From the director...SAMSI Renewed for Another Five Years!

The big news this month is that we have a new award from the National Science Foundation! SAMSI has been renewed for another five years with an increase in funds.

The renewal of our funding is a tribute to the incredible commitment to SAMSI by the whole community – our partner institutions, the national leadership in statistics and applied mathematics, everyone who has served SAMSI as a directorate member or program leader, and more broadly all the people who have visited SAMSI and participated in our research over the years. A big thank you to you all!

This is also a good opportunity to note the fact that SAMSI is now ten years old!

SAMSI officially opened on September 3, 2002, in a ceremony that was attended by Congressman David Price, the leaders of our three partner universities and of NISS, and numerous other dignitaries and officials, all presided over of course by the first director Jim Berger. In September 2002, we also held the opening workshops of our first two programs, on *Inverse Problem Methodology In Complex Stochastic Models* and on *Stochastic Computation*. Completing the first-year programs was one on *Large-Scale Computer Models for Environmental Systems* which was my own first introduction to the complexities of organizing a program at SAMSI.

So what have been the highlights of the last ten years? Any selection must, of course, be subjective, but here are some of mine.

The 2004-05 program on *Data Assimilation for Complex Systems* epitomized SAMSI's mission to "forge a synthesis of the statistical sciences and the applied mathematical sciences with disciplinary science to confront the very hardest and most important data- and model-driven scientific challenges." The topic in this case – data assimilation for numerical weather prediction – raised questions from fundamental science to complex data analyses, all surrounded by challenging computational issues. The program was unusually productive for combining the efforts of statisticians and applied mathematicians, resulting in a special issue of *Physica D* with 15 papers.

Another highlight was *High-Dimensional Inference and Random Matrices* (2006-07). This was also a very successful program for integrating mathematics and statistics, such as a link between multivariate distributions and geometric methods in one working group, or between statistics and integrable systems in another. Another innovation of this program was the successful setting up of a video link with Berkeley, through which two of the working groups were conducted. This program also resulted in a special issue of a journal, in this case *Annals of Statistics*, in which 9 papers were published under Peter Bickel as guest editor.

Turning to more recent times, 2011-12 saw the program on *Uncertainty Quantification* (UQ), which was the culmination of Pierre Gremaud's stint as our first Deputy Director. This was another strong program for collaborations between applied mathematicians and statisticians, for establishing strong interactions with national laboratories, and for SAMSI's high level of involvement in the first national conference on UQ, held in Raleigh.

Throughout the decade, SAMSI has remained faithful to its original mission, but the structure of its programs has evolved considerably. One innovation was to hold summer programs, beginning in 2006 – since then we have held at least one summer program each year except 2011. This year there were two – a program on *Nonlocal Continuum Models for Diffusion, Mechanics, and Other Applications* with a strong applied math flavor, followed by one on *Computational Advertising* that was much more statistical. The latter program led to three working groups analyzing data provided by Yahoo!, one to predict click-through rates on front-page news articles, a second on a recommender system for musical bands, and the third to model bidding behavior in auctions for internet advertisements.

Everyone at SAMSI is looking forward to the next ten years!

Richard Smith
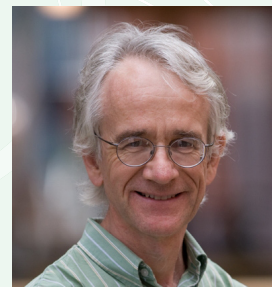
The official opening of SAMSI in 2002. Pictured L-R: James Moeser (UNC), Representative David Price, James Berger (SAMSI), Stuart Cooper (NCSU), Jon Kettenring (NISS) and Nan Keohane (Duke).

SAMSI Announces its 2013-2014 Programs

The two main SAMSI programs for 2013-14 will be Computational Methods in Social Science (CMSS), and Low-Dimensional Structure in High-Dimensional Systems (LDHD). The 2013 summer program will be Neuroimaging Data Analysis (NDA).

CMSS: As in many other disciplines, the social sciences have experienced an explosion in the size and complexity of the data being collected, which in turn, has led social scientists to examine statistical and computational methodology for handling such datasets. In addition, many statisticians and applied mathematicians are also focusing on social sciences in applications of their work, especially in applications to social networks and causal inference. Given these developments, the SAMSI program will focus on three major areas: 1) Social Networks; 2) Agent-Based Models and 3) New Methodology for Censuses and Surveys.

The program leaders for the Computational Methods in Social Science are: Robert Axtell (George Mason University), Elena Erosheva (University of Washington), Doyne Farmer (Oxford University and the Santa Fe Institute), Steve Fienberg (Carnegie Mellon University), Krista Gile (University of Massachusetts, Amherst), Mark Handcock (UCLA), Tian Zheng (Columbia University). Richard Smith is the Directorate Liaison.

LDHD: The program on Low-Dimensional Structure in High-Dimensional Systems (LDHD) is devoted to the development of methodological, theoretical, and computational treatment of high-dimensional mathematical and statistical models. Possibly limited amounts of available data pose added challenges in high dimensions. The program will address these challenges by focusing on low-dimensional structures that approximate or encapsulate given high-dimensional data. Cutting edge methods of dimension reduction will be brought together from probability and statistics, geometry, topology, and computer science. These techniques include variable selection, graphical modeling, classification, dimension reduction in matrix estimation, empirical processes, and manifold learning. Working groups during the program will include theoretical discussions of these tools as well as applications to image and signal analysis, graphs and networks, genetics and genomics, dynamical systems, and machine learning.

The program leaders for the LDHD program are: Florentina Bunea (Cornell), Peter Hoff (Washington), Chris Holmes (Oxford), Peter Kim (Guelph), Vladimir Koltchinskii (Georgia Tech), John Lafferty (U Chicago), Gilad Lerman (Minnesota), Sara van de Geer (ETH Zurich), Marten Wegkamp (Cornell) and Bin Yu (Berkeley). Ezra Miller is the Directorate Liaison.

NDA: Neuroimaging Data Analysis (NDA) encompasses a broad array of methods from imaging, mathematics, and statistics for analysis of high-dimensional, correlated, and complex data from cross-sectional and clustered neuroimaging studies. Such studies collect structural, neurochemical, and functional images as well as clinical and genetic data. Imaging modalities include computed axial tomography (CT), diffusion tensor imaging (DTI), ordinary and functional magnetic resonance imaging (MRI and fMRI), positron emission tomography (PET), and electroencephalography (EEG), among many others. Analyzing neuroimaging data presents both theoretical and computational challenges. This SAMSI summer program will bring together researchers from statistics, mathematics, computer science, biomedical engineering, psychiatry, psychology, neuroscience, radiology, and beyond.

There are several opportunities to participate in these programs. Financial support is available for visiting researchers to be resident at SAMSI for various time periods. Young researchers have special opportunities to participate that typically have a one year appointment. Several postdoctoral positions will also be funded for each SAMSI program. In addition, workshops and working groups provide the opportunity to collaborate with others on research projects and to network with their peers. Dedicated workshops will allow graduate and upper level undergraduate students to learn about the latest research and applications in the statistical and mathematical sciences. All involved researchers will get chances to broaden their interests and skill sets, participate in cutting edge interdisciplinary projects and make new connections. New researchers and members of underrepresented groups are especially encouraged to participate in SAMSI workshops and programs.

To find out more about any of these research programs, or to apply, go to the SAMSI website, www.samsi.info.

2012 IMSM - from solving robot mazes to automating skin allergy tests

Thirty five graduate students from 34 colleges and universities from around the United States spent eight days in July at NC State University working on real-life problems during the 19th Industrial Math/Stats Modeling (IMSM) workshop.

On the first day, industrial and government scientists presented six different problems to the students. Each presenter, along with a faculty mentor, then worked for the next several days with a team of 4-6 students to work on solutions to the presented problems.

This year's presenters included Agustin Calatroni, Jeremy Wildfire, and Herman Mitchell, Rho Inc.; Matthew Farthing, US Army Corps of Engineers; Haris Halilovic, Lord Corporation; Jordan Massad, Sandia National Laboratories; John Peach, MIT Lincoln Laboratory and Mark Wolff, SAS.

Projects varied from looking at saltwater intrusion on fresh water supplies in coastal aquifers to trying to create an automated measurement of allergic reactivity. Participants in the latter group got to collect data on the skin allergy test; this included a field trip to the US Environmental Protection Agency's lab at the University of North Carolina where the students, their mentors and a few other volunteers took allergy skin tests and had their arms photographed for the project.

On the last day of the workshop, students presented their work. From past experience, some of the work presented in these reports eventually led to other research and publications, some of the connections made during the workshop led to future collaborations and some of the students ended up seeking employment at the participating organizations.

Postdoc Profile: Chia Ying Lee
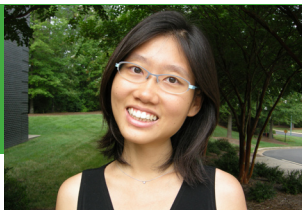
Chia Ying Lee has been busy this summer, advising a group of graduate students for the IMSM *(see story on page 3)* and going to some conferences, but she took some time to tell us more about her background.

Chia grew up in Singapore and received her high school degree there. While she enjoyed mathematics in high school she said that when she started to study math as an undergraduate student at the University of Michigan in Ann Arbor she started to be exposed to more sophisticated and elegant concepts in math, applied math and modelling. "This is when math caught my imagination. It was fun to be able to tackle open problems with powerful mathematical tools, and to try to develop new approaches to the problem," noted Chia.

Chia went to Brown University for her graduate degree and Ph.D. where she majored in applied mathematics. She heard that SAMSI was offering a program in uncertainty quantification last year, and she thought that the program would be a good continuation of her research interest, so she applied to be a postdoctoral fellow.

During her time at SAMSI, Chia has been working on studying techniques for simulating rare events that occur with very small probability. "For example, one can ask, what is the probability to have a very large number of triangles in a random graph that has edges which exist with probability p independently of one another? The challenge in simulating rare probabilities is that the events, being rare, seldom ever "show up." A common technique, importance sampling, makes rare events more frequent, but consequently each event is less "important", and thus requires a weight. I study the importance sampling method in various probabilistic models, including random graphs and stochastic dynamical systems," explained Chia.

Chia appreciates that she can meet researchers of many different areas and interests at SAMSI. Chia commented, "There are always lots of activities happening at once and opportunities to make contacts and begin collaborations. And the people here are incredibly nice!"

Chia's advice to the new postdocs starting in the Fall is to find a good project to work on, but allow themselves to see what other people in different research areas are doing. This academic year Chia will mainly be at UNC Statistics and Operations Research, where she is looking forward to teaching in the Fall. She will also be back at SAMSI from time to time, participating in the postdoc seminars.

When Chia is not working on her research projects, she enjoys playing piano and, after a long hiatus, she started to take part in an adult synchronized skating team at a local skating club!

SAMSI Has a New Blog!

In an effort to increase our communications, SAMSI has developed a blog, which can be found at: www.samsiatrtp.wordpress.com. The blog will feature statisticians and applied mathematicians from around the world who are involved with SAMSI programs and workshops. It will have first-hand accounts from some of the workshops, abstracts of some of the lectures, and highlights from some of our events. We hope this blog will be informative and entertaining . You play a role in this blog as well, because the comments and interactions from entries to the blog will be important to keeping the discussion going.

So, stay tuned to see upcoming blog entries and consider following our RSS feed. Any topics you think we should be covering? Send us your ideas!

Follow us! @NISSSAMSI

Calendar of Events for SAMSI

Opening Workshop on Data-Driven Decisions in Healthcare
August 26-29 2012
Research Triangle Park, NC

Opening Workshop on Statistical and Computational Methodology for Massive Datasets
September 9-12, 2012
Research Triangle Park, NC

Astrostatistics Workshop
September 19-21, 2012
Research Triangle Park, NC

Education and Outreach
Undergraduate Workshop focused on Computational Methodology for Massive Datasets
October 26-27, 2012
Research Triangle Park, NC

SAMSI-FODAVA Workshop on Interactive Visualization and Analysis of Massive Data
December 10-12, 2012
Research Triangle Park, NC

SAMSI-NCAR Workshop on Massive Datasets in Environment and Climate
February 13-15, 2013
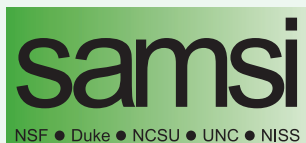National Center for Atmospheric Research, Boulder, CO.

Education and Outreach
Undergraduate Workshop focused on Data-Driven Decisions in Healthcare
February 22-23, 2013
Research Triangle Park, NC

SAMSI-SAVI Workshop on Environmental Statistics
March 4-6, 2013
Research Triangle Park, NC

samsi
NSF ● Duke ● NCSU ● UNC ● NISS

19 T. W. Alexander Drive
P.O. Box 14006
Research Triangle Park, NC 27709-4006