences	NRG
Vukicevic	Tomislava	F	Colorado State U	Cooperative Institute for Research in the Atmosphere	FP

Data Assimilation for Geophysical Systems
Transition Workshop
NISS-SAMSI Building
Workshop Participants
October 5, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Apte	Amit	M	SAMSI and U of North Carolina at Chapel Hill	Mathematics	NRG
Arellano, Jr.	Avelino F.	M	National Center for Atmospheric Research		NRG
Berger	Jim	M	SAMSI		FP
Berliner	Mark	M	Ohio State U	Statistics	FP
Brandt	Achi	M	U of California Los Angeles and Weizmann Institute	Mathematics	FP

Budhiraja	Amarjit	M	U of North Carolina at Chapel Hill	Statistics and Operations Research	NRG
Chin	T. Mike	M	Jet Propulsion Laboratory		FP
Etherton	Brian	M	U of North Carolina at Charlotte	Geography and Earth Sciences	NRG
Foley	Kristen	F	North Carolina State U	Statistics	NRG
Fuentes	Montserrat	F	North Carolina State U	Statistics	FP
Heimbach	Patrick	M	Massachusetts Institute of Technology	Earth, Atmospheric and Planetary Sciences	FP
Hunt	Brian	M	U of Maryland	Institute for Physical Science and Technology	NRG
Ide	Kayo	F	U of California Los Angeles	Institute of Geophysics and Planetary Physics	FP
Ihler	Alexander	M	U of California Irvine	School of Information and Computer Sciences	FP
Jones	Chris	M	SAMSI and U of North Carolina at Chapel Hill	Mathematics	FP
Jun	Mikyoung	F	NCAR and Texas A&M U		NRG
Kalnay	Eugenia	F	U of Maryland	Atmospheric and Ocean Sciences	FP
Khare	Shree	M	National Center for Atmospheric Research	Institute for Mathematics Applied to Geosciences	NRG
Li	Zhijin	M	Raytheon & Jet Propulsion Laboratory		FP
Liu	Fei	F	Duke U	Institute of Statistics & Decision Sciences	NRG
Liu	Liyan	F	U of North Carolina at Chapel Hill	Mathematics	NRG
Majumdar	Sharanya	M	U of Miami	RSMAS, Meteorology and Physical Oceanography	FP
Malmberg	Anders	M	National Center for Atmospheric Research		NRG
Martinsen-Burrell	Neil	M	U of North Carolina at Chapel Hill	Mathematics	NRG
Masters	Wen	F	National Science Foundation		FP
McLaughlin	Dennis	M	Massachusetts Institute of Technology	Civil and Environmental Engineering	FP

Miller	Robert	M	Oregon State U	Oceanic and Atmospheric Sciences	FP
Namachchivaya	Navaratnam Sri	M	U of Illinois at Urbana-Champaign	Aerospace Engineering	NRG
Poje	Andrew	M	City U of New York	Mathematics	FP
Russell	Tom	M	National Science Foundation	Division of Mathematical Sciences	FP
Salman	Hayder	M	U of North Carolina at Chapel Hill	Mathematics	NRG
Smith	Leonard	M	U of Oxford	Pembroke College	FP
Stephens	Monica	F	Spelman U	Mathematics	NRG
Szunyogh	Istvan	M	U of Maryland	Atmospheric and Ocean Sciences	NRG
Tangborn	Andrew	M	U of Maryland-Baltimore County	Joint Center for Earth Systems Technology	NRG
Toth	Zoltan	M	NCEP/NWS/NOAA	Environmental Modeling Center	FP
Vukicevic	Tomislava	F	Colorado State U	Cooperative Institute for Research in the Atmosphere	FP
Wikle	Chris	M	U of Missouri	Statistics	FP
Zhang	Fuqing	M	Texas A&M U	Atmospheric Sciences	FP

Hot Topics Workshop
Collaborations in the Mathematical Geosciences
Radisson Hotel Research Triangle Park
October 6-7, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	0	0	0	0	0	0	0	0	0	0
Unsuppted	84	23	6	68	43	23	33	55	59	27
SAMSI	4	0	0	1	3	1	3	0	NA	NA

Hot Topics Workshop
Collaborations in the Mathematical Geosciences
Radisson Hotel Research Triangle Park
Workshop Participants
October 6-7, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Apte	Amit	M	SAMSI and U of North Carolina at Chapel Hill	Mathematics	NRG
Arellano, Jr.	Avelino F.	M	National Center for Atmospheric Research		NRG
Babu	G. Jogesh	M	Pennsylvania State U	Statistics	FP
Bayarri	M.J.	F	U of Valencia	Statistics	FP
Berliner	Mark	M	Ohio State U	Statistics	FP
Beron-Vera	Francisco	M	U of Miami	RSMAS	FP
Borggaard	Jeffrey	M	Virginia Polytechnic Institute and State U	Mathematics	FP
Brazier	Richard	M	Pennsylvania State U	Mathematics	FP
Budhiraja	Amarjit	M	U of North Carolina at Chapel Hill	Statistics and Operations Research	FP
Chen	Rong	M	National Science Foundation		FP
Chin	Toshio Michael	M	U of Miami	RSMAS	FP
Clemo	Tom	M	Boise State U	Geosciences	NRG
Connor	Charles	M	U of South Florida	Geology	FP
Connor	Laura	F	U of South Florida	Geology	FP
Daniel	Talithia	F	Rice U	Statistics	NRG
Davis	Joshua	M	Duke U	Mathematics	NRG
de Hoop	Maarten V.	M	Purdue U	Mathematics	FP

Dediu	Sava	M	SAMSI and North Carolina State U	Center for Research in Scientific Computing	NRG
DelSole	Timothy	M	George Mason U	Sch of Computational Sciences	FP
Dewar	William	M	Florida State U	Oceanography	FP
Duan	Jinqiao	M	Illinois Institute of Technology	Applied Mathematics	FP
Fairweather	Lindon P.	M	U of Central Florida		NRG
Farthing	Matthew	M	U of North Carolina at Chapel Hill	Environmental Sciences & Engineering	NRG
Forest	Chris E.	M	Massachusetts Institute of Technology	Earth, Atmospheric & Planetary Sciences	FP
Freeston	Michael	M	U of California, Santa Barbara	Institute for Computational Earth System Science	FP
Fuentes	Montserrat	F	North Carolina State U	Statistics	FP
Galluzzo	Benjamin	M	U of Iowa	Mathematics	NRG
Ghosh	Sujit	M	North Carolina State U	Statistics	FP
Glantz	Roland	M	Johns Hopkins U	Geography & Environmental Engineering	NRG
Golden	Kenneth	M	U of Utah	Mathematics	FP
Guddati	Murthy N.	M	North Carolina State U	Civil, Construction & Environmental Engineering	FP
Guillas	Serge	M	Georgia Insistute of Technology	Mathematics	NRG
Guo	Diansheng	M	U of South Carolina	Geography	NRG
Gupta	Sumant K.	M	CFEST, Inc.		FP
He	Xuming	M	U of Illinois at Urbana-Champaign	Statistics	FP
Heimbach	Patrick	M	Massachusetts Institute of Technology	Earth, Atmospheric & Planetary Sciences	NRG
Herbert	Timothy	M	Brown U	Geological Sciences	FP

Hilpert	Markus	M	Johns Hopkins U	Geography & Environmental Eng	FP
Huber	Mark	M	Duke U	Mathematics	FP
Huerta	Gabriel	M	U of New Mexico	Statistics Program	FP
Huybers	Peter	M	Woods Hole Oceanographic Institution		FP
Ide	Kayo	F	U of Califfornia Los Angeles	Atmospheric and Oceanic Sciences	FP
Ihler	Alexander	M	U of California, Irvine	School of Information and Computer Sciences	NRG
Iliescu	Traian	M	Virginia Polytechnic Institute and State U	Mathematics	NRG
Jackson	Charles	M	U of Texas at Austin	Institute for Geophysics	FP
Jones	Christopher	M	SAMSI and U of North Carolina at Chapel Hill	Mathematics	FP
Jun	Mikyoung	F	NCAR and Texas A&M U	Statistics	NRG
Kalnay	Eugenia	F	U of Maryland	Atmospheric and Oceanic Sciences	FP
Kaplan	Alexey	M	Columbia U	Lamont-Doherty Earth Observatory	FP
Karr	Alan F.	M	NISS		FP
Karuri	Stella	F	North Carolina State U	Statistics	NRG
Khatiwala	Samar	M	Columbia U	Lamont-Doherty Earth Observatory	FP
Kramer	Peter	M	Rensselaer Polytechnic Institute	Mathematical Sciences	NRG
Lai	Ming-Jun	M	U of Georgia	Mathematics	FP
Last	Michael	M	NISS		NRG
Li	Yongxiang	M	Pennsylvania State U	Mathematics	NRG
Li	Zhijin	M	Raytheon & Jet Propulsion Laboratory		FP

Liang	Xin-Zhong		U of Illinois at Urbana-Champaign	Atmospheric Sciences	FP
Lin	Enbing	M	U of Toledo	Mathematics	FP
Liu	Liyan	F	U of North Carolina at Chapel Hill	Mathematics	NRG
Lopes	Brian	M	U of North Carolina at Chapel Hill	Statistics & Operations Research	NRG
Lopez	Juan M.	M	Arizona State U	Mathematics and Statistics	FP
Lu	QiQi	F	Mississippi State U	Mathematics and Statistics	NRG
Lund	Robert	M	Clemson U	Mathematical Sciences	FP
Ma	Liyun	F	North Carolina State U	Statistics	NRG
Majumdar	Anandamayee	F	Arizona State U	Mathematics and Statistics	NRG
Malmberg	Anders	M	National Center for Atmospheric Research		NRG
Masters	Wen	F	National Science Foundation		FP
Miller	Casey	M	U of North Carolina at Chapel Hill	Environmental Sciences & Engineering	FP
Miller	Robert	M	Oregon State U	Oceanic and Atmospheric Sciences	FP
Minkoff	Susan	F	U of Maryland, Baltimore County	Mathematics and Statistics	FP
Nail	Amy	F	North Carolina State U	Statistics	NRG
Namachchivaya	N. Sri	M	U of Illinois at Urbana-Champaign	Aerospace Engineering	NRG
Poje	Andrew	M	City U of New York	Mathematics	FP
Reich	Brian	M	North Carolina State U	Statistics	NRG
Restrepo	Juan M.	M	U of Arizona	Mathematics	FP
Routh	Partha	M	Boise State U	Geosciences	NRG
Rundell	William	M	National Science Foundation		FP

Russell	Tom	M	National Science Foundation		FP
Saalfeld	Alan	M	Ohio State U	Geological Sci and Computer Sci & Eng	FP
Salman	Hayder	M	U of North Carolina at Chapel Hill	Mathematics	NRG
Sandu	Adrian	M	Virginia Polytechnic Institute and State U	Computer Science	NRG
Schaffrin	Burkhard	M	Ohio State U	Geological Sciences	FP
Scott	Sherry E.	F	U of North Carolina at Chapel Hill	Mathematics	NRG
Sharma	Surja		U of Maryland, Baltimore County	Astronomy	FP
Shum	C.K.	M	Ohio State U	Geological Sciences	FP
Smith	Leonard	M	U of Oxford	Statistics	FP
Stephens	Monica	F	Spelman College	Mathematics	NRG
Stewart	Aimee	F	U of Kansas	Biodiversity Research Center	FP
Storlie	Curtis	M	SAMSI and North Carolina State U	Statistics	NRG
Sulsky	Deborah	F	U of New Mexico	Mathematics and Statistics	FP
Tangborn	Andrew	M	U of Maryland, Baltimore County	Mathematics and Statistics	NRG
Tissot	Philippe	M	Texas A&M U-Corpus Christi	College of Science & Technology	NRG
van der Hilst	Robert D.	M	Massachusetts Institute of Technology	Earth, Atmospheric & Planetary Sciences	FP
Vera	Francisco	M	NISS and SAMSI		NRG
Wang	Jing	F	Michigan State U	Statistics & Probability	NRG
Wang	Junping	M	National Science Foundation		FP
Wang	Xiaoming		Florida State U	Mathematics	FP
Wanliss	James	M	Embry-Riddle Aeronautical U	Physical Science	FP

Weiss	Jeffrey	M	U of Colorado	Atmospheric and Oceanic Sciences	FP
Welfert	Bruno	M	Arizona State U	Mathematics and Statistics	FP
Wikle	Chris	M	U of Missouri	Statistics	FP
Wolf	Dan	M	North Carolina State U	Statistics	NRG
Woo	Mi-Ja	F	NISS		NRG
Yao	Yonggang	M	Ohio State U	Statistics	NRG
Yuen	Dave	M	U of Minnesota	Geology and Geophysics	FP
Zhan	Mei-Qin		U of North Florida	Mathematics and Statistics	FP
Zhang	Fuqing	M	Texas A&M U	Meteorology	FP
Zhou	Jie	F	U of North Carolina at Chapel Hill	Statistics & Operations Research	NRG
Zhu	Zhengyuan	M	U of North Carolina at Chapel Hill	Statistics & Operations Research	FP
Zimmer	G. Beate	F	Texas A&M U-Corpus Christi	Computing and Mathematical Sciences	FP

Financial Mathematics, Statistics and Econometrics
Workshop on Credit Risk
NISS-SAMSI Building
October 31-November 2, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	8	6	0	7	7	0	4	10	12	5
Unsuppted	20	2	0	8	14	4	8	10	9	3
SAMSI	4	1	0	1	4	1	3	1	NA	NA

Financial Mathematics, Statistics and Econometrics
Workshop on Credit Risk
NISS-SAMSI Building
Supported Workshop Participants
October 31-November 2, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Carmona	Rene	M	Princeton U	Operations Research and Financial Eng	FP
Chen	Nan	M	Columbia U	Industrial Eng and Operations Research	FP
Goldberg	Lisa	F	MSCI Barra	Research	FP
Goldman	Elena	F	Lubin School of Business, Pace U	Finance and Economics	NRG
Goldman	Irina	F	Stevens Institue of Technology	Financial Eng	NRG
Horst	Ulrich	M	U of British Columbia	Mathematics	FP
Ilhan	Aytac	F	U of Oxford		NRG
Lando	David	M	Copenhagen Business School and Princeton U	Operations Research and Financial Eng	FP
Lee	Kiseop	M	U of Louisville	Mathematics	NRG
Papageorgiou	Evan	M	Princeton U	Operations Research and Financial Eng	NRG
Predescu	Mirela	F	U of Toronto	Finance	FP
Sayit	Hasanjan	M	U of Houston	Mathematics	NRG
Sezer	Semih	M	Princeton U	Operations Research and Financial Eng	NRG
Zariphopoulou	Thaleia	F	U of Texas at Austin	Mathematics and Information, Risk & Oper Management	FP

Financial Mathematics, Statistics and Econometrics
Workshop on Credit Risk
NISS-SAMSI Building
Workshop Participants
October 31-November 2, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Bloomfield	Peter	M	North Carolina State U	Statistics	FP
Carmona	Rene	M	Princeton U	Operations Research and Financial Eng	FP
Chen	Nan	M	Columbia U	Industrial Eng and Operations Research	FP
Collin-Dufresne	Pierre	M	Goldman Sachs	Quantitative Strategies	FP
Cotton	Peter	M	Morgan Stanley	Fixed Income	FP
Fouque	Jean-Pierre	M	North Carolina State U	Mathematics	FP
Ghsyels	Eric	M	U of North Carolina at Chapel Hill	Economics	FP
Goldberg	Lisa	F	MSCI Barra	Research	FP
Goldman	Elena	F	Lubin School of Business, Pace U	Finance and Economics	NRG
Goldman	Irina	F	Stevens Institue of Technology	Financial Eng	NRG
Horst	Ulrich	M	U of British Columbia	Mathematics	FP
Hyde	John	M	Duke U	Mathematics	NRG
Ilhan	Aytac	F	U of Oxford		NRG
Lando	David	M	Copenhagen Business School and Princeton U	Operations Research and Financial Eng	FP
Lee	Kiseop	M	U of Louisville	Mathematics	NRG
Mathur	Apoorv	M	North Carolina State U	Financial Mathematics and Operations Research	NRG
Mattingly	Jonathan	M	Duke U	Mathematics	NRG

Miller	Randy	M	Bank of America	Global Portfolio Strategies	FP
Nyamadi	Tsatsu	M	North Carolina State U	Mathematics	NRG
Pang	Tao	M	North Carolina State U	Mathematics	FP
Papageorgiou	Evan	M	Princeton U	Operations Research and Financial Eng	NRG
Pemy	Moustapha	M	SAMSI and North Carolina State U	Center for Research in Scientific Computation	NRG
Pong	DerChuang	M	North Carolina State U	Financial Mathematics	NRG
Predescu	Mirela	F	U of Toronto	Finance	FP
Rodriguez	Jesus	M	SAMSI		NRG
Sayit	Hasanjan	M	U of Houston	Mathematics	NRG
Sezer	Semih	M	Princeton U	Operations Research and Financial Eng	NRG
Sircar	Ronnie	M	Princeton U & SAMSI	Operations Research and Financial Eng	FP
Sloan	Jennifer	F	SAMSI and North Carolina State U	Statistics	NRG
Tu	Chong	M	Duke U	Institute of Statistics and Decision Sciences	NRG
Vestal	Doug	M	SAMSI and North Carolina State U	Mathematics	NRG
Wu	Yichao	M	U of North Carolina at Chapel Hill	Statistics	NRG
Xu	Mingxin	F	U of North Carolina at Charlotte	Mathematics and Statistics	NRG
Yue	Tianyao	M	Duke U	Eng and Operations Research	NRG
Zariphopoulou	Thaleia	F	U of Texas at Austin	Mathematics and Information, Risk & Oper Management	FP
Zhou	Xianwen	M	North Carolina State U	Mathematics	NRG

Latent Variables in the Social Sciences
Transition Workshop
Radisson Hotel Research Triangle Park
November 10-11, 2005

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	13	8	0	12	9	3	1	17	19	10
Unsuppted	34	18	2	24	30	10	1	43	18	7
SAMSI	2	0	0	1	1	1	0	1	NA	NA

Latent Variables in the Social Sciences
Transition Workshop
Radisson Hotel Research Triangle Park
Supported Workshop Participants
November 10-11, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Bentler	Peter	M	U of California, Los Angeles	Psychology and Statistics	FP
Feng	Shibao		Georgetown U Medical Center	Biomathematics and Biostatistics	FP
Gottschall	Amanda	F	Pennsylvania State U	The Methodology Center	NRG
Kamata	Akihito	M	Florida State U	Educational Psychology	FP
Khattree	Ravindra	M	Oakland U/Michigan State U	Epidemiology	FP
Mahmoodshahi	Reza	F	Columbia U	Philosophy, Logic and Methodology	NRG
Meekins	Brian	M	Bureau of Labor Statistics	Office of Survey Methods Research	FP
Miyazaki	Yasuo	M	Virginia Polytechnic Institute & State U	Educational Leadership and Policy Studies	FP
Naik	Dayanand	M	Old Dominion U	Mathematics and Statistics	NRG
Pantoja Galicia	Norberto	M	U of Waterloo		NRG
Park	Min-Hyong	F	Cornell U	Statistics	NRG
Rios-Bedoya	Carlos F	M	Michigan State U	Epidemiology	FP

Root	Tammy	F	Pennsylvania State U	The Methodology Center	NRG
Santamaria	Carolina	F	U of Wisconsin-Madison	Sociology	NRG
Santuzzi	Alecia	F	U of Illinois, Urbana-Champaign	Psychology	NRG
Thomas	Roland	M	Carleton U	Sprott School of Business	FP
Thompson	Mary	F	U of Waterloo	Statistics and Actuarial Science	FP
Visser	Ingmar	M	U of Amsterdam	Psychology	FP
Wang	Liqun	M	U of Manitoba	Statistics	FP
Wu	Yingfeng	M	New York Academy of Medicine	Center for Urban Epidiological Studies	FP
Zavisca	Jane	F	U of Arizona	Sociology	NRG

Latent Variables in the Social Sciences
Transition Workshop
Radisson Hotel Research Triangle Park
Workshop Participants
November 10-11, 2005

Last Name	First Name	Gender	Affiliation	Department	Status
Airoldi	Edoardo	M	Carnegie Mellon U	Computer Sciences	NRG
Baker	Robert	M	MetaMetrics Inc.		FP
Banks	David	M	Duke U	Statistics & Decision Sciences	FP
Banks	H.Thomas	M	North Carolina State U	Center for Research in Scientific Computation	FP
Bauer	Daniel	M	U of North Carolina at Chapel Hill	Psychology	FP
Bayarri	Susie	F	U of Valencia	Statistics & Operations Research	FP
Baydoun	May	F	U of North Carolina at Chapel Hill	Epidemiology	NRG

Bentler	Peter	M	U of California, Los Angeles	Psychology and Statistics	FP
Berger	James	M	SAMSI		FP
Biemer	Paul	M	RTI International & U of North Carolina	Odum Institute	FP
Bollen	Kenneth	M	U of North Carolina at Chapel Hill	Sociology	FP
Burdick	Don	M	MetaMetrics Inc.		FP
Cai	Bo	M	NIEHS	Biostatistics Branch	NRG
Chen	Yi-fu		Univeristy of Georgia	Center for Family Research	NRG
Cheng	Yan	M	Scientific Software International		FP
Christ	Sharon	F	U of North Carolina at Chapel Hill	Odum Institute and Sociology	NRG
du Toit	Mathilda	F	Scientific Software International		FP
du Toit	Stephen	M	Scientific Software International		FP
Dunson	David	M	NIEHS and Duke U	Statistics	FP
Edwards	Lloyd	M	U of North Carolina at Chapel Hill	Biostatistics	FP
Elam	Amanda	F	U of North Carolina at Chapel Hill	Sociology	NRG
Ellis	Alan	M	U of North Carolina at Chapel Hill	Sheps Center for Health Services Research	NRG
Federman	Elizabeth	F	RTI International		FP
Feng	Shibao	M	Georgetown U Medical Center	Biomathematics and Biostatistics	FP
Ghosal	Subhashis	M	North Carolina State U	Statistics	FP
Goodman	Leo	M	U of California, Berkeley	Statistics and Sociology	FP
Gottschall	Amanda	F	Pennsylvania State U	The Methodology Center	NRG
Hardy	Melissa	F	Pennsylvania State U	Sociology & Human Dev and Family Studies	NRG

Herring	Amy	F	U of North Carolina at Chapel Hill	Biostatistics	NRG
Hong	Chung-Chien	M	North Carolina State U	Mathematics	NRG
Huang	Xianzheng		North Carolina State U	Statistics	NRG
Jang	Woncheol	M	Duke U	Statistics & Decision Sciences	NRG
Kamata	Akihito	M	Florida State U	Educational Psychology	FP
Karr	Alan	M	National Institute of Statistical Sciences		FP
Kelly	Christopher	M	U of North Carolina at Chapel Hill	Institute on Aging	NRG
Khattree	Ravindra	M	Oakland U/Michigan State U	Epidemiology	FP
Kinney	Saki	F	Duke U	Statistics & Decision Sciences	NRG
Kolenikov	Stanislav	M	U of Missouri, Columbia	Statistics	NRG
Kovtun	Mikhail	M	Duke U	Center for Demographic Studies	FP
Kuang	Li	F	Columbia U	Sociomedical Sciences	NRG
Lackey	Gerald	M	U of North Carolina at Chapel Hill	Sociology	NRG
Langer	Michelle	F	U of North Carolina at Chapel Hill	Psychology	NRG
MacCallum	Robert	M	U of North Carolina at Chapel Hill	Psychology	FP
Macy	Rebecca	F	U of North Carolina at Chapel Hill	School of Social Work	NRG
Mahmoodshahi	Reza	F	Columbia U	Philosophy, Logic and Methodology	NRG
Meekins	Brian	M	Bureau of Labor Statistics	Office of Survey Methods Research	FP
Miyazaki	Yasuo	M	Virginia Polytechnic Institute & State U	Educational Leadership and Policy Studies	FP
Naik	Dayanand	M	Old Dominion U	Mathematics and Statistics	NRG
Nguyen	Hoan	F	North Carolina State U	Center for Research in Scientific Computation	NRG

Ostbye	Truls	M	Duke U	Community and Family Medicine	FP
Palomo	Jesus	M	National Inst of Env Health Statistics		NRG
Pantoja Galicia	Norberto	M	U of Waterloo		NRG
Park	Min-Hyong	F	Cornell U	Statistics	NRG
Ray	Surajit	M	SAMSI and U of North Carolina at Chapel Hill	Biostatistics	NRG
Reboussin	Beth	F	Wake Forest U School of Medicine	Public Health Sciences -- Biostatistics	NRG
Rios-Bedoya	Carlos F	M	Michigan State U	Epidemiology	FP
Root	Tammy	F	Pennsylvania State U	The Methodology Center	NRG
Santamaria	Carolina	F	U of Wisconsin-Madison	Sociology	NRG
Santuzzi	Alecia	F	U of Illinois, Urbana-Champaign	Psychology	NRG
Saville	Ben	M	U of North Carolina at Chapel Hill	Biostatistics	NRG
Sharpe	Kathryn	F	Duke U	Marketing	NRG
Stenner	Jack	M	MetaMetrics Inc.		FP
Sterba	Sonya	F	U of North Carolina at Chapel Hill	Quantitative Psychology	NRG
Svihula	Judie	F	U of North Carolina at Chapel Hill	Institute on Aging	NRG
Terzian	Mary	F	U of North Carolina at Chapel Hill	School of Social Work	NRG
Thomas	Roland	M	Carleton U	Sprott School of Business	FP
Thompson	Mary	F	U of Waterloo	Statistics and Actuarial Science	FP
Vance	Eric	M	Duke U	Statistics & Decision Sciences	NRG
Vandergrift	Nathan	M	U of North Carolina at Chapel Hill	Frank Porter Graham Institute	NRG
Visser	Ingmar	M	U of Amsterdam	Psychology	FP

Wang	Liqun	M	U of Manitoba	Statistics	FP
Williamson	Gary	M	MetaMetrics Inc.	Research & Development	FP
Wu	Yingfeng	M	New York Academy of Medicine	Center for Urban Epidiological Studies	FP
Yashin	Anatoli	M	Duke U	Center for Demographic Studies	FP
Zavisca	Jane	F	U of Arizona	Sociology	NRG

Education and Outreach
Undergraduate Two-Day Workshop on Financial Mathematics, Statistics and Econometrics
NISS-SAMSI Building
November 18-19, 2005

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Colleges/Univ	Number of Home State
Supported	12	11	0	0	23	14	9	18	16
Unsuppted	6	2	0	1	7	8	0	5	2
SAMSI	3	1	0	1	3	4	0	NA	NA

Education and Outreach
Undergraduate Two-Day Workshop on Financial Mathematics, Statistics and Econometrics
NISS-SAMSI Building
Supported Workshop Participants
November 18-19, 2005

Last Name	First Name	Gender	Affiliation	Major	Status
Canaday	Kylie	F	Indiana U	Economics and Mathematics	S
Celso	Jonathan	M	Loyola College in Maryland	Mathematical Sciences (Operations Research)	S
Criner	Amanda	F	U of Maine	Mathematics	S
Dinh	Ly	F	Connecticut College	Economics	S
Gant	Raymond	M	Benedict College	Mathematics	S

Gupta	Neerja	F	Connecticut College	Mathematics. Economics	S
Han	Stephanie	F	Indiana U	Economics, Political Science	S
Jones	Janelle	F	Spelman College		S
Lamb	Jennifer	F	Northern Kentucky U	Mathematics	S
Luli	Dori	F	North Central College	Applied Math	S
Martin	Elizabeth	F	U of Tennessee	Mathematics (Applied focus)	S
Medina	Eric	M	U of Michigan Ann Arbor	Finance	S
Orbe	Clara	F	Brown U	applied mathematics	S
Pallotta	Michael	M	New Jersey Institute of Technology	Mathematical Science (Finance & Actuarial Sci)	S
Reddinger	Jonathan	M	Montana State U - Bozeman	Mathematics, Economics	S
Romero	Daniel	M	Arizona State U	Mathematics	S
Rudolph	Patrick	M	U of Tennessee Knoxville	Mathematics and Statistics	S
Schaaf	Timothy	M	Grand Valley State U	Economics (B.S.) and Finance (B.B.A)	S
Somen	Bostjan	M	Indiana U	Finance, Mathematics	S
Sopeju	Aduramigba	M	Benedict College	Electrical Eng & Physics	S
Wilkoff	Sean	M	U of California at Berkeley	Mathematics, Statistics, Economics	S
Wix	Jason	M	Middle Tennessee State U	Actuarial Science	S
Wohead	Stacy	F	Southern Methodist U	Mathematics and Economics	S

Education and Outreach
Undergraduate Two-Day Workshop on Financial Mathematics, Statistics and Econometrics
NISS-SAMSI Building
Workshop Participants
November 18-19, 2005

Last Name	First Name	Gender	Affiliation	Major	Status
Alban	Douglas	M	Loyola College in Maryland	Mathematics	S
Anderson	Seth	M	North Carolina State U	Applied Mathematics and Mechanical Eng	S
Appel	Ian	M	Duke U	Mathematics	S
Canaday	Kylie	F	Indiana U	Economics and Mathematics	S
Celso	Jonathan	M	Loyola College in Maryland	Mathematical Sciences (Operations Research)	S
Criner	Amanda	F	U of Maine	Mathematics	S
Dinh	Ly	F	Connecticut College	Economics	S
Gant	Raymond	M	Benedict College	Mathematics	S
Gupta	Neerja	F	Connecticut College	Mathematics. Economics	S
Hains	Kristin	F	U of North Carolina at Wilmington	Mathematics and Statistics	S
Han	Stephanie	F	Indiana U	Economics, Political Science	S
Jones	Janelle	F	Spelman College		S
Lamb	Jennifer	F	Northern Kentucky U	Mathematics	S
Luli	Dori	F	North Central College	Applied Math	S
Martin	Elizabeth	F	U of Tennessee	Mathematics (Applied focus)	S
Medina	Eric	M	U of Michigan Ann Arbor	Finance	S
Orbe	Clara	F	Brown U	applied mathematics	S

Pallotta	Michael	M	New Jersey Institute of Technology	Mathematical Science (Finance & Actuarial Sci)	S
Pemy	Moustapha	M	SAMSI & North Carolina State U	Mathematics	NRG
Reddinger	Jonathan	M	Montana State U - Bozeman	Mathematics, Economics	S
Romero	Daniel	M	Arizona State U	Mathematics	S
Rudolph	Patrick	M	U of Tennessee Knoxville	Mathematics and Statistics	S
Schaaf	Timothy	M	Grand Valley State U	Economics (B.S.) and Finance (B.B.A)	S
Sloan	Jennifer	F	SAMSI & North Carolina State U	Statistics	NRG
Smith	Ralph	M	SAMSI & North Carolina State U	CRSC	F
Somen	Bostjan	M	Indiana U	Finance, Mathematics	S
Sopeju	Aduramigba	M	Benedict College	Electrical Eng & Physics	S
Vestal	Douglas	M	SAMSI & North Carolina State U	Mathematics	NRG
Wilkoff	Sean	M	U of California at Berkeley	Mathematics, Statistics, Economics	S
Wix	Jason	M	Middle Tennessee State U	Actuarial Science	S
Wohead	Stacy	F	Southern Methodist U	Mathematics and Economics	S

Astrostatistics
Opening Workshop and Tutorials
NISS-SAMSI Building (Tutorials) & Radisson Hotel Research Triangle Park (Workshop)
January 18-25, 2006

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	33	11	0	26	18	16	1	27	32	15
Unsuppted	21	10	1	17	15	16	1	15	16	8
SAMSI	2	0	0	0	2	0	1	1	NA	NA

Astrostatistics
Opening Workshop and Tutorials
NISS-SAMSI Building (Tutorials) & Radisson Hotel Research Triangle Park (Workshop)
Supported Workshop Participants
January 18-25, 2006

Last Name	First Name	Gender	Affiliation	Department	Status
Anderes	Ethan	M	U of California, Berkeley	Statistics	NRG
Au	Kinman	M	Carnegie Mellon Universtiy	Statistics	NRG
Barrera-Rojas	Ruth-Stella	F	U Nacional de Colombia	Observatorio Astronomico Nacional	NRG
Bickel	Peter	M	U of California, Berkeley	Statistics	FP
Chen	Jingru	M	Temple U	Statistics	NRG
Chiang	Jim	M	Stanford U	Linear Accelerator Center	NRG
Connors	Alanna	F	Eureka Scientific		FP
de Waal	Daniel	M	U of the Free State	Mathematical Statistics	FP
Dingus	Brenda	F	Los Alamos National Laboratory		FP
Driscoll	Peter	M	San Francisco State U	Physics and Astronomy	NRG
Finn	Lee Samuel	M	Pennsylvania State U	Physics	FP
Fleming	Scott	M	U of Florida	Astronomy	NRG
Genovese	Christopher	M	Carnegie Mellon U	Statistics	FP
Kashyap	Vinay	M	Smithsonian Astrophysical Observatory		FP
Kim	Chunglee	M	Northwestern U	Physics and Astronomy	NRG
Lee	Chun Man	M	Colorado State U	Statistics	NRG
Lyons	Louis	M	Oxford U	Physics	FP

Martinache	Frantz	M	Cornell U	Optical and IR Astronomy	FP
McArthur	Barbara	F	U of Texas at Austin	Astronomy	FP
Narayan	Raman	M	San Francisco State U	Physics and Astronomy	NRG
Pal	Jayanta	M	U of Michigan	Statistics	NRG
Park	Taeyoung	M	Harvard U	Statistics	NRG
Prosper	Harrison	M	Florida State U	Physics	FP
Roe	Byron	M	U of Michigan	Physics	FP
Rojas	Alex	M	Carnegie Mellon U	Statistics	NRG
Scargle	Jeffrey	M	NASA Ames Research Center	Space Science Division	FP
Sen	Bodhisattva	M	U of Michigan	Statistics	NRG
Siemiginowska	Aneta	F	Center for Astrophysics	Chandra X-Ray Center	FP
Snyder-Hugeback	Angela	F	U of Chicago	Statistics	NRG
Wasserman	Larry	M	Carnegie Mellon U	Statistics	FP
Williams	Liliya	F	U of Minnesota	Astronomy	FP
Wolszczan	Alex	M	Pennsylvania State U	Astronomy	FP
Yu	Yaming	M	U of California, Irvine	Statistics	FP

Astrostatistics
Opening Workshop and Tutorials
NISS-SAMSI Building (Tutorials) & Radisson Hotel Research Triangle Park (Workshop)
Workshop Participants
January 18-25, 2006

Last Name	First Name	Gender	Affiliation	Department	Status
Anderes	Ethan	M	U of California, Berkeley	Statistics	NRG
Apte	Amit	M	SAMSI and U of North Carolina	Mathematics	NRG
Arnaud	Keith	M	NASA Goddard Space Flight Center		FP
Au	Kinman	F	Carnegie Mellon U	Statistics	NRG
Babu	Jogesh	M	Pennsylvania State U	Statistics	FP
Banks	David	F	Duke U	Institute of Statistics and Decision Sciences	FP
Barrera-Rojas	Ruth-Stella	F	U Nacional de Colombia	Observatorio Astronomico Nacional	NRG
Bayarri	M.J.	M	U of Valencia	Statistics and Operations Research	FP
Bickel	Peter	F	U of California, Berkeley	Statistics	FP
Bullard	Floyd	F	Duke U	Institute of Statistics and Decision Sciences	NRG
Chen	Jingru	F	Temple U	Statistics	NRG
Chiang	Jim	M	Stanford U	Linear Accelerator Center	NRG
Clyde	Merlise	F	Duke U	Institute of Statistics and Decision Sciences	FP
Connors	Alanna	F	Eureka Scientific		FP
Datta	Gauri	F	U of Georgia	Statistics	FP
de la Cruz	Pablo	F	U of Valencia	Observatori Astronomic	NRG
de Waal	Daniel	M	U of the Free State	Mathematical Statistics	FP

Dingus	Brenda	F	Los Alamos National Laboratory		FP
Driscoll	Peter	M	San Francisco State U	Physics and Astronomy	NRG
Feigelson	Eric	M	Pennsylvania State U	Astronomy & Astrophysics	FP
Finn	Lee Samuel	M	Pennsylvania State U	Physics	FP
Fleming	Scott	M	U of Florida	Astronomy	NRG
Ford	Eric	M	U of California, Berkeley	Astronomy	NRG
Freeman	Peter	M	Carnegie Mellon U	Statistics	FP
Fumana	Marco	M	IASF	Milano, Lambrate	NRG
Genovese	Christopher	M	Carnegie Mellon U	Statistics	FP
Gregory	Philip	M	U of British Columbia	Physics and Astronomy	FP
Jang	Woncheol	M	Duke U	Institute of Statistics and Decision Sciences	NRG
Jefferys	William	F	Universities of Texas and Vermont	Astronomy (UT) & Statistics (UVM)	FP
Jefferys	Thomas	M			NRG
Jennings	Kristofer	M	Purdue U	Statistics	FP
Karr	Alan	M	NISS		FP
Kashyap	Vinay	M	Smithsonian Astrophysical Obsetvty		FP
Keeton	Charles	M	Rutgers U	Physics and Astronomy	FP
Kim	Chunglee	M	Northwestern U	Physics and Astronomy	NRG
Last	Michael	M	NISS		NRG
Lee	Chun Man	M	Colorado State U	Statistics	NRG
Lee	Hyunsook	F	Pennsylvania State U	Statistics	NRG

Li	Kuo-Ping	M	U of North Carolina-Chapel Hill	Biostatistics	NRG
Loh	Ji Meng	M	Columbia U	Statistics	FP
Loredo	Thomas	M	Cornell U	Astronomy	FP
Lyons	Louis	M	Oxford U	Physics	FP
Martinache	Frantz	M	Cornell U	Optical and IR Astronomy	FP
McArthur	Barbara	F	U of Texas at Austin	Astronomy	FP
Medina	Jeff	M	Birkbeck, Univ of London & London Sch of Economics	Philosophy (Birkbeck),Mathematics & Economics (LSE)	FP
Narayan	Raman	M	San Francisco State U	Physics and Astronomy	NRG
Pal	Jayanta	M	U of Michigan	Statistics	NRG
Park	Taeyoung	M	Harvard U	Statistics	NRG
Prosper	Harrison	M	Florida State U	Physics	FP
Qiu	Yang		GlaxoSmithKline	Cheminformatics	FP
Rice	John	M	U of California, Berkeley	Statistics	FP
Riddiough	Christine	F	SAS	Education	FP
Roe	Byron	M	U of Michigan	Physics	FP
Roggenkamp	Kathy	F	U of North Carolina-Chapel Hill	Biostatistics	FP
Rojas	Alex	M	Carnegie Mellon U	Statistics	NRG
Roy	Adam	M	U of California, Irvine	Statistics	NRG
Scargle	Jeffrey	M	NASA Ames Research Center	Space Science Division	FP
Schafer	Chad	M	Carnegie Mellon U	Statistics	FP
Sen	Bodhisattva	M	U of Michigan	Statistics	NRG

Siemiginowska	Aneta	F	Center for Astrophysics	Chandra X-Ray Center	NRG
Sipe	Jeffrey	M	Duke U	Institute of Statistics and Decision Sciences	NRG
Snyder-Hugeback	Angela	F	U of Chicago	Statistics	NRG
van Dyk	David	M	U of California, Irvine	Statistics	FP
Vera	Francisco	M	NISS & SAMSI		NRG
Wasserman	Larry	M	Carnegie Mellon U	Statistics	FP
Weinberg	Martin	M	U of Massachusetts	Astronomy	FP
Willett	Rebecca	F	Duke U	Electrical and Computer Engineering	NRG
Williams	Liliya	F	U of Minnesota	Astronomy	FP
Wittman	David	M	U of California, Davis	Physics	FP
Wolszczan	Alex	M	Pennsylvania State U	Astronomy	FP
Wu	Yichao	M	U of North Carolina-Chapel Hill	Statistics	NRG
Yanchukova	Diana	F	North Carolina State U	Physics	NRG
Young	C. Alex	M	NASA Goddard Space Flight Center		FP
Yu	Yaming	M	U of California, Irvine	Statistics	FP
Zhang	Lingsong	M	U of North Carolina-Chapel Hill	Statistics and Operations Research	NRG
Zhu	Zhengyuan	M	U of North Carolina-Chapel Hill	Statistics and Operations Research	FP

Financial Mathematics, Statistics and Econometrics
Workshop on Model Uncertainty
NISS-SAMSI Building
January 27, 2006

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	8	2	0	6	4	1	0	9	8	6
Unsuppted	16	1	0	9	8	0	2	15	5	3
SAMSI	0	0	0	0	0	0	0	0	NA	NA

Financial Mathematics, Statistics and Econometrics
Workshop on Model Uncertainty
NISS-SAMSI Building
Supported Workshop Participants
January 27, 2006

Last Name	First Name	Gender	Affiliation	Department	Status
Anderson	Evan	M	Northern Illinois U	Economics	FP
Antoine	Bertille	F	U of Montreal and U of North Carolina	Statistics and Operations Research	NRG
Feinberg	Yossi	M	Stanford U	Graduate School of Business	FP
Henry	Marc	M	Columbia U	Economics	NRG
Liu	Hening	M	Northern Illinois U	Economics	NRG
Sargent	Thomas	M	New York U	Economics	FP
Schneider	Martin	M	Federal Reserve Bank of Minneapolis		FP
Uppal	Raman	M	London Business School	Finance	FP
Veldkamp	Laura	F	New York U-Stern	Economics	FP

Financial Mathematics, Statistics and Econometrics
Workshop on Model Uncertainty
NISS-SAMSI Building
Workshop Participants
January 27, 2006

Last Name	First Name	Gender	Affiliation	Department	Status
Aguilar	Michael	M	U of North Carolina-Chapel Hill	Economics	NRG
Anderson	Evan	M	Northern Illinois U	Economics	FP
Antoine	Bertille	F	U of Montreal and U of North Carolina	Statistics and Operations Research	NRG
Chabi-Yo	Fousseni	M	Bank of Canada		NRG
Chen	Xilong	M	U of North Carolina-Chapel Hill	Economics	NRG
Durlauf	Steven	M	U of Wisconsin	Economics	FP
Feinberg	Yossi	M	Stanford U	Graduate School of Business	FP
Fouque	Jean-Pierre	M	North Carolina State U	Math	FP
Gallant	A. Ronald	M	Duke U	Fuqua School of Business	FP
Ghysels	Eric	M	U of North Carolina-Chapel Hill	Economics	FP
Hansen	Lars Peter	M	U of Chicago	Economics	FP
Henry	Marc	M	Columbia U	Economics	NRG
Jahan-Parvar	Mohammad	M	U of North Carolina-Chapel Hill	Economics	NRG
Liu	Hening	M	Northern Illinois U	Economics	NRG
Na	Sungsoo	M	North Carolina State U	Industrial Engineering	NRG
Orhan	Mehmet	M	U of North Carolina-Chapel Hill	Finance	NRG
Pang	Tao	M	North Carolina State U	Mathematics	FP

Renault	Eric	M	U of North Carolina-Chapel Hill	Economics	FP
Robinson	Rhett	M	U of North Carolina-Chapel Hill	Economics	NRG
Rondina	Giacomo	M	U of Wisconsin, Madison	Economics	NRG
Rossi	Barbara	F	Duke U	Economics	FP
Sargent	Thomas	M	New York U	Economics	FP
Schneider	Martin	M	Federal Reserve Bank of Minneapolis		FP
Sinko	Arthur	M	U of North Carolina-Chapel Hill	Economics	NRG
Tauchen	George	M	Duke U	Economics	FP
Uppal	Raman	M	London Business School	Finance	FP
Veldkamp	Laura	F	New York U-Stern	Economics	FP

National Defense and Homeland Security
Mid-Year Meeting on Anomaly Detection
National Center for Health Statistics, Maryland
February 3, 2006

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	7	0	0	6	1	3	2	2	7	5
Unsuppted	17	6	5	19	9	3	3	22	12	6
SAMSI	0	0	0	0	0	0	0	0	NA	NA

National Defense and Homeland Security
Mid-Year Meeting on Anomaly Detection
National Center for Health Statistics, Maryland
Supported Workshop Participants
February 3, 2006

Last Name	**First Name**	**Gender**	**Affiliation**	**Department**	**Status**
Banks	David	M	Duke U	Statistics	FP

Datta	Gauri	M	U of Georgia	Statistics	FP
Dickey	David	M	North Carolina State U	Statistics	FP
Gill	Ryan	M	U of Louisville	Mathematics	FP
Lynch	Jim	M	U of South Carolina	Mathematics	FP
Karr	Alan	M	NISS		FP
Porter	Michael	M	U of Virginia	Engineering	NRG

National Defense and Homeland Security
Mid-Year Meeting on Anomaly Detection
National Center for Health Statistics, Maryland
Workshop Participants
February 3, 2006

Last Name	First Name	Gender	Affiliation	Department	Status
Banks	David	M	Duke U	Statistics	FP
Barbari	Daniel	M	George Mason U		NRG
Brown	Donald	M	U of Virginia	Systems & Information Engineering	FP
Burkom	Howard	M	Johns Hopkins U	Applied Physics Laboratory	FP
Choi	Jai		Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Cox	Lawrence	M	Centers for Disease Control and Prevention		FP
Datta	Gauri	M	U of Georgia	Statistics	FP
Dickey	David	M	North Carolina State U	Statistics	FP
Edgerton	James	M	Johns Hopkins U	Applied Physics Laboratory	FP
Gill	Jaime	F	U of Louisville		NRG

Gill	Ryan	M	U of Louisville	Mathematics	FP
Gonzalez	Joe Fred	M	Centers for Disease Control and Prevention		FP
Grady	Nancy	F	SAIC		FP
Huang	Lan		National Institutes of Health	National Cancer Institute	FP
Karr	Alan	M	NISS		FP
Kedem	Ben	M	U of Maryland	Mathematics	NRG
Kelly	Doug	M	U of North Carolina-Chapel Hill	Statistics and Operations Research	FP
Kim	Jay	M	Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Last	Michael	M	NISS		NRG
Lin	Chien-Chih		George Mason U		NRG
Lin	Jessica	F	George Mason U	ISE	NRG
Lotze	Thomas	M	U of Maryland	Mathematics	FP
Louie	Mary	F	Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Lynch	Jim	M	U of South Carolina	Mathematics	FP
Mei	Yajun	M	Geogia Tech	Industrial and Systems Engineering	NRG
Nandram	Bal		Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Porter	Michael	M	U of Virginia	Engineering	NRG
Priebe	Carey	M	Johns Hopkins U	Mathematics and Statistics	FP
Reeves	Denise	F	MITRE		FP
Said	Yasmin	F	George Mason U	Computational Statistics	NRG
Stinchcomb	David	M	National Institutes of Health	National Cancer Institute	FP

Vera	Francisco	M	NISS		NRG
Wegman	Ed	M	George Mason U	Computational Statistics	FP
Wei	Rong		Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Wouhib	Abera	M	Centers for Disease Control and Prevention		FP

Financial Mathematics, Statistics and Econometrics
Transition Workshop
Radisson Hotel Research Triangle Park
February 27-28, 2006

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	18	4	0	13	9	1	12	9	19	12
Unsuppted	38	10	0	19	29	16	8	24	13	3
SAMSI	3	1	0	0	4	1	3	0	NA	NA

Financial Mathematics, Statistics and Econometrics
Transition Workshop
Radisson Hotel Research Triangle Park
Supported Workshop Participants
February 27-28, 2006

Last Name	First Name	Gender	Affiliation	Department	Status
Ait-Sahalia	Yacine	F	Princeton U	Economics	NRG
Bayraktar	Erhan	M	U of Michigan	Mathematics	NRG
Bhatti	Chad R	M	Rice U	Statistics	NRG
Dupire	Bruno	M	Bloomberg	Quantitative Research	FP
El Karoui	Nicole	F	Ecole Polytechnique	Centre de Mathematiques Appliquees	FP
Fan	Jianqing	M	Princeton Univesity	Operations Research and Financial Eng	FP
Fan	Yanqin	F	Vanderbilt U	Economics	FP

Figueroa-Lopez	Jose	M	Purdue U	Mathematics	FP
Henderson	Vicky	F	Princeton U	Operations Research and Financial Eng	FP
Hillebrand	Eric	M	Louisiana State U	Economics	NRG
Hobson	David	M	U of Bath	Operations Research and Financial Eng	FP
Khaliq	Abdul	M	Middle Tennessee State U	Mathematical Sciences	FP
Lee	Kiseop	M	U of Louisville	Mathematics	NRG
Liu	Ruihua	M	U of Dayton	Mathematics	FP
Protter	Philip	M	Cornell U	Operations Research	FP
Rezaei	Mahmoud	M	Clemson U	Mathematics Department	NRG
Rogers	Chris	M	U of Cambridge	Statistics Laboratory	FP
Sayit	Hasanjan	M	U of Houston	Mathematics	NRG
Schmidt	Thorsten	M	U of Leipzig	Mathematics and Computer Sciences	FP
Taqqu	Murad	M	Boston U	Mathematics	FP
Weber	Stefan	M	Cornell U	Operations Research and Industrial Eng	FP
Xia	Qing	M	U of Maryland-College Park	Mathematics	NRG
Zhang	Bing	M	U of Maryland	Mathematics	NRG

Financial Mathematics, Statistics and Econometrics
Transition Workshop
Radisson Hotel Research Triangle Park
Workshop Participants
February 27-28, 2006

Last Name	First Name	Gender	Affiliation	Department	Status
Agrawal	Gunjan	M	U of North Carolina-Chapel Hill	Statistics and Operations Research	NRG
Aguilar	Mike	M	U of North Carolina-Chapel Hill	Economics	NRG
Ait-Sahalia	Yacine	F	Princeton U	Economics	NRG
Allen	Jackie	F	SAS Institute, Inc.		FP
Bayraktar	Erhan	M	U of Michigan	Mathematics	NRG
Bhatti	Chad R	M	Rice U	Statistics	NRG
Bloomfield	Peter	M	North Carolina State U	Statistics	FP
Bolia	Nomesh	M	U of North Carolina-Chapel Hill	Statistics and Operations Research	NRG
Bollerslev	Tim	M	Duke U	Economics	FP
Busch	Thomas	M	Aarhus U & Duke U	Economics	NRG
Cao	Hongyuan	F	U of North Carolina-Chapel Hill	Statistics and Operations Research	NRG
Dedov	Maxym	M	Duke U	Economics	NRG
Dowla	Arif	M	Stochastic Logic Ltd.		FP
Dupire	Bruno	M	Bloomberg	Quantitative Research	FP
El Karoui	Nicole	F	Ecole Polytechnique	Centre de Mathematiques Appliquees	FP
Fackler	Paul	M	North Carolina State U	Statistics	FP
Fan	Jianqing	M	Princeton Univesity	Operations Research and Financial Eng	FP

Fan	Yanqin	F	Vanderbilt U	Economics	FP
Figueroa-Lopez	Jose	M	Purdue U	Mathematics	FP
Fouque	Jean-Pierre	M	U of California-Santa Barbara	Statistics and Applied Probability	FP
Gallant	Ronald	M	Duke U		FP
Ghysels	Eric	M	U of North Carolina-Chapel Hill	Economics	FP
Henderson	Vicky	F	Princeton U	Operations Research and Financial Eng	FP
Hillebrand	Eric	M	Louisiana State U	Economics	NRG
Hobson	David	M	U of Bath	Operations Research and Financial Eng	FP
Hoe	Chee Liang	M	Duke U	Physics	NRG
Hoe	SingRu	F	U of Texas-Arlington	Finance and Real Estate	NRG
Holmfeldt	Mia	F	Lund U & Duke U	Economics	NRG
Huang	Xin	M	Duke U	Economics	NRG
Ji	Chuanshu	M	U of North Carolina-Chapel Hill	Statistics and Operations Research	FP
Karuri	Stella	F	North Carolina State U	Statistics	FP
Khaliq	Abdul	M	Middle Tennessee State U	Mathematical Sciences	FP
Larson	Marcus	M	Lund U & Duke U	Economics	NRG
Lee	Kiseop	M	U of Louisville	Mathematics	NRG
Liu	Ruihua	M	U of Dayton	Mathematics	FP
Meyer	Frank	M	RMIC Corporation	Actuarial R&D, Pricing Group	FP
Na	Sungsoo	M	North Carolina State U	Industrial Engineering	NRG
Narayanan	Sriram	M	U of North Carolina-Chapel Hill	Kenan-Flagler Business Sch, Oper Management	FP

Nguyen	Bao	M	North Carolina State U	Mathematics	NRG
Pang	Tao	M	North Carolina State U	Mathematics	FP
Pelletier	Denis	M	North Carolina State U	Economics	NRG
Pemy	Moustapha	M	SAMSI & North Carolina State U	Mathematics	NRG
Protter	Philip	M	Cornell U	Operations Research	FP
Rambharat	Ricky	M	Duke U	Institute of Statistics and Decision Sciences	NRG
Renault	Eric	M	U of North Carolina-Chapel Hill	Economics	FP
Rezaei	Mahmoud	M	Clemson U	Mathematics Department	NRG
Rodriguez	Jesus	M	SAMSI		NRG
Rogers	Chris	M	U of Cambridge	Statistics Laboratory	FP
Ross	Kevin	M	U of North Carolina-Chapel Hill	Statistics and Operations Research	NRG
Rossi	Barbara	F	Duke U	Economics	FP
Sayit	Hasanjan	M	U of Houston	Mathematics	NRG
Schmidt	Thorsten	M	U of Leipzig	Mathematics and Computer Sciences	FP
Scroggs	Jeffrey	M	North Carolina State U	Mathematics	FP
Shabalin	Andrey	M	U of North Carolina-Chapel Hill	Statistics	NRG
Shaliastovich	Ivan	M	Duke U	Economics	NRG
Sinko	Arthur	M	U of North Carolina-Chapel Hill	Economics	NRG
Sizova	Natalia	F	Duke U	Economics	NRG
Sloan	Jennifer	F	North Carolina State U & SAMSI	Statistics	NRG
Taqqu	Murad	M	Boston U	Mathematics	FP

Tauchen	George	M	Duke U	Economics	FP
Toivanen	Jari	M	North Carolina State U	Center for Research in Scientific Computation	NRG
Vestal	Doug	M	North Carolina State U & SAMSI	Mathematics	NRG
Weber	Stefan	M	Cornell U	Operations Research and Industrial Eng	FP
Wolpert	Robert	M	Duke U	Institute of Statistics and Decision Sciences	FP
Wu	Yichao	M	U of North Carolina-Chapel Hill	Statistics and Operations Research	NRG
Xia	Qing	M	U of Maryland-College Park	Mathematics	NRG
Xu	Mingxin	F	U of North Carolina-Charlotte	Mathematics and Statistics	NRG
Yoon	Jungyeon	F	U of North Carolina-Chapel Hill	Statistics	NRG
Zhang	Bing	M	U of Maryland	Mathematics	NRG
Zhou	Xianwen	M	North Carolina State U	Mathematics	NRG

National Defense and Homeland Security
Mid-Year Meeting on Social Networks
Carnegie Mellon University
March 2, 2006

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	5	0	0	4	1	2	1	2	4	3
Unsuppted	12	5	0	7	10	1	4	12	5	4
SAMSI	0	0	0	0	0	0	0	0	NA	NA

National Defense and Homeland Security
Mid-Year Meeting on Social Networks
Carnegie Mellon University
Supported Workshop Participants
March 2, 2006

Last Name	First Name	Gender	Affiliation	Department	Status
Banks	David	M	Duke U	Statistics	FP
Karr	Alan	M	NISS		FP
Kolaczyk	Eric	M	Boston U	Mathematics	FP
Vance	Eric	M	Duke U	Statistics	NRG

National Defense and Homeland Security
Mid-Year Meeting on Social Networks
Carnegie Mellon University
Workshop Participants
March 2, 2006

Last Name	First Name	Gender	Affiliation	Department	Status
Airoldi	Edoardo	M	Carnegie Mellon U	Computer Science	NRG
Banks	David	M	Duke U	Statistics	FP
Blei	David	M	Princeton U	Computer Science	FP
Cohen	William	M	Carnegie Mellon U	Computer Science	FP
Fienberg	Stephen	M	Carnegie Mellon U	Statistics	FP
Frantz	Terrill	M	Carnegie Mellon U	Computer Science	FP
Goldenberg	Anna	F	Carnegie Mellon U	Computer Science	NRG
Gross	Justin	M	Carnegie Mellon U	Public Policy and Management	NRG
Hanneke	Stephen	M	Carnegie Mellon U	Computer Science	NRG

Hong	Chung-Chien	M	North Carolina State U	Mathematics	NRG
Karr	Alan	M	NISS		FP
Katzoff	Myron	M	Centers for Disease Control and Prevention		FP
Kolaczyk	Eric	M	Boston U	Mathematics	FP
Krackhardt	David	M	Carnegie Mellon U	Public Policy and Management	FP
Lee	Ann	F	Yale U	Mathematics	NRG
Mills	Shirley	F	Carleton U	Mathematics & Statistics	FP
Norminton	Ted	M	Carleton U	Mathematics & Statistics	FP
Sarkar	Purnamrita	F	Carnegie Mellon U	Computer Science	NRG
Vance	Eric	M	Duke U	Statistics	NRG
Xing	Eric	M	Carnegie Mellon U	Computer Science	NRG
Xuezhong	Chen	M	Carnegie Mellon U	Computer Science	NRG
Zheng	Alice	F	Carnegie Mellon U	Computer Science	NRG

Education and Outreach
Undergraduate Two-Day Workshop on National Defense and Homeland Security

NISS-SAMSI Building
March 3-4, 2006

Participants	Male	Female	Unspec-ified	Faculty	Student	Stat/Math Majors	Other/Unspecified	Number of Colleges/Univ	Number of Home State
Supported	12	6	0	0	18	13	5	14	11
Unsuppted	2	2	0	0	4	4	0	2	1
SAMSI	2	1	0	1	2	2	1	NA	NA

Education and Outreach
Undergraduate Two-Day Workshop on National Defense and Homeland Security
NISS-SAMSI Building
Supported Workshop Participants
March 3-4, 2006

Last Name	First Name	Gender	Affiliation	Major	Status
Bartholomew	Randyn	M	Cornell U	Mathematics	S
Carney, Jr.	Dan	M	U of Cincinnati	Mathematics & Physics	S
Fonstad	Rachel	F	Winona State U	Mathematics & Statistics	S
Greengard	Daniel	M	Wesleyan U	Mathematics	S
Helland	Robert	M	St. Cloud State U	Statistics	S
Hogue	Douglas	M	U of Cincinnati	Mathematics	S
Jain	Nikhil	M	Dartmouth College	Physics & Economics	S
Neyer	Mark	M	Xavier U	Physics, Computer Science, Mathematics	S
Omodunbi	Oluwaropo	M	New Jersey Institute of Technology	Mathematics & Actuarial Science	S
Osborn	Jacob	M	Xavier U	Computer Science & Mathematics	S
Pentico	Cassidy	F	Utah State U	Statistics	S
Phelps	William	M	Middle Tennessee State U	Actuarial Science & Financial Eng	S
Rehm	Keri	F	Meredith College	Mathematics & Religion	S
Ringer	Laura	F	Mississippi State U	Mathematics & Psychology	S
Traud	Amanda	F	North Carolina State U	Applied Mathematics	S
VanMeter	Julie	F	U of Cincinnati	Mathematics & Secondary Education	S
Whitehead	Mark	M	U of Louisville	Mathematics	S

Wix	Jason	M	Middle Tennessee State U	Actuarial Science	S

Education and Outreach
Undergraduate Two-Day Workshop on National Defense and Homeland Security
NISS-SAMSI Building
Workshop Participants
March 3-4, 2006

Last Name	First Name	Gender	Affiliation	Major	Status
Anderson	Seth	M	North Carolina State U	Mathematics & Engineering	S
Baker	Breeanne	F	Meredith College	Mathematics & French	S
Bartholomew	Randyn	M	Cornell U	Mathematics	S
Carney, Jr.	Dan	M	U of Cincinnati	Mathematics & Physics	S
Denogean	Lisa	F	SAMSI & North Carolina State U	Statistics	NRG
Fonstad	Rachel	F	Winona State U	Mathematics & Statistics	S
Greengard	Daniel	M	Wesleyan U	Mathematics	S
Harness	Chad	M	North Carolina State U	Statistics	S
Helland	Robert	M	St. Cloud State U	Statistics	S
Hogue	Douglas	M	U of Cincinnati	Mathematics	S
Jain	Nikhil	M	Dartmouth College	Physics & Economics	S
Neyer	Mark	M	Xavier U	Physics, Computer Science, Mathematics	S
Omodunbi	Oluwaropo	M	New Jersey Institute of Technology	Mathematics & Actuarial Science	S
Osborn	Jacob	M	Xavier U	Computer Science & Mathematics	S
Pentico	Cassidy	F	Utah State U	Statistics	S

Phelps	William	M	Middle Tennessee State U	Actuarial Science & Financial Eng	S
Rehm	Keri	F	Meredith College	Mathematics & Religion	S
Ringer	Laura	F	Mississippi State U	Mathematics & Psychology	S
Smith	Ralph	M	SAMSI & North Carolina State U	CRSC	F
Traud	Amanda	F	North Carolina State U	Applied Mathematics	S
Tripoli	Jennifer	F	Meredith College	Mathematics	S
VanMeter	Julie	F	U of Cincinnati	Mathematics & Secondary Education	S
Vera	Francisco	M	NISS & SAMSI		NRG
Whitehead	Mark	M	U of Louisville	Mathematics	S
Wix	Jason	M	Middle Tennessee State U	Actuarial Science	S

National Defense and Homeland Security
Mid-Year Meeting on Data Confidentiality
National Center for Health Statistics, Maryland
March 13, 2006

Participants	Male	Female	Unspec-ified	Faculty/ Professional	New Researcher/ Student	Stat	Math	Other	Number of Home Institution	Number of Home State
Supported	5	5	0	3	7	2	2	6	6	3
Unsuppted	14	3	8	24	1	3	0	22	9	4
SAMSI	0	1	0	0	1	1	0	0	NA	NA

National Defense and Homeland Security
Mid-Year Meeting on Data Confidentiality
National Center for Health Statistics, Maryland
Supported Workshop Participants
March 13, 2006

Last Name	First Name	Gender	Affiliation	Department	Status
Denogean	Lisa	F	SAMSI & North Carolina State U	Statistics	NRG

Ghosh	Joyee	F	Duke U	Statistics	NRG
Karr	Alan	M	NISS		FP
Kim	Se Hee	F	U of North Carolina	Biostatistics	NRG
Lin	Xiaodong	M	U of Cincinnati	Mathematics	NRG
Lynch	Jim	M	U of South Carolina	Mathematics	FP
Oganyan	Anna	F	NISS		NRG
Qaqish	Bahjat	M	U of North Carolina	Biostatistics	FP
Vera	Francisco	M	NISS		NRG
Woo	Mi-Ja	F	NISS		NRG

National Defense and Homeland Security
Mid-Year Meeting on Data Confidentiality
National Center for Health Statistics, Maryland
Workshop Participants
March 13, 2006

Last Name	**First Name**	**Gender**	**Affiliation**	**Department**	**Status**
Beyene	Negasi		Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Bose	Jonaki		Department of Transportation		FP
Bournazian	Jacob	M	US Department of Energy		FP
Cai	Rong		Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Chase	David	M	US Department of Housing and Urban Dev	Policy Development & Research	FP
Choi	Jai		Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Chiu	Pei-Lu		Centers for Disease Control and Prevention	National Center for Health Statistics	FP

Cohen	Stephen	M	Bureau of Labor Statistics		FP
Cox	Lawrence	M	Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Denogean	Lisa	F	SAMSI & North Carolina State U	Statistics	NRG
Ghosh	Joyee	F	Duke U	Statistics & Decision Sciences	NRG
Gonzalez, Jr.	Joe Fred	M	Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Hadden	Wilbur C.	M	Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Harris	Kenneth W.	M	Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Hawala	Sam	M	US Census Bureau		FP
Karr	Alan	M	NISS		FP
Katzoff	Myron	M	Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Khare	Meena	F	Centers for Disease Control and Prevention		FP
Kim	Jay	M	Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Kim	Se Hee	F	U of North Carolina	Biostatistics	NRG
Lin	Xiaodong	M	U of Cincinnati	Mathematics	NRG
Lu	Ruey-Pyng		US Department of Energy		FP
Lynch	Jim	M	U of South Carolina	Mathematics	FP
Massell	Paul B.	M	US Census Bureau		FP
Mitra	Robin	F	Duke U	Statistics & Decision Sciences	NRG
Oganyan	Anna	F	NISS		NRG
Qaqish	Bahjat	M	U of North Carolina	Biostatistics	FP
Reiter	Jerome	M	Duke U	Statistics & Decision Sciences	FP

Russell	Neil	M	Department of Education	National Center for Education Statistics	FP
Slavkovic	Aleksandra	F	Pennsylvania State U	Statistics	FP
Vera	Francisco	M	NISS		NRG
Wei	Rong		Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Woo	Mi-Ja	F	NISS		NRG
Wouhib	Abera	M	Centers for Disease Control and Prevention	National Center for Health Statistics	FP
Xu	Zeyu		Department of Education	National Center for Education Statistics	FP

APPENDIX B – Workshop Programs and Abstracts

I. NATIONAL DEFENSE AND HOMELAND SECURITY

A. Opening Workshop Program & Abstracts
September 11-14, 2005

Sunday--September 11, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

12:00-1:00 PM	Registration
1:00-2:00 PM	"The Collapse of the World Trade Center: Statistical Approaches" **James Filliben**, National Institute of Standards and Technology
2:00-2:30 PM	Break
2:30-3:30 PM	"New OR Models for Homeland Security" **Richard Larson**, Massachusetts Institute of Technology
3:30-4:00 PM	Break
4:00-5:00 PM	"Privacy and Confidentiality in a Homeland Security Context: Matching Disclosure Limitation" **Stephen Fienberg**, Carnegie Mellon University
5:00 PM	Adjourn

Monday--September 12, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

8:30-9:00 AM	Registration and Continental Breakfast
9:00-9:30 AM	Welcome and Introductions **Jim Berger**, SAMSI **Alan F. Karr**, NISS and SAMSI **Lawrence Cox**, National Center for Health Statistics **Nell Sedransk**, NISS
9:30-10:30 AM	"Automated Metadata" **Edward Wegman**, George Mason University
10:30-11:00 AM	Break
11:00-12:00 PM	"Statistics Can Count in National Security" **Nancy Spruill**, Department of Defense

12:00-1:00 PM Lunch

1:00-2:45 PM	**New Researchers Session** (15 minutes each)
1:00-1:15	"Using High Dimensional Geographic Feature Space and Classification Trees in the Detection of Changes in Criminal Point Processes" **Michael Porter,** University of Virginia
1:15-1:30	"Agent-based Model Applicable to Homeland Security and Disease Control" **Yasmin Said,** George Mason University
1:30-2:00	"Co-clustering of Social Networks" **J.T. Rigsby,** Naval Surface Warfare Center
2:00-2:30	"Timing & Synchronization: The Statistics that Facilitate this Defense Enabler" **Lara Schmidt,** RAND Corporation
2:30-2:45	"Estimation of New Epidemic Trends Using a Poisson Gradual Change Model" **Ryan Gill,** University of Louisville
2:45-3:00 PM	Break
3:00-4:30 PM	**Panel on Risk**

- "Game-Theoretic Methods in National Defense and Homeland Security"
  Vicki Bier, University of Wisconsin
- "Agricultural Disasters, Natural or Not: Risks and Readiness"
  Barrett Slenning, North Carolina State University
- **Richard Picard,** Los Alamos National Laboratory

4:30-5:00 PM	Poster Sales Talks
6:30-8:30 PM	**Poster Session and Reception at NISS-SAMSI Building** 19 TW Alexander Drive Research Triangle Park, NC 27709 (919) 685-9300

Poster Session Transportation Details:

Presenters:
A shuttle will leave the Radisson Hotel at 5:30pm to bring you to SAMSI in order to setup your posters. If you miss this shuttle, please see someone at the front desk to request the Radisson Shuttle. All presenters should be at SAMSI no later than 6:00pm to ensure enough time for setup.

All other participants:
Continuous shuttle service between the Radisson and SAMSI will be provided. The first shuttle will leave the Radisson at 6:25pm and the last shuttle departs from SAMSI at 8:30pm.

Tuesday--September 13, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

8:30-9:00 AM Registration and Continental Breakfast

9:00-10:00 AM "Terrorism Risk Analysis"
Detlof von Winterfeldt, University of Southern California

10:00-10:30 AM Break

10:30-12:00 PM **Panel on Syndromic Surveillance**

- "BioSense Overview and Statistical Research Needs"
  Henry Rolka, Centers for Disease Control and Prevention
- "Wavelet-Based Monitoring for Modern Biosurveillance"
  Galit Shmueli, University of Maryland
- "Joinpoint Logistic Regression Model for Monitoring Trends in Occurrence Data"
  Greg Rempala, University of Louisville

12:00-1:30 PM Lunch

1:30-3:00 PM **Panel on Cybersecurity**

- **Mike Gallaher,** RTI International
- "Detecting Computer Masquerade Attacks"
  Roy Maxion, Carnagie Mellon University
- "Some Statistical Applications in Computer and Network Security" **David Marchette**, Naval Surface Warfare Center

3:00-3:30 PM Break

3:30-5:00 PM Birds of a Feather Sessions

Wednesday--September 14, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

8:30-9:00 AM Continental Breakfast

9:00-9:45 AM "Statistical Approaches to Authorship Attribution"
David Madigan, Rutgers University

9:45-10:30 AM "Pattern Discovery in Massive Social Networks"
Hugh Chipman, Acadia University
Shirley Mills, Carleton University

10:30-11:00 AM Break

11:00-12:30 PM **Panel on New Applications**

- "Some Approaches and Challenges in Secure Statistical Analyses"
  Jerome Reiter, Duke University
- "The Human Sensor"
  Dale Anderson, Pacific Northwest National Laboratory
- **Myron Katzoff**, National Center of Health Statistics

12:30-1:30 PM Lunch

1:30-2:30 PM Reports from Birds of a Feather Sessions; Discussion

2:30 PM Working Group Meetings

Dale Anderson and **Deborah K Carlson**
Pacific Northwest National Laboratory
Department of National Security
dale.anderson@pnl.gov

"The Human Sensor"

Many significant historical events were a direct result of alert and perceptive people simply doing their job. The German invasion of Belgium, prior to World War II, was initially thwarted because an alert soldier investigated sounds of an airplane crash and found plans of the invasion on the person of two German officers. The attempt to deliver and detonate an explosive device, somewhere in the city of Seattle during the year 2000 celebration, was intercepted by an alert and perceptive US Customs officer. The safety of nuclear reactors, as measured by the presence of cracks in cooling systems under pressure, has been assessed by human inspectors for many years. Numerous other examples can be cited. In all these case studies, the human has properties that are remarkably similar to physical sensors. For example, the GeigerMueller counter is an effective nuclear emissions detector if the activity rate of the source is not to high. Otherwise, the detector becomes saturated akin to a lightbulb and is ineffective. Such saturation is the analog to human work load and work stress. Acoustic sensors can be tuned to be optimally sensitive to narrow frequency windows. Such narrow band tuning has an analog in police protocols. Officers may be required by policy to draw a weapon when responding to a call, sharply changing the sensitivity and focus of the officer. The potential tactical power of the human sensor is significant. However to fully utilize human perception as a sensor, the response must be represented mathematically, ideally with a probability model. There is a reasonable expectation that this can be accomplished because in most operational settings, strict protocols guide the actions of the human sensor. Developing probability models of human sensors will facilitate the construction of the full likelihood of human and physical sensors. From this likelihood, the human sensor becomes integral to realtime Bayesian decision methods. This presentation offers examples and ideas on the probabilistic representation of human senor response. The concept of a human senor, while provocative offer a grand challenge for the broad mathematical statistics discipline.

Vicki Bier
University of Wisconsin-Madison
Industrial and Systems Engineering
bier@engr.wisc.edu

"Game-Theoretic Methods in National Defense and Homeland Security"

In dealing with rare and extreme events (such as disasters or failures of highly redundant engineered systems), for which empirical data are likely to be sparse, classical statistical methods have been of relatively little use. Instead, risk analysis is commonly used to decompose complex systems into elements for which larger amounts of empirical data may be available. However, the routine application of risk analysis by itself is not adequate in the defense and security domain. Protecting against intentional attacks is fundamentally different from protecting against accidents or acts of nature. In particular, an intelligent and adaptable adversary may adopt a different offensive strategy to circumvent or disable our protective security measures. Game theory provides a way of taking this into account analytically. Thus, national defense and homeland security can benefit from a combination of techniques that have not usually been used in tandem. I will discuss approaches for applying risk analysis and game theory to the problem of defending complex systems against attacks by knowledgeable and adaptable adversaries. The results of such work can yield insights into the nature of optimal defensive investments, to obtain the best trade-off between cost and security.

Hugh Chipman
Acadia University
Department of Mathematics & Statistics

Shirley Mills
Carleton University
School of Mathematics and Statistics

"Pattern Discovery in Massive Social Networks"

Time-stamped communication data between nodes in a network can provide invaluable information on the network organization, subgroup structure and membership, and "interesting" nodes. Such data can arise from a variety of sources, both public domain and classified. A typical example would be email communications, with each node in the network corresponding to an email address, and an edge being a message. A public domain example of such data is the Enron email corpus, consisting of over 600,000 emails between 153 people over 40 months. Challenges include time-varying patterns, heterogeneous node behaviour, sequential cause-and-effect patterns, non-persistent nodes, and a desire to identify patterns in massive data streams. We will discuss approaches for filtering data, models for detecting patterns just described, and connections to models for social networks.

Stephen Fienberg
Carnegie Mellon University
Department of Statistics
fienberg@stat.cmu.edu

"Privacy and Confidentiality in a Homeland Security Context: Matching and Disclosure Limitation"

The events of September 11, 2001 led to heightened attention in the United States and elsewhere regarding the use of multiple government and private databases for the identification of possible perpetrators of future terrorist attacks. This talk will present an overview of some proposals that have surfaced for the searching multiple databases without compromising possible pledges of confidentiality to the individuals whose data are included. Many of these proposals are linked to the computer-science literature on privacy-preserving datamining. In particular, we focus on the matching problem across databases and the concept of “selective revelation” and their confidentiality implications. We explain why, even with encryption and nominally-protected individual data bases, one still needs to be concerned with the violation of privacy through linkage across data bases.

James Filliben
National Institute of Standards and Technology
filliben@nist.gov

“The Collapse of the World Trade Center: Statistical Approaches”

The Federal Building and Fire Safety Investigation of the World Trade Center Disaster is currently essentially completed. The pre-collapse progression was extremely complicated, with structural, thermal, dynamic, and stochastic interdependencies across both time and space. Four pre-collapse stages (a simplification of reality) will be discussed: aircraft impact, fire spread, thermal propagation through insulation, and structural deformation. Engineering issues and the statistical methodologies to address these issues will be discussed. A major challenge in the statistical analysis of the World Trade Center was the relatively meager amount of data--little physical evidence remained that could shed light on important events occurring in the core of the WTC buildings. In this regard, the study was simultaneously assisted—and complicated--by reliance on computational engineering virtual data—primarily in the form of NIST's FDS (Fire Dynamics Simulator) and phase-specific FEA (finite element analysis) computational models. As analyses progress from component to sub-assembly to global models, such computational models require characterization and validation--it will be shown how experiment design played an important role in this regard. Various other statistical analysis techniques (e.g., complex demodulation for assessing post-impact building oscillation frequency and--indirectly--building damage) will also be discussed. This paper will emphasize the methodologies employed. Conclusions and recommendations resulting from the Federal Building and Fire Safety Investigation of the World Trade Center Disaster are presented in the investigation final report, due to be released in draft form in the spring of 2005.

Keywords: Graphical Data Analysis, Experiment Design, Orthogonality, Time Series Analysis, Sensitivity Analysis, Problem-Solving Framework, Exploratory Data Analysis, Inference Feasibility, Statistical Analysis

Mike Gallaher
RTI International
Technology Economics & Policy
mpg@rti.org

Title and Abstract: TBA

Ryan Gill
Department of Mathematics
University of Louisville

"Estimation of New Epidemic Trends Using a Poisson Gradual Change Model"

This talk discusses a gradual change model for the Poisson distribution with emphasis on the identifiability, uniqueness, and existence of estimators produced by maximum likelihood. This model is used to analyze data from the SARS epidemic of 2003. In addition, procedures for detecting multiple changes are discussed and applied to this data set.

Myron Katzoff
Centers for Disease Control and Prevention
National Center for Health Statistics
mjk5@cdc.gov

Title and Abstract: TBA

Richard Larson
Massachusetts Institute of Technology
Department of Civil and Environmental Engineering
rclarson@mit.edu

"New OR Models for Homeland Security"

Abstract: TBA

David Madigan
Rutgers University
Department of Statistics
dmadigan@rutgers.edu

"Statistical Approaches in Authorship Attributions"

Abstract: TBA

David Marchette
Naval Surface Warfare Center
dmarchette@gmail.com

"Some Statistical Applications in Computer and Network Security"

I will discuss some areas of computer and network security that would benefit from the application of modern statistical techniques. In particular I will focus on modeling normal activity for the purposes of detecting attacks. This is referred to as "anomaly detection" in the security world, and has had very mixed results. Perhaps statisticians can provide some suggestions for improving the performance of such systems.

Roy Maxion
Carnegie Mellon University

Computer Science Department
maxion@cs.cmu.edu

"Detecting Computer Masquerade Attacks"

A masquerade attack, in which one on-line user impersonates another, constitutes a serious form of computer abuse. The seriousness is compounded when the masquerader is an insider -- a person who is recognized, both by the computer and by its administrators, as an accepted or legitimate member of the organization sponsoring the host or network under attack. Automatic discovery of masqueraders is sometimes undertaken by detecting significant departures from normal user behavior, as represented, for example, by a user profile formed from system audit data. While the success of this approach has been limited, the reasons for its unsatisfying performance are not obvious, possibly because most reports do not elucidate the origins of errors made by the detection mechanisms. This work takes as its point of departure a series of experiments framed by Schonlau et al. (2001), in which truncated user-command lines served as profile data. In extending Schonlau's work with a new classification algorithm (naive Bayes), a 56% improvement in masquerade detection was achieved at a corresponding false-alarm rate of 1.3%. Further improvement, up to 82% correct masquerader detection, was achieved using enriched command-line data, and additional improvements up to 92% were garnered using false-alarm mitigation techniques. A detailed error analysis, based on an alternative data configuration, reveals why some users are good masqueraders and others are not.

Richard Picard
Los Alamos National Laboratory
picard@lanl.gov

Title and Abstract: TBA

Michael Porter
University of Virginia
Department of Systems and Information Engineering
mdp2u@virginia.edu

"Using High Dimensional Geographic Feature Space and Classification Trees in the Detection of Local Changes in Criminal Point Processes"

A method is presented for detecting changes in the behavior or activity level of a criminal or terrorist point process. The locations of past events and an associated vector of geographic feature values are employed in the analysis. Classification trees are implemented to partition the high dimensional feature space, which can include mixed variables. A generalized likelihood ratio test with Monte Carlo simulations is used to test for significance. Local changes can be detected in both the intensity of the point process and distribution of point locations. An example is provided of Breaking and Entering crime locations of two types and over two-time periods to demonstrate the use of this technique in detecting local regions of change.

Jerome Reiter
Duke University
Institute of Statistics & Decision Sciences

jerry@stat.duke.edu

"Some Approaches and Challenges in Secure Statistical Analyses"

Two or more agencies may seek to "pool" their data without revealing their individual values. In this talk, I review some proposed methods for accomplishing secure statistical analyses. I also discuss some challenges with the aim of fostering research in this area.

Greg Rempala
University of Louisville
Department of Mathematics
g.rempala@louisville.edu

"Joinpoint Logistic Regression Model for Monitoring Trends in Occurrence Data"

Identification of changes in recent trends of diseases is a crucial problem in many areas of epidemiology and public health research. Most recently researchers faced that issue while working on identifying patterns in a recent world-wide outbreak of SARS epidemic. In many similar situations, in order to obtain a consistent characterization of population trends relevant to prevention, early detection or treatment of a disease, the statistical methodology based on the joinpoint or (graduate change–point) regression is of interest. This methodology characterizes a trend using joint linear segments and may be viewed as a special case of the non-parametric spline regression with variable number of knots.

In the current presentation the model based on this general principle but specifically designed for analyzing trends in clustered binary data is discussed. The model can be viewed as an extension of the simple multiple joinpoint regression to the generalized linear model with a canonical (logit) link in a quasi-binomial family. The procedure for selecting the number of joinpoints (knots) in the model is presented and a bootstrap-based method for testing the validity of the final selection of the set of knots is outlined. The procedure avoids the computationally intense and not always reliable grid search typically used in the similar setting. The presentation shall also point to some possible applications of the technique to monitoring trends in occurrence data and, in particular, detecting outbreaks of epidemics.

J.T. Rigsby
Naval Surface Warfare Center
john.rigsby@navy.mil

"Co-clustering of Social Networks"

This presentation will cover spectral decomposition based co-clustering and how it is useful for clustering two-mode social networks. The Davis, Gardner, and Gardner data set will be used as a specific example.

Henry Rolka
Centers for Disease Control and Prevention
National Center for Public Health Informatics
HRolka@CDC.Gov

"BioSense Overview and Statistical Research Needs"

The purpose of the BioSense Initiative is to support enhanced early event detection and quantified situational awareness of possible bioterrorism attacks and other disease outbreaks. It is designed to provide public health officials with timely information about their communityâ€™s health status and to support an early and effective public health response. The BioSense System is a surveillance tool that utilizes daily inputs of health-related data sources at the zip code level, maps the health indicator codes to syndrome categories, and enables jurisdictional accessibility to the data by authorized public health officials. BioSense currently includes data from ambulatory care visits and laboratory test orders. Data analysis in an empirical surveillance system that uses temporal and geographic context as a baseline against which to compare for early public health event detection is complicated by dynamic changes in reporting volumes and other data flow factors. An overview of BioSense will be provided that includes data acquisition characteristics, system implementation issues and a statistical analytic strategy. A status update on the current level of implementation and future development plans will be also be given. The focus will be to inform research agenda ideas that are consistent with practical operational needs.

Yasmin Said
George Mason University
Center for Computational Statistics
ysaid99@hotmail.com

"Agent-based Model Applicable to Homeland Security and Disease Control"

This talk discusses a prototype for developing a model framework for alcohol use and abuse that provides an assessment for interventions, which are meant to minimize acute outcomes (intentional and unintentional injuries/death). The goal is to assess the effectiveness of interventions without causing a financial or social burden, and without imposing interventions that are ultimately ineffective (or even simply not cost effective). Our framework is ecological (individual agents and interactions are represented), stochastic (neither individual behavior nor consequences of interventions are certain) and very flexible. Constructing the framework raises issues in the domain science of alcohol, statistics, mathematics, and computer science. We have developed a time and space dependent stochastic digraph model of alcohol use and abuse. The model is intended as a social network model that captures the dynamics of alcohol abuse and in particular the acute outcomes associated with alcohol abuse. The intent is to study potential interventions and investigate their effectiveness at reducing the overall prevalence of acute outcomes. Current interventions focus on one outcome at a time rather than simultaneously considering all outcomes. The work involves sophisticated mathematics (stochastic digraphs) as well as intensive data collection. It is clear that a similar model structure of social networks can be applied to terrorist networks with the same ability to examine interventions in order to assess their effectiveness.

Lara Schmidt
RAND Corporation
laras@rand.org

"Timing & Synchronization: The Statistics that Facilitate this Defense Enabler"

Precision timing drives our high-tech world. Large manufacturing processes, public utilities, communication networks, and even financial systems demand accurate and precise timing. Along with much of corporate America, the military ultimately relies on the timescale of the Department of Defense. High performance atomic clocks measure time on the order of a few billionths of a second, allowing navigation, targeting, and surveillance at unprecedented precisions. However, as we strive to pack more data across communication lines at ever-higher rates of speed, or track and navigate on interplanetary scales, nanosecond precision will not suffice. In addition to designing more stable atomic clock hardware, the development of more powerful statistical algorithms paves the way to a sub-nanosecond future.

This work shows that atomic clock data are well described by long-memory models. Used extensively in econometrics and other application areas, long-memory time series models describe processes exhibiting strong serial correlation persisting over long time spans. Their application to atomic timescales yields a new approach to clock modeling based upon a fractional difference prewhitening strategy. The fractional differencing transformation is developed and tested via simulations and live data tests. This technique yields estimators of atomic clock parameters that are both more efficient and more powerful than traditional estimators. Finally, the application of this technique to the world's largest ensemble of atomic clocks validates the use of long-memory models in atomic timekeeping. As a result of the full implementation of long-memory clock models, the task of designing a sub nanosecond timescale becomes feasible.

Galit Shmueli
University of Maryland
Decision & Information Technologies, Smith School of Business
gshmueli@rhsmith.umd.edu

“Wavelet-Based Monitoring for Modern Biosurveillance”

Current biosurveillance relies on classical statistical control charts for detecting disease outbreaks in syndromic data. However, these are not always suitable in this context. Assumptions of normality, independence, and stationarity are typically violated in syndromic data. Furthermore, outbreak signatures in such data are of unknown patterns, and therefore call for “general detectors”. Background noise is not easily modeled, and thus parametric modeling is challenging. We propose non-parametric wavelet-based methods for outbreak detection, which make less assumptions and are suitable for detecting abnormalities of unknown form. Wavelets have been widely used for data denoising and compression, but little work exists on using them for prospective monitoring. We discuss monitoring-based issues and illustrate them using data on military clinic visits.

Barrett Slenning
North Carolina State University, College of Veterinary Medicine
Population Health & Pathobiology
barrett_slenning@ncsu.edu

“Agricultural Disasters, Natural or Not: Risks and Readiness”

Agriculture is mobile, concentrated in ownership and geography, and operates through complex production and marketing networks. Its high economic efficiencies also make it brittle to perturbations anywhere in the system, a fact known by both those protecting agriculture and those wishing it ill. Lessons learned from previous natural disasters may not be helpful, and could hinder efforts to minimize agroterror threats. This presentation will walk through examples of vulnerabilities within modern agriculture, their ripple effects through to public health, and will highlight areas where applied mathematics solutions are needed.

Nancy Spruill
Office of the Under Secretary of Defense for Acquisition, Technology & Logistics
Nancy.Spruill@osd.mil

"Statistics Can Count in National Security"

Statistics can - and do - count in national security. I will present some insight into what the Department of Defense is doing with respect to material acquisition for national defense. I'll talk about the process of identifying, developing, producing and delivering the right weapon system to the right war fighter at the right place and at the right time. I'll also talk about tools that are available to war fighters to help them ensure they're getting - and maintaining - what they need.

Similarly, Business Managers have a key role in this process and they have the added burden of operating within cost, performance and schedule parameters. They, too, have tools available to help them execute their programs in a cost effective and efficient manner and I'll discuss some of them. I'll close by discussing some challenges facing the Department and suggest some homework that perhaps some in the workshop might be able to help with.

Detlof von Winterfeldt
University of Southern California
Department of Public Policy and Management
detlof@aol.com

"Terrorism Risk Analysis"

Abstract: TBA

Edward Wegman
George Mason University
Center for Computational Statistics
ewegman@galaxy.gmu.edu

"Automated Metadata"

Homeland security implies searching massive databases for information involving possible terrorists and the threats they are likely to bring. Many of these databases include free-form text such as intercepted emails and transcripts of phone calls. The implication is that these massive databases are sufficiently large that they cannot be thoroughly examined by humans. Generally metadata involve information about the format of the data, but not necessarily the actual content of the data. The concept of automated metadata is to use data mining tools to extract features from the data and to attach the features to the data as digital objects in the form of metadata. Thus an investigator could search for specific datasets having some desired features. We have employed this notion with a dataset involving 16,000 articles gathered from CNN and Reuters. This work is joint with **Faleh Alshameri** and is part of his Ph.D. dissertation work.

Poster Abstracts:

Deepak Agarwal
AT&T Labs-Research
Department of Statistics
dagarwal@research.att.com

"KFC: A Kalman Filtering Approach to Monitor Massive Contingency Tables"

Consider a computational model of streaming data where a block of records are simultaneously added to the database at regular time intervals (e.g. daily, hourly etc). Our focus is on detecting anomalous behaviour by comparing data in the current block to some baseline model based on historic data. However, we are more interested in detecting anomalous patterns rather than detecting unusual records. One way to accomplish this is to monitor counts in contingency tables corresponding to combinations of categorical variables. We employ a sequential procedure where a)cell forecasts are produced using cell specific Kalman Filters b)residuals of observed from predicted are adjusted for uninteresting and known sources of variation and c)statistically significant residuals are detected using a Bayesian shrinkage technique which automatically builds in appropriate penalty for conducting multiple tests. A decision theoretic approach is used to detect outliers and gradual changes spanning more than one time point. The method is illustrated in a syndromic surveillance context using ten years of emergency room admissions data in the NY/NJ area.

Erhan Bayraktar
University of Michigan
Department of Mathematics
erhan@umich.edu

"Minimizing the Lifetime Ruin Probability with Constrained Borrowing"

We determine the optimal investment strategy of an individual who targets a given rate of consumption and who seeks to minimize the probability of going bankrupt before she dies, also known as lifetime ruin. We impose two types of borrowing constraints: First, we do not allow the individual to borrow money to invest in the risky asset nor to sell the risky asset short. However, the latter is not a real restriction because in the unconstrained case, the individual does not sell the risky asset short. Second, we allow the individual to borrow money but only at a rate that is higher than the rate earned on the riskless asset.

We consider two forms of the consumption function: (1) The individual consumes at a constant (real) dollar rate, and (2) the individual consumes a constant proportion of her wealth. The first is arguably more realistic, but the second is closely connected with Merton's model of optimal consumption and investment under power utility. We demonstrate that connection in this paper, as well as include numerical examples to illustrate our results.

This is a joint work with **Virginia R. Young**.

Michal Czajkowski
University of Louisville
Department of Mathematics
m.czajkowski@louisville.edu

"Modeling Excess Mortality with Logistic Regression"

We consider a logistic regression model with clustered binary responses and a set of covariates which includes variable having a pre-assumed, fixed coefficient. We argue that the model is appropriate for instance when assessing excess mortality in multiple temporal clusters of disease occurrences. An example is given in which the method is applied retrospectively to analyzing temporal pattern of cancer mortality in a cohort of now retired workers from the chemical plant in Louisville, KY. The results are compared with that obtained from standard mortality analysis.

Kevin Ward Drummey
U. S. Federal Government
Department of Defense
DR_KEV21144@MSN.COM

"Predictive Stream Mining Challenges at DOD"

In this poster presentation, we describe some of the unique challenges faced by the DOD when mining data in a real-time environment for "items of interest." Generally, we regard an item as a collection of metadata associated with an event of some type, where the metadata is generated from multiple, multimedia, raw data sources. Our problem is then to build and deploy predictive models to classify incoming items as "interesting" (to process) or not interesting (to discard). We are motivated by our customers' limited time, resources, and data storage, all of which make it impossible for our customers to examine every item. We must also rely on customer feedback in order to build predictive models. However, customer feedback is often inconsistent, being provided by multiple "labelers," and feedback is inherently biased, being based on only the examined items. One of our goals, though, is real-time automation of the predictive modeling process, and this process must incorporate feedback in some way. Other important predictive modeling challenges include: real-time data cleansing and feature selection; processing predominantly categorical metadata containing hundreds to thousands of categories; modeling highly non-stationary metadata while detecting "changepoints" in model performance; and predicting an extremely rare "target" class, whose definition actually changes over time due to the temporal nature of events and customer requirements.

Jianping Fan
University of Carolina at Charlotte
Department of Computer Science
jfan@uncc.edu

Xiaodong Lin
University of Cincinnati
Department of Mathematical Sciences
linxd@math.uc.edu

"Privacy-Preserving Integration of Surveillance Videos for Homeland Security Application"

In this presentation, we have developed a novel approach to enable privacy-preserving integration of surveillance videos from multiple groups for homeland security application. First, the salient objects are detected automatically from surveillance video. Second, privacy-sensitive salient objects and video events are filtered out from the surveillance videos. Third, the filtered surveillance videos are shared among groups for discovering interesting events for homeland security applications.

This is joint work with **Hangzai Luo**.

Kathryn Mills
Government of Canada
Department of Defense
kmbox7-travel@yahoo.ca

Dr. Ted Norminton and **Dr. Shirley Mills**
Carleton University
School of Mathematics and Statistics
norminton@rogers.com,
smills@math.carleton.ca

"Poster-Visualization of Network Scanning"

Computer software to visualize network scanning activity in real-time.

Chad Schafer
Carnegie Mellon University
Department of Statistics
cschafer@stat.cmu.edu

"A Method for Constructing Confidence Sets of Optimal Expected Size"

In the physical sciences it is common that a complex stochastic model relates unknown parameters to observable data. I present a method for approximating minimax expected size confidence procedures for these situations.

This work was motivated by a problem in cosmology. Observations of the Cosmic Microwave Background Radiation (CMB) are an important source of information regarding the values of physical parameters. Recent experiments, such as WMAP, observe the CMB with high precision, yet estimating cosmological parameters remains difficult due to the complexity of the stochastic model.

Plume models, such as that developed at NARAC at Lawrence Livermore National Laboratory, describe the behavior of atmospheric releases given the location and nature of the source. The inverse problem, "event reconstruction" given observed data, is important, but challenging. This method presents a potential way of utilizing these large simulation models for inference.

Galit Shmueli
University of Maryland
Decision & Information Technologies, Smith School of Business
gshmueli@rhsmith.umd.edu

"Wavelet Methods for Syndromic Surveillance"

Current biosurveillance relies on classical statistical control charts for detecting disease outbreaks in syndromic data. However, these are not always suitable in this context.

Assumptions of normality, independence, and stationarity are typically violated in syndromic data. Furthermore, outbreak signatures in such data are of unknown patterns, and therefore call for “general detectors”. Background noise is not easily modeled, and thus parametric modeling is challenging. We propose non-parametric wavelet-based methods for outbreak detection, which make less assumptions and are suitable for detecting abnormalities of unknown form. Wavelets have been widely used for data denoising and compression, but little work exists on using them for prospective monitoring. We discuss monitoring-based issues and illustrate them using data on military clinic visits.

B. *Mid-Year Meeting on Anomaly Detection Program and Abstracts*
February 3, 2006

9:30-9:45 AM	Welcome **Alan Karr**, NISS **Lawrence Cox**, NCHS
9:45-10:15 AM	Scan Statistics **Gauri Datta**, University of Georgia
10:15-11:15 AM	*Keynote Presentation* Scan Statistics on Error Graphs **Carey Priebe**, Johns Hopkins University
11:15-11:30 AM	Break
11:30-12:30 PM	***Keynote Presentation*** Adaptations of Data Modeling and Process Control for Prospective Biosurveillance **Howard Burkom**, Johns Hopkins University
12:30-1:30 PM	Lunch
1:30-2:00 PM	Logistic Joinpoint Methods for Occupational Cohort Data **Ryan Gill**, University of Louisville
2:00-3:00 PM	*Keynote Presentation* Discrete Choice Models for Incident Prediction **Donald Brown**, University of Virginia
3:00-3:15 PM	Break
3:15-3:45 PM	Introduction to Time Series and Intervention **David Dickey**, North Carolina State University
3:45-4:15 PM	Feature Based Density Estimation **Francisco Vera**, NISS
4:15-4:45 PM	Detecting Jumps in Piece-wise Locally Stationary Time Series. **Michael Last**, NISS

4:45-5:00 PM Break

5:00-5:30 PM Group Discussion of Future Work

Donald E. Brown, Keynote Presentation
Discrete Choice Models for Incident Prediction
Terrorist and criminal incidents take a number of forms, for instance, suicide bombings, explosive devices, sniper shootings and arson. Conventional approaches to mathematically understanding and predicting the locations of such attacks are to model them as point patterns and apply techniques from the extensive literature of point pattern analysis. Such methods, including density estimation and clustering techniques, typically focus on the spatial relationships of the attacks. While these approaches do occasionally reveal important spatial consistencies they often fail as predictive models. Their failure stems from an incorrect model of the underlying causal mechanism which is not a physical or natural phenomenon but a choice process by a motivated attacker. To capture this choice process requires an approach to understanding the preferences of the attackers. This presentation shows how this can be done using utility functions. These functions then form the basis for predictions of future attacks. This presentation discusses this approach and shows its usefulness to several problems in incident prediction.

Howard S. Burkom, Keynote Presentation
Adaptations of Data Modeling and Process Control for Prospective Biosurveillance
Syndromic surveillance involves the monitoring of available data sources for early warning of outbreaks of unspecified disease or of specified disease before the confirmation of identifying symptoms, with the objective to complement physician sentinel surveillance with false alarm rates acceptable to the public health infrastructure. Data sources include clinical data such as counts of syndrome-specific emergency department visits or physician office visits, and nonclinical data such as over-the-counter remedy sales and school/work absentee rates. A common approach among system developers has been to adapt chart-based methods from the field of statistical process control. Major obstacles to this approach are the evolving and often nonstationary input data streams, the uncertainty of the nature of the signal to be detected, and the presence of systematic or periodic behavior in the data background. Thus, robust detection performance, measured by timeliness and sensitivity at controlled alert rates, requires a combination of modeling and process control suitable to the characteristics of the monitored data. This presentation investigates such combination methods using standard evaluation techniques as well as focused techniques specific to the biosurveillance context. A discussion of generalized exponential smoothing methods will be given along with a performance comparison against a recently published regression modeling approach.

David A. Dickey
Introduction to Time Series and Intervention
In this shortened version of my workshop presentation, I will briefly talk about time series structures, show the time series representation for the exponentially weighted moving average operator that we have seen in some of our syndromic surveillance papers, and show some examples of intervention analysis, including a new example detailing the impact of the 9/11 terrorist strike on passenger volume at our local RDU airport.

Ryan Gill
Logistic Joinpoint Methods for Occupational Cohort Data
We describe logistic joinpoint methods for modeling incidence data. The model is well-suited for data where temporal changes in the functional pattern are suspected. We report the current progress and future plans in applying the methods to mortality data for the Louisville VC cohort of chemical workers.

Michael Last
Detecting Jumps in Piece-wise Locally Stationary Time Series
Finding jumps in piece-wise stationary time series is a problem considered by many authors. This model is too rigid for some applications, such as seismology. I will discuss a method for detecting jumps in piece-wise locally stationary time series, and look at applications to work in seismology and speech recognition.

Carey E. Priebe, Keynote Presentation
Scan Statistics on Enron Graphs
We introduce a theory of scan statistics on graphs and apply the ideas to the problem of anomaly detection in a time series of Enron email graphs.

Francisco Vera
Feature Based Density Estimation
Ideas to extend upon Donald E. Brown's criminal prediction model.

C. *Mid-Year Meeting on Social Networks Program and Abstracts*
March 2, 2006

9:00-9:30 AM	**Alan Karr**, NISS
9:30-9:45 AM	Discussion
9:45-10:15 AM	**Purnamrita Sarkar**, CMU
10:15-10:30 AM	Discussion
10:30-11:00 AM	Coffee
11:00-11:30 AM	**Edo Airoldi** and **Eric Xing**, CMU
11:30-11:45 AM	Discussion
11:45-1:00 PM	Lunch
1:00-1:30 PM	**Negash Medhin**, NC State
1:30-1:45 PM	Discussion
1:45-2:15 PM	**Eric Kolaczyk**, Boston University
2:15-2:30 PM	Discussion

2:30-2:45 PM	Coffee
2:45-3:15 PM	**Stephen Hanneke**, CMU
3:15-3:30 PM	Discussion
3:30-4:00 PM	**Eric Vance**, Duke
4:00-4:15 PM	Discussion
4:15-4:45 PM	**Goldenberg** and **Zheng**, CMU
4:45-5:00 PM	Discussion

Latent Space Mixture Models for Networks
Edoardo Airoldi and **Eric Xing**
Carnegie Mellon University

Projecting observed interactions onto a low-dimensional latent space is a convenient way to visualize structure that possibly underlies the data. Popular methods, however, tend to separate the projection task from the quantitative analysis of the structure that is possibly present in the latent space. For example, latent space models project interactions onto a latent space by inverting some function of data and latent positional elements, whereas stochastic block models seek specific structural regularities in the form of latent clusters and cluster-to-cluster connection patterns.

We develop a new methodology that integrates the task of projecting data onto a latent space with that of seeking structure. We explicitly posit a parametric version of the regularities we wish to infer, e.g., soft clusters, as structural elements of the latent space onto which we project that observed interactions. We present a specific model that extends the latent space model of Hoff et al. (2002) by positing a mixture of Gaussians in the latent space, and we derive a convenient variational approximation to solve the Bayes problem for this model.

Dynamic Contextual Friendship Networks
Anna Goldenberg and **Alice Zheng**
Carnegie Mellon University

We live in a society built upon the complex web of interpersonal relationships. For decades, researchers have been fascinated with the characterization of such social networks, examples of which include academic research paper co-authorships, film actor co-star relationships, and many more. With the recent rise of large online user communities, the study of these networks seems more relevant than ever.

In this talk, we focus on developing a generative model of evolving social networks. The social actors in our model have evolving distributions over spheres of interaction, which we term "contexts." The model allows for the birth and death of social ties and addition of new actors. We study the robustness of our model by examining the statistical properties of simulated networks in comparison to those of real networks. We conclude with computational issues of parameter learning in this model.

Chain Graph Temporal Models of Social Networks

Stephen Hanneke
Carnegie Mellon University

We propose a family of markov statistical models for social network evolution over time, which represents an extension of Exponential Random Graph Models (ERGMs). Many of the methods and theorems for ERGMs are readily adapted for this model, including MCMC maximum likelihood estimation algorithms. We discuss example models of this type along with empirical results for estimation and prediction tasks.

Optimization and differential games approaches for the analysis of social networks

Chung-Chien Hong, N. G. Medhin
North Carolina State University

We present two approaches to the study of social networks. The first method is based on nonlinear programming and the second on differential games. In the nonlinear programming approach we consider a social group where each actor of the social group has limited statistical information on each of the other actors. Each actor also has a set of preferred values and attributes. Further, the likelihood of a link from one actor to another is likely to be higher in the case of perceived reciprocity. A non-linear programming problem is constructed to obtain a social matrix, where an entry of 1 in the *ij*-th entry indicates the presence of a link from actor *i* to actor *j*, whereas an entry of zero indicates the absence of a link. We also construct a probability social matrix where the entries represent the probability of a link. The model is made dynamic by varying the preferred values of each actor. Finally, we study the movements of actors leading to the identification of cliques. In this model, an actor need not know every member of the group. However, there is an increasing probability to get acquainted with more actors as time passes.
The differential games approach starts with a dynamical model where the players/actors have a set of strategies reflecting their values and preferences. The differential game is studied to understand the social network and its time evolution.

On Network Sampling and Inference of Network Structure: A Case Study Using Trace route and the Internet

Eric Kolaczyk
Boston University

Empirical network measurements have been at the heart of a variety of discoveries of both commonality and differences in network structure across disciplines. Among other things, such discoveries have inspired a keen interest in the development of generative network graph models that can reproduce observed characteristics. However, recent work in a number of fields in the past few years has found that the method by which such measurements are obtained can have important implications on the extent to which the observed characteristics accurately reflect those of the 'true' underlying network. In this talk I will review some examples of such work and discuss the problem of reliable estimation of certain network characteristics. I will concentrate on the context of traceroute sampling in the Internet and the challenge of attacking various `species' problems.

Dynamic Social Network Analysis using Latent Space Models

Purnamrita Sarkar

Carnegie Mellon University

This work explores two aspects of social network modeling. First, we generalize a successful static model of relationships into a dynamic model that accounts for friendships drifting over time. Second, we show how to make it tractable to learn such models from data, even as the number of entities n gets large. The generalized model associates each entity with a point in p-dimensional Euclidian latent space. The points can move as time progresses but large moves in latent space are improbable. Observed links between entities are more likely if the entities are close in latent space. We show how to make such a model tractable (sub-quadratic in the number of entities) by the use of appropriate kernel function for similarity in latent space; the use of low dimensional kd-trees; a new efficient dynamic adaptation of multidimensional scaling for a first pass of approximate projection of entities into latent space; and an efficient conjugate gradient update rule for non-linear local optimization in which amortized time per entity during an update is O(log n). We use both synthetic and real-world data on upto 11,000 entities which indicate linear scaling in computation time and improved performance over four alternative approaches. We also illustrate the system operating on twelve years of NIPS co-publication data.

Social Networks in Elephants
Eric Vance
Duke University

Wild female African elephants live in a matriarchal society and form persistent family groups. However, within these groups the elephants frequently split into subgroups in a process known as fission/fusion, and these patterns of affiliation are not well understood. In this talk I use a bilinear mixed effects model proposed by Peter Hoff (2005) to isolate several key components of elephant social behavior. This model incorporates the key notion of an unobserved latent social space to better describe the interactions between elephants. The model is flexible enough to include predictors of pairwise affiliation, such as kinships, which allows large-mammal ecologists to test assumptions about elephant social structure, and to develop new theories of why and how elephants interact.

D. *Mid-Year Meeting on Data Confidentiality Program and Abstracts*
March 13, 2006

8:45 AM Welcome and Introductions

9:00 AM Session I: Mathematical Tools for SDL

9:00 Mathematical Models For Data Confidentiality Problems
Lawrence Cox, NCHS

10:00 Break

10:30 Algebraic Statistics for Statistical Disclosure Limitation in Contingency Tables
Aleksandra Slavkovic, Penn State

11:15 A Stochastic Process Approach to the Analysis of Swapping for Categorical Variables

Lisa Denogean, SAMSI

11:45 Discussion
James Lynch, University of South Carolina

12:15 PM Lunch

1:30 PM Session II: Inference and Quality

1:30 Effect of Rounding on Data Quality
Jay Kim, NCHS

2:15 Combinations of SDC Methods for Continuous Microdata
Anna Oganyan, NISS

2:45 New Measures of Data Utility
Mi-Ja Woo, NISS

3:15 Break

3:45 Secure Statistics Software for Horizontally Partitioned Data
Francisco Vera, NISS

4:15 Discussion
Jerome Reiter, Duke University

4:45 PM Adjourn

Mathematical Models for Data Confidentiality Problems
Lawrence H. Cox, Ph.D.
National Center for Health Statistics
LCOX@CDC.GOV

I will summarize mathematical models for data rounding, data perturbation, complementary cell suppression, and disclosure audit in two-way tables. A related method, quality-preserving controlled tabular adjustment, will not be discussed due to limitations of time. I will illustrate how these problems may be efficiently solved (to optimality) for two-way tables and related classes of tables, but are extremely difficult to impossible to solve for other classes.

A Stochastic Process Approach to the Analysis of Swapping for Categorical Variables
Lisa Denogean
Statistical and Applied Mathematical Sciences Institute
lisa.denogean@gmail.com

Data swapping can be used by government agencies to protect the confidentiality of publicly released data files. We study the stochastic process generated by data swapping applied to a data file of categorical variables. The purpose is to understand the effect of swapping and to help the original data owners to determine which variables to swap and

how much to swap. We discuss various utility measures and introduce the idea of measuring distance from the limit rather than from the original file. In addition, we introduce a new type of swapping that we propose is superior to current methods.

Effects of Rounding on Data Quality

Jay J. Kim
National Center for Health Statistics
pzk3@cdc.gov

Integer data such as frequency counts may be rounded to an integer base to several purposes including disclosure limitation. Similarly, it is sometimes necessary to round noninteger data to integer data (viz., base 1 rounding) for statistical purposes, e.g., rounding expected sample counts (noninteger) to actual sample counts (integer). We evaluate the effects of four methods of rounding data on data quality and utility in two ways: (1) bias and variance (increase in total mean square error) and (2) effects on the underlying distribution of the data. The four rounding rules are conventional rounding, modified conventional rounding, zero-restricted 50/50 rounding, and unbiased rounding.

Combinations of SDC methods for continuous microdata

Anna Oganian, NISS

Algebraic Statistics for Statistical Disclosure Limitation in Contingency Tables

Aleksandra B. Slavkovic
Penn State
sesa@stat.psu.edu

Categorical data collected by federal agencies and non-government survey organizations are often summarized in tabular form. The release of partial information is of public utility and typically involves reporting collections of marginal and conditional tables. Given this information, tools from algebraic geometry can be used to characterize discrete distributions for contingency tables, and to determine a disclosure risk. Algebraic statistics exploits the use of polynomial algebra and algebraic geometry for statistical inference. We demonstrate how the tools from algebraic geometry are used to represent the tables of counts and describe the locus (T) of all possible tables under the given constraints. We discuss some practical implication of using algebraic statistics for data privacy and confidentiality problems.

Secure Statistics Software for Horizontally Partitioned Data

Francisco Vera
National Institute of Statistical Sciences
Statistical and Applied Mathematical Sciences Institute
veraf@niss.org

There are several methods for secure computations of statistics on distributed data. At the National Institute of Statistical Sciences a software to do some of these computations is under development. A snap shot of some of its current and future capabilities is given in this talk.

New measures of data utility
Mi-Ja Woo, NISS
mjwoo@niss.org

When data is released to the public, it is important to find the data alteration method with high confidentiality that provides satisfactory data quality. I will focus on developing methods of measuring data quality when the distribution of data is unknown. We treat the data utility as a problem of evaluating similarities of original data structure to masked data structure. The data utilities we present here are rooted in the cumulative distribute function, clustering and propensity score approaches. When the distribution is departed from normal, simulations for a wide variety of data structures show how these measures can be used for evaluating disclosure limitation procedures.

II. FINANCIAL MATHEMATICS, STATISTICS AND ECONOMETRICS

A. *Workshop on Stochastic Modeling in Financial Mathematics* (co-sponsored with and held at CRM, U of Montreal) *Program*
June 1-5, 2005

Wednesday--June 1, 2005

9:00-9:30 AM Registration and Continental Breakfast

9:30-10:15 AM Atlas Models for Large Equity Markets
I. Karatzas

10:15-10:30 AM Coffee Break

10:30-11:15 AM PDE Approach to Valuation and Hedging of Credit Derivatives.
M. Jeanblanc

11:15-11:30 AM Coffee Break

11:30-12:15 PM Gamma Constraints and Second Order Backward SDE's
N. Touzi

12:15-2:30 PM Lunch

2:30-3:15 PM Sensitivity Analysis of Utility Based Prices and Risk-Tolerance Wealth Processes
D. Kramkov

3:15-3:30 PM Coffee Break

3:30-4:15 PM Variance Dynamics: Joint Evidence from Options and High-Frequency Returns
L. Wu

4:15-4:30 PM Coffee Break

4:30-5:15 PM Efficient Derivative Pricing by Extended Method of Moments

E. Renault

5:15-6:00 PM Postdoc Student Session

5:15-5:35 About Some Modification of the Constant Elasticity of Variance Diffusion Model **R. Makarov**

5:35-6:00 Optimal Static-Dynamic Hedges for Employee Stock Options **A. Ilhan**

Thursday--June 2, 2005

900-9:30 AM Registration and Continental Breakfast

9:30-10:15 AM P. Schonbucher

10:15-10:30 AM Coffee Break

10:30-11:15 AM An Equilibrium Economy with Irreversible Investment
U. Haussmann

11:15-11:30 AM Coffee Break

11:30-12:15 PM **R. Sircar**

12:15-2:30 PM Lunch

2:30-3:15 PM **T. Zariphopoulou**

3:15-3:30 PM Coffee Break

3:30-4:15 PM **T. Bielecki**

4:15-4:30 PM Coffee Break

4:30-5:15 PM Non-ergodic Stock Price Dynamics in Financial Markets with Heterogeneous Agents
U. Horst

Friday--June 3, 2005

900-9:30 AM Registration and Continental Breakfast

9:30-10:15 AM Option Pricing for Pure Jump Processes with Markov Switching Compensators
R.J. Elliott

10:15-10:30 AM Coffee Break

10:30-11:15 AM Portfolio Optimization without Forecasts
R. Almgren

11:15-11:30 AM Coffee Break

Saturday--June 4, 2005

900-9:30 AM Registration and Continental Breakfast

9:30-10:15 AM Additive and Multiplicative Duals for America Option Pricing.
P. Glasserman

10:15-10:30 AM Coffee Break

10:30-11:15 AM An ARCH Model for Several Time Scales
E. Hillebrand

11:15-11:30 AM Coffee Break

11:30-12:15 PM Discretely Sampled Barrier and Bermudan Options
S. Howison

12:15-2:30 PM Lunch

2:30-3:15 PM **J. Fan**

3:15-3:30 PM Coffee Break

3:30-4:15 PM Credit Spread, Endogenous Default, and Implied Volatility with Jump Risk
S. Kou

4:15-4:30 PM Coffee Break

4:30-5:15 PM Postdoc Student Session

4:30-4:50 A Control Variate Method to Evaluate Option Prices Under Stochastic Volatility Models **S. Han**

4:50-5:15 Some Comments on Statistical Modeling in Finance **Y. Jin**

Sunday--June 5, 2005

900-9:30 AM Registration and Continental Breakfast

9:30-10:15 AM Classical and Bayesian Analysis of Dynamic Correlation Models. **D. Pelletier**

10:15-10:30 AM Coffee Break

10:30-11:15 AM Solutions to Hedging Problems with Interacting Ito and Point Processes
D. Becherer

11:15-11:30 AM	Coffee Break
11:30-12:15 PM	Dynamic Principal-Agent Problems with Perfect Information. **A. Cadenillas**

B. *Opening Workshop Program & Abstracts*
September 18-21, 2005

Sunday--September 18, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

8:15-900 AM		Registration and Continental Breakfast
9:00-5:00 PM		**Tutorials**
	9:00-10:30	Financial Mathematics **Ronnie Sircar**, Princeton University
	10:30-11:00	Coffee Break
	11:00-12:30	Financial Mathematics *(continued)*
	12:30-1:30	Lunch
	1:30-3:00	Financial Econometrics **Bas Werker**, Tilburg University
	3:00-3:30	Coffee Break
	3:30-5:00	Financial Econometrics *(continued)*
6:30-8:30 PM		**Poster Session and Reception** Radisson, Room ABC (2nd floor) Poster Presenters should arrive no later than 6:00pm to ensure they have sufficient time for setup.

Monday--September 19, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

8:30-9:00 AM	Registration and Continental Breakfast
9:00-9:15 AM	Introduction and Welcome **Jim Berger**, SAMSI **Jean-Pierre Fouque**, North Carolina State University **Eric Ghysels**, University of North Carolina at Chapel Hill **Nell Sedransk**, National Institute of Statistical Sciences

9:15-10:15 AM *Inaugural Lecture:* "The Spline GARCH Model for Unconditional Volatility and its Global Macroeconomic Causes"
Robert Engle, New York University
2003 Nobel Laureate

10:15-10:45 AM Coffee Break

10:45-12:15 PM SESSION 1: *Model Uncertainty*
Organizers: **Eric Ghysels** and **Lars Hansen**
Moderator: **Eric Renault**, University of North Carolina at Chapel Hill

- "Recursive Robust Estimation and Control Without Commitment"
  Lars Peter Hansen, University of Chicago
- "The Impact of Risk and Uncertainty on Expected Returns"
  Jennifer Juergens, Arizona State University

12:15-1:45 PM Lunch

1:45-400 PM SESSION 2: *Statistics and Finance*
Organizer: **Ruey Tsay**
Moderators: **Daren Cline**, Texas A&M University
Zongwu Cai, University of North Carolina at Charlotte

- "Regression Splines and Singular Stochastic Control in a Modified Black-Scholes Theory Incorporating Transactions Costs" **T.L. Lai,** Stanford University
- "Random Field and Affine Models for Interest Rates: An Empirical Comparison" **Alan Bester,** University of Chicago
- "Reduced–form models of corporate default: An Empirical Analysis" **Antje Berndt,** Cornell University

4:00-4:30 PM Coffee Break

4:30-5:45 PM Panel Discussion*: Models of Risk and Uncertainty: What is the Future?*
Organizer: **Eric Ghysels**
Moderator: **George Tauchen,** Duke University

- **Ole Barndorff-Nielsen**, Aarhus University
- **Robert Engle**, New York University
- **Lars Hansen**, University of Chicago
- **Ruey Tsay**, Uiversity of Chicago

Tuesday--September 20, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

8:00-8:30 AM Registration and Continental Breakfast

8:30-10:00 AM SESSION 3: *Derivatives*

Moderators: **Nicole Elkaroui**, Ecole Polytechnique
Ronnie Sircar, Princeton University

- "Dispersion Trading in US Equity Markets" **Marco Avellaneda**, New York University
- "Consistency of Indifference Pricing in Incomplete Markets" **Thaleia Zariphopoulou**, University of Texas-Austin

10:00-10:30 AM Coffee Break

10:30-11:30 AM SESSION 4: *Energy Markets*
Moderator: **Glen Snider**, Progress Energy (Raleigh, NC)

- "Mathematical Challenges if the Electricity Markets" **Rene Carmona**, Princeton University

11:30-12:15 PM SESSION 5: *Extreme Events*
Organizers and Moderators:
Eric Renault, University of North Carolina at Chapel Hill
Richard Smith, University of North Carolina at Chapel Hill
George Tauchen, Duke University

- "Robust Variation Estimation Using Kernels in Financial Econometrics" (joint work with Ole Barndorff-Nielsen, Peter Hansen and Asger Lunde) **Neil Shephard**, University of Oxford

12:15-2:00 PM Lunch

2:00-3:30 PM SESSION 5: *Extreme Events* (continued)

- "Hedging Options in Presence of Jumps" **Rama Cont,** Ecole Polytechnique
- "Financial Data and the Hidden Semimartingale Model" **Per Mykland**, University of Chicago

3:30-4:00 PM Coffee Break

4:00-5:30 PM SESSION 6: *Credit Risk*

Moderators: **Monique Jeanblanc**, University of Evry
Jean-Pierre Fouque, North Carolina State University

- "Modeling the Recovery Rate in a Reduced Form Model" **Robert Jarrow**, Cornell University
- "Unified Valuation of Corporate Liabilities, Equity Derivatives and Credit Derivatives" **Vadim Linetsky**, Northwestern University

Wednesday--September 21, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

8:00-8:30 AM Registration and Continental Breakfast

8:30-10:00 AM SESSION 7: *Computational Issues*
Moderator: **Ron Gallant**, Duke University

- "Agent-Based Financial Markets with Heterogeneous Memory as a Generator for Long Memory in Asset Return Volatility" **Blake LeBaron**, Brandeis University
- "Applications of High-Performance Computing in Finance" **Stathis Tompaidis**, University of Texas-Austin

10:00-10:30 AM Coffee Break

10:30-12:00 PM SESSION 8: *Practioners Session*
Moderator: **Peter Cotton**, Morgan Stanley

- "Market Diversity and the Distribution of Capital in Equity Markets" **Adrian Banner**, INTECH
- "Modelling Mortaged Backed Securities" **Yakov Kanter**, Morgan Stanley

12:00-1:30 PM Lunch

1:30-3:00 PM SESSION 9: *Portfolio Optimization*
Moderator: **Thaleia Zariphopoulou**, University of Texas-Austin

- "Risk Minimizing Static-Dynamic Hedges for Exotic Options" **Aytac Ilhan**, University of Oxford
- "Mutual Fund Portfolio Choice in the Presence of Dynamic Flows" **Ron Kaniel**, Duke University

3:00-3:30 PM Coffee Break

3:30-5:30 PM Discussion and Continuation of Working Groups

Inaugural Lecture:

Robert Engle, 2003 Nobel Laureate
New York University, Stern School of Business
Salomon Center
rengle@stern.nyu.edu

"The Spline GARCH Model for Unconditional Volatility and its Global Macroeconomic Causes"

We introduce a new model to measure unconditional volatility, the Spline-GARCH. The model is applied to equity markets for 50 countries for up to 50 years of daily data. Macroeconomic determinants of unconditional volatility are investigated. It is found that volatility in macroeconomic factors such as GDP growth, inflation and short term interest rates are important explanatory variables that increase volatility. There is evidence that high inflation and low growth of output are positive determinants. Volatility is higher for emerging markets and for markets with small numbers of listings but also for large economies.

Marco Avellaneda
New York University
Courant Institute of Mathematical Sciences
avellane@cims.nyu.edu

"Dispersion Trading in US Equity Markets"

Dispersion trading (DT) consists of a strategy where one simultaneously sells index options and buys options on the index components or, vice-versa, buys index options and sells options on the underlying stocks. The fundamental theme associated with DT is "correlation trading." This presentation gives stylized facts about DT in U.S. equity markets and some of the main indicators of correlation that can be used in dispersion trading. We derive some statistical properties of these indicators and outline a somewhat crude but effective approach to price-discovery using the method of steepest descent (RISK Magazine, October 2002). The talk also discusses hedging and risk-management considerations for the resulting option portfolios.

Adrian Banner
INTECH
adrian@enhanced.com

"Market Diversity and the Distribution of Capital in Equity Markets"

Antje Berndt
Carnegie Melon University
Tepper School of Business
aberndt@andrew.cmu.edu

"Reduced-Form Models of Corporate Default: An Empirical Analysis"

We test five reduced-form models of corporate default, using as data Moody's KMV estimates of actual default probabilities (EDFs) for 103 firms in four industries: automobiles, broadcasting and entertainment, healthcare, and oil and gas. Based on the nonparametric specification test of Hong and Li (2005), we strongly reject several popular affine models of default arrival intensities. Assuming that the logarithm of the default intensity is specified as an Ornstein-Uhlenbeck process, we present maximum-likelihood estimates of the term structure of corporate default probabilities, allowing for a joint distribution of EDFs across firms in a given industry sector. Additional data on default swap (CDS) market rates allow us to establish a strong link between actual and risk-neutral default probabilities, and to obtain maximum-likelihood estimates of the time-series behavior of recent default risk premia for U.S. corporate debt. We find dramatic variation over time in risk premia, from peaks in the third quarter of 2002, dropping by roughly 50% to late 2003. Portions of this talk are based on joint work with **R. Douglas, D. Duffie, M. Ferguson, and D. Schranz**.

Alan Bester
University of Chicago
Graduate School of Business
cbester@gsb.uchicago.edu

"Random Field and Affine Models for Interest Rates: An Empirical Comparison"

Traditional affine models of the term structure are eminently tractable, but suffer from empirical difficulties. Random field models offer great flexibility in fitting term structure data, but are widely considered non-implementable unless they are approximated by a low-dimensional system. I develop a state-space estimation framework where both random field and affine models can be estimated by MCMC using the same panel of forward rate data. I find that random field models are much better able to fit the patterns of volatility and correlation in a long historical sample of U.S. Treasury forward rates.

Rene Carmona
Princeton University
ORFE & Bendheim Center for Finance
rcarmona@princeton.edu

"Mathematical Challenges of the Electricity Markets"

The first part of the talk will offer a non-technical review of some of the challenging mathematical problems faced by financial engineers involved in the energy markets. We will concentrate our attention to the idiosyncrasies of the electricity markets.

The second part of the talk will concentrate on applications of optimal switching methods to the valuation of power plants and gas storage facilities.

Rama Cont
Ecole Polytechnique
Department of Mathematics
Rama.Cont@polytechnique.fr

"Hedging Options in Presence of Jumps"

The presence of large, sudden moves in market prices and the empirical shortcomings of diffusion models in describing patterns observed in market prices of options have made processes with jumps increasingly popular in option pricing applications.

We presence empirical evidence for presence of jumps in price series using an approach, jointly developed with C. Mancini, which enables, in a non-parametric setting, to disentangle jumps from the continuous component of price evolution.

The presence of such discontinuities in asset prices has important consequences for the hedging of options and leads us to question various intuitions about hedging conveyed by diffusion models. We compare various hedging strategies using the underlying asset and a set of liquid options and examine the properties of the hedging error, both theoretically and through numerical experiments. We illustrate in particular that using sensitivities to compute (Delta-neutral and Gamma-neutral) hedge ratios can lead to a large hedging error and illustrate how such strategies can be improved by using a risk-minimizing approach to hedging and by taking positions in options.

References:

- Cont, R. and Tankov, P. (2004) Nonparametric calibration of jump-diffusion option pricing models, Journ. Computational Finance, Vol. 7 (3), 1 - 49.
- Cont, R. and Voltchkova, E. (2005) Integro-differential equations for option prices in exponential Levy models, Finance and Stochastics, Vol. 9 (3), 299-325.
- Cont R., Tankov P. and Voltchkova, E. (2005) Hedging options in presence of jumps, Working Paper.
- Cont, R. and Mancini, C. (2005) Detecting jumps in asset prices, Working paper.
- Cont, R. and Tankov, P. (2003) Financial modelling with jump processes, CRC Press.

Lars Peter Hansen
University of Chicago
Department of Economics
l-hansen@uchicago.edu

"Recursive Robust Estimation and Control without Commitment"

In a Markov decision problem with hidden state variables, a posterior distribution serves as a state variable and Bayes' law under an approximating model gives its law of motion. A decision maker expresses fear that his model is misspecified by surrounding it with a set of alternatives that are nearby when measured by their expected log likelihood ratios (entropies). Martingales under the approximating model represent alternative models. A decision maker constructs a sequence of robust decision rules by pretending that a sequence of minimizing players choose increments to a martingale and distortions to the prior over the hidden state. A risk sensitivity operator induces robustness to perturbations of the approximating model conditioned on the hidden state. Another risk sensitivity operator induces robustness to the prior distribution over the hidden state. We use these operators to extend our earlier work to problems that contain hidden states. The worst case martingale is over determined, expressing an intertemporal inconsistency of worst case beliefs about the hidden state, but not about observables.

Aytac Ilhan
University of Oxford
Mathematical Institute
ilhan@maths.ox.ac.uk

"Risk Minimizing Static-Dynamic Hedges for Exotic Options"

We study optimal hedging of exotic options using a combination of a static position in vanilla options and dynamic trading of the underlying asset, where the risk of a strategy is assessed by a convex risk measure. We give conditions guaranteeing differentiability and strict convexity of the risk measure in the hedging quantity, therefore existence of a unique solution to the hedging problem. This is illustrated in a model where the investor holds a European option written on a nontraded asset, and accepts positions that have expected losses under a certain threshold.

Robert A. Jarrow
Cornell University
Business School
raj15@cornell.edu

"Modeling the Recovery Rate in a Reduced Form Model"

This paper provides a model for the recovery rate process in a reduced form model. After default, a firm continues to operate, and the recovery rate is determined by the value of the firm's assets relative to its liabilities. The debt recovers a different magnitude depending upon whether or not the firm becomes insolvent. Although this recovery rate process is similar to that used in a structural model, the reduced form approach is maintained by utilizing information reduction in the sense of Guo, Jarrow and Zeng (2005). Our model is able to provide analytic expressions for a firm's default intensity, insolvency intensity, and zero-coupon bond prices.

Jennifer L. Juergens
Arizona State University
Department of Finance
jennifer.juergens@asu.edu

"The Impact of Risk and Uncertainty on Expected Returns"

We study asset pricing when a representative agent faces model uncertainty and empirically demonstrate that model uncertainty matters for asset pricing both in reduced from regressions and in a dynamic representative agent model with microfoundations. We link the amount of model uncertainty in the economy to measurable quantities via the predictions of professional forecasters. In particular, we construct aggregate measures of model uncertainty based on the disagreement of professional forecasters about future expected returns and demonstrate that the measure has a significant effect on excess returns. We run regressions representing the typical risk-return trade-off, where risk is represented by conditional volatility and augment these regressions with the measure of model uncertainty attributing different weights to each professional forecaster. The weighting scheme is estimated via a method that is inspired by the recent work on MIDAS regressions. We find stronger empirical evidence for a uncertainty-return trade-off than a traditional risk-return trade-off. We also construct a dynamic economic model in which the beliefs of agents are given by the predictions of professional forecasters and show that the model provides a better explanation of asset prices than models in which agents have rational expectations. When agents have beliefs related to the beliefs of professional forecasters, we show there exists a reasonable level of risk-aversion and a reasonable time-discount factor that can match the equity premium and risk-free rates observed in data.

Ron Kaniel
Duke University
Fuqua School of Business
ron.kaniel@duke.edu

"Mutual Fund Portfolio Choice in the Presence of Dynamic Flows"

We analyze the implications of the widely used fixed fraction of funds fees on a mutual fund manager's portfolio decisions. In our model, a log utility investor is allowed to dynamically allocate capital between an actively managed mutual fund and a locally riskless bond. The optimal fund portfolio is shown to be the one that maximizes the market value of the fees received, and is independent of the manager's utility function. The presence of dynamic flows induces "flow hedging" portfolio distortions on the part of the fund, even though the investor is myopic. We predict a positive relationship between a fund's proportional fee rate and its volatility. This is a consequence of higher fee funds holding more extreme equity positions. However, the overall dollar amount of equity held by a fund can be independent of the fee rate. While both the fund portfolio and the investor's trading strategy depend on the proportional fee rate, the equilibrium value functions do not. Implications related to the measured performance-fundflow relationship and its dependence on the fee rate are derived. Finally, we show that our results hold even if in addition to trading the fund and the bond the investor is allowed to directly trade some of the risky securities, but not all.

Yakov Kanter
Morgan Stanley
Interest Rate Analytical Modelling
yakov.kanter@morganstanley.com

"Modelling Mortgage Backed Securities"

I will discuss some of the challenges of modelling MBS as interest rate derivatives and what can be done to overcome them.

Tze Leung Lai
Stanford University
Department of Statistics
lait@stat.stanford.edu

"Regression Splines and Singular Stochastic Control in a Modified Black-Scholes Theory Incorporating Transactions Costs"

Blake LeBaron
Brandeis University
International Business School
blebaron@brandeis.edu

"Agent-Based Financial Markets with Heterogeneous Memory as a Generator for Long Memory in Asset Return Volatility"

This talk explores some theoretical mechanisms for volatility persistence. In particular, it is shown that an agent-based model with traders assessing strategies using different

amounts of past data (heterogeneous memory) can generate persistent return volatility. Furthermore, for certain distributions of agents this volatility appears to be long memory which is similar to results from actual financial time series.

Vadim Linetsky
Northwestern University
Department of Industrial Engineering and Management Science
linetsky@iems.northwestern.edu

"Unified Valuation of Corporate Liabilities, Equity Derivatives and Credit Derivatives"

We consider the problem of developing a unified framework for the valuation of corporate liabilities, equity derivatives, and credit derivatives. Theory and empirical evidence suggest that default indicators such as credit default swap (CDS) spreads and corporate bond yields are positively related to historical volatility and implied volatilities of equity options. Theory and empirical evidence also suggest that a stock's realized volatility is negatively related to its price (leverage effect) and that implied volatilities are decreasing in the option's strike price (implied volatility skew). We propose and study several parsimonious reduced-form models of defaultable stock which capture these fundamental relationships and enable us to price equity and credit derivatives in a unified fashion. In particular, corporate credit spreads and equity option volatility skews are closely related in this class of models.

Per Mykland
University of Chicago
Department of Statistics
mykland@galton.uchicago.edu

"Financial Data and the Hidden Semimartingale Model"

The availability of high frequency data for financial instruments has opened the possibility of accurately determining volatility in small time periods, such as one day. Recent work on such estimation indicates that it is necessary to treat the data with a hidden semimartingale model, typically by the addition of measurement error. We review the emerging theory on this subject, including two- and multiscale sampling. We also consider broader error schemes, through Makrov kernels and such phenomena as rounding due to discreteness of prices.

Neil Shephard
University of Oxford
Nuffield College
neil.shephard@nuf.ox.ac.uk

"Robust Variation Estimation Using Kernels in Financial Econometrics" (joint work with **Ole Barndorff-Nieldsen**, **Peter Hansen** and **Asger Lunde**)

Realised bipower variation consistently estimates the quadratic variation of the continuous component of prices. In this paper we generalise this concept to realised bipower covariation, study its properties, illustrate its use, derive its asymptotic distribution and use it to test for jumps in multivariate price processes.

Ronnie Sircar
Princeton University
Department of Operations Research & Financial Engineering
sircar@princeton.edu

"Tutorial on Financial Mathematics"

This tutorial will attempt to give a broad overview of topics of interest in Financial Mathematics for newcomers to the field. We will introduce some of the issues to be discussed in detail at the Opening Workshop and highlight some classical successes in the field, for example in option pricing, portfolio optimization and (if time allows) credit risk.

Stathis Tompaidis
University of Texas at Austin
McCombs School of Business
Tompaidis@mccombs.utexas.edu

"Applications of High-Performance Computing in Finance"

We discuss applications of parallel computing in Finance. We describe two examples where parallel computing has been used succesfully, one in optimal asset allocation, and the other in derivative pricing.

Bas Werker
Tilburg University
Finance and Econometrics Groups
Werker@uvt.nl

"A Primer in Financial Econometrics"

In this tutorial, we give an overview of the basic stochastic volatility (Finance)/conditional heteroskedasticity (Econometrics)/time-varying scale models (Statistics) models, with a special emphasis on statistical inference (estimation and testing). The second part of the tutoral will focus on so-called semiparametric specifications with an emphasis on rank-based inference.

Thaleia Zariphopoulou
University of Texas at Austin
Department of Mathematics and Red McCombs Business School
zariphop@math.utexas.edu

"Consistency of Indifference Pricing in Incomplete Markets"

Indifference pricing offers a way to valuate and manage risks in incomplete market environments. Several questions arise related, among others, to numeraire consistency and behavior of indifference prices across different maturities. These issues as well as the structure of the indifference-hedging strategies will be discussed.

Poster Titles:

Bertille Antoine (Department of Statistics, Universite de Montreal and University of North Carolina at Chapel Hill) *"Portfolio Selection with Estimation Risk: A Test Based Approach"*

Erhan Bayraktar (Department of Mathematics, University of Michigan) *"Minimizing the Lifetime Ruin Probability with Constrained Borrowing"* (Joint work with **Virginia R. Young**)

Francisco Chamu (Department of Statistics, University of North Carolina at Chapel Hill) *"Financial Risk Assessment with Max-Stable Processes"*

Maxym Dedov (Department of Economics, Duke University) *"Joint Pricing of Bonds and Stocks: A Quadratic Model"*

Elena Goldman (Department of Finance and Economics, Pace University) *"Asymmetric Adjustment of Realized Volatility"*

Irina Goldman (Department of Financial Engineering, Stevens Institute of Technology) *"On the Distribution of Market Capitalization and Enterprise Value over Rank: Analytical Treatment and Empirical Findings"*

Martin Groth (Centre of Mathematics for Applications, University of Oslo) *"Indifference pricing in the Barndorff-Nielsen - Shepard model; A PDE approach."*

Sanggohn Han (Department of Economics and Statistics, North Carolina State University) *"Inference Regarding Multiple Structural Changes in Linear Models Estimated Via Instrumental Variables"*

Eric Hillebrand (Department of Economics, Louisiana State University) *"Neglecting Parameter Changes in Autoregressive Models"*

Xin Huang (Department of Economics, Duke University) *"A Semiparametric Framework for Modeling and Forecasting Jumps and Volatility in Speculative Prices"*

Alec Kercheval (Department of Mathematics, Florida State University) *"Optimal Covariances in Risk Model Aggregation"*

Abdul Khaliq (Department of Mathematical Sciences, Middle Tennessee State University) *"Numerical PDE Approach for the Valuation of Multi-Asset Exotic Options"*

Uta Kretschmer (University of Bonn, Germany) *"The Dynamics of Bipower Variation, Realized Volatility, and Returns"*

Kostas Kyriakoulis (Department of Economics, North Carolina State University) *"Second Order Approximations to GMM Statistics"*

Michael Levine (Department of Statistics, Purdue University) "Nonparametric Estimation of Volatility Models with Serially Dependent Innovations"

Feng Liu (Department of Statistics, North Carolina State University)
"The t Copula with Correlation Structure of EGARCH- DCC for Market VaR"

Xiaofang Ma (Department of Computer Science, University of Toronto)
"Loss Distribution Evaluation for Synthetic CDOs" (Joint work with **Ken Jackson** (Department of Computer Science) and **Alex Kreinin** (Algorithmics Inc.)

Linyan Miao (Department of Financial Analytics, Stevens Institute of Technology)
"Empirical Study of Value-at-Risk and Expected Shortfall Models with Heavy Tails"

Mehmet Aras Orhan (Department of Finance, University of North Carolina at Chapel Hill) *"The Term Structure of Trading Activity"*

Dmitry Ostrovsky (Department of Mathematics, Lehigh University)
"Option Pricing in Random Time: Black-Scholes-Merton and CEV"

Alessandro Palandri (Department of Economics, Duke University)
"Sequential Conditional Correlations: Inference and Evaluation"

Moustapha Pemy (Department of Mathematics, North Carolina State University)
"Optimal Stock Liquidation in a Regime Switching Model with Finite Time Horizon"

Caio Ibsen Rodrigues de Almeida (Statistics/Business School, Ibmec-RJ, Brazil) *"Do Options Contain Information About Excess Bond Returns?"* (Joint work with **Scott Joslin** and **Jeremy Graveline**)

Kevin Ross (Department of Statistics, University of North Carolina at Chapel Hill)
"Convergent Numerical Scheme for Singular Control with State Constraints in a Portfolio Selection Problem"

Arthur Sinko (Department of Economics, University of North Carolina at Chapel Hill)
"*Estimation of Large Covariance Matrices for Risk Management Purpose"*

Viktor Todorov (Department of Economics, Duke University)
"Econometric Analysis of Jump Driven Stochastic Volatility Models"

Jules van Binsbergen (Fuqua School of Business, Duke University)
"Portfolio Choice and Premium Policy for Pension Funds"

Li Wang (Department of Statistics and Probability, Michigan State University)
"Confidence Bands for Regression Curve Under Weak Dependence"

Mingxin Xu (Department of Mathematics and Statistics, University of North Carolina at Charlotte) *"Risk Measure Pricing and Hedging in Incomplete Markets"*

Jun Yan (Department of Statistics and Actuarial Science, University of Iowa)

"The R Package Copula"

Bing Zhang (Department of Mathematics, University of Maryland)
"Price Barrier Options in FX Under Stochastic Skew Model via MC Simulation and Finite Differences"

Peng Zhang (Department of Statistics, University of Pennsylvania, Wharton School)
"Bayesian Inference for Random Coefficient Dynamic Panel Data Models"

Stephen Zhou (Department of Applied Mathematics, North Carolina State University)
"A Martingale Control Variate Method for LSM in Pricing American Options"

C. *Workshop on Credit Risk Program and Abstracts*
October 31-November 2, 2005

Monday--October 31, 2005
NISS Building, Room 104

1:30-2:00 PM Registration

2:00-2:45 PM "Optimal Bank Loan Portfolio Hedging: A Practitioner's Perspective"
Randy Miller, Bank of America

3:00-3:45 PM "The Valuation of Correlation-Dependent Credit Derivatives Using a Structural Model"
Mirela Predescu, University of Toronto

4:00-4:45 PM "t-Statistics for Weighted Means in Credit Risk Modelling"
Kiseop Lee, University of Louisville

5:00-5:45 PM "Credit Spreads, Capital Structure, and Implied Volatility with Endogenous Default and Jump Risk"
Nan Chen, Columbia University

5:45-6:45 PM **Reception**

Tuesday -- November 1, 2005
NISS Building, Room 104

8:30-9:15 AM Registration and Continental Breakfast

9:15-10:00 AM "Endogenous Time of Default"
Rene Carmona, Princeton University

10:15-11:00 AM "Decomposing Swap Spreads"
David Lando, Copenhagen Business School & Princeton University

11:15-12:00 PM "Credit Spreads and Credit Risk Premia"

Pierre Collin-Dufresne, Goldman Sachs

12:00-1:30 PM Lunch

1:30-2:15 PM "Default and Stochastic Volatility"
Jean-Pierre Fouque, North Carolina State University

2:30-3:15 PM "The Performance of Structural Models of Default for Firms with Liquid CDS Spreads"
Mirela Predescu, University of Toronto

3:30-4:15 PM "The Statistical Mechanics of Credit Contagion"
Ulrich Horst, University of British Columbia

Wednesday--November 2, 2005
NISS Building, Room 104

8:30-9:15 AM Continental Breakfast

9:15-10:00 AM "A Top Down Approach to Multi-Name Credit"
Lisa Goldberg, MSCI Barra

10:15-11:00 AM "Credit Ratings"
Peter Bloomfield, North Carolina State University

11:15-12:00 PM "Indifference Prices and Stochastic Risk Preferences"
Thaleia Zariphopoulou, University of Texas at Austin

12:00-1:30 PM Lunch

Peter Bloomfield
North Carolina State University
Department of Statistics
bloomfield@stat.ncsu.edu

"Credit Ratings"

The systems of credit ratings maintained by agencies such as Moody's and Standard & Poor's play an important role in evaluating credit risk. The agencies' historical data are valuable in interpreting the rating of a given obligor. Two issues that arise in using those historical data are: how to describe the dependence among rating changes for several obligors; and how to relate those changes to the broader economy. Both issues can be addressed by using an appropriate copula.

Rene Carmona
Princeton University
Department of Operations Research & Financial Engineering
rcarmona@princeton.edu

"Endogenous Time of Default"

We review several forms of the classical Leland's model of the capital structure of a leveraged firm and we give numerical evidence of the changes in the optimal time of default when options are added to the initial mix equity/debt.

Nan Chen
Columbia University
Department of Industrial Engineering and Operations Research
nc2005@columbia.edu

"Credit Spreads, Capital Structure, and Implied Volatility with Endogenous Default and Jump Risk"

We propose a two-sided jump model for credit risk by extending the Leland-Toft endogenous default model based on the geometric Brownian motion. The model shows that jump risk and endogenuous default can have significant impacts on credit spreads, optimal capital structure, and implied volatility of equity options: (1) The jump and endogenuous default can produce a variety of non-zero credit spreads, including upward, humped, and downward shapes; interesting enough, the model can even produce, consistent with empirical findings, upward credit spreads for speculative grade bonds. (2) The jump risk leads to much lower optimal debt/equity ratio; in fact, with jump risk, highly risky firms tend to have very little debt. (3) The two-sided jumps lead to a variety of shapes for the implied volatility of equity options, even for long maturity options; and although in general credit spreads and implied volatility tend to move in the same direction under exogenous default models, but this may not be true in presence of endogenuous default and jumps. In terms of mathematical contribution, we give a proof of a version of the "smooth fitting" principle for the jump model, justifying a conjecture first suggested by Leland and Toft under the Brownian model.

Pierre Collin-Dufresne
Goldman Sachs
Department of Quantitative Strategies
pierre.collin-dufresne@gs.com

"Credit Spreads and Credit Risk Premia"

We examine whether 'large' historical credit spreads can be explained within a standard asset pricing framework in the face of low historical default rates. For this to be the case, we show that the pricing kernel must covary strongly and negatively with asset prices -- a characteristic which is also needed to explain the equity premium puzzle. As such, we explore whether those pricing kernels that have been successful at capturing historical equity returns (e.g., Campbell and Cochrane (CC 1999) and Bansal and Yaron (BY 2004)) can also explain the 'credit spread puzzle'. We find this to be the case if risk premia are strongly time-varying (i.e., not just large and constant) and the default boundary is counter-cyclical. Finally, to investigate the time-series implications, we feed historical consumption innovations into the CC model (with counter-cyclical default boundaries) and show that the predicted credit spreads fit both the level and dynamics of historical credit spreads quite well.

Peter Cotton
Morgan Stanley
Department of Fixed Income

peter.cotton@morganstanley.com

"There are Hard Problems in Finance"

Most sensible scientific approaches to the management and trading of portfolio credit products, hybrids and complex products run up against reasonably tough mathematical or computational hurdles. Furthermore, it is not difficult to speculate on the future directions for sell side financial engineering when such problems will be all the more important. I will argue that attempts to circumvent these are naive, that progress can be made, and that meaningful cooperation between industry and academia is possible - albeit difficult - and essential. Two methods for achieving this cooperation are considered. The first is a collection of non-trivial applied mathematical problems of enormous importance to both industry and academia. The second is the establishment of a testbed for rigorous model assessment.

Jean-Pierre Fouque
North Carolina State University
Department of Mathematics
fouque@math.ncsu.edu

"Default and Stochastic Volatility"

We analyze the effects of stochastic volatility on defaults and correlation of defaults.

Lisa Goldberg
MSCI Barra
Department of research
lisa.goldberg@mscibarra.com

"A Top Down Approach to Multi-Name Credit"

We describe a top down approach to forecasting economy-wide defaults and to pricing multi-name credit derivatives from a top down perspective. A self-exciting process is used to model events in the economy and random thinning is used to generate single name default processes.

Ulrich Horst
University of British Columbia
Department of Mathematics
horst@math.ubc.ca

"The Statistical Mechanics of Credit Contagion"

We study a model of credit contagion with an infinite number of interacting firms. Individual ratings depend on the ratings of some designated business partners and the average rating throughout the whole economy. Based on new convergence results for a class of particle systems with macroscopic interaction we show that the dynamics of average ratings follows a deterministic measure-valued ODE. If the interaction between different firms is sufficiently weak, the ODE converges to a unique steady state. Ergodicity breaks down when the interaction becomes too strong.

David Lando
Copenhagen Business School & Princeton University
Operations Research & Financial Engineering
dlando@princeton.edu

"Decomposing Swap Spreads"

We analyze a six-factor model for Treasury bonds, corporate bonds, and swap rates and decompose swap spreads into three components: A convenience yield from holding Treasuries, a credit risk element from the underlying LIBOR rate, and a factor specific to the swap market. In the later part of our sample, the swap-specific factor is strongly correlated with hedging activity in the MBS market. The model further sheds light on the relationship between AA hazard rates and the spread between LIBOR rates and GC repo rates and on the level of the riskless rate compared to swap and Treasury rates.

Joint work with Peter Feldhütter

Kiseop Lee
University of Louisville
Department of Mathematics
kiseop.lee@louisville.edu

"t-Statistics for Weighted Means in Credit Risk Modelling"

We present a generalization of the two-sample t-test for equality of the means to the case where the sample values are to be given unequal weights. This is a natural situation in financial risk modelling where some samples are considered more reliable than others in predicting a common mean. We describe pooled and unpooled weighted t-tests, and show with an example of real credit data that using the standard unweighted t-test can lead to the poor statistical conclusion.

Randy Miller
Bank of America
Department of Global Portfolio Strategies
randy.j.miller@bankofamerica.com

"Optimal Bank Loan Portfolio Hedging: A Practitioner's Perspective"

Banks are actively using credit derivative markets for managing default and migration risks in their loan portfolios. The presentation will review experience with an implementation of a CVAR minimization method developed by Uryasev, Rockafellar, et.al. The optimization results help formulate portfolio rebalancing strategy. The loan portfolio rebalancing with derivatives generates a related problem of hedging P&L volatility in the credit derivative book. The static hedging problem features include incomplete markets, jump-diffusions, basis risk, and model risk. Results of some alternative hedging methods will be reviewed. Also, the two problems are interdependent. They set up a potential risk-reward trade-off between the best loan rebalancing strategies and the best P&L volatility hedging strategies.

Mirela Predescu
University of Toronto

Department of Finance
mirela.predescu01@rotman.utoronto.ca

Monday’s presentation --
“The Valuation of Correlation-Dependent Credit Derivatives Using a Structural Model”

In 1976 Black and Cox proposed a structural model where an obligor defaults when the value of its assets hits a certain barrier. In 2001 Zhou showed how the model can be extended to two obligors whose assets are correlated. In this paper we show how the model can be extended to a large number of different obligors. The correlations between the assets of the obligors are determined by one or more factors. We examine the dynamics for credit spreads implied by the model and explore how the model price tranches of collateralized debt obligations (CDOs). We compare the model with the widely used Gaussian copula model of survival time and test how well the model fits market data on the prices of CDO tranches. We consider two extensions of the model. The first reflects empirical research showing that default correlations are positively dependent on default rates. The second reflects empirical research showing that recovery rates are negatively dependent on default rates.

Tuesday’s presentation --
“The Performance of Structural Models of Default for Firms with Liquid CDS Spreads”

This paper investigates the performance of structural models using a sample of companies with very liquid credit default swap (CDS) contracts. First, I analyze the models’ ability to predict future credit spreads and compare their performance with that of more naïve alternatives that take into account just past CDS spreads. The structural models are implemented in two ways: one is using historical equity data along with balance sheet information to estimate the model, the other is using both equity and past CDS spreads data, in addition to balance sheet information. I find that, on average, a naïve method outperforms the structural models, regardless of their implementation procedure. There is, however, a sub-sample of firm weeks in which the structural models outperform the naïve approach, and these are weeks with significant changes in CDS spreads. Second, I investigate the structural models’ value in predicting credit ratings migrations. I find that long-term changes in the default probabilities implied by structural models have incremental value in anticipating rating downgrades above and beyond CDS spreads changes. However, this is not the case for short-term changes. Over short periods, the incremental information of structural models disappears once I control for CDS spreads.

Thaleia Zariphopoulou
University of Texas at Austin
Department of Mathematics and IROM
zariphop@math.utexas.edu

“Indifference Prices and Stochastic Risk Preferences”

In this talk I will discuss certain aspects of utility based prices related to their behavior across numeraires and investment horizons. I will also analyze the optimal behavior and risk monitoring in the case of stochastic risk preferences. Examples from the Credit Risk area will be presented.

D. *Workshop on Model Uncertainty Program and Abstracts*
January 27, 2006

Friday -- January 27, 2006
NISS Building, Room 104

8:30-8:50 AM	Registration
8:50-9:00 AM	Welcome and Opening Remarks
9:00-9:45 AM	"Recursive Robust Estimation and Control Without Commitment" **Lars Hansen**, University of Chicago and **Tom Sargent,** New York University
9:45-10:30 AM	"Design Limits and Optimal Policy Evaluation" **William Brock**, **Steve Durlauf,** and **Giacomo Rondina,** University of Wisconsin
10:30-11:00 AM	Coffee Break
11:00-11:45 AM	"Ambiguity, Information Quality and Asset Pricing" **Larry Epstein**, University of Rochester and **Martin Schneider,** New York University
11:45-12:30 PM	"To Hold Familiar Assets or To Diversify? Keynes Meets Markowitz" **Phelim Boyle,** University of Waterloo, **Lorenzo Garlappi,** University of Texas-Austin, **Raman Uppal,** London Business School, and **Tan Wang,** University of British Columbia
12:30-1:30 PM	Lunch
1:30-2:15 PM	"Small Noise Methods for Risk Sensitive/Robust Economies" **Evan Anderson,** Northern Illinois University and **Lars Hansen,** University of Chicago
2:15-3:00 PM	"The Impact of Risk and Uncertainty on Expected Returns" **Evan Anderson** (Northern Illinois Univ.), **Eric Ghysels** (UNC-Chapel Hill), and **Jennifer Juergens** (Arizona State Univ.)
3:00-3:30 PM	Coffee Break
3:30-4:15 PM	"Non-Bayesian Testing of a Stochastic Prediction" **Eddie Dekel**, Northwestern University and Tel-Aviv University and **Yossi Feinberg**, Stanford University
4:15-5:00 PM	"Information Acquisition and Portfolio Under-Diversification" **Stijn Van Nieuwerburgh** and **Laura Veldkamp,** New York University
5:00-5:45 PM	"Testing Non-identifying Restrictions"

Marc Henry, Columbia University

5:45 PM Closing Comments

Evan Anderson
Northern Illinois University
Department of Economics
ewanderson@niu.edu

Lars Peter Hansen
University of Chicago
Department of Economics
l-hansen@uchicago.edu

"Small Noise Methods for Risk Sensitive/Robust Economies"

We study small noise expansions for discrete-time infinite horizon control problems with risk sensitivity or equivalently with a concern about robustness. We follow Epstein and Zin (1989) and model the preferences of the decision-maker recursively. As shown by Hansen and Sargent (1995) for linear-quadratic, Gaussian control problems, the recursive formulation of risk sensitivity preserves the tractability of risk-sensitive control theory. The resulting risk-sensitive control problem has a solution that is identical to that of a particular type of robust control problem. Our focus is on using small noise expansions for three different purposes:

- To provide a fast method for solving dynamic stochastic problems.
- To quantify the affect of uncertainty on optimal control laws.
- To quantify how the introduction of risk sensitivity or robustness alters optimal control laws.

Fousseni Chabi-Yo
Bank of Canada
fchabiyo@bankofcanada.ca

"Disentangling the Effects of Heterogeneous Beliefs and Preferences on Asset Prices"

Co-authors: **Eric Ghysels** and **Eric Renault** (University of North Carolina-Chapel Hill)

Steven Durlauf
University of Wisconsin, Madison
Department of Economics
sdurlauf@ssc.wisc.edu

Giacomo Rondina
University of Wisconsin, Madison
Department of Economics
grodina@wisc.edu

"Design Limits and Optimal Policy Evaluation"

This paper has two goals. First, we attempt to characterize fundamental (i.e. unavoidable) tradeoffs between the volatility of state variables at different frequencies in dynamic economic models. This abstract characterization, which is based upon and in turn generalizes a set of interesting results in the control theory literature, will be used to understand how different monetary policy rules engage in frequency by frequency specific stabilization of the economy. Our analysis allows one to compute frequency by frequency Phillips curves (which trade off output and inflation variance at a given frequency) for alternative policy rules and thereby generalize the sorts of inflation/output volatility tradeoffs that are conventionally studied.

Second, we introduce model uncertainty into our analysis to evaluate the frequency by frequency
robustness of rules when such uncertainty is present. With the seminal work of Hansen and Sargent, much recent work in macroeconomics has focused on the analysis of contexts in which economic actors face model uncertainty. Our paper attempts to extend this research in a number of directions using the ideas of fundamental limits. We consider single input/multiple output systems (as occur when the federal funds rate is used to stabilize inflation and output) in which backwards and forward looking elements are present.

Yossi Feinberg
Stanford University
Graduate School of Business
yossi@gsb.stanford.edu

"Non-Bayesian Testing of a Stochastic Prediction"

We propose a method to test a prediction of the distribution of a stochastic process. In a non-Bayesian non-parametric setting, a predicted distribution is tested using a realization of the stochastic process. A test associates a set of realizations for each predicted distribution, on which the prediction passes. So that there are no type I errors, a prediction assigns probability 1 to its test set. Nevertheless, these sets are "small", in the sense that "most" distributions assign it probability 0, and hence there are "few" type II errors. It is also shown that there exists such a test that cannot be manipulated, in the sense that an uninformed predictor who is pretending to know the true distribution is guaranteed to fail on an uncountable number of realizations, no matter what randomized prediction he employs. The notion of a small set we use is category I, described in more detail in the paper. **JEL Classification**: K9

Co-author: **Eddie Dekelyand** (Northwestern University and Tel-Aviv University)

Lars Peter Hansen
University of Chicago
Department of Economics
l-hansen@uchicago.edu

Thomas Sargent
New York University
Department of Economics
ts43@nyu.edu

"Recursive Robust Estimation and Control without Commitment"

In a Markov decision problem with hidden state variables, a posterior distribution serves as a state variable and Bayes' law under an approximating model gives its law of motion. A decision maker expresses fear that his model is misspecified by surrounding it with a set of alternatives that are nearby when measured by their expected log likelihood ratios (entropies). Martingales %under the approximating model represent alternative models. A decision maker constructs a sequence of robust decision rules by pretending that a sequence of minimizing players choose increments to a martingale and distortions to the prior over the hidden state. A risk sensitivity operator induces robustness to perturbations of the approximating model conditioned on the hidden state. Another risk sensitivity operator induces robustness to the prior distribution over the hidden state. We use these operators to extend the approach of Hansen and Sargent (IEEE) to problems that contain hidden states. The worst case martingale is overdetermined, expressing an intertemporal inconsistency of worst case beliefs about the hidden state, but not about observables.

Marc Henry
Columbia University
Department of Economics
mh530@columbia.edu

"Testing Non-Identifying Restrictions"

We propose a conservative testing procedure for the validity of restrictions in a structural model without identifying assumptions. The model is defined as a binary relation between latent and observable variables, coupled with a hypothesized family of distributions for the latent variables. The objective of the testing procedure is to determine whether this hypothesized family of latent variable distributions has a non-empty intersection with the set of distributions compatible with the observable data generating process and the binary relation defining the model.

Martin Schneider
Federal Reserve Bank of Minneapolis
ms1927@nyu.edu

"Ambiguity, Information Quality and Asset Pricing"

When ambiguity-averse investors process news of uncertain quality, they act as if they take a worst-case assessment of quality. As a result, they react more strongly to bad news than to good news. They also dislike assets for which information quality is poor, especially when the underlying fundamentals are volatile. These effects induce ambiguity premia that depend on idiosyncratic risk in fundamentals as well as skewness in returns. Moreover, shocks to information quality can have persistent negative effects on prices even if fundamentals do not change.

Co-author: **Larry Epstein** (University of Rochester)

Raman Uppal
London Business School
Department of Finance
ruppal@london.edu

Lorenzo Garlappi
University of Texas-Austin
Department of Finance
lorenzo.garlappi@mccombs.utexas.edu

"To Hold Familiar Assets or To Diversify? Keynes Meets Markowitz"

Co-authors: **Phelim Boyle** (University of Waterloo) and **Tan Wang** (University of British Columbia)

Laura Veldkamp
New York University-Stern
Department of Economics
lveldkam@stern.nyu.edu

"Information Acquisition and Portfolio Under-Diversification"

We develop a rational model of investors who choose which asset payoffs to acquire information about, before forming portfolios. Scale economies in information acquisition

lead investors to specialize in learning about a set of highly-correlated assets. Knowing more about these assets makes them less risky and more desirable to hold. Benefits to specialization compete with benefits to diversification. The resulting asset portfolios appear under-diversified from the perspective of standard theory, but are optimal. In equilibrium, information is a strategic substitute because assets that many investors learn about have low expected returns. Increasing returns, combined with strategic substitutability leads ex-ante identical investors to specialize in different information, and hold different portfolios. Information choice rationalizes investing in a diversified fund and a set of highly-correlated assets, an allocation observed in the data but usually deemed anomalous.

Co-author: **Stijn Van Nieuwerburgh** (New York University-Stern)

E. *Transition Workshop Program and Abstracts*
February 27-28, 2006

Monday -- February 27, 2006
Radisson Hotel Research Triangle Park
3rd Floor, Room H

8:45-9:15 AM Registration and Continental Breakfast

9:15-9:30 AM	Welcome and Opening Remarks **Jim Berger**, SAMSI **Jean-Pierre Fouque**, University of California-Santa Barbara **Eric Ghysels**, University of North Carolina-Chapel Hill
9:30-10:15 AM	"Risk Measures and Optimal Risk Transfer under Interest Rate Ambiguity" **Nicole El Karoui**, Ecole Polytechnique
10:15-10:30 AM	Coffee Break
10:30-11:15 AM	"Model Free Results on Volatility Derivatives" **Bruno Dupire**, Bloomberg
11:30-12:15 PM	"Explaining Exercise Patterns for Executive Stock Options: Risk Aversion and Costly Exercise" **Vicky Henderson**, Princeton University

12:15-2:00 PM Lunch

2:00-2:45 PM	"Portfolio Choice with a Large Number of Assets: Jumps and Diversification" **Yacine Ait-Sahalia**, Princeton University
2:45-3:00 PM	Coffee Break
3:00-3:45 PM	"A Semiparametric Framework for Modeling and Forecasting Jumps and Volatility in Speculative Prices" **Xin Huang**, Duke University

4:00-4:45 PM "Detecting and Predicting Forecast Breakdowns"
Barbara Rossi, Duke University

Tuesday -- February 28, 2006
Radisson Hotel Research Triangle Park
3rd Floor, Room H

9:00-9:30 AM Continental Breakfast

9:30-10:15 AM "Deterministic Stochastic Optimal Control"
Chris Rogers, Cambridge University

10:15-10:30 AM Coffee Break

10:30-11:15 AM "Recent Results in Liquidity Risk"
Philip Protter, Cornell University

11:30-12:15 PM "Optimal Stopping, Utility Maximisation and Time Consistency"
David Hobson, University of Bath

12:15-2:00 PM Lunch

2:00-2:45 PM "Aggregation of Nonparametric Estimators for Volatility Matrix"
Jianqing Fan, Princeton University

2:45-3:00 PM Coffee Break

3:00-3:45 PM "Sharp Correlation Bounds and Their Applications"
Yanqin Fan, Vanderbilt University

4:00-4:45 PM "Evaluating Value-at-Risk Models with Desk-Level Data"
Denis Pelletier, North Carolina State University

Yacine Ait-Sahalia
Princeton University
Department of Economics
yacine@princeton.edu

"Portfolio Choice with a Large Number of Assets: Jumps and Diversification"

We analyze the portfolio selection problem of an investor facing both Brownian and jump risks. By decomposing the two types of risks on a well-chosen basis, we provide a new methodology for determining the optimal solution in closed form, up to a constant. We show that the optimal solution is for the investor to focus on controlling his exposure to the jump risk, while exploiting differences in the asset returns diffusive characteristics in the orthogonal space. We then examine the solution to the portfolio problem as the number of assets available to the investor increases, and study the asymptotic distribution of the investor's wealth and optimal portfolio.

Joint work with Julio Cacho-Diaz and Tom Hurd.

Bruno Dupire
Bloomberg
Department of Quantitative Research
bdupire@bloomberg.net

"Model Free Results on Volatility Derivatives"

Markets have witnessed a surge of volatility related instruments, which calls for an understanding of the associated pricing and hedging issues. We review the available instruments, then investigate the arbitrage links between the different vehicles to trade volatility before inspecting various pricing models. We then exploit links with the Skorohod embedding problem to obtain lower bounds for the price of options on realized variance. A core feature of this approach is that it makes use of the skew information (dependency of implied volatility on the strike) as opposed to traditional models models that purely pay attention to the variance swap dynamics.

Nicole El Karoui
Ecole Polytechnique
Centre de Mathematiques Appliquees
elkaroui@cmapx.polytechnique.fr

"Risk Measures and Optimal Risk Transfer under Interest Rate Ambiguity"

Monetary risk measures is a way to calculate the amount of cash reserve for hedging risky position. This process requires that the riskiness of the future position and the reserve amount are evaluated in the same num´eraire. In the theory presented in the book of Foellmer and Schied, risky position are "implicitly" discounted to evaluate actual riskiness. Such discounting process presents additional (and different) source of risks, with ambiguity on the different scenarios of interest rates evolution. Similar problems appear when cash-flows are depending on default risk. A suitable measure of risk should account and disentangle the risk deriving from the future financial position pers´e and risk/ambiguity deriving from the discount factor. To do that, we relax the assumption of cash invariance into sub-additive cash invariance, and give a dual characterization of these new functionals in terms of sub-linear probability measures. Classical problems as optimal risk transfer or optimal hedging are studied in this new context, using inf-convolution technics in static or dynamic framework.

Joint work with Claudia Ravanelli (ISB, University of Zurich, Switzerland)

Jianqing Fan
Princeton Univesity
Department of Operations Research and Financial Engineering
jqfan@princeton.edu

"Aggregation of Nonparametric Estimators for Volatility Matrix"

An aggregated method of nonparametric estimators based on time-domain and state-domain estimators is proposed and studied. To attenuate the curse of dimensionality, we propose a factor modeling strategy. We first investigate the asymptotic behaviors of

nonparametric estimators of the volatility matrix in the time domain and in the state domain. The asymptotic normality is separately established for nonparametric estimators in the time domain and state domain. These two estimators are asymptotically independent. Hence, they can be combined, through a dynamic weighting scheme, to improve the efficiency of the estimated volatility matrix. The optimal dynamic weights are derived and it is shown that the aggregated estimator uniformly dominates the volatility matrix estimators using time-domain or state-domain smoothing alone. A simulation study, based on an essentially affine model for the term structure, is conducted and it demonstrates convincingly that the newly proposed procedure outperforms both time- and state-domain estimators. Empirical studies endorse further the advantages of our aggregated method.

Yanqin Fan
Vanderbilt University
Department of Economics
yanqin.fan@vanderbilt.edu

"Sharp Correlation Bounds and Their Applications"

In this paper, we establish asymptotic properties, including the consistency and asymptotic normality, of nonparametric estimators of the sharp bounds on the correlation between two random variables. We demonstrate both theoretically and numerically that the sharp bounds may differ from the traditionally used bounds [-1,1] and the nonparametric estimators of the sharp bounds shed light on the strength of the type of dependence, linear or nonlinear, between two random variables. To facilitate inference on the true sharp bounds, we provide easy-to-compute estimators of the asymptotic variances of the nonparametric estimators of the sharp bounds. Using the sharp correlation bounds on the unobserved covariates, we derive sharp bounds on the correlation of durations in bivariate hazard rate models with unobserved heterogeneity and the correlation of dependent variables in bivariate log-linear regression models with unobserved covariates. These results provide insight on the selection of distributions of the unobserved heterogeneity in bivariate hazard rate models and unobserved covariates in log-linear regression models.

Vicky Henderson
Princeton University
Department of Operations Research and Financial Engineering
vhenders@princeton.edu

"Explaining Exercise Patterns for Executive Stock Options: Risk Aversion and Costly Exercise"

David Hobson
University of Bath
Department of Operations Research and Financial Engineering
dhobson@princeton.edu

"Optimal Stopping, Utility Maximisation and Time Consistency"

Xin Huang
Duke University

Department of Economics
xinhuang@duke.edu

"A Semiparametric Framework for Modeling and Forecasting Jumps and Volatility in Speculative Prices"
Based on the recent theoretical results of non-parametric jump detection, we turn the latent jump process into observable realized measure and separate it from the continuous volatility. We build models for different components of the jump process, which can be combined with the existing model for the continuous volatility to forecast future return variability better. Using more than a decade of high-frequency five-minute returns for futures contracts on equity and fix-income, we detect time varying jump intensity, serial correlation in jump occurrence and jump size, in contract to most exiting empirical results. Evidence from simulated data confirms such serial correlation to be bona-fide, and not induced by the noise in the jump detection test statistics. Our model for the jump duration effectively purges such serial correlation in the jump occurrence process, and brings the unconditional distribution of the durations closer to the exponential distribution. In view of the big contribution from the overnight return variance to the total return variance, we construct an augmented GARCH model for the overnight returns, and find significant GARCH pattern, and strong asymmetric predictive power from the same-day trading-time continuous volatility.

Denis Pelletier
North Carolina State University
Department of Economics
denis_pelletier@ncsu.edu

"Evaluating Value-at-Risk Models with Desk-Level Data"

We present new evidence on disaggregated profit and loss and VaR forecasts obtained from a large international commercial bank. Our dataset includes daily P/L generated by four separate business lines within the bank. All four business lines are involved in securities trading and each is observed daily for a period of at least two years. We also collected the corresponding daily, 1-day ahead VaR forecasts for each business line. Given this rich dataset, we provide an integrated, unifying framework for assessing the accuracy of VaR forecasts. Our approach includes many existing backtesting techniques as special cases. In addition, we describe some new tests which are suggested by our framework. A thorough Monte Carlo comparison of the various methods is conducted to provide guidance as to which of these many tests have the best finite-sample size and power properties.

Philip Protter
Cornell University
Department of Operations Research
pep4@cornell.edu

"Recent Results in Liquidity Risk"

Classical theories of financial markets assume an infinitely liquid market and that all traders act as price takers. This theory is a good approximation for highly liquid stocks, although even there it does not apply well for large traders or for modelling transaction costs. We extend the classical approach by formulating a new model that takes into account illiquidities. This turns out to be important for hedging options and for calculating the consequent liquidity, in the course of following a hedging strategy to replicate an option (approximately). Our approach hypothesizes a stochastic supply curve for a security's price as a function of trade size. This leads to a new definition of a self-financing trading strategy, and additional restrictions on hedging strategies. Book data provided by Morgan Stanley is used to test the model, and to describe the supply curves of highly liquid, somewhat liquid, and relatively illiquid stocks.

Chris Rogers
University of Cambridge
Statistics Laboratory
l.c.g.rogers@statslab.cam.ac.uk

"Deterministic Stochastic Optimal Control"

This paper approaches optimal control problems for discrete-time controlled Markov processes by representing the value of the problem in a dual Lagrangian form. This approach is a completely novel way to look any stochastic optimal control problem, independent of (but complementing) the classical dynamic-programming/value-function approach. The representation obtained opens up the possibility of numerical methods based on Monte Carlo simulation which may be advantageous in high-dimensional problems, or in problems with complicated constraints.

Barbara Rossi
Duke University
Department of Economics
brossi@econ.duke.edu

"Detecting and Predicting Forecast Breakdowns"

We propose a theoretical framework for assessing whether a forecast model estimated over one period can provide good forecasts over a subsequent period. We formalize this idea by defining a forecast breakdown as a situation in which the out-of-sample performance of the model, judged by some loss function, is significantly worse than its in-sample performance. Our framework, which is valid under general conditions, can be used not only to detect past forecast breakdowns but also to predict future ones. We show that main causes of forecast breakdowns are instabilities in the data generating process and relate the properties of our forecast breakdown test to those of existing structural break tests. The main differences are that our test is robust to the presence of unstable regressors and that it has greater power than previous tests to capture systematic forecast errors caused by recurring breaks that are ignored by the forecast model. As a by-product, we show that our results can be applied to forecast rationality tests and provide the appropriate asymptotic variance estimator that corrects the size distortions of previous forecast rationality tests. The empirical application finds evidence of a forecast breakdown in the Phillips' curve forecasts of U.S. inflation over the past three decades, and links it to inflation volatility and to changes in the monetary policy reaction function of the Fed.

III. ASTROSTATISTICS

A. *Planning Workshop Agenda*

July 14-16, 2005
LOCATION: The meeting will be at NASA-AMES (see directions below). Those needing a hotel should immediately reserve a room (rooms will be held for only a short while longer) at the Larkspur Landing Home Suite Hotel in Sunnyvale, California. See www.larkspurhotels.com/larkspur/locations.html for more information. To make a reservation, call 408-733-1212 and mention that you are attending the SAMSI meeting.

Thursday will be devoted primarily to scientific discussion, including learning about the wide variety of research interests of the participants. Each participant will have roughly 30 minutes in which to describe his/her interests or applications, although a significant portion of this time should be reserved for questions and discussion. As this is a *planning meeting*, it is of less importance to present research results than to give an indication as to the importance of an area and the types of problems that might be particularly appropriate for the SAMSI program to consider. Friday morning will be devoted mostly to discussion of the SAMSI program itself, especially discussion of potential participants and the planning of workshops and events that will be held during the program.

Thursday, July 14

8:30 AM A van and cars will leave from the hotel to NASA-AMES.

9:00-9:15 AM Welcome by Jim Berger, Jogesh Babu, and Jeff Scargle

9:15-9:45 AM Louis Lyons – "Setting Limits on Rates of Physical Processes in Particle Physics"

9:45-10:15 AM David van Dyk – "Topics in Astrostatics"

10:15-10:45 AM Fabrizia Guglielmetti – "Background-Sources Separation in Xray Astronomical Images"

10:45-11:15 AM Coffee Break

11:15-11:45 AM Jeff Scargle – "Optimal Segmentation of Data"

11:45-12:15 PM Discussion of Potential Scientific Themes – Merlise Clyde, Moderator

12:15-1:15 PM Lunch

1:15-1:45 PM Tom Loredo – "Topics in Astrostatistics"

1:45-2:15 PM Fionn Murtagh – "Status Report on COST Action"

2:15-2:45 PM Phil Gregory – "Bayesian Extrasolar Planet Detection"

2:45-3:15 PM Coffee Break

3:15-3:45 PM Andrew Connolly – "Variability in Time Domain Surveys or Clustering and the Time Domain with Massive Astronomical Data Sets"

3:45-4:15 PM	Bill Jefferys – "White Dwarfs, Galaxies, and Hierarchical Bayes"
4:15-4:45 PM	Donald Richards – "Incomplete Data"
4:45-5:30 PM	Discussion of Potential Scientific Themes – Larry Wasserman, Moderator

Friday morning, July 15

9:00-10:00 AM	Goals, Structure and Resources of the Program – Jim Berger
10:00-10:30 AM	Discussion of possible participants – Alanna Connors, moderator
10:30-11:00 AM	Coffee Break
11:00-11:30 AM	Tutorials and Workshops – Jogesh Babu, moderator
11:30-12:00 PM	Wrap-up On Potential Scientific Themes – Peter Bickel, Moderator
12:00-	Lunch followed by a one-hour tour of the new Visualization Facility (one of the top visualization facilities in the country). *Let Jeff Scargle know if you will be attending this tour, as he needs to make preparations!*

Participants:

Jogesh Babu, Program Leader; Penn State University (babu@stat.psu.edu)
James Berger, Director of SAMSI (berger@samsi.info)
Peter Bickel, NAC Co-Chair; Univ. Calif. Berkeley (bickel@stat.berkeley.edu)
Floyd Bullard, Statistics, Duke University (floyd@stat.duke.edu)
Merlise Clyde, Statistics, Duke University (clyde@stat.duke.edu)
Alanna Connors, Eureka Scientific (aconnors@comptel.sr.unh.edu)
Andrew Connolly, Astronomy, Carnegie-Mellon University (ajc@tiamat.phyast.pitt.edu)
Phil Gregory, Astronomy, Univ. British Columbia (gregory@physics.ubc.ca)
Fabrizia Guglielmetti, Astronomy, Max-Planck Institute (fag@ipp.mpg.de)
Bill Jefferys, Astronomy, University of New Hampshire (bill@astro.as.utexas.edu)
Tom Loredo, Astronomy, Cornell University (loredo@astro.cornell.edu)
Louis Lyons, Physics, Oxford University (l.lyons1@physics.ox.ac.uk)
Fionn Murtagh, Computer Science, University of London (fionn@cs.rhul.ac.uk)
Don Richards, Statistics, Penn State Univ. (richards@stat.psu.edu)
Jeff Scargle, Astronomy, NASA-AMES (jeffrey@cosmic.arc.nasa.gov)
Megan Sosey, Space Telescope Science Institute (sosey@stsci.edu)
David van Dyk, Univ. California, Irvine (dvd@ics.uci.edu)
Larry Wasserman, Statistics, Carnegie-Mellon Univ. (larry@stat.cmu.edu)

Transportation and Directions:

From airports to hotel: Those flying into San Jose airport can just take a shuttle or taxi to the Larkspur Landing Hotel. Those flying into San Francisco can take the Super Shuttle (www.supershuttle.com/htm/cities/sfo.htm) or share taxis with others flying in at the same time. (If you are flying in on Wednesday at around 5:30, contact Jeff Scargle for a possible ride.)

From hotel to NASA-Ames/Moffett Field: North on Mathilda 0.3 miles to US 101, take the North bound entrance. After 2 miles, take the Moffett Field/NASA Ames exit (this is one exit PAST the Moffett South Gate exit). This puts you on Moffett Blvd. headed to the Main Gate. At or just before the guards at the gate, turn right into a small parking lot in front of the Visitor Registration and Pass Office. Park and enter this office to get a badge.

From other locations to NASA-Ames: The directions are pretty much the same. If you are coming south on 101, get off at the exit marked Moffett Exits (it is one past the Highway 85 exit) which puts you on Moffett Blvd. as above. There are professional directions at: http://www.nasa.gov/centers/ames/home/Directions.html

B. *Opening Tutorials and Workshop Programs and Abstracts*
January 18-25, 2006

TUTORIAL 1: *BAYESIAN ASTROSTATISTICS*
Leader **-- Tom Loredo,** Cornell University

Wednesday – January 18, 2006
NISS Building, Room 104

8:30-9:00 AM	Registration and Continental Breakfast
9:00-9:15 AM	Welcome **Jim Berger**, SAMSI
9:15-10:30 AM	Fundamentals **Tom Loredo**, Cornell University
10:30-10:45 AM	Coffee Break
10:45-12:00 PM	Basic Counting and Point Process Models **Tom Loredo**, Cornell University
12:00-1:30 PM	Lunch
1:30-2:45 PM	Probability and Frequency **Tom Loredo**, Cornell University
2:45-3:00 PM	Coffee Break
3:00-5:00 PM	Statistical Computing, with Python I **Tom Loredo**, Cornell University

Thursday – January 19, 2006
NISS Building, Room 104

8:45-9:15 AM	Continental Breakfast
9:15-10:30 AM	Models with Gaussian Uncertainties I **Philip Gregory**, University of British Columbia
10:30-10:45 AM	Coffee Break
10:45-12:00 PM	Models with Gaussian Uncertainties II **Philip Gregory**, University of British Columbia
12:00-1:30 PM	Lunch
1:30-2:45 PM	Basic Bayesian Computation **Tom Loredo**, Cornell University
2:45-3:00 PM	Coffee Break
3:00-5:00 PM	Statistical Computing, with Python II **Tom Loredo**, Cornell University

Friday – January 20, 2006
NISS Building, Room 104

8:45-9:15 AM	Continental Breakfast
9:15-10:30 AM	Introduction to Markov Chain Monte Carlo, with R **William Jefferys**, Universities of Texas and Vermont
10:30-10:45 AM	Coffee Break
10:45-12:00 PM	Hierarchical Bayesian Modeling, with R **William Jefferys**, Universities of Texas and Vermont
12:00-1:30 PM	Lunch
1:30-2:45 PM	Model Selection and Model Averaging, with R **William Jefferys**, Universities of Texas and Vermont
2:45-3:00 PM	Coffee Break
3:00-4:15 PM	Parallel Chain MCMC, with Mathematica **Philip Gregory**, University of British Columbia
4:15-4:30 PM	Distribute and Complete Feedback Forms
4:30-5:30 PM	Bayesian Miscellany **Tom Loredo**, Cornell University

TUTORIAL 2: *NONPARAMETRIC INFERENCE*
Leaders – **Chad Schafer** and **Larry Wasserman,** Carnegie Mellon University
January 21-22, 2006

Saturday – January 21, 2006
NISS Building, Room 104

9:30-10:00 AM Registration and Continental Breakfast

10:00-12:00 PM Lecture and Practicum 1

12:00-2:00 PM Lunch

2:00-4:00 PM Lecture and Practicum 2

Sunday – January 22, 2006
NISS Building, Room 104

9:30-10:00 AM Continental Breakfast

10:00-12:00 PM Lecture and Practicum 3

12:00-2:00 PM Lunch

2:00-4:00 PM Lecture and Practicum 4

TUTORIAL 3: *ASTRONOMY FOR STATISTICIANS*
Leaders – **Eric Feigelson**, Pennsylvania State University
William Jefferys, Universities of Texas and Vermont
January 21-22, 2006

Saturday – January 21, 2006
NISS Building, Room 203

8:30-9:00 AM Registration and Continental Breakfast

9:00-10:00 AM Movie and Overview
William Jefferys, Universities of Texas and Vermont

10:00-11:00 AM Solar System
William Jefferys, Universities of Texas and Vermont

11:00-11:30 AM Coffee Break

11:30-12:00 PM Exoplanets
Eric Feigelson, Pennsylvania State University

12:00-1:30 PM Lunch

1:30-2:00 PM Star and Planet Formation
Eric Feigelson, Pennsylvania State University

2:00-3:00 PM Stellar Physics and HR Diagram
William Jefferys, Universities of Texas and Vermont

3:00-3:30 PM Coffee Break

3:30-4:30 PM Stellar Evolution
William Jefferys, Universities of Texas and Vermont

Sunday – January 22, 2006
NISS Building, Room 203

8:30-9:00 AM Continental Breakfast

9:00-10:00 AM Multiwavelength Telescopes and Surveys
Eric Feigelson, Pennsylvania State University

10:00-10:45 AM Milky Way and Stellar Populations
William Jefferys, Universities of Texas and Vermont

10:45-11:00 AM Coffee Break

11:00-12:00 PM Galaxies (including Clustering and AGN)
Eric Feigelson, Pennsylvania State University

12:00-1:30 PM Lunch

1:30-2:30 PM Classic Cosmology
Eric Feigelson, Pennsylvania State University

2:30-3:00 PM Recent Developments in Cosmology
Eric Feigelson, Pennsylvania State University

3:00-3:15 PM Coffee Break

3:15-3:45 PM Discussion
Eric Feigelson, Pennsylvania State University
William Jefferys, Universities of Texas and Vermont

3:45-4:30 PM Summary
Eric Feigelson, Pennsylvania State University

Astrostatistics Opening Workshop and Tutorials
January 23-25, 2006

Monday – January 23, 2006
Radisson Hotel Research Triangle Park
3rd Floor, Room H

8:30-9:00 AM Registration and Continental Breakfast

9:00-9:10 AM	Welcome and Introductions **Jim Berger**, SAMSI **G. Jogesh Babu**, Pennsylvania State University
9:10-11:55 AM	Session on Source Detection Chair: **David van Dyk**, University of California, Irvine
9:10-9:55	"Detection of Gamma Ray Sources" **Brenda Dingus**, Los Alamos National Laboratory
9:55-10:40	"Star Counts and Galaxy Properties from Surveys: Past, Current and Future Trends" **Martin Weinberg**, University of Massachusetts
10:40-11:10	Coffee Break
11:10-11:55	Panel/Group Discussion Panelists: **Brenda Dingus**, Los Alamos National Laboratory **Vinay Kashyap**, Smithsonian Astrophysical Observatory **David van Dyk**, University of California, Irvine **Martin Weinberg**, University of Massachusetts **Rebecca Willett**, Duke University
11:55-12:30 PM	Introduction to Poster Presentations
12:30-1:30 PM	Lunch
1:30-3:45 PM	Session on Exoplanets Chair: **William Jefferys**, Universities of Texas and Vermont
1:30-2:15	"Searches for Radio Pulsars and Planets Around Them" **Alex Wolszczan**, Pennsylvania State University
2:15-3:00	"Analysis of Radial Velocity and Astrometric Signals in the Detection of Multi-Planet Extrasolar Planetary Systems" **Barbara McArthur**, University of Texas
3:00-3:45	Panel/Group Discussion Panelists: **William Jefferys**, Universities of Texas and Vermont **Michael Lavine**, Duke University **Barbara McArthur**, University of Texas **Alex Wolszczan**, Pennsylvania State University
3:45-4:15 PM	Coffee Break
4:15-5:15 PM	**5-Minute Madness** Anyone who wishes can give a 5-minute talk describing their research and interests, an interesting problem, or whatever!

Transparencies or slides are limited to four, and slides must be put into the session computer over the break.

6:30-8:30 PM **Poster Session and Reception**
Radisson Hotel Research Triangle Park, 2nd Floor

SAMSI will provide poster presentation boards and removable adhesive squares. The board dimensions are 4 ft. wide by 3 ft. high. They are tri-fold with each side being 1ft. wide and the center 2 ft. wide. Please make sure your poster fits the board. The boards can usually accommodate up to 16 pages of letter-sized paper (landscaped).

Poster Presenters: Report to the 2nd floor by 6pm to setup your poster. We will also have materials available if you would like to put your slides on the poster boards during lunch.

Tuesday – January 24, 2006
Radisson Hotel Research Triangle Park
3rd Floor, Room H

8:30-9:00 AM Continental Breakfast

9:00-11:45 AM Session on Surveys
Chair: **Tom Loredo**, Cornell University

9:00-9:45 "Large Optical Surveys"
David Wittman, University of California, Davis

9:45-10:30 "Censoring and Truncation in Astronomical Surveys"
Eric Feigelson, Pennsylvania State University

10:30-11:00 Coffee Break

11:00-11:45 Panel/Group Discussion
Panelists: **Eric Feigelson**, Pennsylvania State University
Tom Loredo, Cornell University
Larry Wasserman, Carnegie Mellon University
David Wittman, University of California, Davis

11:45-12:15 PM Introduction to Particle Physics
Louis Lyons, University of Oxford

12:15-1:15 PM Lunch

1:15-3:30 PM Session on Particle Physics
Chair: **Louis Lyons**, University of Oxford

1:15-2:00 "Bayesian Inside the Gate: A View from Particle Physics"
Harrison Prosper, Florida State University

2:00-2:45 "Some Current Statistical Considerations in Particle Physics"
Byron Roe, University of Michigan

2:45-3:30 Panel/Group Discussion
Panelists: **Peter Bickel**, University of California, Berkely
Louis Lyons, University of Oxford
Harrison Prosper, Florida State University
Byron Roe, University of Michigan

3:30-4:00 PM Coffee Break

4:00-5:15 PM New Researchers Session

- "Statistics for Astronomy: Inference about the Population of Quasars"
  Angela Snyder-Hugeback, University of Chicago
- "An Approach to Detecting Non-Gaussianity in the Cosmic Microwave Background"
  Kristofer Jennings, Purdue University
- "Measuring Clustering of Absorbers"
  Ji Meng Loh, Columbia University
- "Statistical Modeling of Sunspot Cycles"
  Yaming Yu, University of California, Irvine

Wednesday – January 25, 2006
Radisson Hotel Research Triangle Park
3rd Floor, Room H

8:30-9:00 AM Continental Breakfast

9:00-11:45 AM Session on Gravitational Lensing
Chair: **Arlie Petters**, Duke University

9:00-9:45 "Stochastic Gravitational Lensing and the Nature of Dark Matter"
Charles Keeton, Rutgers University

9:45-10:30 "Gravitational Lensing: Mass Reconstruction Methods and Results"
Liliya Williams, University of Minnesota

10:30-11:00 Coffee Break

11:00-11:45 Panel/Group Discussion
Panelists: **Charles Keeton**, Rutgers University
Arlie Petters, Duke University
Liliya Williams, University of Minnesota
Zhengyuan Zhu, University of North Carolina-Chapel Hill

11:45-1:30 PM Formation of Discussion Groups & Discussion Over Lunch

1:30-3:45 PM Session on Other Topics
Chair: **Alanna Connors**, Eureka Scientific

1:30-2:15 "Estimating the Proportion of False Null Hypotheses"
John Rice, University of California, Berkeley

2:15-3:00 "Nonparametric Inference in Astrophysics and Cosmology: Biases and Variants"
Christopher Genovese, Carnegie Mellon University

3:00-3:45 Panel/Group Discussion

Panelists:
Keith Arnaud, NASA Goddard Space Flight Center.
Alanna Connors, Eureka Scientific.
Christopher Genovese, Carnegie Mellon University.
John Rice, University of California, Berkeley

3:45-4:15 PM Coffee Break

4:15-5:15 PM Discussion Group Reports and Working Group Formation

Brenda Dingus
Los Alamos National Laboratory
dingus@lanl.gov

"Detection of Gamma Ray Sources"

Eric Feigelson
Pennsylvania State University
Department of Astronomy & Astrophysics
edf@astro.psu.edu

"Censoring and Truncation in Astronomical Surveys"

A common mode of studying astronomical populations is to scan the sky in a waveband and develop a flux-limited (i.e. truncated) catalog of sources. Numerous such catalogs now exist each with around 5-20 measured quantities for logN=4-9 objects. Major statistical questions include classification into different types of objects (e.g. stars, quasars, galaxies), estimation of distances, and establishment of the distributions of luminosities (luminosity functions, LFs). A serious challenge is treating heteroscedastic measurement errors with known variances for each cell in the multivariate database. When catalogs at different wavebands are merged, many cells have nondetections which, when measurement errors are considered, can be treated as left-censored values. This creates another methodological challenge: to analyze multivariate databases with censoring in all variables.

Christopher Genovese
Carnegie Mellon University
Department of Statistics
genovese@cmu.edu

"Nonparametric Inference in Astrophysics and Cosmology: Biases and Variants"

The goal of nonparametric methods is to make sharp inferences about unknown functions with a minimum of assumptions. The explosion of available astronomical and cosmological data in recent years has made new and sophisticated nonparametric methods both practical and effective. Using two cosmological problems as examples -- inference for the Cosmic Microwave Background spectrum and for the Dark Energy equation of state -- I will review recent developments in nonparametric inference and illustrate some key considerations for constructing good procedures. I will discuss ways to determine the appropriate amount of smoothing and will analyze the "biases" that affect inferential performance. I will show how to construct confidence sets for the function of interest and discuss recent results on how confidence sets adapt to unknown smoothness.

Kristofer Jennings
Purdue University
Department of Statistics
jennings@stat.purdue.edu

"An Approach to Detecting Non-Gaussianity in the Cosmic Microwave Background"

Charles Keeton
Rutgers University
Department of Physics and Astronomy
keeton@physics.rutgers.edu

"Stochastic Gravitational Lensing and the Nature of Dark Matter"

The gravitational deflection of light provides a unique probe of dark matter around galaxies. Anomalies in the properties of observed gravitational lens systems are thought to reveal small-scale structure in galaxy dark matter halos, and constrain the physical nature of the (still mysterious) dark matter. With this inspiration, we are developing an exciting synergy between astrophysics, mathematics, and statistics. To create the only rigorous method of identifying the "anomalies," we have extended catastrophe theory to new limits and applied it to real astrophysical data. To properly describe the physical phenomenon, which is inherently stochastic, we are building a probabilistic theory of gravitational lensing. The theory raises new mathematical issues, such as the roots of random polynomials. We will then use our theory to develop new statistical tools for analyzing the astrophysical data. The goal of this talk is to present the astrophysical motivation and discuss why a multi-disciplinary approach is both crucial and compelling.

Ji Meng Loh
Columbia University
Department of Statistics
meng@stat.columbia.edu

"Measuring Clustering of Absorbers"

Barbara McArthur
University of Texas at Austin
Department of Astronomy
mca@barney.as.utexas.edu

"Analysis of Radial Velocity and Astrometric Signals in the Detection of Multi-Planet Extrasolar Planetary Systems"

Harrison Prosper
Florida State University
Department of Physics
harry@hep.fsu.edu

"Bayesian Inside the Gate: A View from Particle Physics"

During the past decade, Bayesian methods have been deployed with considerable success in a few high-profile analyses in particle physics. This, along with the field's ongoing, concerted, effort to understand statistical issues better, has led to the narrowing of the frequentist/Bayesian divide and the recognition that both sides have useful things to say. In this talk, after a brief account of why Bayesian methods are being embraced, I illustrate how Bayesian methods have been, and are being used, in particle physics, and I end with a statement of outstanding issues that would benefit from a broad discussion.

John Rice
University of California, Berkeley
Department of Statistics
rice@stat.berkeley.edu

"Estimating the Proportion of False Null Hypotheses"

I will discuss the problem of estimating the proportion of false null hypotheses among a very large number of independently tested null hypotheses, focusing on the scenario in which the proportion is very small. Concrete motivation comes from the Taiwanese-American Occultation Survey of the Kuiper Belt.

Byron Roe
University of Michigan
Department of Physics
byronroe@umich.edu

"Some Current Statistical Considerations in Particle Physics"

Two of the statistical problems that have recently concerned particle physicists are discussed. The first problem is the question of nuisance parameters, those parameters which are not the subject of the experiment, but must considered to get to the parameter(s) of interest. The second problem is that of classification, using a number of variables, called here particle identification variables (PID) to separate signal from background in data. The emphasis for classification will be on modern methods, successors to neural nets.

Angela Snyder-Hugeback
University of Chicago
Department of Statistics
angiehugeback@gmail.com

"Statistics for Astronomy: Inference about the Population of Quasars"

I am interested in the problem of density estimation when there is a known bias in the data collection procedure. This interest is motivated by a research problem that was brought to my attention by Sebastian Jester, formerly of the Experimental Astrophysics Group at FermiLab and currently a researcher at Oxford. My advisor, Marc Coram, and I have been working with Sebastian to estimate the density of quasars at various redshift distances from Earth. A major complicating factor in our analysis is that the quasars in our dataset were not selected uniformly at random from the population all quasars in the universe. Rather, the selection probabilities for each quasar are dependent on that quasar's apparent brightness when viewed by a telescope from Earth as well as on its redshift distance from Earth. To date we have constructed a likelihood-based parametric model for the number density of quasars, in terms of the absolute optical magnitude and the redshift distance from earth.

In my thesis work, I would like to further develop our applied research by expanding our model to include additional "extended" quasars, as well as to incorporate both the optical and radio luminosity of each quasar into the model. More broad issues that I would like to explore include assessing goodness of fit, building more flexible parameterizations, and penalizing to avoid overfitting.

Martin Weinberg
University of Massachusetts
Department of Astronomy
weinberg@astro.umass.edu

"Star Counts and Galaxy Properties from Surveys: Past, Current and Future Trends"

The increasing wealth of astronomical data promises detailed constraints on scientific theories. I will describe two representative statistical problems motivated by the scientific promise of the 2MASS database: inferring Milky Way structure from star counts and galaxy properties from atlas images. The standard approach for each along with challenges and pitfalls will be outlined. Recent work at UMass attempts to repose these as Bayesian inference problems and we have developed a parallel MCMC-based software platform for this purpose. This system is designed to overcome many of the computational difficulties that pose a barrier common usage. I will illustrate our successes and failures as a case study of applied statistical computation in the real world (and look forward to suggestions from experts at this meeting). Finally, I will throw out a few thoughts on new directions for knowledge acquisition from survey data.

Liliya Williams
University of Minnesota
Department of Astronomy
llrw@astro.umn.edu

"Gravitational Lensing: Mass Reconstruction Methods and Results"

Knowledge of the detailed distribution of dark matter in galaxies and clusters of galaxies can provide clues about the physical nature of dark matter. Strong gravitational lensing, whereby distant sources are multiply-imaged by foreground galaxies and clusters is widely used to reconstruct sky-projected mass distribution in these objects. Lensing mass reconstruction using primarily the observed properties of multiply-imaged sources is an example of an under-constrained problem, and there is a variety of specific methods in use. I will discuss parametric and non-parametric techniques, error estimation, as well as some recent results concerning dark matter.

David Wittman
University of California, Davis
Department of Physics
dwittman@physics.ucdavis.edu

"Large Optical Surveys"

Many of today's astrophysical questions, such as the nature of dark energy, require massive data sets to answer, and the astronomy community is fostering ever larger surveys in response. I will review current and future large optical surveys: their scientific motivation, implementation, datasets, and challenges. I will particularly focus on the largest of them all, the proposed Large Synoptic Survey Telescope (LSST). LSST will repeatedly survey the entire sky visible from its site, generating a several hundred petabyte dataset with a time dimension in addition to the usual dimensions of a static survey. This survey has incredible potential for discovery, and also offers numerous challenges for astrostatistics.

Alex Wolszczan
Pennsylvania State University
Department of Astronomy
alex@astro.psu.edu

"Searches for Radio Pulsars and Planets Around Them"

I will review the methodology of pulsar surveys and searches for neutron star planets with the precision pulse timing method.

Yaming Yu
University of California, Irvine
Department of Statistics
yamingy@ics.uci.edu

"Statistical Modeling of Sunspot Cycles"

Poster Abstracts:

Ethan Anderes
University of California, Berkeley
Department of Statistics
anderes@stat.berkeley.edu

"Estimating Deformations of Isotropic Gaussian Random Fields on the Plane"

We present a new approach to the estimation of the deformation of an isotropic Gaussian random field on $\Bbb R^2$ based on dense observations of a single realization of the deformed random field. Under this framework we investigate the identification and estimation of deformations. We also present a complete methodological package---from model assumptions to algorithmic recovery of the deformation---for the class of non-stationary processes obtained by deforming isotropic Gaussian random fields.

Kinman Au
Carnegie Mellon University
Department of Statistics
kinman@stat.cmu.edu

"Inferring Galaxy Morphology Through Texture Analysis"

We give an approach to estimate galaxy morphology from digital images. In particular, our algorithm extracts orientation information of the texture at difference scales, and merges the multiscale information into an unified representation. By fitting a morphological model based on the textural information, we derive an quantitative and physically meaningful description of galaxy morphology. Such description will help scientists to study how galaxy morphology evolve over time, and the effect of environment toward the evolution. The answers will provide important clues about the origin of the Universe.

Antonio Cava
Milano University & INAF-OAPD
Department of Physics
cava@pd.astro.it

"The Wide-field Imaging Nearby Galaxy Survey"

Peter Driscoll
San Francisco State University
Department of Physics and Astronomy
driscoll@stars.sfsu.edu

"A Comparison of Least-Squares and Bayesian Techniques in Fitting the Orbits of Extrasolar Planets"

Eric Ford
University of California, Berkeley
Department of Astronomy
eford@berkeley.edu

"Developing a Bayesian Toolbox for Detection and Orbit Determination of Extrasolar Planets"

Radial velocity surveys have now detected over 150 extrasolar planets around nearby main sequence stars. Many of these planets are clearly detected and have well characterized orbits, thanks to a large ratio of the velocity amplitude to measurement precision and observations spanning many orbital periods. However, a growing number of planets have orbital periods comparable to the duration of observations and/or induce radial velocity variations not much larger than the measurement precision (Fig. 1). For such planets, there are often large uncertainties in the orbital parameters. In the most extreme cases, even establishing the reality of a periodic signal is difficult. These difficulties become even more severe for multiple planet systems which require simultaneously fitting numerous model parameters. So far, most analyzes of extrasolar planets have relied on frequentist methods such as maximum likelihood. I review recent progress in developing the necessary computational tools for implementing such analyzes. I demonstrate these techniques with a Bayesian analysis of a recent triple planet system orbiting HD 37124.

Peter Freeman
Carnegie Mellon University
Department of Statistics
pfreeman@cmu.edu

"Non-Parametric Analysis of Supernova Data and the Dark Energy Equation of State"

Philip Gregory
University of British Columbia
Department of Physics and Astronomy
gregory@phas.ubc.ca

"A Bayesian Multi-Planet Kepler Periodogram for Exoplanet Detection"

Woncheol Jang
Duke University
Institute of Statistics and Decision Sciences
wjang@stat.duke.edu

"Density Estimation and Clustering in Astronomical Sky Surveys"

William Jefferys
Universities of Texas and Vermont
Department of Astronomy (UT) & Statistics (UVM)
bill@astro.as.utexas.edu

"A Bayesian Approach to Analyzing Star Cluster Parameters"

Hyunsook Lee
Pennsylvania State University
Department of Statistics
hlee@stat.psu.edu

Title: TBA

Jayanta Pal
University of Michigan
Department of Statistics

jpal@umich.edu

"Estimating a Decreasing Density for the Dark Matter in Nnearby Dwarf Galaxies"

Taeyoung Park
Harvard University
Department of Statistics
tpark@stat.harvard.edu

"Fitting Narrow Emission Lines in X-ray Spectra"

Alex Rojas
Carnegie Mellon University
Department of Statistics
arojas@stat.cmu.edu

"Characterization of Galaxy Evolution as function of Local Environment"

Adam Roy
University of California, Irvine
Department of Statistics
ajroy@uci.edu

"Highly Structured Models and Statistical Computation in High-Energy Astrophysics"

David Valls-Gabaud
Canada-France-Hawaii Telescope
dvg@cfht.hawaii.edu

"Genetic Algorithms for Gravitational Lenses"

Diana Yanchukova and **Don Ellison**
North Carolina State University
Department of Physics
dyanchu@ncsu.edu

"High Energy Photon Emission in Young Supernova Remnants"

Young SNRs are believed to produce cosmic ray ions and electrons, but direct evidence for ion acceleration in SNRs remains illusive. An important key to the solution of this problem concerns the relative efficiency for producing inverse Compton radiation vs. gamma-ray production via pion decay. We outline elements of this problem and compare results to recent H.E.S.S. TeV observations of young SNRs.

Ruth-Stella Barrera-Rojas
Universidad Nacional de Colombia
Observatorio Astronomico Nacional
barrera_ruth@yahoo.com

"Multivariate Analysis of Gamma-Ray Bursts from BATSE 4B"

We present here a multivariate analysis of gamma-ray burst (GRB) properties for discriminating between distinct classes of GRBs. We are applying the methodology proposed by (Mukherjee, Feigelson et al, 1998), so we use a multivariate clustering procedure on a sample of 1637 bursts from the Forth BATSE Catalog, this is a parametric maximum likelihood model-based clustering procedure assuming multinormal populations calculated with the EM Algorithm and validated with the Bayesian Information Criterion, the software for the calculations is EMMIX. Joint work with Antonio Uribe.

Floyd Bullard
Duke University
Institute of Statistics and Decision Sciences
floyd@stat.duke.edu

"Finding Smaller Exoplanets"

Arbitrarily small exoplanets can be detected using radial velocity measurements even when the velocity measurements' margins of error are substantially greater than the velocity function's amplitude. The profile log likelihood function will always show a clear peak at the true period with enough observations.

Hyunsook Lee
Pennsylvania State University
Department of Statistics
hlee@stat.psu.edu

"Convex Hull Peeling: Nonparametric Multivariate Data Analysis Tools"

An ad hoc device on multidimensional massive data is in demand. However, multivariate data analysis tools not imposing multivariate normal distribution exist rarely. We introduce convex hull peeling algorithms as a such device for the analysis of multidimensional massive data. Only the convexity of data sets is assumed. These convex hull peeling algorithms are designed to estimate quantiles, detect outliers, and measure distribution shapes of multidimensional data. Additionally, the algorithms are exemplified with Monte Carlo simulations and SDSS DR4 Quasars.

Taeyoung Park
Harvard University
Department of Statistics
tpark@stat.harvard.edu

"Efficient X-ray Spectral Fitting with Narrow Emission Lines"

From a statistical point of view, spectral analysis is the modeling of the distribution of photon energies, a distribution that can be formulated as a finite mixture of two photon groups, a continuum term and a set of emission lines. While the continuum describes a general shape of a spectrum, each emission line represents a positive aberration from the continuum in a narrow band of energies. Here, we focus on a single emission line that can be modeled with a Gaussian distribution or a delta function. Spectral data are contaminated by several non-trivial physical processes including non-homogeneous stochastic censoring, blurring of photon energies, and background contamination. To account for these processes, we consider a hierarchical structure of missing data under a Bayesian perspective. To fit the resulting highly structured multilevel spectral models, we devise efficient Gibbs sampling strategies. As an illustration, we apply our strategies to the X-ray spectrum of the high redshift quasar, PG 1634+706.

IV. EDUCATION AND OUTREACH

A. *PREP Workshop on Mathematics Meets Biology: Epidemics, Data Fitting and Chaos Program* -- May 25-28, 2005

Wednesday – May 25, 2005
University of Louisiana

7:45 AM	Bus departs hotel to workshop
8:00-8:45 AM	Continental Breakfast
8:45-9:00 AM	Welcoming Remarks **Dean Clark**
9:00-10:15 AM	An Overview of HIV and modeling in general, derivation of specific HIV equations
10:15-11:50 AM	Introduction to Matlab, model simulation, examine model behavior and steady states
11:50-1:30 PM Lunch	
1:30-2:00 PM	Development of Models
2:00-2:45 PM	Introduction to inverse problems
3:00-4:30 AM	Experiments on inverse problems
4:45 PM	Bus departs workshop for hotel

Thursday – May 26, 2005
University of Louisiana

8:00 AM	Bus departs hotel to workshop
8:15-9:00 AM	Continental Breakfast
9:00-10:45 AM	Lecture on MCMC methods including worksheets and examples, hands-on experience
10:45-11:00 AM	Coffee Break
11:00-11:50 AM	Experiments
11:50-1:30 PM Lunch	

1:30-5:00 PM Tour of swamps

Friday – May 27, 2005
University of Louisiana

8:00 AM Bus departs hotel to workshop

8:15-9:00 AM Continental Breakfast

9:00-9:50 AM Review of classical differential equation models

9:50-10:00 AM Coffee Break

10:00-10:50 AM Recent models on competitive exclusion and coexistence

10:50-11:00 AM Coffee Break

11:00-11:50 AM Hands-on experience for participants

11:50-1:30 PM Lunch

1:30-4:30 PM continue on projects

4:45 PM Bus departs workshop for hotel

Saturday – May 28, 2005
University of Louisiana

8:00 AM Bus departs hotel to workshop

8:15-9:00 AM Continental Breakfast

9:00-9:50 AM Review of difference equation models

9:50-10:00 AM Coffee Break

10:00-12:00 PM Hands-on experience for participants

B. *SAMSI-CRSC Undergraduate Workshop*
May 30-June 3, 2005

Monday – May 30, 2005

8:30-9:15 AM Breakfast

9:15-9:30 AM Welcome
H.T. Banks, North Carolina State University

9:30-10:30 AM Introduction and Background

H.T. Banks

10:30-10:45 AM Break

10:45-11:45 AM Introduction to the *forward problem:* solving the harmonic oscillator equation

11:45-12:45 PM Transportation to SAMSI

12:45-1:30 PM Lunch at SAMSI

1:30-4:00 PM Presentation of Latent Variable in the Social Sciences
Jane Zavisca, SAMSI

4:00 PM Depart SAMSI for Lake Crabtree

5:00 PM Dinner at Lake Crabtree

Tuesday – May 31, 2005

9:00-10:15 AM Brief introduction to the computing system and Matlab

10:15-10:30 AM Break

10:30-12:00 PM Basic statistical concepts and some probability essentials.

12:00-1:00 PM Lunch

1:00-2:45 PM Linear inverse problems: a Matlab tutoiral

2:45-3:15 PM Break

3:15-5:00 PM Statistical view of linear least squares: a Matlab tutorial

Wednesday – June 1, 2005

9:00-11:30 AM *Vibrating beam* data collection at CRSC laboratory

11:30-12:30 PM Lunch

12:30-1:45 PM Reflection on the data collection and modeling experiences

1:45-2:00 PM Break

2:00-3:30 PM Solving the *vibrating beam* inverse problem

3:30-3:45 PM Break

3:45-5:00 PM Teams work on their inverse problem

Thursday – June 2, 2005

9:00-10:30 AM Statistical analysis for the *vibrating beam* inverse problem

10:30-10:45 AM Break

10:45-12:00 PM Teams work on their inverse problem

12:00-1:00 PM Lunch

1:00-2:30 PM What could we do better? Alternative models/statistical methods.

2:30-3:00 PM Break

3:00-4:30 PM Teams work on their inverse problem and begin to prepare reports

Friday – June 3, 2005

9:00-10:30 AM Presentations and discussion

10:30-10:45 AM Break

10:45-11:45 AM Presentations and discussion

11:45-12:00 PM Closing remarks and workshop evaluations

12:00-1:00 PM Lunch

C. SAMSI-CRSC Industrial Mathematical & Statistical Modeling Workshop for Graduates Program and Problem Abstracts
July 25-August 2, 2005

Sunday – July 24, 2005

Participants arrive at NC State University

Monday – July 25, 2005
Harrelson Hall, NC State University

9:00-9:15 AM **Opening Remarks**
Prof. Mansoor Haider

Prof. Hien Tran Associate Head, Department of Mathematics, NCSU

Prof. Ralph Smith Associate Director, Statistical & Applied Mathematical Sciences Inst. Associate Director, Center for Research in Scientific Computation, NCSU

9:20-9:40 AM	*Problem 1 – Optimal Mission Planning* Greg Angelides, MIT Lincoln Lab
9:45-10:05 AM	*Problem 2 – Mathematical Modeling of Enzyme Action in Fuel Alcohol Production* Eric Allain, Guillermo Coward-Kelly, Novozymes Inc.
10:10-10:30 AM	*Problem 3 – Control of Dynamic System Response with Directional Forcing Sources* Anna Howard, Lord Corporation
10:30-10:45 AM	Break & Refreshments – Harrelson Hall, Room 245
10:50-11:10 AM	*Problem 4 – Studying Mathematical Models of Signal Transduction: Model Reduction and Parameter Observability* Anna Georgieva, Novartis Pharma
11:15-11:35 AM	*Problem 5 – Power Preallocation and Control in a Heavy Traffic Queueing Model for Wireless Systems* Priya Ranjan, Intelligent Automation
11:40-12:00 PM	*Problem 6 – First-order Modeling of a Fluid Pump Driven by a Piezoceramic Disk Actuator.* William Ott, PAR Technologies
12:05-12:25 PM	*Problem 7 – Modeling and Forecasting of Electric Power Load.* Mike Settlage, Glen Snider, Progress Energy

Tuesday – July 26, 2005
Harrelson Hall, NC State University

8:30-5:00 PM Working Session

Wednesday – July 27, 2005
Harrelson Hall, NC State University

8:30-5:00 pm Working Session

Thursday – July 28, 2005
Harrelson Hall, NC State University

8:30-5:00 pm Working Session

Friday – July 29, 2005
Harrelson Hall, NC State University

8:30-12:00 pm Working Session

12:00-1:00 pm Lunch

1:00-5:00 pm Tour of Centennial Campus & Math Lab

Saturday – July 30, 2005
Harrelson Hall, NC State University

8:30-11:00 am Working Session

11:00 am Free Afternoon

Sunday – July 31, 2005
Harrelson Hall, NC State University

Free Day

Monday – August 1, 2005
Harrelson Hall, NC State University

8:30-5:00 pm Working Session

Tuesday – August 2, 2005
Harrelson Hall, NC State University

8:30-12:30 pm Working Session

12:30-1:30 pm Lunch

1:30-5:00 pm Formal Presentation of Results

Wednesday – August 3, 2005

Departure of participants from NC State University

Problem Abstracts:

Problem 1. **Optimal Mission Planning**

Presenter: Greg Angelides
Lincoln Laboratory
Massachusetts Institute of Technology
244 Wood Street
Lexington, MA 02420-9108
(781) 981-5500
gregangelides@ll.mit.edu

Team Members: Andy Bartlett, Andre Berger, Craig Lipkin, Nsoki Mavinga, Elizabeth Perez, Eamonn Tweedy, Erik Wheeler

Faculty Consultants: Zhilin Li, Amy Langville

Working Session: **Room 130** Harrelson

Planning the route of U.S. aircraft through hostile air spaces is an important problem for all of the military services. Many countries possess integrated air defense systems, networks of early warning radars and other sensors, data fusion centers and surface-to-air missile (SAM) systems whose integrated function is to detect, track and direct fire on adversary aircraft. Mission planners attempt to route through these threats, minimizing radar tracking and SAM engagement opportunities while obeying constraints on aircraft speed, altitude and turn capability, mission range, mission time, fuel, deconfliction with other aircraft and locations of ingress and egress from the threat region. Complicating this problem is the fact that mission planners must do their routing with incomplete and uncertain information on threat number and location, and must complete their route determination for multiple aircraft in a timely manner (typically much less than a day). This project will evaluate the types of approaches most applicable to this problem, including classical cost minimization techniques, neural networks, genetic algorithms, combinations of these techniques and any other techniques that the team chooses to consider. The team will initially research an optimal approach, assuming that the disposition of the threat is known. The team will then expand its analysis to consider the impact of threat uncertainty, first assuming that the locations of the threats are known to within a given geolocation uncertainty, then including incompleteness in the threat by assuming that each threat has a certain likelihood of being in historically determined locations. Time permitting, the team will then consider the general case of mission planning in the face of uncertainties in threat number and location.

Problem 2. Mathematical Modeling of Enzyme Action in Fuel Alcohol Production

Presenters: Eric J. Allain, Ph.D.
Guillermo Coward-Kelly, Ph.D.
Novozymes North America, Inc.
77 Perry Chapel Church Rd.
Franklinton, NC 27525
(919) 494-3195
GCK@novozymes.com

Team Members: Mihaela Froehlich, Mary Gruber, Kamyar Hazaveh, April Stepp, Chris Vogl, Xiaohai Wan, Xiaofeng Xia

Faculty Consultants: Mette Olufsen, Charlie Smith

Working Session: **Room 136** Harrelson

The production of ethanol from renewable resources (corn, wheat, cellulosic biomass etc.) is of great interest to countries all around the world. This is especially true in the US where the availability of feedstocks for fuel ethanol production could reduce dependence on foreign oil. Ethanol is also a more ecologically sound fuel than gasoline both in terms of being renewable as well as being cleaner burning. All of this has created a rapidly growing fuel alcohol industry as well as highly active research in this area. However, if ethanol is to be economically competitive with gasoline, its production process must be highly efficient. Enzymes play a key role in this regard by catalytically hydrolyzing feedstock polymers into glucose which can then be converted to ethanol by fermentation. As such, a great deal of effort has been aimed at developing highly efficient, stable enzymes that are able to work in the conditions required for fuel alcohol production. At Novozymes, finding and engineering improved enzymes for fuel alcohol production is pursued through two complementary paths 1) a high throughput screening approach and

2) a rational approach based on our understanding of how enzymes are working in the process. Recently, we have been focusing on the use of mathematical modeling as a tool to help us in the second approach.

A model has been proposed to describe the enzymatic hydrolysis of granular starch to glucose. The model is based on Michaelis-Menten kinetic equations describing the rate of change of five chemical species present in the reaction. Problems and questions regarding this model include:

- We currently don't have a convenient method to determine the concentration of one of the chemical species (available starch non-reducing ends). Is there a way to make this model useful without this information?
- At the beginning of the reaction, the substrate (starch) is in granular form. This means that most of the bonds are unavailable for enzymatic attack. As the reaction proceeds more and more of these bonds become available. What is the best way to take this into account in the model?

It is fairly easy for us to get dynamic data at various enzyme dosages for most of the chemical species of interest. Is there a better model we could use to take advantage of this fact?

Problem 3. Control of Dynamic System Response with Directional Forcing Sources.

Presenter: Anna Howard, Ph.D.
Lord Corporation
Thomas Lord Research Center
110 Lord Dr.
Cary, NC 27511
(919) 469-2500
akthoward@gmail.com

Team Members: Robert Benim, Mike Bowman, John Gonzalez, Emek Kose, Morteza Mahyari, Wei Zhang

Faculty Consultants: Dmitry Zenkov, Kazi Ito

Working Session: **Room 221** Harrelson

Adaptive control of dynamic system response to disturbances using actuators and sensors is fairly well understood. For example, such systems are in production to control vibration and noise on various aircraft. This project will expand upon such systems by considering the additional degree-of-freedom of directionality of the actuator force. Participants will be asked to explore the existance and uniqueness of an optimal solution, and then to propose an adaptive algorithm such as gradient descent that will seek the optimal solution. Early focus will be on simple systems and then move toward generalization.

Problem 4. Studying Mathematical Models of Signal Transduction: Model Reduction and Parameter Observability

Presenter: Anna Georgieva, Ph.D.

Modeling and Simulation, Biology
Novartis Pharma
One Health Plaza
East Hanover, NJ 07936-1080
(862) 778-8300
anna.georgieva@novartis.com

Team Members: Ananda Bandulasiri, Chad Gonzales, Janine Haugh, Bing Jian, Sarah Lynn Joyner, Nesli Saglanmak, Thobile Thukani, Guohua Zhou

Faculty Consultants: Sharon Lubkin, Alina Chertock

Working Session: **Room 346** Harrelson

Cellular signal transduction pathways are characterized by their high complexity. Mathematical models describing these processes might be of great help to gain qualitative and, most importantly, quantitative knowledge about such complex systems. Detailed mathematical description of such systems leads to very large systems of ordinary differential equations. The situation becomes even more complicated when we consider two or more pathways simultaneously. Therefore, it becomes necessary to simplify the models, but in such a way that the reduced models still capture essential dynamics. In addition, due to the sparseness of appropriate experimental data, a subset of the model parameters needs to be determined using mathematical optimization. We are interested to find the best possible way to characterize the accuracy/observability of the model parameters. Models of signal transduction will be considered in the context of biomarker identification for drug therapy. In particular, model predictions will be used to contribute to better understanding of molecular mechanisms of cancer.

Problem 5. Power Preallocation and Control in a Heavy Traffic Queueing Model for Wireless Systems

Presenter: Priya Ranjan, Ph.D.
Intelligent Automation, Inc.
15400 Calhoun Drive, Suite 400
Rockville, MD 20855
(301) 294-5200
pranjan@i-a-i.com

Team Members: Norou Diawara, Xiaoying Han, Bret Hanlon, Chuan Lin, Menaka Navaratna, Amith Vijayat, Anastasia Wong

Faculty Consultants: Robert Buche, Jim Zhang

Working Session: **Room 168** Harrelson

The heavy traffic method for queueing analysis assumes the system is operating at near capacity in the sense (loosely) that there is “just enough” control resources to balance the mean arrivals rate and departure rate. Under this assumption, the method obtains via weak limits an approximation to the queuing dynamics, modeled as a stochastic differential equation. The method has been used with wireline systems and shows promise for wireless systems; the main difference being the random operating environment in wireless systems.

We will consider a heavy traffic wireless model where the control resource is power. In particular the total power available can be divided into two types of uses: most of the total will be *preallocated* power used for balancing the mean arrivals and the remaining will be *reserve* power used to handle the stochastic variations about the mean dynamics due to, for example, the variations in the arrival process to the queues and channel process. Given a fixed amount of the total power used for the preallocated power, the preallocation to the individual queues in a particular channel state is not uniquely determined from balancing the mean dynamics. Furthermore, we wish to consider being able to increase the amount used for preallocated power so that we can increase the mean arrival rates. In both cases we wish to consider optimality criteria (e.g. in the latter case, the objective function reflects the benefit of increasing the arrival rates), obtaining the preallocated power for the (queue state, channel state)-components via (static) optimization.

When setting the level of the total power to be assigned as preallocated power, one must also consider, of course, that this affects the amount of reserve power available for handling the stochastic variations. The effect of the reserve power is quantified in a cost function of an associated stochastic control problem using the stochastic differential equation modeling the queueing dynamics obtained under the heavy traffic method. This cost can be computed numerically.

Finally, it is hoped that some heuristics and possibly quantifications for assigning the amount of the total power to be used for preallocation can be obtained.

Problem 6. First-Order Modeling of a Fluid Pump Driven by a Piezoceramic Disk Actuator

Presenter: William Ott
PAR Technologies LLC
1000 Lucas Way Suite B
Hampton, VA 23666
(757) 766-4525
WOtt@partechinc.com

Team Members: Chandia Affane Aji, Tom Braun, Seshadri Damu, Thomas Deems, Ingunn Gunnarsdottir, Sidharth Rupani, Necibe Tuncer

Faculty Consultants: Ralph Smith, William Oates

Working Session: **Room 124** Harrelson

One of the applications of the PAR RLP™ disk actuator is using the actuator as the driving diaphragm in liquid or gas pumps. A typical PAR pump consists of a disk actuator that is mounted along its periphery and a system of one way valves. Since this is a resonant device it is important to be able to predict the resonant frequency of the pump including any effect of the external system to which it is attached.

An important tool in the design of these devices is an accurate model that will predict the system resonant frequency and describe the fluid flow as a function of a prescribed diaphragm deflection. This model would allow quick comparisons to be made which can explore the effects on flow of the variation in geometric and material parameters of the pump and its components. The pump model should include the effects of flow through the valves as well as backflow due to mismatches in the timing of the inlet and outlet valves. The model should also include the effects of backpressure on the flow. Data will be available to verify the model developed.

Problem 7. Modeling and Forecasting of Electric Power Load

Presenters: Glen Snider
Mike Settlage
Progress Energy
P.O. Box 1551
Raleigh, NC 27602-1551
(919) 546-2154
glen.snider@pgnmail.com

Team Members: Jennifer Geis, Zheng Li, Danny Modlin, Eduardo Osorio, Min-Hyong Park, Jyothsna Prabhakaran, Troy Tingey

Faculty Consultants: Tao Pang, Denis Pelletier

Working Session: **Room 375** Harrelson

Progress Energy is a Fortune 250 diversified energy company with more than 24,000 megawatts of generation capacity. Company holdings include two electric utilities serving more than 2.8 million customers in North Carolina, South Carolina, Georgia, and Florida. The Competitive Commercial Operations Department manages long term electric and gas contracts with other utilities, municipalities, marketing companies, and co-ops. This management involves various financial products (options, forwards, futures, swaps) for power generation from six merchant generation plants for several major customers including Duke Energy and Florida Power and Light. Participants in the power market include gas producers, pipeline companies, local delivery companies, consumers and marketers. Consumption in the power market is broken down as roughly 38% by industrial users, 37% by residential users, and the rest for generation of electricity. Residential consumption is heaviest during winter months as reflected by the statistic that, according to NYMEX data, 43% of natural gas is consumed through December to the next March.

The goals of the workshop project are to:

- Analyze historical weather and electric power load data with mathematical tools (e.g., regression/time series methods) to explore the correlation between power load and: Area or area combinations. Time periods. Weather conditions.
- Verify proposed correlation with historical data and provide relating quality indicators, e.g., R^2.
- Simulate/forecast future power load and compare with existing load models to include: An hourly simulation that could forecast the next 24 hours of load (T + 24). A simulation of hourly load for the next seven days (T + 168).

- Research power load forecasting techniques and apply varying time series and regression modeling techniques to project next day (up to next week) load consumption based on forecasted weather.
- Present modeling results for varying techniques providing a recommended model that has the best predictive power given all models and functional forms tested as part of the project.

D. *Undergraduate Two-Day Workshop on National Defense and Homeland Security*
March 3-4, 2006

Friday – March 3, 2006
NISS-SAMSI Building, Room 104

8:45 AM	Carolina Livery Shuttle departs Hawthorn Suites
9:00-9:30 AM	Arrival at SAMSI and Continental Breakfast
9:30-10:00 AM	Welcome and Introduction to SAMSI **Alan Karr**, Director – National Institute of Statistical Sciences (NISS), Associate Director of SAMSI
10:00-10:30 AM	Session on Social Networks **David Banks**, Duke University **Negash Medhin**, North Carolina State University **Hoan Nguyen**, North Carolina State University **Eric Vance**, Duke University
10:30-11:00 AM	Coffee Break
11:00-12:00 PM	Social Networks (continued)
12:00-1:00 PM Lunch	
1:00-2:30 PM	Session on Agricultural Systems **Ping Bai**, University of North Carolina-Chapel Hill **Sava Dediu**, SAMSI and North Carolina State University **Anjela Govan**, North Carolina State University
2:30-3:00 PM	Coffee Break
3:00-4:30 PM	Session on Anomaly Detection **David Dickey**, North Carolina State **Francisco Vera**, NISS and SAMSI
4:30 PM	Carolina Livery Shuttle back to Hawthorn Suites
5:30-7:30 PM	Pizza Party in Hawthorn's Breakfast Room

Saturday – March 4, 2005
NISS-SAMSI Building, Room 104

8:30 AM	Carolina Livery Shuttle departs Hawthorn Suites
8:45-9:30 AM	Arrival at SAMSI and Breakfast
9:30-11:30 AM	Session on Data Confidentiality **Alan Karr**, National Institute of Statistical Sciences
11:30-12:00 PM	Wrap-up
12:00 PM	Adjournment and Departure Carolina Livery will depart SAMSI for the airport and the Hawthorn.

E. *Undergraduate Two-Day Workshop on Financial Mathematics, Statistics and Econometrics Program* November 18-19, 2005

Friday – November 18, 2005
NISS-SAMSI Building, Room 104

9:15 AM	Carolina Livery Shuttle departs Candlewood Suites
9:30-10:00 AM	Arrival at SAMSI and Continental Breakfast
10:00-10:10 AM	Welcome and Introduction to SAMSI **Ralph Smith**, North Carolina State University Associate Director of CRSC and SAMSI
10:10-11:00 AM	Introduction to Financial Mathematics and Discrete Trees **Jean-Pierre Fouque**, North Carolina State University
11:00-12:00 PM	Binomial Tree Models **Jeff Scroggs**, North Carolina State University
12:00-1:00 PM Lunch	
1:00-2:20 PM	Financial Markets and Empirical Regularities **Mike Aguilar**, University of North Carolina at Chapel Hill
2:20-2:40 PM	Coffee Break
2:40-4:20 PM	Portfolios **Mingxin Xu**, University of North Carolina at Charlotte
4:30 PM	Carolina Livery Shuttle back to Candlewood Suites
5:20 PM	Carolina Livery Shuttle to Park Diner
5:30-7:30 PM	Dinner Party at Park Diner 5410 Hwy 55 East, Durham, NC
7:30 PM	Carolina Livery back to Candlewood Suites

Saturday – November 19, 2005
NISS-SAMSI Building, Room 104

8:30 AM	Carolina Livery Shuttle departs Candlewood Suites
8:45-9:15 AM	Arrival at SAMSI and Breakfast
9:15-10:30 AM	Credit Risk **Jean-Pierre Fouque**, North Carolina State University
10:30-10:45 AM	Coffee Break
10:45-12:00 PM	Credit Risk -- *continued*
12:00 PM	Adjournment and Departure

V. LATENT VARIABLES IN THE SOCIAL SCIENCES

A. *GLAMM Seminars Program & Abstracts*

April 13-15, 2005

The seminars will be held 1:00pm until 5:00pm at the National Institute of Statistical Sciences (NISS) Building, Room 104. There will be a short break halfway through the session.

Instructors: **Sophie Rabe-Hesketh**, University of California Berkeley
Anders Skrondal, London School of Economics

Registered participants:

Jim Berger, SAMSI
Ken Bollen, University of North Carolina
Lloyd Edwards, University of North Carolina
Subhashis Ghosal, North Carolina State University
Jiezhun Gu, North Carolina State University
John Hipp, University of North Carolina
Aki Kamata, Florida State University
Alan Karr, NISS
Saki Kinney, Duke University
Negash Medhin, North Carolina State University
Hoan Nguyen, SAMSI
Jesus Palomo, SAMSI
John Samuels, North Carolina State University
Ingmar Visser, University of Amsterdam

B. *Working Group Year-End Summaries*

May 19, 2006

Thursday – May 19, 2005
Radisson Hotel RTP, Room FG (3rd floor)

8:30-9:00 am Registration and Continental Breakfast

9:00-10:00 am Categorical Observed Variable Group

10:00-10:15 am Coffee Break

10:15-11:15 am Multilevel and Structural Equation Models Group

11:15-11:30 am Coffee Break

11:30-12:30 pm Longitudinal Data Group

12:30-1:45 pm Lunch

1:45-2:45 pm Model Uncertainty Group

2:45-3:00 pm Coffee Break

3:00-4:00 pm Social Networks Group

4:00-5:00 pm Planning of Fall Conference

C. *Transition Workshop Program*
November 10-11, 2005

Thursday – November 10, 2005
Radisson Hotel Research Triangle Park, Room FG

8:30-9:00 AM Registration and Continental Breakfast

9:00-9:15 AM Welcome and Introduction
Jim Berger, SAMSI
Kenneth Bollen, University of North Carolina at Chapel Hill

9:45-10:45 AM **Multilevel Structural Equation Models**

- "Overview of Existing Approaches for Multilevel Measurement Models" **Akihito Kamata**, Florida State University
- "Limitations of Current Approaches and Possible Directions for Multilevel Measurement Models" **Daniel Bauer**, University of North Carolina at Chapel Hill
- "Latent Effect Modeling" **Yasuo Miyazaki**, Virginia Polytechnic Institute and State University

10:45-11:00 AM Coffee Break

11:00-12:15 PM **Keynote Presentation**
"Maximal Reliability of Unit-Weighted Composites"
Peter Bentler, University of California Los Angeles

12:15-1:15 PM Lunch

1:15-2:00 PM **Longitudinal Models**

- "Computing Confidence Intervals for Predicting New Observations in the Linear Mixed Model" **Lloyd Edwards**, University of North Carolina at Chapel Hill. **Gary Williamson**, MetaMetrics, Inc.
- "Determinants of Nursing Home Regulatory Activity in the 50 States: An Analysis from the Political Economy Perspective". **Christopher Kelly**, University of North Carolina at Chapel Hill. **Lloyd Edwards**, University of North Carolina at Chapel Hill

2:00-2:30 PM **Categorical Dependent Variables with Error in Covariates**

"A Consistent Instrumental Variable Estimator for Errors in Covariates in Limited Dependent Variable Models"
Kenneth Bollen, University of North Carolina at Chapel Hill. **Roland Thomas**, Carleton University. **Liqun Wang**, University of Manitoba

2:30-2:45 PM Coffee Break

2:45-3:45 PM **Latent Class Analysis**

- "Estimating the Level of Underreporting of Expenditures Among Expenditure Reporters: A Further Micro-Level Latent Class Analysis" **Brian Meekins**, Bureau of Labor Statistics
- "Latent Class Analysis of Rotation Group Bias: The Case of Unemployment" **Jane Zavisca**, University of Arizona
- "Model-Based Estimation of Drug Use$_4$ Prevalence Using Item Count Data" **Paul Biemer**, RTI International and University of North Carolina at Chapel Hill

3:45-5:00 PM **Keynote Presentation**
"Some Simple Latent-Structure Models, Based on Rasch-Type Latent-Trait Models and on Latent-Class Models, for the Analysis of Cross-Classified Data" **Leo Goodman**, University of California Berkeley

Friday – November 11, 2005
Radisson Hotel Research Triangle Park
Room FG

8:00-8:30 AM Registration and Continental Breakfast

8:30-9:30 AM **Social Networks**

- "Optimal Noise Variance of a Social Network Continuum Model" **Hoan Nguyen**, North Carolina State University. **H.T. Banks**, North Carolina State University
- "Agent-Based Methods for Dynamic Social Networks" **Eric Vance**, Duke University. **David Banks**, Duke University
- "Non-Linear Programming Methods for Dynamic Social Networks" **Chung-Chien Hong**, North Carolina State University

9:30-10:30 AM Model Uncertainty – Part I

- "Bayesian Model Selection and Averaging in Structural Equation Models" **David Dunson**, National Institute of Environmental Health Sciences and Duke University. **Jesus Palomo**, National Institute of Environmental Health Sciences **Jane Zavisca**, University of Arizona
- "Bayesian Covariance Selection in Nonparametric Random Effects and Generalized Linear Mixed Models" **Bo Cai**, National Institute of Environmental Health Sciences. **David Dunson**, National Institute of Environmental Health Sciences and Duke University

10:30-10:45 AM Coffee Break

10:45-11:45 AM Model Uncertainty – Part II

- "A Scaled Unit Information Prior Approximation to the Bayes Factor"
  Kenneth Bollen, University of North Carolina at Chapel Hill.
  Surajit Ray, University of North Carolina at Chapel Hill.
  Jane Zavisca, University of Arizona
- "Generalized Bayesian Information Criterion" **Ingmar Visser**, University of Amsterdam. **Surajit Ray**, University of North Carolina at Chapel Hill

11:45-12:45 PM Panel Discussion:
"Latent Variable Models: Unanswered Questions"
Peter Bentler, University of California Los Angeles
Kenneth Bollen, University of North Carolina at Chapel Hill.
Leo Goodman, University of California Berkeley

VI. GENOMES TO GLOBAL HEALTH: COMPUTATIONAL BIOLOGY OF INFECTIOUS DISEASES

A. *Transition Workshop Program*
May 22-24, 2005

Sunday--May 22, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

8:30-9:00 AM Registration and Continental Breakfast.

9:00-9:15 AM Welcome
Jim Berger, SAMSI

9:15-9:30 AM Collective Intelligence
Tom Kepler, Duke University Medical Center

9:30-10:30 AM The Tropical Disease Initiative
Stephen Maurer, University of California Berkeley

10:30-10:45 AM Break

10:45-11:30 AM The Science of the Tropical Disease Initiative
Marc-Marti-Renom, University of California San Francisco

11:30-12:45 PM Lunch

12:45-1:35 PM The Malaria Capers
Bob Desowitz, University of North Carolina

1:35-2:25 PM Genome-Wide Inferences Of Recombination, positive And Negative Selection IN THE Agent Of Malaria, Plasmodium Falciparum-Towards A Malaria
Phillip Awadalla, North Carolina State University

2:25-2:40 PM Break

2:40-3:30 PM Proteomics for Malaria Drug Vaccine Target Discovery
Tim Haystead, Duke University

3:30-4:20 PM PlasmoDB: The Plasmodium Genome Resurce
Jessica Kissinger, University of Georgia

4:20-5:10 PM Utilizing the Plasmodium falciparum Genome to Combat Malaria
Rachael Isokpehi, Jackson State University

5:10-6:00 PM Future Direction for Malaria Research
Victoria McGovern, Burroughs-Welcome Foundation

6:00 PM **Closing Remarks**
The Organizers

Monday – May 23, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

8:30-9:00 AM Registration and Continental Breakfast

9:00-10:45 AM Legal and Social Aspects of Malaria Vaccine Development

10:45-11:00 AM Coffee Break

11:00-12:45 PM Legal and Social Aspects *(continued)*

12:45-2:00 PM Lunch

2:00-3:45 PM Scientific and Technical Aspects of Malaria Vaccine Development

3:45-4:00 PM Coffee Break

4:00-5:45 PM Scientific and Technical Aspects *(continued)*

5:45-6:00 PM Closing Remarks

Tuesday – May 24, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

8:30-9:00 AM Registration and Continental Breakfast

9:00-10:45 AM Legal and Social Aspects of Malaria Drug Development

10:45-11:00 AM Coffee Break

10:45-12:00 PM Legal and Social Aspects *(continued)*

12:45-2:00 PM Lunch

2:00-3:45 PM Scientific and Technical Aspects of Malaria Drug Development

3:45-4:00 PM Coffee Break

4:00-5:45 PM Scientific and Technical Aspects *(continued)*

5:45-6:00 PM Closing Remarks

VII. DATA ASSIMILATION FOR GEOPHYSICAL SYSTEMS

A. *Summer School on Fusing Geophysical Models with Data: From Theory to Practice to Theory*
June 13-17, 2005

Sunday – June 12, 2005

7:00-9:00 PM Reception at the Millennium

Monday – June 13, 2005

8:00 AM	University of Colorado (CU) bus will depart promptly to transport participants from the Millennium to NCAR
8:30-10:00 AM	Practical issues for data assimilation and weather forecasting. **Xiang-Yu (Hans) Huang**, NCAR
10:00-10:30 AM	Break
10:30-12:00 PM	Statistics for Model/Data Fusion **Chris Wikle**, University of Missouri
12:00-12:15 PM	Informal Discussion
12:15-1:15 PM	Lunch
1:15-2:45 PM	DART Tutorial
2:45-3:30 PM	Break and Afternoon vignette: **Chris Snyder**, NCAR
3:30-4:45 PM	DART Tutorial
5:00 PM	CU Bus back to the Millennium

Tuesday – June 14, 2005

8:10 AM	CU Bus departs for NCAR
8:30-10:00 AM	Building Climate Models. **Andrew Gettleman**, NCAR
10:00-10:30 AM	Break
10:30-12:00 PM	Estimating parameters in dynamical systems **Jonathan Stroud**, University of Pennsylvania
12:00-12:15 PM	Informal Discussion
12:15-1:15 PM	Lunch
1:15-2:45 PM	DART Tutorial
2:45-3:30 PM	Break and Afternoon vignette: **Peter Sullivan**, NCAR
3:30-4:45 PM	DART Tutorial
5:00 PM	CU Bus back to the Millennium

Wednesday – June 15, 2005

8:10 AM	CU Bus departs for NCA

8:30-9:45 AM	Targeting observations **Carolyn Reynolds**, Naval Research Laboratory
9:45-10:15 AM	Break
10:15-11:45 AM	Adaptive observation and observing system design **Shree Khare**, SAMSI
11:45-12:15 PM	Informal Discussion
12:15-1:15 PM	Lunch
1:15-2:45 PM	DART Tutorial
2:45-3:30 PM	Break and Afternoon vignette: **Doug Nychka**, NCAR
3:30-4:45 PM	DART Tutorial
5:00 PM	CU Bus back to the Millennium

Thursday – June 16, 2005

8:10 AM	CU Bus departs for NCAR
8:30-10:00 AM	Particle filters and ensembles **Thomas Bengtsson**, University of California, Berkeley
10:00-10:30 AM	Break
10:30-12:00 PM	De-Fusing perfect model expectation from theory in practice. **Leonard Smith**, University of Oxford
12:00-12:15 PM	Informal Discussion
12:15-1:15 PM	Lunch
1:15-2:45 PM	DART Tutorial
2:45-3:30 PM	Break and Afternoon vignette
3:30-4:45 PM	DART Tutorial
5:00 PM	CU Bus to the Millennium

Friday – June 17, 2005

8:10 AM	CU Bus departs for NCAR
8:30-10:00 AM	Microtalks

10:00-10:30 AM Break

10:30-11:15 AM Breakout groups

11:15-12:15 PM Group reports

B. *Mini-Workshop on Bridging Statistical Approaches and Sequential Data Assimilation Program*
June 27, 2005

Monday – June 27, 2005
NISS Building, Room 104

9:30-10:00 AM Continental Breakfast and Welcome

10:00-11:00 AM *Introduction to Statistical Data Assimilation*
Mark Berliner, Ohio State University

11:00-12:00 PM *Data Assimilation Through Particle Filters*
Mike Kouritzin, University of Alberta

12:00-1:00 PM Lunch

1:00-2:00 PM *Data Assimilation Using Multi-Resolution Spatio-Temporal Models*
Noel Cressie, Ohio State University

2:00-2:30 PM Coffee Break

2:30-3:30 PM *An Ensemble Kalman Filter for Real-Time Assimilation of Satellite Data*
Jonathan Stroud, University of Pennsylvania

3:30-4:00 pm Open Discussion

C. *Transition Workshop Program and Abstracts*
October 5, 2005

Wednesday – October 5, 2005
NISS Building, Room 104

8:00-8:30 AM Registration and Continental Breakfast

8:30-8:45 AM Welcome
Jim Berger, SAMSI

Opening Remarks
Christopher Jones, SAMSI and University of North Carolina at Chapel Hill

8:45-9:30 AM	"Data Assimilation in the Earth Sciences: Building on and Going Beyond Classic Estimation Theory" **Dennis McLaughlin**, Massachusetts Institute of Technology
9:30-10:15 AM	15-Minute Talks
9:30-9:45	"Analysis and Prediction of a Noisy Nonlinear Ocean" **Robert N. Miller**, Oregon State University
9:45-10:00	"Estimation of Clouds in Atmospheric Models" **Tomi Vukicevic**, Colorado State University
10:00-10:15	"Incorporation of Dynamic Balance in Data Assimilation and Application to Coastal Ocean" **Zhijin Li**, NASA Jet Propulsion Laboratory **Kayo Ide**, University of California Los Angeles
10:15-10:45 AM	Coffee Break
10:45-11:30 AM	"How to Use Imperfect Models in Assimilating Data" **Zoltan Toth**, National Centers for Environmental Protection
11:30-12:15 PM	15-Minute Talks
11:30-11:45	"A Local Ensemble Transform Kalman Filter: Perfect Model Results with the Lorenz-95 Model" **Brian Hunt**, University of Maryland
11:45-12:00	"Statistical Data Assimilation for Hurricane Forecasting" **Montserrat Fuentes**, North Carolina State University **Kristen Foley**, North Carolina State University **Lian Xie**, North Carolina State University
12:00-12:15	"Lagrangian Data Assimilation and Overcoming the Saddle Effect" **Christopher Jones**, SAMSI and University of North Carolina at Chapel Hill **Kayo Ide**, University of California Los Angeles **Liyan Liu**, University of North Carolina at Chapel Hill
12:15-1:30 PM	Lunch -- Featuring "SAMSI DA Program, Spring 2005"
1:30-2:15 PM	"Multiscale Methods of Data Assimilation" **Achi Brandt**, University of California Los Angeles & Weizmann Institute
2:15-3:00 PM	15-Minute Talks
2:15-2:30	"Assessing Predictability with a Local Ensemble Kalman Filter" **Istvan Szunyogh**, University of Maryland

2:30-2:45 "Chemical Source Inversion Using Assimilated Constituent Observations" **Andrew Tangborn**, University of Maryland-Baltimore County

2:45-3:00 "Dynamical Systems Perspective of Observing System Design for Lagrangian Data Assimilation" **Kayo Ide**, University of California Los Angeles. **Christopher Jones**, SAMSI and University of North Carolina at Chapel Hill
Hayder Salman, University of North Carolina at Chapel Hill

3:00-3:30 PM Refreshments & Open Discussion

3:30-4:15 PM "Issues and Suggestions for Particle Filtering in High Dimensions"
Mark Berliner, Ohio State University

4:15-5:00 PM 15-Minute Talks

4:15-4:30 "Graphical Models in Data Assimilation"
Alex Ihler, University of California Irvine

4:30-4:45 "Directed Drifter Launch Strategies for Lagrangian Data"
Andrew Poje, City University of New York
Anne Molcard, Italian National Research Council
Tamay Ozgokmen, University of Miami

4:45-5:00 "Will the Future of Data Assimilation be 4D-Var or EnKF?"
Eugenia Kalnay, University of Maryland

5:00-5:30 PM Open Discussion

Mark Berliner
Ohio State University
Department of Statistics
mb@stat.ohio-state.edu

Chris Wikle
University of Missouri
Department of Statistics
wiklec@missouri.edu

"Issues and Suggestions for Particle Filtering in High Dimensions"

Though Monte Carlo or ensemble based approximations to Bayesian sequential updating are well-known and readily implemented, in principle, they may exhibit problematic behavior in very high-dimensional settings. Specifically, the weights associated with ensemble members tend to over-concentrate on a few or even one ensemble member, leading to inefficient, highly variable results. Suggestions for dealing with this problem are discussed. The basic idea is to develop usable weights that avoid over-concentration, yet maintain reasonable value in indicating high and low probability ensemble members. To that end, we seek procedures based on the "informative" part of the data. Some potential strategies are reviewed.

Achi Brandt
University of California Los Angeles and Weizmann Institute
Department of Mathematics

abrandt@math.ucla.edu

"Multiscale Methods of Data Assimilation"

Following an introduction to general nonlinear multigrid algorithms, their many potential benefits for solving inverse PDE problems are explained, focusing on the problem of atmospheric data assimilation. The equations of very stable and adaptable implicit time steps can be solved at a cost comparable to that of explicit steps. The multiscale computation allows the data assimilation to account for correlation at all scales, at a cost again just comparable to solving the direct PDEs. Such computations can also facilitate full (not just initial-condition) control (which is more sensitive and accurate), yield flexible multiscale representation and fast inversion of full-matrix covariances, improve regularization (e.g., exploiting scale-dependent statistical theories), continuously fast-assimilate new observations, organize observational data in efficient hierarchical structures, and allow scale-dependent data types.

Montserrat Fuentes, Kristen Foley
North Carolina State University
Department of Statistics
fuentes@stat.ncsu.edu
kmfoley@ncsu.edu

Lian Xie
North Carolina State University
Marine Earth & Atmospheric
Sciences Department
lian_xie@ncsu.edu

"Statistical Data Assimilation for Hurricane Forecasting"

We present a model-uncertainty estimation procedure and a data assimilation method for a coastal ocean prediction and assessment system (COPAS), to provide timely and accurate assessments and predictions of coastal marine environment. Gridded wind fields are used to spin up and force these ocean numerical models. Currently, these wind field drivers are specified by deterministic models, that are a function of the central pressure and location of the storm center and parameters to define the shape of the pressure profile of the storm. While these equations incorporate important physical knowledge about the structure of hurricane wind fields, they cannot always capture the asymmetric and dynamic nature of a hurricane. We present a statistical framework to account for variability not captured by the wind physical model. A linear model of coregionalization is used to account for spatial variability in the horizontal and vertical wind components as well as the covariance between components at the same location. A Bayesian framework allows for estimation of the parameters of the multivariate spatial model and the physically based wind data from buoys, ships, aircraft and satellite. The proposed statistical model is used to create an ensemble of wind fields for a sequential ensemble data assimilation method to improve hurricane forecasting using COPAS.

This methodology is applied to a couple of case studies in the Eastern US coast, one in September 1989 when we had Hurricane Hugo, and the other in September 1999 when we had Hurricane Floyd.

Brian Hunt
University of Maryland
Institute for Physical Science and Technology
bhunt@ipst.umd.edu

"A Local Ensemble Transform Kalman Filter: Perfect Model Results with the Lorenz-95 Model"

I will describe briefly a general-purpose approach to data assimilation that combines elements of the Ensemble Transform Kalman Filter (Bishop et al., 2001) and the Local Ensemble Kalman Filter (Ott et al., 2004), and present results obtained with John Harlim using a model proposed by Lorenz in 1995 that illustrate some practical considerations in using this approach.

Kayo Ide
University of California Los Angeles
Inst. of Geophysics & Planetary Physics
kayo@atmos.ucla.edu

Chris Jones and Hayder Salman
SAMSI & UNC, Chapel Hill
Department of Mathematics
ckrtj@amath.unc.edu

"Dynamical Systems Perspective of Observing System Design for Lagrangian Data Assimilation"

We present the Lagrangian data assimilation (LaDA) method due to Ide and collaborators (Ide et al 2002, Kuznetsov et al 2003). We invoke an ensemble Kalman filter in order to estimate and forecast the (ocean) state using the shallow-water model (Salman et al, 2005). Based on the augmented state representation, the LaDA eliminates the need for any conventionally used approximation in assimilating the Lagrangian information. This augmentation also allows us to use dynamical systems theory for the design of a comprehensive observing system. We show how deploying drifters in the flow near the (Lagrangian) saddle point enhances the information content of the (Eulerian) flow dynamics extracted from the Lagrangian data using LaDA.

Alexander Ihler
University of California Irvine
School of Information and Computer Sciences
ihler@ics.uci.edu

"Graphical Models in Data Assimilation"

In data assimilation of a time series, one incorporates past and current observations with a model of system dynamics, so as to improve the model's subsequent simulation or predictions. From a statistical point of view, this can be regarded as a process of estimating a collection of random variables (representing the state of the physical system) which are related both spatially and temporally. Given values for some of these variables (typically times past) we require estimates of others (typically corresponding to future times, or to variables which cannot be directly observed). Graphical models have emerged as an effective formalism for assisting in these types of inference tasks, particularly for large numbers of random variables. Graphical models provide a means of representing the structure of dependencies among the variables. This structure can be used to construct efficient algorithms for optimal or approximate estimation and other inference tasks. We describe several examples of how graphical models have been applied to data assimilation problems, including Markov chains and Kalman filtering for optimal estimation in time series, multi-resolution models for tomography and image processing, and Markov random fields for modelling rainfall patterns in space and time.

Chris Jones and Liyan Liu

Kayo Ide

SAMSI & UNC at Chapel Hill UCLA
Department of Mathematics Institute of Geophysics & Planetary Physics
ckrtj@amath.unc.edu kayo@atmos.ucla.edu

"Lagrangian Data Assimilation and Overcoming the Saddle Effect"

By removing the necessity for a commonly used approximation, in terms of velocity, in assimilating Lagrangian data, a recently developed Lagrangian data assimilation (LaDA) method has been shown to offer a variety of advantages (Ide et al 2002, Kuznetsov et al 2003, Salman et al, 2005).The success of LaDA depends on properly estimating the error correlation between the ocean model variables and drifter positions within the augmented state space. The estimation process may, however, fail when the Lagrangian instrument passes near to a (Lagrangian) saddle point of the ocean flow. This leads to a sudden divergence of the filter, termed the "saddle effect." We present strategies for addressing this problem using tracer control and an ensemble filter.

Eugenia Kalnay
University of Maryland
Department of Atmospheric and Ocean Sciences
ekalnay@atmos.umd.edu

"Will the Future of Data Assimilation be 4D-Var or EnKF?"

We consider the advantages and disadvantages of EnKF and 4D-Var, in view of simple experiments with the Lorenz (1963) model, with the SPEEDY primitive equations model (using both perfect model and reanalysis "observations"), and in view of recent results with both perfect models and real observations. We point out some advantages of the Local Ensemble Transform Kalman Filter, and its extension to 4 dimensions, which brings to EnKF the main advantage of 4D-Var. A table summarizing the pros and cons of the two methods by Lorenc (2004) is adapted with additions and comments.

Zhijin Li
Raytheon & Jet Propulsion Laboratory
Zhijin.li@jpl.nasa.gov

Kayo Ide
University of California Los Angeles
Institute of Geophysics & Planetary Physics
kayo@atmos.ucla.edu

"Incorporation of Dynamic Balance in Data Assimilation and Application to Coastal Oceans"

Data assimilation in meteorology and oceanography is commonly described as the process through which all the observed and predicted information are used in order to estimate as accurately as possible the state of atmospheric or oceanic flow and the algorithm is rooted in optimal estimation theory. However, the estimated state should be constrained to be close to or on slow manifolds or dynamic attractors, and current data assimilation algorithms do not incorporate this capability in the framework of optimal estimation theory. We are exploring a theoretical framework to address this issue and suggesting a practical method.

Dennis McLaughlin

Massachusetts Institute of Technology
Department of Civil and Environmental Engineering
dennism@mit.edu

"Data Assimilation in the Earth Sciences: Building on and Going beyond Classic Estimation Theory"

The increase in data available from remote sensing platforms and *in situ* sensor networks together with recent improvements in earth science modeling capabilities have prompted much interest in the field of data assimilation. Data assimilation/data fusion methods characterize environmental systems by using distributed models to merge uncertain data taken over different scales, with different accuracies, frequencies, and coverage. Data assimilation problems can be conveniently posed as stochastic estimation problems. The details depend on the objectives of the characterization (e.g forecasting vs. retrospective analysis) and on the focus of the estimation procedure (e.g. parameter vs. state estimation). In any case, classical estimation theory provides an attractive way to formulate and solve problems in a wide range of disciplines. Most texts and papers in the field rely on the assumptions and perspectives of this theory, including i) a least-squares perspective which often implicitly assumes that system states are Gaussian (i.e. fully characterized by their first two moments) , ii) a reliance on linear or quasi-linear theory, iii) assumptions that model and measurement errors are additive and independent of the states and iv) a tendency to neglect the computational issues that arise when the system state size is very large (as is often the case in earth science applications). Since all of these assumptions are likely to be problematic in real applications there is a growing awareness in the field that we need to go beyond classical estimation theory and develop a new set of methods that are appropriate for problems of realistic size and complexity. This talk highlights some of the critical conceptual/methodological issues facing environmental data assimilation today and briefly discusses some promising directions for future research.

Robert Miller
Oregon State University
Department of Oceanic and Atmospheric Sciences
miller@coas.oregonstate.edu

"Analysis and Prediction of a Noisy Nonlinear Ocean"

Most of our data assimilation systems were developed for linear systems, or at least for systems whose linearized versions had significant skill. We seek ways to understand nonlinear phenomena in the ocean through application of dynamical systems theory and extension of techniques developed for linearized systems. We present preliminary results of an application of a variational data assimilation technique to the Kuroshio off Japan, and comparisons to data and to a detailed and finely resolved model that does not allow direct application of advanced data assimilation.

Andrew C. Poje
City University of New York
Department of Mathematics
poje@math.csi.cuny.edu

"Directed Drifter Launch Strategies for Lagrangian Data Assimilation"

The dependence of the fidelity of a Lagrangian data assimilation scheme on the initial launch locations of the observation is studied in the context of a reduced gravity, primitive equation model of the midlatitude ocean circulation. We develop a directed drifter launch strategy based on tracking the Lagrangian manifolds emanating from strongly hyperbolic regions in a given flow field. In a series of twin assimilation experiments, the convergence of the data assimilating scheme to model truth is shown to be consistently and significantly improved by such directed launches when compared to similar, but randomly chosen, initial configurations. In general, the performance of the assimilation scheme is shown to depend strongly on the independence of the Lagrangian observations and on the temporal persistence of the velocity field corrections provided by the data. Both quantities are naturally maximized by the directed launch scheme.

This is joint work with **Anne Molcard** and **Tamay Ozgokmen**.

Istvan Szunyogh
University of Maryland
Department of Atmospheric and Ocean Sciences
szunyogh@ipst.umd.edu

“Assessing Predictability with a Local Ensemble Kalman Filter”

In this presentation, the spatio-temporally changing nature of predictability in the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) is discussed. Atmospheric predictability is assessed for the perfect-model scenario, where forecast uncertainties are entirely due to uncertainties in the estimates of the initial states. The imperfect initial conditions are obtained by assimilating simulated noisy observations of the “true” states with the Local Ensemble Kalman Filter (LEKF) data assimilation scheme.

For this specific choice of the model and data assimilation system, the forecast errors grow exponentially in the extra-tropics and linearly in the tropics. The analysis errors are the smallest in the regions, the extratropical storm tracks, where the growth of the forecast errors is the fastest. This seemingly paradoxical result can be explained by the strong anti-correlation between the local dimensionality and the error variance explained by the LEKF ensemble. This strong anti-correlation makes the LEKF algorithm extremely efficient in estimating the analysis and forecast uncertainties in the regions of local low dimensionality, which coincide with the regions fastest error growth. The efficient estimation of the space of uncertainties enables the LEKF to produce very accurate analyses and very accurate estimates of the forecast uncertainties. It is conjectured that the results presented here could be reproduced with any suitably formulated ensemble-based Kalman filter data assimilation scheme.

Andrew Tangborn
University of Maryland-Baltimore County
Joint Center for Earth Systems Technology
tangborn@umbc.edu

“Chemical Source Inversion Using Assimilated Constituent Observations”

We present a comparison of source inversion for chemical constituent sources using assimilated constituent observations with direct use of the observations. In this model problem, a two-dimensional spectral transport model is combined with a Kalman filter. Inversion is carried out using a Green's function method. Observations are simulated from a "true" state with added Gaussian noise. The forecast state uses the same spectral model, but differs by an unbiased error. Two different observing systems are used, in situ and satellite.

Source inversion is carried out by either direct use of the observations in the Green's function inversion, or by first assimilating the observations and using the analysis as if they were observations. We have conducted 20 twin experiments for each case and find that in the limiting cases of very few localized observations, or an extremely large observation network there is little advantage to carrying out assimilation first. However, in intermediate observation densities, there is a significant increase in the accuracy of the source inversion standard deviation using the Kalman filter algorithm followed by Green's function inversion.

Zoltan Toth
NCEP/NWS/NOAA
Environmental Modeling Center
Zoltan.Toth@noaa.gov

"How to Use Imperfect Models in Assimilating Data"

This presentation will briefly review existing approaches to the use of imperfect models in the assimilation of observational data. Shortcomings of the currently used approaches, as well as attempts to ameliorate some of the problems, will also be discussed. A new forecast initialization method, based on a distinction between initial value and model related errors, will also be presented, along with a demonstration of the method with a simple dynamical system.

Tomislava Vukicevic
Colorado State University
Cooperative Institute for Research in the Atmosphere
tomi@cira.colostate.edu

"Estimation of Clouds in Atmospheric Models"

Accurate estimates of cloud properties are required for improved understanding of feedbacks within the atmospheric system and the system's predictability on wide range of spatial and temporal scales from individual storms to climate. Quantitative observations of clouds and precipitation are typically obtained by indirect, remote sensing methods. Although considerable progress has been made in remotely sensing and retrieving the cloud properties, complex 3D cloud structure and interaction with the atmospheric environment is not well specified from observations alone. We propose that accurate estimates of evolving 3D cloudy atmosphere on cloud resolving scales could be derived by assimilation of remote sensing observations of the cloudy atmosphere into a cloud resolving dynamical model such that they optimally constrain what controls the model solution. The model solution is controlled by initial and boundary conditions and model parameterizations.

The complex problem is initially approached by applying a new 4D-variational (4DVAR) research data assimilation system with a cloud resolving model to the assimilation of the GOES (Geostationary Operational Environmental Satellites) imager observations. The study results show that the observations can significantly improve modeled cloud, resulting in a 3D distribution of the hydrometeor mixing ratio and number concentration parameters with nearly zero mean error and a small standard deviation in the observation space. The strength of constraint by the selected observations depends upon conditions in the model to support the cloud formation and on the information content of the observations.

The sensitivity of data assimilation results to the assimilation technique was also studied. The experiments with the model error indicate that this error in the form of a generic linear forcing in the governing equations, which was adopted from other 4DVAR data assimilation studies, is not suitable for the cloud resolving data assimilation. This is because the actual, unknown error, is likely highly nonlinear in the state space and consequently cannot be well represented with Gaussian statistics. The estimation of so called physical parameters may be more appropriate in this case. An illustrative example is produced with the Lorenz 3-component chaos model.

VIII. DATA MINING AND MACHINE LEARNING

A. *Technology Short Course Program*

July 25-29, 2005

Monday -- July 25, 2005
NISS-SAMSI Building, Room 104

8:30-9:00 AM Registration and Continental Breakfast.

9:00-12:00 PM	Introduction, Cross-Validation, the Bootstrap, Search Strategies, and Smoothing
12:00-1:30 PM	Lunch
1:30-3:30 PM	Computer Lab on G-Gobi Visualization and Smoothing
3:30-3:45 PM	Coffee Break
3:45-5:15 PM	"Visualization and Data Mining for Microarrays" **Jack Liu,** GlaxoSmithKline

Tuesday -- July 26, 2005
NISS-SAMSI Building, Room 104

8:30-9:00 AM	Continental Breakfast
9:00-12:00 PM	Review and Comparison of Nonparametric Regression Methods: AM, GAM, PPR, ACE, AVAS, MARS, CART, Neural Nets; the Backfitting Algorithm
12:00-1:30 PM	Lunch

1:30-3:30 PM	Computer Lab on the DRAT Package for Multivariate Nonparametric Regression
3:30-3:45 PM	Coffee Break
3:45-5:15 PM	"Issues with High Dimension, Low Sample Size Data" **J.S. Marron**, University of North Carolina

Wednesday -- July 27, 2005
NISS-SAMSI Building, Room 104

8:30-9:00 AM	Continental Breakfast
9:00-12:00 PM	Classification and Clustering: SVMs, Random Forests, Boosting
12:00-1:30 PM	Lunch
1:30-3:30 PM	Computer Lab on Classification and Boosting
3:30-3:45 PM	Coffee Break
3:45-5:15 PM	"Model Complexity and Regularization" **Feng Liang**, Duke University

Thursday -- July 28, 2005
NISS-SAMSI Building, Room 104

8:30-9:00 AM	Continental Breakfast
9:00-12:00 PM	Bases and Wavelets
12:00-1:30 PM	Lunch
1:30-3:30 PM	Computer Lab on SVMs and Random Forests
3:30-3:45 PM	Coffee Break
3:45-5:15 PM	"Bayesian Model Averaging" **Merlise Clyde**, Duke University

Friday -- July 29, 2005
NISS-SAMSI Building, Room 104

8:30-9:00 AM	Continental Breakfast
9:00-12:00 PM	PAC Bounds and VC Classes
12:00-1:30 PM	Lunch
1:30-3:30 PM	Computer Lab on Wavelets (Decimated and Non-Decimated)

3:30-3:45 PM Coffee Break

3:45-5:15 PM "Survey of New Ideas in Data Mining"
David Banks, Duke University

IX. WORKSHOPS CO-SPONSORED BY SAMSI

A. *Random Graphs and Stochastic Computation*
June 13-14, 2005

Monday – June 13, 2005

8:45-9:00 AM	Welcome and Introduction **Jim Berger**, SAMSI **Mike West**, Duke University
9:00-9:50 AM	Hyper Markov Priors for Parameters of Gaussian Graphical Models. **Helene Massam**, York University
9:50-10:40 AM	Fitting and Interpreting Sparse Gaussian Graphical Models. **Beatrix Jones,** Massey University
10:40-11:10 AM	Coffee Break
11:10-12:00 PM	Random Generation of Regular Graphs **Mark Huber**, Duke Unversity
12:00-12:50 PM	Laplacian Eigenfunctions in Random Graphs **Mikhail Belkin**, University of Chicago

12:50-1:50 PM Lunch

1:50-2:40 PM	Structural Learning in Graphical Models with Mixed Variables. **Adrian Dobra**, Duke University
2:40-3:30 PM	MCMC Methods for Discrete Graphical Models **Claudia Tarantola**, University of Pavia
3:30-4:10 PM	Coffee Break
4:10-5:00 PM	Random Graphs and Combinatorial Rigidity **Brigitte Servatius**, Worcester Polytechnic Institute
5:00-6:30 PM	Evening Break of Informal Discussion, Room 203
6:30-8:30 PM	Poster Session and Reception

Tuesday – June 14, 2005

9:00-9:50 AM Statistical Models for Social Networks

Mark Handcock, University of Washington

9:50-10:40 AM Sampling in Graphs. **Steven Thompson**, Pennsylvania State University

10:40-11:10 AM Coffee Break

11:10-12:00 PM Estimating Models for Protein-Protein Interaction Graphs
Alun Thomas, Unversity of Utah

12:00-12:50 PM Connectivity, Component Sizes and Distances in the Power Law Random Graphs
Dmitri Znamenskiy, Eurandom

12:50-1:50 PM Lunch

1:50-2:40 PM Bayesian Covariance Selection
David Dunson, NIEHS & Duke University

2:40-3:30 PM Variational Methods for Inference in Graphical Models
Eric Xing, Carnegie Mellon University

3:30-4:10 PM Coffee Break

4:10-5:00 PM DAGs of Binary Variables
Carlos Rodriguez, SUNY Albany

5:00-5:30 PM Wrap-up Discussion

A. *Collaborations in the Mathematical Geosciences*
October 6-7, 2005

Thursday--October 6, 2005
Radisson Hotel Research Triangle Park
Room H (3^{rd} Floor)

8:00-8:45 AM Registration and Continental Breakfast

8:45-9:00 AM Introduction and Welcome

9:00-10:00 AM Paired Presentations I: *Atmospheric Science*

- **Eugenia Kalnay**, University of Maryland
- **Christopher Wikle**, University of Missouri

10:00-10:30 AM Coffee Break

10:30-11:30 PM Paired Presentation II: *Earth Science*

- **Robert van der Hilst**, Massachusetts Institute of Technology

- **Maarten de Hoop**, Purdue University

11:30-12:00 PM Poster Sales Talks (2 minutes maximum)

12:00-1:30 PM Lunch

1:30-2:30 PM Paired Presentation III: *Ocean Science*

- **William Dewar**, Florida State University
- **Juan Restrepo**, University of Arizona

2:30-3:00 PM Coffee Break

3:00-4:00 PM Two-Minute Madness: All Participants May Speak

4:00-4:15 PM Coffee Break

4:15-4:30 PM Charge to Breakout Discussion Groups

4:30-6:00 PM Breakout Discussions by GEO Topic

- *Atmospheric Science* -- Moderator: **Mark Berliner**, Ohio State University
- *Earth Science* -- Moderator: TBA
- *Ocean Science* -- Moderator: **Christopher Jones**, University of North Carolina at Chapel Hill
- *Polar Science* -- Moderator: TBA
- *Space Science* -- Moderator: **Jogesh Babu**, Pennsylvania State University

6:30-8:30 PM Poster Session and Refreshments
Radisson Hotel RTP, Room FG – 3rd Floor

Friday--October 7, 2005
Radisson Hotel Research Triangle Park
Room H (3rd Floor)

8:00-8:30 AM Continental Breakfast

8:30-10:00 AM Reports from GEO Breakout Discussions

10:00-10:30 AM Coffee Break

10:30-11:30 PM New Researcher Session (10-minute talks)

11:30-1:30 PM Breakout Discussion by Math/Stat Topic (continuing over lunch)

- *Multiscale Modeling* -- Moderator: TBA
- *Massive Databases* -- Moderator: TBA
- *Representing* -- Moderator: TBA

- *Data Assimilation* -- Moderator: **Kayo Ide**, University of California Los Angeles

1:30-3:00 PM Reports from Math/Stat Breakout Discussions

3:00-3:30 PM Coffee Break

3:30-4:30 PM Question and Answer Session with NSF Program Officers

4:30 PM Workshop Adjourns

APPENDIX G – Workshop Evaluation Summaries

At every SAMSI workshop participants were given an evaluation questionnaire to complete. A sample questionnaire is at the end of this appendix. Summaries of the participant evaluations are presented below.

The evaluations of scientific content are presented in three graphs: i) SAMSI 2005-6 Program workshops, ii) Follow-on workshops to previous SAMSI Programs and iii) Student workshops, at both graduate and undergraduate levels. At least 80% of the participants' at each SAMSI Program workshop, whether contemporaneous or follow-on, rated the scientific content Very Good to Excellent. For undergraduate workshops, the ratings were more varied, with fewer generally rating the workshops Excellent and a higher proportion rating them Good to Very Good. Judging from the undergraduates' written comments, the satisfaction with the science of the workshops depended on the level of the individual student's preparation as well as the quality of the workshop itself. However it is also noteworthy that some students who volunteered that the technical level of the workshop was beyond their current capability also wrote enthusiastically about their participation.

SAMSI staff and facilities for workshops have been very highly rated every since SAMSI opened its doors. The continuing satisfaction with SAMSI staff is a point of pride. Minor problems with transportation to/from SAMSI continue to receive attention to keep the workshops running smoothly. Note that undergraduates are housed on campus at NC State for the Interdisciplinary Undergraduate Workshop, so that their primary activities are within walking distance.

In 2005-6 SAMSI workshops attracted from 16 to 157 participants; several of these (both large and small) were oversubscribed for the available space and/or for the workshop goals. Keys to abbreviations on the graphical summaries follow.

2005-6 Programs

NDHS:	National Defense and Homeland Security		
		Opening workshop	93 participants
FMSE:	Financial Mathematics, Statistics and Econometrics		
		Opening workshop	157 participants
		Model Uncertainty	27 participants
		Transition workshop	70 participants
Astro:	Astrostatistics		
		Tutorial #1 (pre-Opening)	34 participants
		Tutorial #2 (pre-Opening)	24 participants
		Tutorial #3 (pre-Opening)	29 participants
		Opening workshop	67 participants

Hot Topics & 2002-5 Programs

LVSS:	Latent Variables in the Social Sciences		
		Transition workshop	75 participants
CompBio:	Computational Biology of Infectious Disease		

Closing workshop 25 participants

Data A: Data Assimilation for Geophysical Systems
Transition workshop 39 participants

RGSC: Random Graphs and Stochastic Computation
Hot Topic 36 participants

Data Mining: Data Mining and Machine Learning
Technology short courser 16 participants

MAA PREP: PREP Workshop, joint with MAA
Teachers' workshop 29 participants

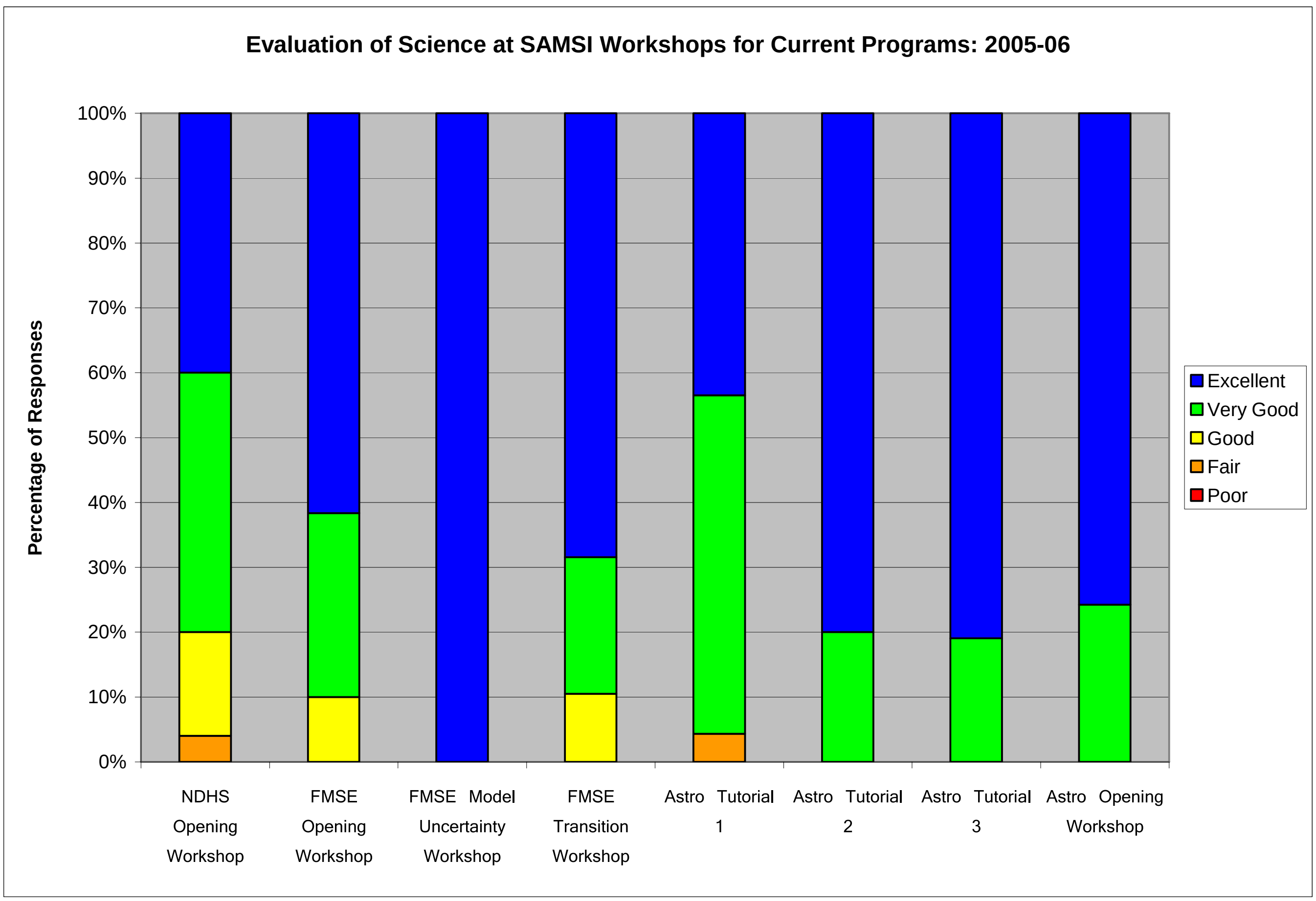
Student Workshops and Programs

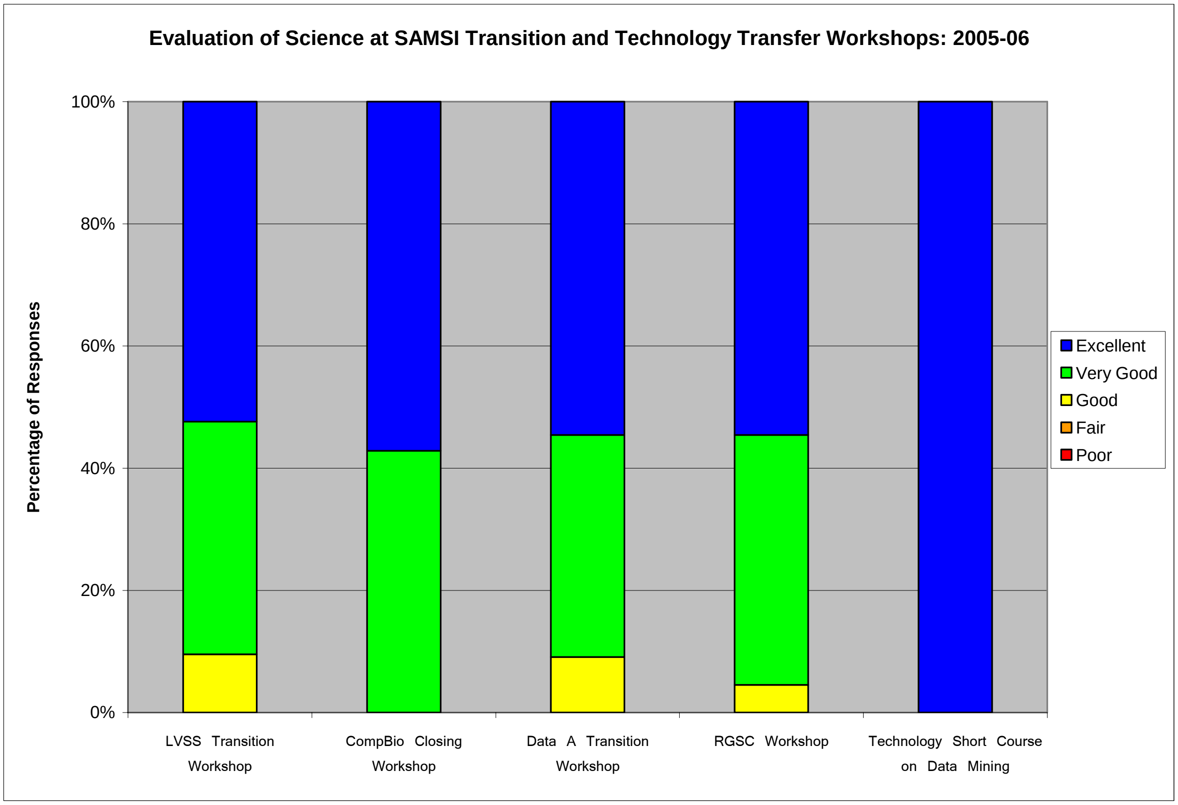
Undergraduates:

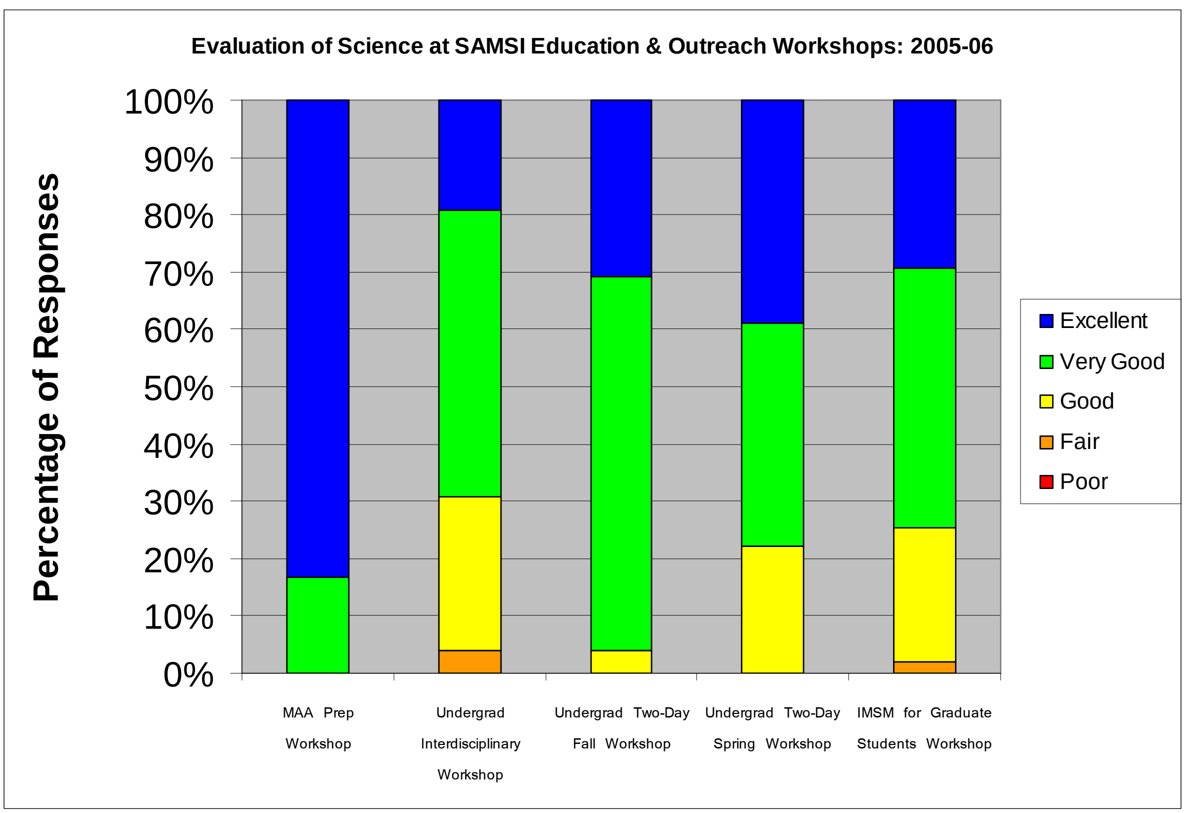
Summer Interdisciplinary workshop 36 participants
Fall Undergraduate two-day workshop 31 participants
Spring Undergraduate two-day workshop 25 participants

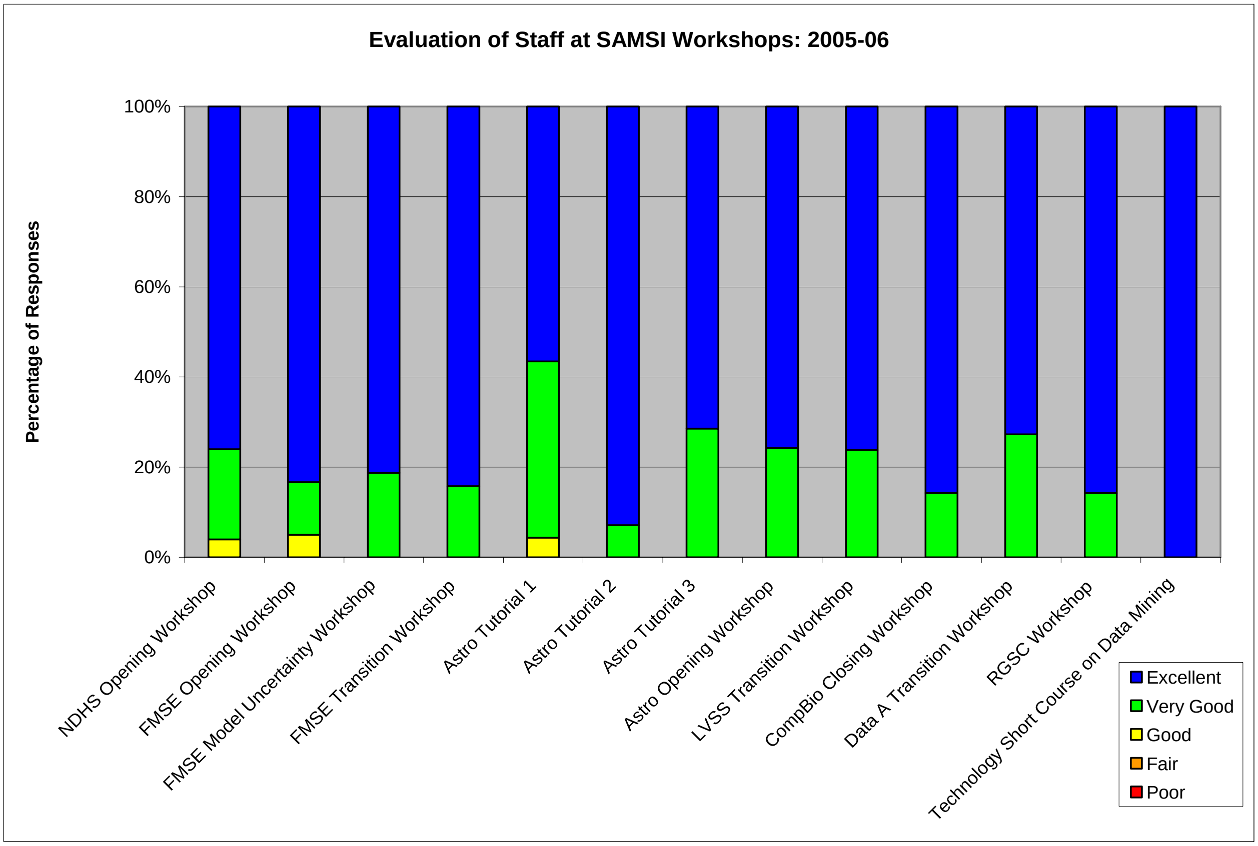
Graduate students:

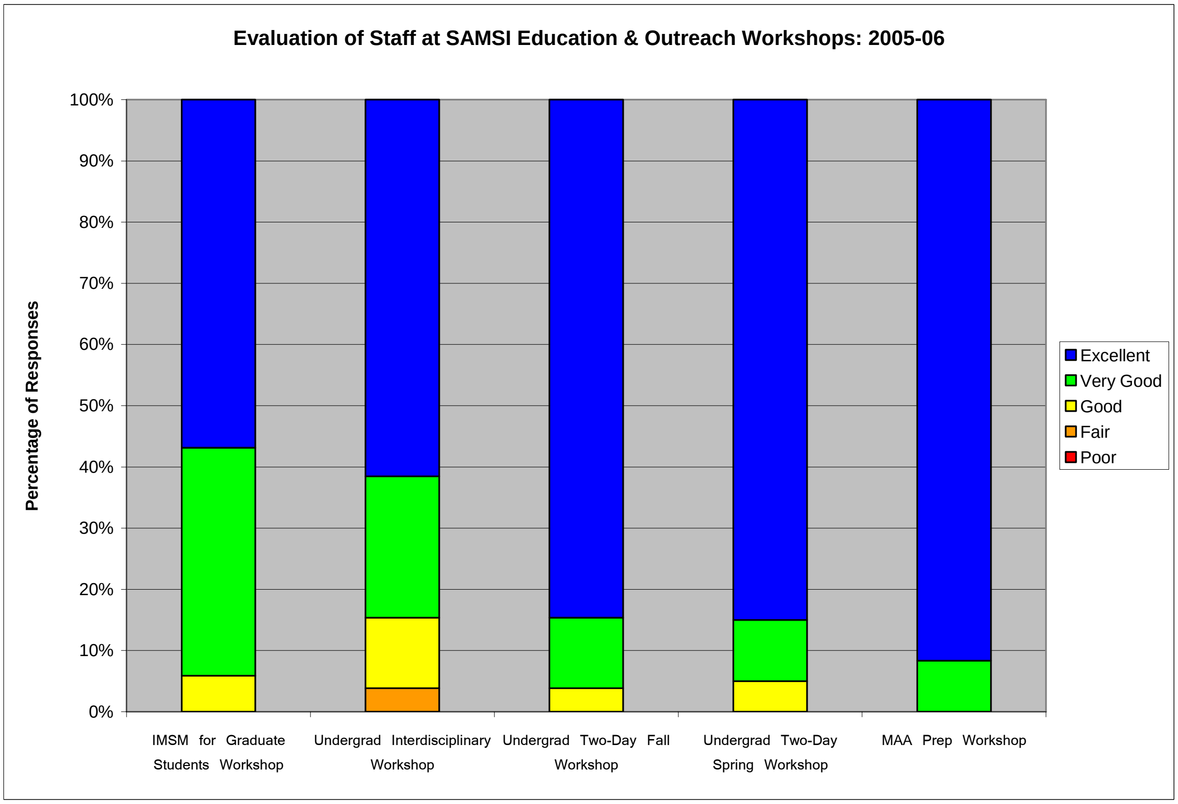
IMSM: Industrial Mathematical, Statistical Modeling 60 participants

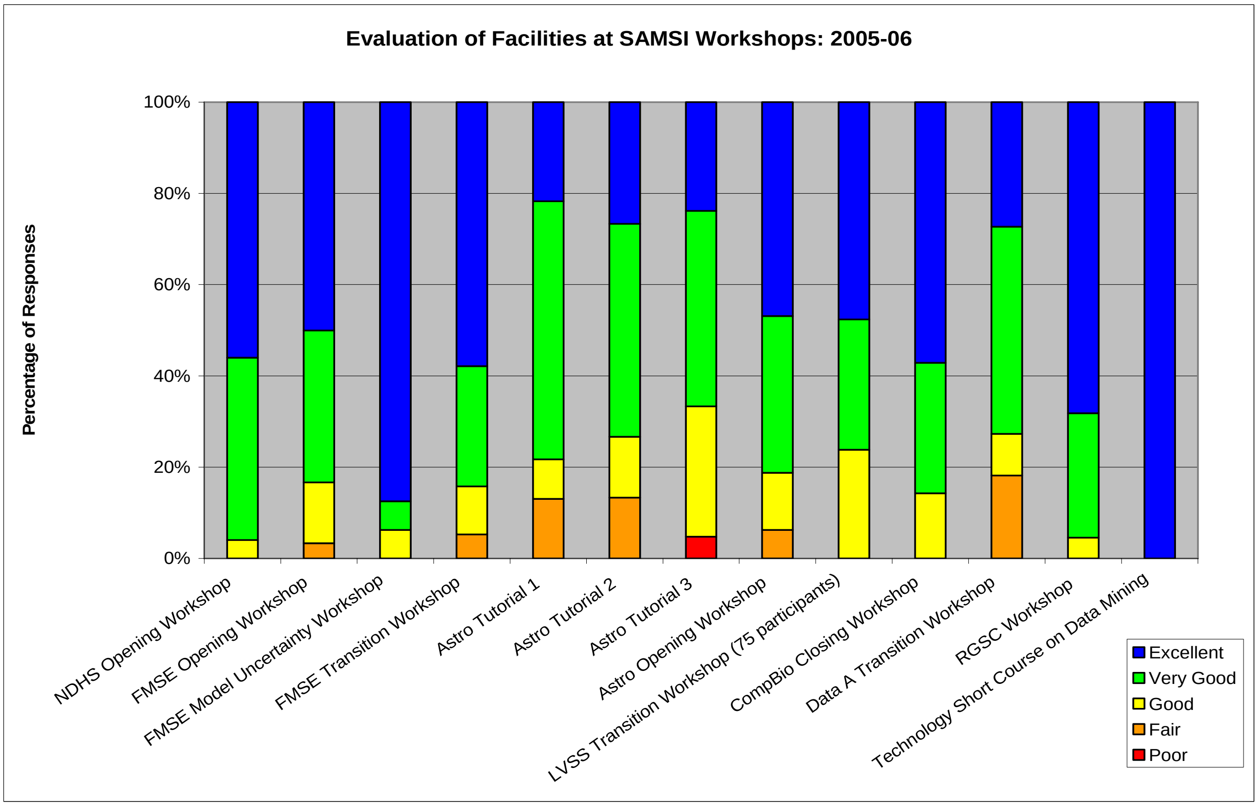


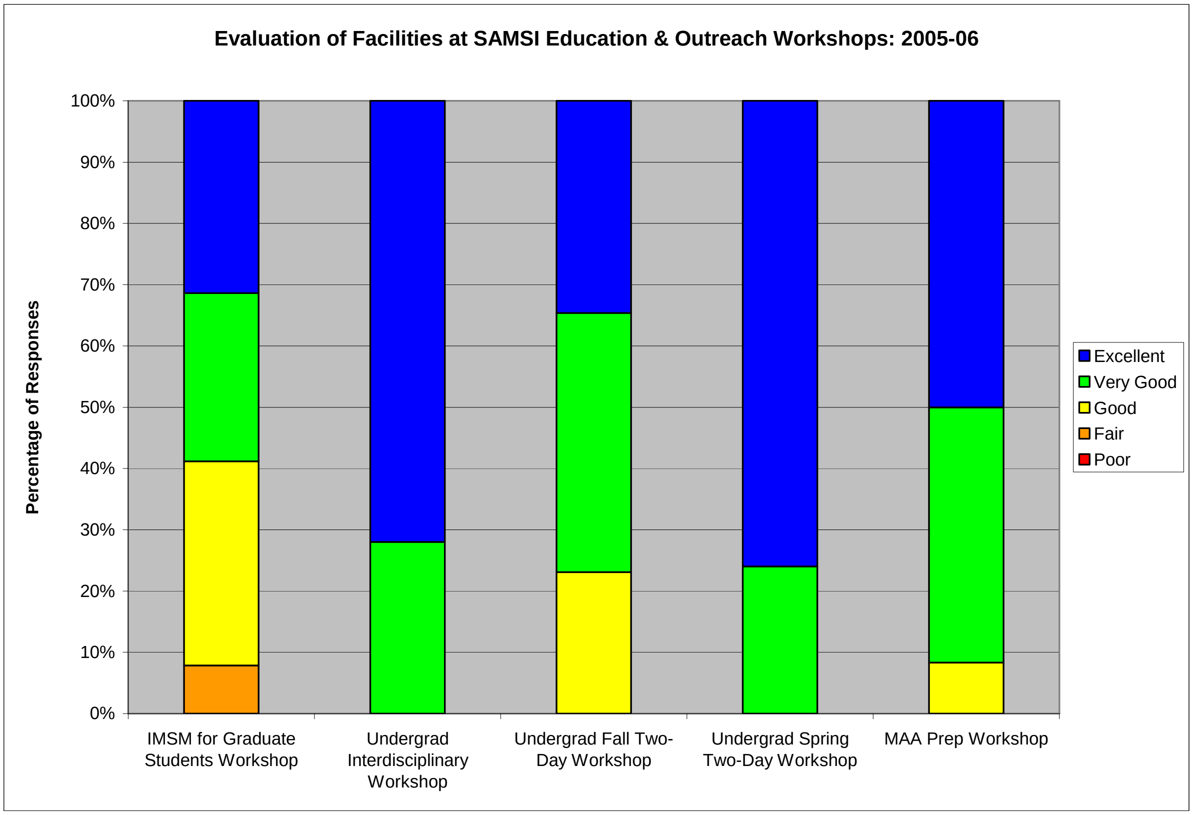

Evaluation of Science at SAMSI Transition and Technology Transfer Workshops: 2005-06
Percentage of Responses
100%
80%
60%
40%
20%
0%
LVSS Transition Workshop
CompBio Closing Workshop
Data A Transition Workshop
RGSC Workshop
Technology Short Course on Data Mining
Excellent
Very Good
Good
Fair
Poor

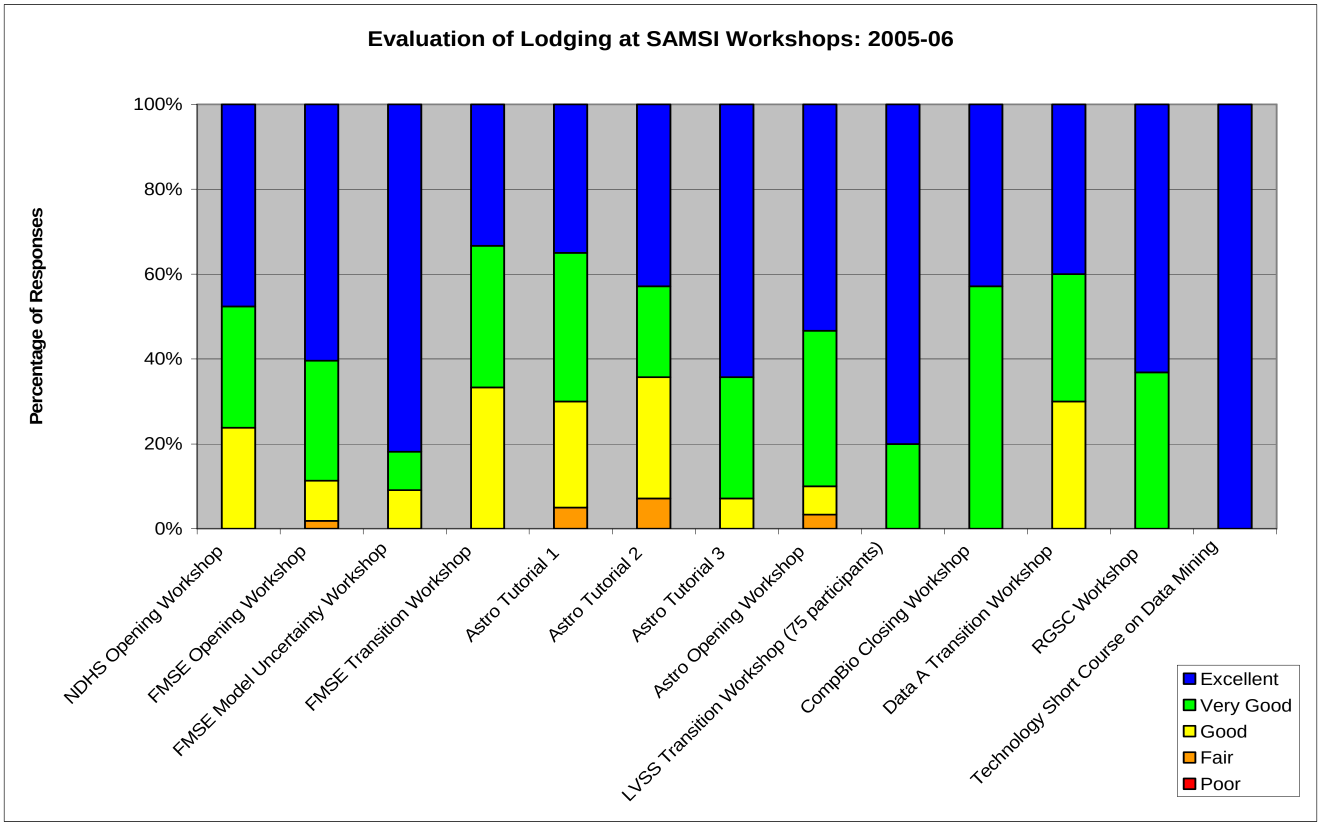

Evaluation of Science at SAMSI Education & Outreach Workshops: 2005-06
Percentage of Responses
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
MAA Prep Workshop
Undergrad Interdisciplinary Workshop
Undergrad Two-Day Fall Workshop
Undergrad Two-Day Spring Workshop
IMSM for Graduate Students Workshop
Excellent
Very Good
Good
Fair
Poor


Evaluation of Staff at SAMSI Workshops: 2005-06
Percentage of Responses
100%
80%
60%
40%
20%
0%
NDHS Opening Workshop
FMSE Opening Workshop
FMSE Model Uncertainty Workshop
FMSE Transition Workshop
Astro Tutorial 1
Astro Tutorial 2
Astro Tutorial 3
Astro Opening Workshop
LVSS Transition Workshop
CompBio Closing Workshop
Data A Transition Workshop
RGSC Workshop
Technology Short Course on Data Mining
Excellent
Very Good
Good
Fair
Poor

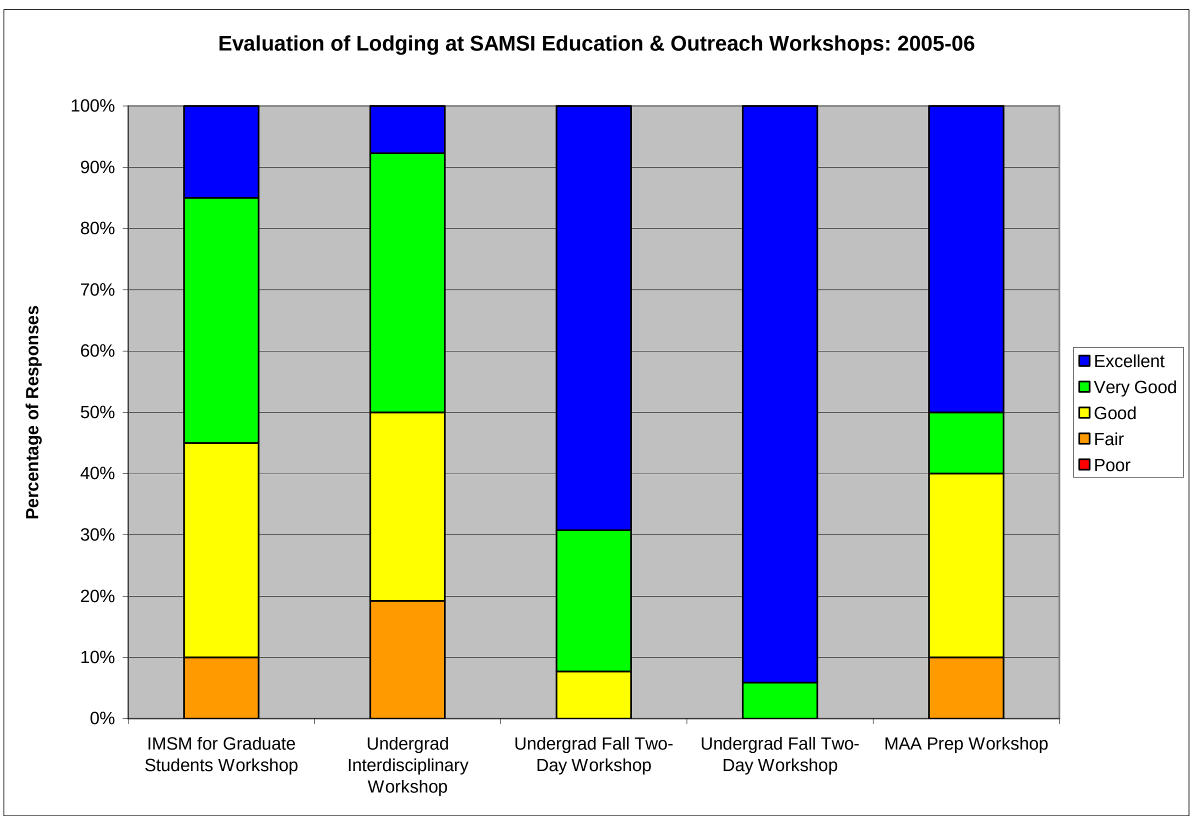

Evaluation of Staff at SAMSI Education & Outreach Workshops: 2005-06
Percentage of Responses
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
IMSM for Graduate Students Workshop
Undergrad Interdisciplinary Workshop
Undergrad Two-Day Fall Workshop
Undergrad Two-Day Spring Workshop
MAA Prep Workshop
Excellent
Very Good
Good
Fair
Poor


Evaluation of Facilities at SAMSI Workshops: 2005-06
Percentage of Responses
100%
80%
60%
40%
20%
0%
NDHS Opening Workshop
FMSE Opening Workshop
FMSE Model Uncertainty Workshop
FMSE Transition Workshop
Astro Tutorial 1
Astro Tutorial 2
Astro Tutorial 3
Astro Opening Workshop
LVSS Transition Workshop (75 participants)
CompBio Closing Workshop
Data A Transition Workshop
RGSC Workshop
Technology Short Course on Data Mining
Excellent
Very Good
Good
Fair
Poor

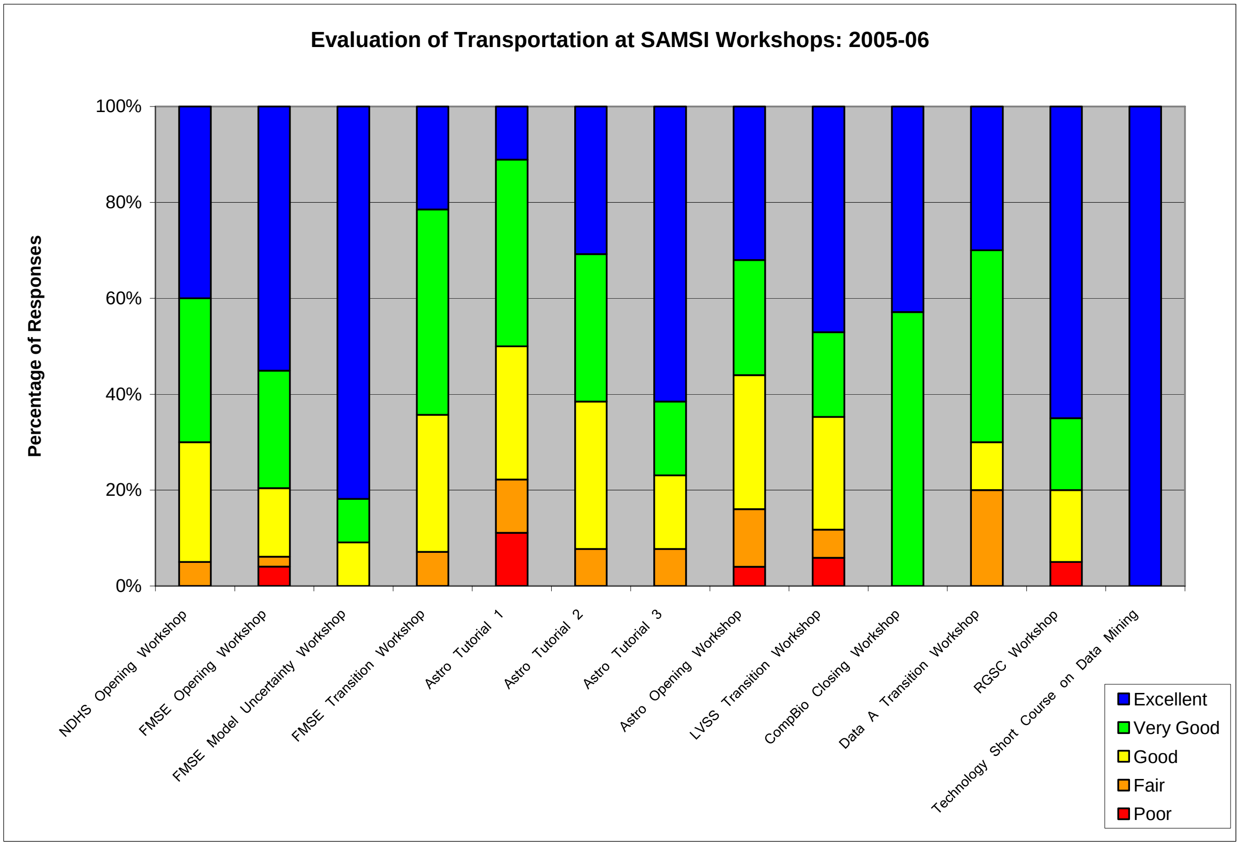

Evaluation of Facilities at SAMSI Education & Outreach Workshops: 2005-06
Percentage of Responses
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
IMSM for Graduate Students Workshop
Undergrad Interdisciplinary Workshop
Undergrad Fall Two-Day Workshop
Undergrad Spring Two-Day Workshop
MAA Prep Workshop
Excellent
Very Good
Good
Fair
Poor

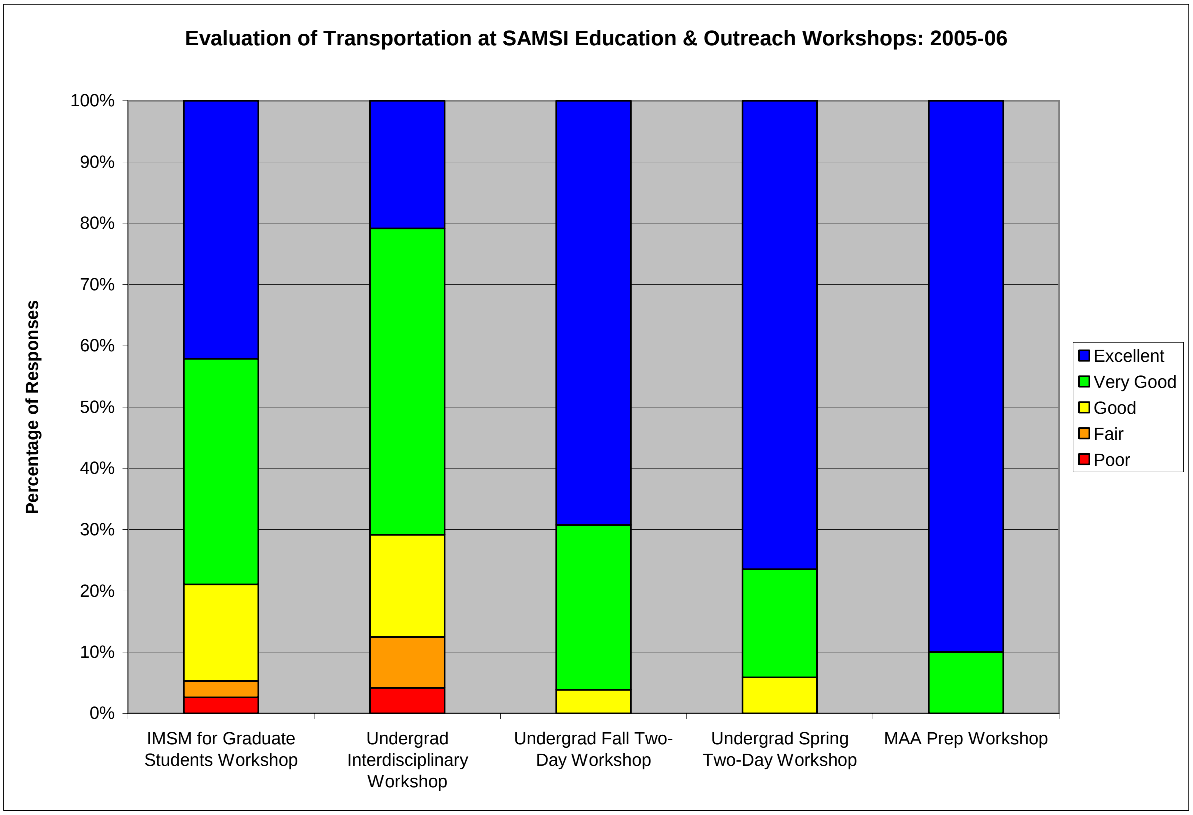


Evaluation of Lodging at SAMSI Workshops: 2005-06
Percentage of Responses
100%
80%
60%
40%
20%
0%
NDHS Opening Workshop
FMSE Opening Workshop
FMSE Model Uncertainty Workshop
FMSE Transition Workshop
Astro Tutorial 1
Astro Tutorial 2
Astro Tutorial 3
Astro Opening Workshop
LVSS Transition Workshop (75 participants)
CompBio Closing Workshop
Data A Transition Workshop
RGSC Workshop
Technology Short Course on Data Mining
Excellent
Very Good
Good
Fair
Poor



Evaluation of Lodging at SAMSI Education & Outreach Workshops: 2005-06
Percentage of Responses
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
IMSM for Graduate Students Workshop
Undergrad Interdisciplinary Workshop
Undergrad Fall Two-Day Workshop
Undergrad Fall Two-Day Workshop
MAA Prep Workshop
Excellent
Very Good
Good
Fair
Poor


Evaluation of Transportation at SAMSI Workshops: 2005-06
Percentage of Responses
100%
80%
60%
40%
20%
0%
NDHS Opening Workshop
FMSE Opening Workshop
FMSE Model Uncertainty Workshop
FMSE Transition Workshop
Astro Tutorial 1
Astro Tutorial 2
Astro Tutorial 3
Astro Opening Workshop
LVSS Transition Workshop
CompBio Closing Workshop
Data A Transition Workshop
RGSC Workshop
Technology Short Course on Data Mining
Excellent
Very Good
Good
Fair
Poor


Evaluation of Transportation at SAMSI Education & Outreach Workshops: 2005-06
Percentage of Responses
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
IMSM for Graduate Students Workshop
Undergrad Interdisciplinary Workshop
Undergrad Fall Two-Day Workshop
Undergrad Spring Two-Day Workshop
MAA Prep Workshop
Excellent
Very Good
Good
Fair
Poor