Annual Scientific Report

2002-2003

May 1, 2003

SAMSI Annual Scientific Report for 2002-2003

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, 2002 – June 30, 2003. Past and future activities of SAMSI are also discussed.

Executive Summary

SAMSI held its inauguration ceremony on September 3, 2002, attended by Congressman David Price, the leadership of Duke University, North Carolina State University, the University of North Carolina at Chapel Hill, the National Institute of Statistical Sciences and numerous local dignitaries and officials. The subsequent eight months have seen the initiation of major programs and SAMSI activities, as well as significant planning for the future. The following is a summary of the highlights from this first year and of the plans for the future.

First Year Programs and Activities

Focused Study Programs

- Inverse Problem Methodology in Complex Stochastic Models (Fall, 2002)
  - Tutorials and Opening workshop (9/21/02-9/24/02)
  - Closing workshop (5/14/03-5/15/03)
- Large Scale Computer Models for Environmental Systems (Spring, 2003)
  - Tutorials and workshop on Multiscale Modeling for Environmental Systems (2/2/03-2/7/03)
  - Workshop on Simulation and Optimization (4/28/03-4/40/03)
  - Workshop on Spatial-Temporal Modeling (6/1/03-6/6/03 at NCAR)

Synthesis Program

- Challenges in Stochastic Computation (Full year)
  - Tutorials and Opening workshop (9/25/02-10/1/02)
  - Midterm Special workdays (1/23/03, 1/30/03, 2/13/03, 2/20/03)
  - Closing workshop (6/26/03-6/28/03)

Education and Outreach

- Outreach Days for Undergraduates (November 9, 2002, and February 1, 2003)
- Undergraduate Interdisciplinary Workshop (June 9-13, 2003)
- Industrial Mathematical and Statistical Modeling Workshop for Graduate Students (7/21/03-7/30/03)
- Graduate Courses at SAMSI
  - Inverse problem methodology in complex stochastic models, Fall 2002
  - Multiphase transport phenomena, Spring 2003
  - Environmental statistics, Spring 2003

Distinguished Lecture Series

- Held the first Tuesday of each month

Program Planning Workshops

- Data Mining and Machine Learning (11/11/02)
- Multiscale Model Development and Control Design (12/17/02)

Future Program Schedule

Focused Study Programs

- Network Modeling for the Internet (Fall, 2003)
  - Tutorials and Workshop on Measurement, Modeling and Heavy Traffic (9/17/03-9/20/03)
  - Internet Tomography and Sensor Networks (10/12/03-10/15/03)
- Multiscale Model Development and Control Design (Spring, 2004)
  - Tutorials and opening workshop (1/17/04-1/21/04)
  - Workshop on Multi-scale Phenomena, Theory and Computation in Soft-matter, Nano-materials (2/15/04-2/17/04)
  - Workshop on Fluctuations and Continuum Equations for Granular Flow (4/15/04-4/17/04)

Synthesis Program

- Data Mining and Machine Learning (Full year)
  - Tutorials and Opening workshop (9/6/03-9/10/03)
  - Closing workshop (5/20/04-5/22/04)

Education and Outreach

- Graduate Courses at SAMSI
  - Internet Traffic, Fall 2003
  - Data Mining and Machine Learning, Fall 2003
  - Heavy Tailed Distributions, Fall 2003
  - Multiscale Modeling, Spring 2004

Tentative Plans for Later Years

- Computational Biology of Infectious Disease (Fall, 2004)
- Data Assimilation (Spring, 2005)
- Social Sciences (all year)
- Random Matrices (Spring, 2005)

Developments and Initiatives

First-Year Developments

- It was decided that one-semester programs should also have a significant level of activity in the other semester, either to continue the established activity of a Fall semester program, or to provide significant lead-in for a Spring program. Thus the Inverse program continued through the Spring, with closing workshop in May.
- Planning workshops were instituted as a way to provide focus for future programs.
- Program leaders were required to provide a Goals and Outcomes document, at the beginning of the program, to ensure that a clear focus has been established.

- It was judged beneficial to sometimes hold workshops away from SAMSI. Two examples are the Workshop on Spatial-Temporal Modeling, to be held at NCAR,and a Planning Workshop on Random Matrices to be held at ARCC.
- The NISS University Affiliates program was converted to a joint NISS/SAMSI University Affiliates program, and a number of new departments (including mathematics departments) were enrolled.
- Building modifications were made to create a better environment for visitors.
- A technical report series was started, and an agreement reached with SIAM/ASA to have a SAMSI series of publications.
- One-day outreach programs were initiated, and proved to be extremely popular.

Planned Future Developments

- The website will be redesigned.
- For Year 2, it was decided to move away from conventional workshops, and try a variety of activities more focused on interaction and discussion.
- Wireless networking will be installed at SAMSI and at the Governors Inn.
- Video taping of lectures for web dissemination will be implemented, if the technology has become simple and inexpensive enough.
- Another full-time staff person will be added, to deal with the increasing technical needs of SAMSI.

Table of Contents

I. Annual Progress Report

A. Program Personnel

1. Program and Activity Organizers

Program Organizers

Program	Name	Affiliation	Field
Inverse Problem Methodology in Complex Stochastic Models	Richard Albanese	AFRL, Brooks AFB	Medicine
	H.T. Banks (Co-Chair)	NCSU	Applied Math
	Marie Davidian (Co-Chair)	NCSU	Statistics
	Sarah Holte	Hutchinson	Biostatistics
	Joyce McLaughlin	Rensselaer Poly	Applied Math
	Alan Perelson	LANL	Biology
	George Papanicolaou	Stanford, NAC	Mathematics
	John Rice	UC Berkeley	Statistics
	Robert Wolpert	Duke	Statistics
Challenges in Stochastic Computation	Merlise Clyde (Co-Chair)	Duke	Statistics
	Jean-Pierre Fouque	NCSU	Mathematics
	Alan Gelfand	Connecticut	Statistics
	David Heckerman	Microsoft, NAC	CS and Stat
	Mark Huber	Duke	Probability
	Greg Lawler	Cornell	Probability
	Jun Liu	Harvard	Statistics
	John Monahan	NCSU	Statistics
	Michael Newton	Wisconsin Madison	Bioinformatics
	Scott Schmidler	Duke	Bioinformatics
	Mike West (Co-Chair)	Duke	Statistics
Large-Scale Computer Models for Environmental Systems	Mark Berliner	Ohio State	Statistics
	Montserrat Fuentes	NCSU	Statistics
	William Gray	Notre Dame	Geosciences
	Gabriele Hegerl	Duke	Meteorology
	Sallie Keller-McNulty	LANL, NAC	Statistics
	C. Tim Kelley	NCSU	Applied Math
	Andrew Madja	Courant Institute	Applied Math
	Richard McLaughlin	UNC	Applied Math
	Cass T. Miller	UNC	Environment
	Doug Nychka	NCAR	Geostatistics
	Richard L. Smith (Chair)	UNC	Statistics
Education & Outreach	H.T. Banks (Chair)	NCSU	Applied Math
	Johnny Houston	Elizabeth City State	Math and CS
	Rachel Levy	NCSU	Mathematics
	J. Blair Lyttle	Enloe HS, Raleigh	Statistics
	Negash Medhin	Clark Atlanta	Mathematics
	Daniel Teague	NC School Science and Math, Durham	Mathematics
	Wei Feng	UNC-Wilmington	Math and Stat

Activity Organizers

Activity	**Name**
Inverse Opening Workshop - *September 21-24, 2002*	H.T. Banks Marie Davidian
Stochastic Computation Opening Workshop - *September 25-October 4, 2002*	Mike West Merlise Clyde
Education & Outreach One-Day Workshop - *November 9, 2002*	H.T. Banks Negash Medhin
StoCom One-Day Workshop on Contingency Tables - *January 23, 2003*	Ian Dinwoodie
StoCom One-Day Workshop on Model Selection - *January 30, 2003*	Merlise Clyde
Education & Outreach One-Day Workshop - *February 1, 2003*	H.T. Banks Negash Medhin
Environment Workshop on Multi-Scale Modeling - *February 2-7, 2003*	R. McLaughlin, Cass Miller, Richard Smith
StoCom One-Day Workshop on Graphical Models - *February 13, 2003*	Mike West
StoCom One-Day Workshop on Financial Modeling - *February 20, 2003*	Jean-Pierre Fouque
Environment One-Day Workshop on Data Assimilation - *March 25, 2003*	Richard Smith
Environment Workshop on Simulation & Optimization - *April 28-30, 2003*	William Gray, C.T. Kelley, Cass Miller
Inverse Closing Workshop - *May 14-15, 2003*	H.T. Banks Marie Davidian
Environment One-Day Workshop on Porous Media - *May 16, 2003*	W. Gray, C. Miller
Workshop on Spatial-Temporal Statistics – *June 1-6, 2003,* at NCAR	Montse Fuentes, Doug Nychka, Richard Smith
Education & Outreach Interdisciplinary Workshop for Undergraduates – *June 9-13, 2003*	H.T. Banks, Negash Medhin
Stochastic Computation Closing Workshop - *June 26-28, 2003*	Mike West Merlise Clyde
Education & Outreach Industrial Mathematical and Statistical Modeling Workshop for Graduate Students – *July 21-30, 2003*	H.T. Banks, Pierre Gremaud, Alan Karr, Negash Mehin, Ralph Smith

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:

F – *Faculty Release Person*, defined as an individual from a partner university of SAMSI who is accorded release time for participation in the SAMSI program.

G – *Graduate Student*, receiving a research assistantship from SAMSI.

P – *Postdoctoral Fellow*, receiving support 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 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.

Grey - is used to indicate individuals that were funded by partner university cost sharing.

I. Inverse Problems Methodology in Complex Stochastic Models

Name	Gender	Affiliation	Department	Designation
Ackleh, Azmy	M	Univ of Louisiana	Mathematics	T
Anderssen, Robert	M	CSIRO		V
Bardsley, Johnathan	M	North Carolina State U & SAMSI	Mathematics	P
Bayarri, M.J.	F	U of Valencia	Statistics	V
Bendsoe, Martin	M	Tech U of Denmark	Mathematics	T
Chiswell, Karen	F	North Carolina State U	Statistics	G
Davidian, Marie	F	North Carolina State U	Statistics	F
Genton, Yanyuan Ma	F	North Carolina State U	Mathematics	P
Gunzburger, Max	M	Florida State U	Mathematics	T
Holte, Sarah	F	Hutchinson Cancer Cen	Public Health Sciences	T
ter Horst, Enrique	M	Duke U	Statistics	G
Kojima, Fumio	M	Kobe U	CS & Engineering	V
Kunisch, Karl	M	U of Graz	Mathematics	V
Madsen, Kristen	F	North Carolina State U	Mathematics	G
Mueller, Peter	M	Anderson Cancer Cen	Biostatistics	T
Ring, Wolfgang	M	Institute of Math	Mathematics	V
Rosner, Gary	M	Anderson Cancer Cen	Biostatistics	T
Walsh, Danny	M	SAMSI	Mathematics	P

II. Challenges in Stochastic Computation

Name	Gender	Affiliation	Department	Designation
Caffo, Brian	M	Johns Hopkins U	Biostatistics	T
Carmona, Rene	M	Princeton U	Operations Research & Financial Engineering	T
Carter, Christopher	M	Hong Kong U of Science & Tech		U
Chen, Ming-hui	M	U of Connecticut	Statistics	T
Clyde, Merlise	F	Duke U	Statistics	F
Cripps, Ed	M	U of New South Wales	Mathematics	V
Dinwoodie, Ian	M	Tulane U	Mathematics	U
Dobra, Adrian	M	Duke, NISS & SAMSI	Statistics	P
Doucet, Arnaud	M	Cambridge U	Engineering	T
Fouque, Jean-Pierre	M	North Carolina State U	Mathematics	F
George, Edward	M	U of Pennsylvania	Statistics	T
Godsill, Simon	M	Cambridge U	Engineering	V
Hans, Christopher	M	Duke U	Statistics	G
Huber, Mark	M	Duke U	Mathematics	F
Ibrahim, Joe	M	U of North Carolina	Biostatistics	F
Jones, Beatrix	F	SAMSI		P
Kohnen, Christine	F	Duke U	Statistics	G
Lee, Beom	M	U of North Carolina	Statistics	G

Madigan, David	M	Rutgers U	Statistics	T
Mallick, Bani	M	Texas A&M U	Statistics	V
Molina, German	M	Duke U	Statistics	G
Natvig, Bent	M	U of Oslo	Mathematics	V
Nicholas, Michael	M	Duke U	Mathematics	G
Paulo, Rui	M	SAMSI & NISS		P
Polson, Nicholas	M	U of Chicago	Graduate School of Business	T
Sebastiani, Paola	F	U of Massachusetts	Mathematics & Statistics	T
Tavare, Simon	M	U of Southern California	Computational & Molecular Biology	V
West, Mike	M	Duke U	Statistics	F
Wolfe, Patrick	M	Cambridge U	Engineering	V
Yoshida, Ruriko	F	U of California, Davis	Mathematics	T

III. Large-Scale Computer Models in Environmental Systems

Name	Gender	Affiliation	Department	Designation
Anderson, Dan	M	George Mason U	Mathematical Sciences	T
Berliner, Mark	M	Ohio State University	Statistics	T
Bourlioux, Anne	F	U of Montreal	Mathematics & Statistics	T
Bronski, Jared	M	University of Illinois	Mathematics	T
Caragea, Petruta	F	U of North Carolina	Statistics	G
Dawson, Clint	M	U of Texas at Austin	Aerospace Eng & Engineering Mechanics	T
Dennis, John	M	Rice U	Computational & Applied Mathematics	T
Fuentes, Montserrat	F	North Carolina State U	Statistics	F
Gray, William	M	U of Notre Dame	Civil Eng & Geological Sciences	U
Higdon, Dave	M	Los Alamos National Laboratory	Statistical Sciences Group	V
Hilpert, Markus	M	Johns Hopkins U		T
Ide, Kayo	F	U of CA, Los Angeles	Atmospheric Sciences	T
Kavanagh, Kathleen	F	North Carolina State U	Mathematics	G
Kelley, C. Tim	M	North Carolina State U	Mathematics	F
Kuensch, Hans	M	ETH Zentrum	Statistics	V
Laurie, Henri	M	U of Cape Town		V
Mattingly, Jonathan	M	Inst for Advanced Study		T
Mayer, Alex	M	Michigan Tech U	Geology	T
McLaughlin, Richard	M	U of North Carolina	Applied Mathematics	F

Mezic, Igor	M	U of CA, Santa Barbara	Mechanical & Environmental Eng	T
Miller, Casey	M	U of North Carolina	Environmental Sciences & Engineering	F
Negrete, Diego del Castillo	M	Oak Ridge National Laboratory	Complex Systems Dynamics	T
Nychka, Doug	M	NCAR		T
Pan, Doris	F	U of North Carolina	Statistics	G
Poje, Andrew	M	College of Staten Island	Mathematics	T
Russell, Thomas	M	U of Colorado	Mathematics	V
Sahu, Sujit	M	U of Southampton	Mathematical Studies	V
Sampson, Paul	M	U of Washington	Statistics	T
Smith, Richard	M	U of North Carolina	Statistics	F
Stein, Michael	M	U of Chicago	Statistics	T
Tebaldi, Claudia	F	NCAR		T
Wikle, Christopher	M	U of Missouri	Statistics	T
Yu, Jie	F	SAMSI	Mathematics	P
Zheng, Xiaoyu	F	U of North Carolina	Mathematics	G
Zidek, Jim	M	U of British Columbia	Statistics	V

3. Summary of Activity Participants*

		Underrepresented Groups		
Activity	**# Participants**	**# Female**	**# African-American**	**# Hispanic**
Inverse Opening Workshop	122	39	5	11
Stochastic Computation Opening Workshop	129	40	2	8
Education & Outreach One-Day Workshop - Nov 2002	31	14	16	0
StoCom One-Day Workshop on Contingency Tables	23	no formal registration		
StoCom One-Day Workshop on Model Selection	27	no formal registration		
Education & Outreach One-Day Workshop - Feb 2003	30	13	20	0
Environment Workshop on Multi-Scale Modeling	84	26	2	4
StoCom One-Day Workshop on Graphical Models	31	no formal registration		
StoCom One-Day Workshop on Financial Modeling	22	no formal registration		
Environment One-Day Workshop	26	no formal registration		
Environmental Workshop on Simulation & Optimization	49	18	3	0
Inverse Closing Workshop	scheduled for May 14 & 15: will be reported next year			
Environment One-Day Workshop in Porous Media	scheduled for May 16: will be reported next year			
Stochastic Computation Closing Workshop	scheduled for June 26-28: will be reported next year			

*Participant lists for workshops are given in Appendix A.

B. Postdoctoral Fellows

The SAMSI Postdoctoral Fellows, for 2002-2003, with a brief synopsis (details available in Sections B.2 and B.3 below) of their activities were:

Johnathan Bardsley, participated in the Inverse Problems Program, collaborated with H. T. Banks on inverse problems in electro-magnetics on uncertainty in estimation of polarization, gave six talks, and has two papers in progress.

Adrian Dobra, participated in the Categorical Data and Graphical Modelling subgroups of the Stochastic Computation Program, collaborated with Y. Chen, I. Dinwoodie, E.A. Erosheva, S.E. Fienberg, M. Huber, B. Jones, A. F. Karr, A. P. Sanil, C. Tebaldi and M. West, gave two talks and has seven papers at various stages of development.

Beatrix Jones, participated in the Large Scale Graphical Models group of the Stochastic Computation Program, collaborated with M. West, A. Dobra, C. Carter and C. Hans on stochastic algorithms, and has 3 papers at various stages.

Yanyuan Ma, participated in the Inverse Problems Program, collaborated with H. T. Banks, M. Davidian, M. Genton, L. Potter and A. Tsiatis on MCMC on PBPK models and on semiparametric methods, gave several talks and has 4 papers in progress or submitted.

Rui Paulo, participated in the Model Selection subgroup of the Stochastic Computation Program, collaborated with M. Clyde on comparing methods of computing marginal likelihoods and Bayes factors, gave a talk, and has several papers in progress.

Daniel Walsh, participated in Inverse Problems Program, collaborated with H. T. Banks, J. O. Berger and J. Sacks, on Proper Orthogonal Decomposition for PDE and Bayesian model dimension reduction, gave a talk and has 3 papers at various stages.

Jie Yu, participated in the Environmental Program, collaborated with Y. Chao, C. Jones, K. Ide, S. Lozier and R. MacLaughlin on Lagrangian Data Assimilation in oceanographic studies and flows in porous media, gave three talks, and has a paper in progress.

The SAMSI Postdoctoral Fellowship experience included opportunities for collaboration in the SAMSI spirit of bringing together Statisticians and Applied Mathematicians. These opportunities came during the SAMSI Workshops, and during the Working Groups that met weekly at SAMSI, and from the SAMSI courses.

To enhance contact between SAMSI and NISS Postdoctoral Fellows, particularly those participating in different programs, as well as to provide a forum for informal communication with the Directorate, a monthly pizza lunch was held. These lunches started with conversation on SAMSI questions, cares and concerns (on the parts of both the Postdoctoral Fellows, and also the Directorate). The discussion frequently then moved on to a wide range of “scientific folklore” matters such as scientific cultural differences between statistics and applied mathematics, how the academic job market works, how to write papers and grant proposals, and how to make a high quality verbal presentation (e.g. relative merits of methods used by various SAMSI Distinguished Lecturers were discussed). Finally, there was a presentation by one of the postdoctoral fellows on their own current research.

Effective mentoring of postdoctoral fellows is an important SAMSI goal. A SAMSI mechanism to ensure that each postdoctoral fellow had at least two people with whom they could personally discuss any concerns that might come up, was "double coverage" of mentoring assignments. This was done by assigning both a "scientific mentor" (usually the senior scientist most connected with the research) and an "administrative mentor" (a member of the Directorate, different from the scientific mentor), to each postdoctoral fellow. The mentoring assignments for 2002-2003 are given in Section B.1.

To assess performance of SAMSI in terms of the overall postdoctoral experience, a Postdoctoral Questionnaire was used in April 2003. This was constructed with input from the postdoctoral fellows (during some of the postdoc pizza lunches). The questions and answers from each postdoctoral fellow can be found in Section B.4 below. The single clearest impression from these is that overall the postdoctoral fellows were very happy with their SAMSI experience, and feel that it has given substantial value added to their careers. However, as expected, there was some variation in the experience. One postdoctoral fellow was slow to get started (apparently due to some uncertainty caused by splitting the mentoring role in too many ways). This problem will be addressed in the future by ensuring that each postdoctoral fellow has a single primary scientific mentor, who understands and accepts the responsibility that this entails. This evaluation process turned up a second area where improvements will be made next year: the Administrative Mentor system. The questionnaire revealed that the majority of postdoctoral fellows either did not list an Administrative Mentor, or else listed the wrong person. This did not present a serious problem this year, because the scientific mentoring quality was generally very high, but a clear lesson is that for this to be a useful safety net for keeping people well mentored, improvements are needed. The plan for next year is to provide reminders to each administrative mentor to meet twice each semester with their mentees.

As another means of assessing the quality of the SAMSI experience, the Scientific Mentors were asked to comment on each of the Postdoctoral Fellows. These reports are in Section B.5 below. Again the overall impression is very positive. It is clear that the Postdoctoral Fellows have made very important contributions to SAMSI. Another good indication is that the situation of the Postdoctoral Fellow with initial difficulties was recognized, and corrective steps were taken.

An important part of the SAMSI Postdoctoral Fellowship Program is a stated goal of providing two year positions. This is important because for positions of one year or less, the postdoctoral fellow must start looking for the next at nearly the same time as starting the present one. However, because of the one half to one year lifetime of SAMSI programs, this has taken some effort to implement. It has been done by sharing addition postdoctoral fellowship years, in a variety of different ways, with other institutions around the Research Triangle.

In summary, the SAMSI Postdoctoral Program has been generally very successful in this first year. The postdoctoral fellows have been making well appreciated contributions to their programs, and been gaining valuable career skills for themselves. As expected with any new program there are areas in need of improvement, but these have been identified, and plans for addressing them next year have been made.

Action items for improvement of the SAMSI Postdoctoral Program in upcoming years:

- Remind Administrative Mentors to meet twice each semester with assigned postdoctoral fellows.
- Always have a primary scientific mentor (who understands the responsibility), even when it is scientifically sensible for several mentors to be involved.
- Discuss these topics, together with an overview of what SAMSI is all about, at the first SAMSI postdoc lunch each year.

1. Postdoctoral Mentoring Assignments

	Scientific Mentor	Administrative Mentor
Johnathan Bardsley	H. T. Banks	J. S. Marron
Adrian Dobra	M. West	M. Clyde
Beatrix Jones	M. West	A. F. Karr
Yanyuan Ma	M. Davidian	H. T. Banks
Rui Paulo	M. Clyde	J. O. Berger
Daniel Walsh	J. O. Berger	A. F. Karr
Jie Yu	C. Jones	J. S. Marron

2. PostDoctoral Fellows Mid-Year Activity Reports

These reports were written by each postdoctoral fellow, in December 2002.

Johnathan Bardsley: SAMSI post-doc Johnathan M. Bardsley is a part of the Inverse Problems group at SAMSI. He is working with Dr. H. T. Banks on problems in Electromagnetics. A portion of the work he has done this fall culminated in a paper with Dr. Banks titled "Wellposedness for a Second Order System Arising in a Time Domain Electromagnetic Scattering Problem Using Perfectly Matched Layers". Final changes are being made to this paper, which will be submitted in the first part of 2003. In October, Johnathan gave a talk on this work at SAMSI during a meeting with researchers in electromagnetics from around the world. The goal of his current work on this problem is to use perfectly matched layers for solving two dimensional electromagnetics inverse problems. He gave a talk on this work on December 12 in San Antonio, Texas, at a meeting with the electromagnetics group at Brooks Air Force Base headed by Dr. Richard Albanese.

On December 3, Johnathan gave a Colloquium Talk in the Departments of Mathematics and Computer Science at Wake Forest University on the importance of statistical considerations in astronomical imaging. The title of the talk was "Statistics, Constrained Optimization, and Astronomical Imaging".

Adrian Dobra: My research activities at SAMSI in the fall semester of 2002 can be classified in three categories:

- I continued and finalized the projects I was involved in at the National Institute of Statistical Sciences. These projects were focused on disclosure limitation methods for categorical data and have generated the publications [2, 4, 3] as a result of my e
orts in this period.
- I was involved in the \Categorical Data" group of the SAMSI Stochastic Computation project. I have contributed to the publications [5, 1] that have emerged for the activities in this group. Moreover, together with Ian Dindoodie, I have developed a program to compute a Markov basis for an irreducible symmetric Markov chain on multi-way table of counts constrained with arbitrary fixed marginals and with a set of cells having fixed counts. This program is available online at http://www.stat.duke.edu/ adobra/MarkovBasesAlgebra/.
- I was also involved in the \Graphical Models" group of the SAMSI Stochastic Computation project. Together with Mike West, I have developed a stochastic algorithm for determining relevant decomposable graphical models with thousands of variables. We have applied this search strategy on gene expression datasets. In the Spring 2003 I will continue the activities I mentioned above.

References

[1] I. Dinwoodie, M. Huber, Y. Chen, and A. Dobra. Simulation from triangular tables with hardy-weinberg constraints. Manuscript, 2002.

[2] A. Dobra, E.A. Erosheva, and S.E. Fienberg. Disclosure limitation methods based on bounds for large contingency tables with application to disability data. In H. Bozdogan, editor, Proceedings of Conference on the New Frontiers of Statistical Data Mining. CRC Press, 2002.

[3] A. Dobra, A.F. Karr, and A.P. Sanil. Preserving confidentiality of high-dimensional tabulated data: Statistical and computational issues. Statistics and Computing, 2003. to appear.

[4] A. Dobra, A.F. Karr, A.P. Sanil, and S.E. Fienberg. Software systems for tabular data releases. International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems, 10:529{544}, 2002.

[5] A. Dobra, C. Tebaldi, and M. West. Distribution functions over spaces of tables. Manuscript, 2002.

Beatrix Jones: I've been working with Mike West, Adrian Dobra, Chris Carter and Chris Hans on stochastic algorithms to explore the space of Gaussian graphical models in high dimensional data sets. A "Gaussian graphical model" is just a Gaussian model for a set of variables where the conditional independence structure of the model is represented by a graph; each node corresponds to a variable, and if 2 nodes do not have an edge between them, those 2 variables are conditionally independent given the rest of the variables. Our research focuses on inferring the graphical structure: additional parameters that describe the covariance matrix conditional on the graphical structure are marginalized over. We have also restricted our research to decomposable graphical models. A decomposable graph does not have any chordless cycles larger than size 3. This subset of models is particularly computationally tractable.

My specific program related activities have been to translate (from matlab to C++) a Gibbs sampler that moves over the space of decomposable graphical models, and to develop a Metropolis-Hastings sampler for the same purpose. We are comparing these two samplers with other stochastic search algorithms that do not sample from the posterior of model space. We are also giving considerable attention to selection of prior distributions that favor simple graphical structures. A uniform prior over decomposable graphs favors "medium sized" graphs (if there are N nodes, roughly N^2/4 edges). When a high dimensional data set is being considered, this is often much too complex. We are exploring strategies for specifying priors that penalize the number of edges, encouraging on the order of N edges per graph. (Tree structures correspond to N-1 edges).

In non-program activities, I have prepared a revisions of papers submitted to Biometrics and Genetics. The Genetics paper, "Bayesian Sperm Competition Estimates," written with A. G. Clark, has been accepted. I have also acted as a reviewer for the journal "Mathematical Biosciences".

Yanyuan Ma: People I have been working with: HT Banks, Marie Davidian, Marc Genton, Laura Potter. I have been working on using MCMC on PBPK models, and studying the departure from normality of the random effects.

Paper in Progress: " A Simulation Based Comparison Between Parametric methods, and Semiparametric Methods in a PBPK model". With HT Banks and Laura Potter.

Paper submitted: "A Semiparametric Class of Generalized Skew-Elliptical Distributions". With Marc Genton, submitted to Biometrika.

Work in Progress: try to implement Generalized Skew Elliptical Distributions to random effects models. It will be with Marie Davidian and Marc Genton if it finally works out.

SAMSI Duties: attending SAMSI seminars and classes.

Rui Paulo: As a SAMSI Postdoctoral Fellow, I have been mainly working in the Stochastic Computation program, more specifically in the Model Selection group. The leader of the group is Prof. Merlise Clyde, and as such I have been working essentially under her guidance. Nonetheless, the group as a whole has been extremely active, and as a consequence the orientation of the research has been a joint effort where everyone, from graduate students to project leaders, has been playing a determinant role.

We have decided to focus attention initially on the Linear Multiple Regression problem, because it is perhaps one of the most widely used models in practice, it is mathematically tractable to a certain extent, and it is simple enough so that one can expect to deeply understand the problems and proposed solutions. Not surprisingly, it has been extensively addressed in the literature, and therefore it makes a perfect ground for the synthesis goals of the group.

The main aspects of this problem as far as model selection and model averaging go are prior specification on the model-specific parameters and search algorithms designed to explore the model space when the number of covariates is large.

We started by surveying the types of prior distributions on the model-specific parameters that have been used so far in the literature. The main goal of this research is to address stochastic computation issues, but we believe that it makes little sense to compare algorithms if they are computing a solution that is not the most appropriate one.

We noted that the majority of the papers utilizes some variation or another on a conjugate prior, and in particular Arnold Zellner's g-prior is perhaps the most widely 1 used. The reasons for that popularity are clear: the conjugate form allows for an explicit formula for the marginal under each competing model, which substantially decreases the amount of computation needed, and the form of the prior covariance matrix, mimicking that of the likelihood, is extremely convenient. Nonetheless, this type of prior is subject to a series of criticisms, one of them being the lack of consistent as the evidence against the null model becomes overwhelming.

An alternative to the g-prior that overcomes this paradox is the one developed by Zellner and Siow (1986). It is a Cauchy prior, and therefore can be written as a scale mixture of Normals, the mixing distribution being an Inverse-Gamma. As it turns out, this prior specification can be seen as a placing a g-prior on the regression parameters and putting an additional prior on the hyperparameter g, in this case an Inverse-Gamma.

This observation prompted the idea of using different mixing distributions on this hyperparameter g, and thus defining alternatives to the g-prior and to the Zellner-Siow prior that perhaps achieve better performance.

As we mentioned above, the availability of closed form expressions for the marginals under each of the competing models is a very attractive property. Simple manipulations make it clear that if we utilize mixing densities of the form (1 + g)_a=2, the resulting marginal can be expressed using the hypergeometric function, for which very efficient numerical routines exist. The properties of the resulting priors were extensively studied.

The seemingly disadvantageous aspect of the Zellner-Siow prior of not allowing for closed form expressions for the marginals has been alleviated by the derivation of a very accurate Laplace-type approximation that can be used as an alternative to computing the one-dimensional integral numerically.

Parallely, different search algorithms are being used to compare the behavior of the considered priors, and the algorithms themselves are being studied and compared at the same time.

I have been keeping track of the research efforts by maintaining a set of working notes that have been distributed within the group and that are constantly growing and being perfectionated.

Daniel Walsh: I've been working with Tom Banks, investigating the use and implementation of Principal Orthogonal Decomposition (POD) techniques with regard to reduced order modeling of Partial Differential Equations.

I have also been working with Jim Berger and Jerry Sacks investigating the applicability of the aforementioned POD techniques to reduce the computational demands of fitting deterministic computer models in a Bayesian fashion.

In non-SAMSI related activities I have submitted a paper on using Partial Bayes factors to distinguish between different point process models.

3. Postdoctoral Fellows Year-End Activity Reports

These reports were written by each postdoctoral fellow, in May 2003

Johnathan Bardsley: The focus of my research while I have been at SAMSI has been inverse problems in electromagnetics. The solution of such problems requires the ability to obtain fast and accurate forward solutions of Maxwell's equations. It is important to know that forward solutions are accurate and stable. In the fall of 2002, I worked on proving that a particular weak formulation for solving Maxwell's equations in two dimensions was well-posed. This work is contained in a paper in progress with H. T. Banks, "Wellposedness for a Second Order System Arising in a Time Domain Electromagnetic Scattering Problem Using Perfectly Matched Layers".

Since the beginning of this year, my focus has changed. I spent the months of January and February developing a forward solver for Maxwell's equations in two dimensions. I used the finite difference time domain method (FDTD) and perfectly matched layer (PML) absorbing boundary conditions. Using this forward solver, I have since then worked on the inverse problem of parameter identification for a Debye medium with unknown parameters. This work is near completion and will be contained in the forth coming manuscript with H. T. Banks, "Electromagnetic Parameter Identification for a Debye Polarization Model".

I hope to wrap up all of this work some time this summer.

TALKS:
- "Time Domain Electromagnetic Scattering Problem Using Perfectly Matched Layers," October 2002, SAMSI, Electromagnetics Meeting.
- "Statistics, Constrained Optimization, and Astronomical Imaging," Colloquium Talk, Departments of Mathematics and Computer Science, Wake Forest University, Winston-Salem, NC, December 2002

- "The Banks, Browning PML Formulation: Should we use it to solve the Inverse Problem," San Antonio, Meeting with Electromagnetics Group at Brooks Air Force Base, December 2002.
- "An Inverse Problem in Atmospheric Imaging," Colloquium Talk, Departments of Mathematical Sciences, University of Montana, Missoula, MT, April 2003.
- "The Scattering of Electromagnetic Radiation by Dispersive Dielectrics: The FDTD Method," University of Paris VI, Paris, France, April 2003.
- "2D Electromagnetic Parameter Identification for a Debye Polarization Model," SAMSI Inverse Problems Final Workshop, May 2003.

SERVICE: I met with several of the post-doc candidates, but I can't remember their names. I met with the statistician who is faculty at Ohio State (a black fellow), the grad student from Iowa State (for the control program), your post-doc at UNC from S. Korea, the Chinese fellow from Princeton (for the control program), and I will be meeting with the candidate next week.

JOB: I will be starting a job this coming fall at the University of Montana as a tenure track faculty member. This job was at the very top of my wish list, so I am very excited. Undoubtedly, my connection with SAMSI was an important factor in me getting this job. Thank you SAMSI and SAMSI directorate!!

Adrian Dobra: My research activities at SAMSI in the spring semester of 2003 were as follows:
- On January 23, 2003, I gave the talk \Bayesian Inference in Incomplete Multi-way Tables," at the Mid-term Workday for Contingency Tables, Stochastic Computation Program.
- On February 13, 2003, I gave the talk \Stochastic search and optimization approaches," at the Mid-term Workday for Graphical Models, Stochastic Computation Program.
- I was involved in the \Categorical Data" group of the SAMSI Stochastic Computation project. We have finalized and submitted to Biometrika the paper \Monte-Carlo Algorithms for Hardy-Weinberg Proportions." We are in the process of writing another paper about exact methods for assessing the fit of models for contingency tables with ordered categories.
- I was also involved in the \Graphical Models" group of the SAMSI Stochastic Computation project. We have finalized and submitted to Journal of Multivariate Analysis the paper “Sparse Graphical Models for Exploring Gene Expression Data." We are in the process of writing and finalizing two other papers describing new results as well as computational issues and solutions for graphical model search in large, sparse datasets.
- I have actively participated in various activities at SAMSI.

Beatrix Jones: I have continued working with the large scale graphical models group of the stochastic computation program. This semester we have expanded the focus to include non-decomposable as well as decomposable graphical models. I gave a talk on our work with decomposable models at the mid-term workshop in February. At the end of June, I will speak in our closing workshop about the additional computational difficulties of considering non-decomposable models. We have also started preparing a paper on our work. Our research involves "scaling up" previous approaches to this problem so it relies heavily on approaches already in the literature. Consequently, we are planning a paper that both reviews this literature, talks about our experiences applying these methods to higher dimensional datasets, and the new methods we have developed to deal with the computational and interpretational challenges of high dimensional data. This review paper will be submitted to Statistical Science. Finally, I have set up a contract (to begin in August) with the National Marine Fisheries Service to apply Small Area Estimation approaches to some of their data. This contract grew out of a contact made at a SAMSI meeting with Scott Nichols of NMFS.

Yanyuan Ma: Paper submitted: A semiparametric class of generalized skew-elliptical distributions with Marc Genton

Papers under preparation:
- Linear mixed effect models with skew elliptical random effects with Marc Genton and Marie Davidian
- A simulation based comparison between parametric and semiparametric methods in PBPK models, with Laura Potter and Tom Banks
- Semiparametric Efficiency bound for a type of selection models with Marc Genton and Butch Tsiatis

Talks: SAMSI postdoc lunch talk

Interview: Interviewed one postdoc candidate

Rui Paulo: After the one-day workshop that was held at SAMSI, and where we presented a talk entitled Zellner-Siow Priors in Variable Selection" summarizing some aspects of the first part of the research we have been developing, we decided to start addressing another set of issues in Stochastic Computation in the context of Model Selection.

In the face of several alternative statistical models, one in usually interested in the question of which is the best model, according to some criterion. In a decision-theoretic scenario, this would imply full specification of the loss structure but, in most cases, in a Bayesian framework, one is usually faced with the problem of computing the posterior probability of each of the competing models. This turns out to be a hard problem for which several solutions have been proposed in the literature.

We have been implementing most of these methods in the context of variable selection for the probit regression model. By doing so, we have been gaining a better understanding of the basic ideas behind each method, which has been allowing us to further uncover their limitations and advantages. We will next repeat the exercise for a more complicated model, possibly including random effects.

In June, we will have the opportunity to present some of the work developed during this year at an international workshop to be held in France and entitled “Fourth International Workshop on Objective Prior Methodology”.

Daniel Walsh: I have been working on the computer model validation project. I modified the software package GaSP to run on our system, in order to supply MLE's for our MCMC software. Attended a meeting in Detroit with GM engineers to discuss a new computer model dataset. I am currently investigating extending the computer model validation framework to extend to multivariate/functional data. I am giving a talk in May at the IPAM Applied Inverse Problems conference.

Jie Yu: My research activity is primarily on development and implementation of new ideas for Lagrangian Data Assimilation in oceanographic studies (with Prof. Chris Jones at UNC and Dr. Kayo Ide at UCLA). Because of the use of Lagrangian data, there are certain special features in the formulation of EKF (Extended Kalman Filter), which may lead to a considerable reduction of the computational cost. We have identified the possible ways to do so. Numerical experiments to test the ideas are on the way.

As a SAMSI postdoc, I have attended most of the seminars for the environmental program, as well as SAMSI distinguished lectures. I was on duty to take care of the equipment and set up poster session during the workshop (Feb. 2 - 7 and April 28 - 30).
I have also interacted with SAMSI visitors, such as meetings with postdoc candidate (Marcus Calhoun-Lopez) and with the distinguished lecture speaker (Grace Wahba).

In the past four months, I was invited to give three talks on my previous research. I acknowledged the full support of SAMSI to my present position. If you want the details, I will provide them.

C. Graduate Student Participation

I. Inverse Problem Methodology In Complex Stochastic Models

Graduate students Brian Adams, Brandy Benedict, Stacey Beun, Karen Chiswell, and Jeff Hood from North Carolina State University, and Enrique ter-Horst from Duke University participated in the Inverse program activities, including

- Regular research meetings every Tuesday, 10-12 AM at SAMSI;
- The SAMSI course on Inverse Problems;
- Assisting with the SAMSI Education & Outreach one day programs on November 9, 2002 and February 1, 2003.
- Assisting with the the SAMSI Education and Outreach Undergraduate Workshop, June 9-13, 2003: The students and postdocs, under the direction and supervision of Banks and Davidian, are currently preparing tutorial lectures in statistics and applied mathematics and designing experiments for the undergraduates who are attending the workshop.

The students who were supported by SAMSI, and their activities are as follows:

Karen Chiswell (NCSU Statistics, 100% SAMSI graduate student) is a third year graduate student in statistics at NCSU who is beginning the research phase of thesis work toward a Ph.D. She played a fundamental role in assessing uncertainty in HIV Structured Treatment Interruptions (STI) data the group obtained and analyzed during the year. As a result of her involvement in the SAMSI program, she will have both strong modeling/applied math and statistical components in her thesis work on PBPK models. Banks and Davidian will serve on her thesis committee.

Jeff Hood (NCSU Mathematics, SAMSI graduate student) is a first year graduate student in applied mathematics who took 4 graduate courses each semester in addition to participating in the activities above. His research activities focused on nonlinear ordinary differential equation models for the progression of HIV at the cellular level. He also carried major responsibilities for developing a set of preliminary notes (to eventually be published as a monograph) from the course taught by Banks and Davidian.

Enrique ter-Horst (Duke Statistics, 100% SAMSI graduate student) is a fourth year graduate student finishing his dissertation in financial mathematics. He worked with one of the program leaders, Robert Wolpert, on functional inverse problems, as well as participating in the activities listed above.

II. Challenges in Stochastic Computation

Chris Hans (Duke Statistics, 50% SAMSI graduate student): Chris is a full participant in all StoCom workshops, and is fully engaged in collaborative research on statistical graphical modelling and stochastic computation, and in presentations at workshops and undergraduate outreach meetings. Beyond the period of SAMSI support (September, 2002-May, 2003) he will be supported on non-SAMSI Duke research funds to continue research in this areas throughout the summer 2003 (and perhaps beyond, depending on his interests).

Christine Kohnen (Duke Statistics, 100% SAMSI graduate student): Christine is involved in variable selection for linear and generalized linear models. She has implemented a deterministic Bayesian best subset procedure in R that overcomes variable limitations of the Volinsky/Raftery S-Plus search, and that extends the previous program to include priors that are scale mixtures of normals. She has implemented several new Metropolis-Hastings algorithms for stochastic search, as well a novel orthogonal data augmentation algorithm for Bayesian model averaging.

Beom Lee (UNC Statistics, 100% SAMSI graduate student): Beom is interested in several areas such as Monte Carlo simulation, MCMC methods, and sequential importance sampling. He applies those statistical methods to a variety of stochastic volatility models on stock prices, interest rate term structures, and option prices. He is working with StoCom researchers, led by Chuanshu Ji, on an approximation method using central limit theorems in MCMC computation of option prices.

German Molina (Duke Statistics, 50% SAMSI graduate student): German is a full participant in both the model selection and financial mathematics working groups. In the former he has developed an adaptive stochastic search algorithm for model selection based on rejection sampling version of sampling without replacement. In the financial models component German is working on MCMC algorithms for estimation of multiple volatility scales. His research on these projects will be part of his dissertation and he expects to continue working on these topics both in the summer and after his defense in July.

Michael Nicholas (Duke Mathematics, 100% SAMSI graduate student): Nicholas is interested in probability in sampling, Monte Carlo methods, and complexity. He has developed a MATLAB implementation of our new sampling algorithm for Hardy-Weinberg equilibrium. Part of this involves implementing a known, efficient algorithm for hypergeometric sampling called the ``ratio-of-uniforms" approach by Stadlober and Zechner. This code should be made available to the public soon.

A number of non-SAMSI funded graduate students, in RTP and from elsewhere in the US and abroad were involved in research arising from the program. These include

- Carlos Carvalho (Duke Statistics), graphical models and methods of stochastic computation in nondecomposable models;

- Ruriko Yoshida and Raymond Hemmecke (U.C. Davis Mathematics), working with Professor Jesus DeLoera on lattice point enumeration arising from the contingency tables component through collaboration with core program faculty;
- Sean Han (NCSU Mathematics), working with StoCom researchers German Molina and Professor Fouque to use MCMC method for estimating stochastic volatility models with multiple stochastic volatility scales;
- Ingunn Tvete (University of Oslo, Norway, Mathematics), who participated in several working groups as a visiting graduate student working with her visiting professor Bent Natvig.

III. Large-scale Computer Models for Environmental Systems

Doris Pan (UNC Environmental Sciences, 100% SAMSI graduate student): Doris submitted two manuscripts for publication so far this year and defended her dissertation in April. She attended the course on multiphase transport phenomena taught jointly by Gray and Miller, as well as the multiscale and simulation and optimization workshops. She will be a speaker at the one-day closure workshop on 16 May 2003. Doris also made an outreach presentation and will help prepare materials for the June outreach meeting. Doris was active interacting with visitors and gained much from her SAMSI experience.

Kristen Madsen (NCSU Statistics, 100% SAMSI graduate student): Kristen has just started working on her dissertation research under the supervision of Montse Fuentes and Doug Nychka. Her dissertation topic has been inspired by discussions at SAMSI about the scientific and mathematical challenge of weather forecasting. She attended two of the SAMSI workshops; Multi-scale modeling and space-time modeling; she also took Richard Smith's course on environmental statistics. This course has helped her to understand better the methodology of spatial statistics that she will use in her thesis. She participated in the Combining Networks group and attended all the environmental seminars on Wednesdays.This Summer she will spend sometime at NCAR to continue the research on probabilistic ensemble forecasting that she started under the SAMSI umbrella.

Petrutza Caragea (UNC Statistics, 100% SAMSI graduate student): Dr. Caragea completed her PhD thesis entitled "Approximate likelihoods for spatial processes" during the Spring 2003 semester, and also went on the job market (she will be starting as an assistant professor of statistics at Iowa State in the fall). During the course of her graduate student career she has also been extensively involved in EPA collaboration with Dr. David Holland, on the analysis of spatial and temporal trends in atmospheric sulfate data. Her involvement with the program at SAMSI has benefited both sides of her work, methodological research about spatial processes and applications to a variety of environmental areas. Specifically, she has attended the weekly group meetings, will attend the Boulder workshop, and has also been actively involved with two of the working groups: Spatial-Temporal Statistics (including Dave Holland's subgroup on network design) and Combining Networks.

Xiayu Zheng (UNC Mathematics, 100% SAMSI Graduate Student): Xiaoyu's thesis topic is concerned with multiple scales in complex fluids, which shares many common themes with the activities at SAMSI. First, Xiaoyu has benefited from the course of Gray and Miller on closure schemes for porous media predictions. Xiaoyu has been studying kinetic and mesoscopic mean-field models for flowing macromolecules, where one has the similar problem of closure schemes and the inherent sensitivity associated with them.
Second, Xiaoyu has benefited from the statistical and probabilistic concepts and methods represented all semester by some of the leading scientists in the world. These ideas and tools have already influenced her research; this semester she has undertaken a statistical study of expected time of convergence to bulk attractors in simple shear flow of nematic polymers. She has developed numerical codes for stochastic molecular orientational simulations in shear flow, as well as kinetic simulations for the probability distribution function, and illustrated their consistency.

A number of non-SAMSI funded graduate students were involved in activities of the program. These include

- **Amy Nail** (NCSU Statistics): Ms. Nail is a first-year PhD student but came to the program with extensive past knowledge of EPA environmental monitoring work having been herself worked at the EPA for a couple of years. She is working with Marc Genton and Sujit Ghosh as joint advisors at NCSU. At SAMSI, she attended Richard Smith's course on environmental statistics and all three of the workshops. She has taken a very active role in the weekly group meetings including giving one of the 15-minute presentations and helping coordinate the spatial-temporal statistics working group that is being led by Jim Zidek. Her future work will benefit from SAMSI through the exposure both to a wide range of environmental statistics problems and some of the top experts in the methodology of spatial statistics.
- **Darryl Cooney** (NCSU Statistics): Mr. Cooney is a student of Montse Fuentes and has worked with Montse and Peter Finkelstein in the Combining Networks group, where his contribution has included an initial analysis of the EPA data. He attended the Workshop on Multi-Scale Modeling. His dissertation topic has been motivated by the scientific and statistical problems presented and discussed during the Combining Networks group meetings. EPA has offered him support to continue the research initiated with this SAMSI group.
- **Brian Lopes** (UNC Statistics): Mr. Lopes has just started his PhD dissertation research under the direction of Richard Smith. He attended Richard Smith's course in Environmental Statistics. He has attended the "source apportionment" working group and completed an initial spatial statistics analysis of some of the data studied in that group. He will give a poster presentation about this at the ISI Environmental Statistics meeting in Spain in July. He has also attended discussions with Dr. David Holland of the EPA about particulate matter modeling and associated network design problems. He is likely to write his dissertation in one of these areas though at the time of writing it is not clear which. Either way, the SAMSI program has been extremely valuable in exposing him to problems and techniques useful to his research.
- **Li Chen** (NCSU Statistics): Ms. Chen is working on space-time models for environmental processes. She attended two of the SAMSI workshops; Multi-scale

modeling and space-time modeling, she also attended Richard Smith's course on environmental statistics. This course has helped her to better understand the concepts, theories, and modeling techniques needed for her dissertation research. She presented a poster at each one of the two workshops mentioned above and that gave her the opportunity to share and discuss her current thesis research with top experts in the methodology of spatial statistics.

- **Arin Chaudhuri** (NCSU statistics): Mr. Chaudhuri has worked with Montse Fuentes and Dave Holland in the Network design group, where his contribution has included an analysis of EPA monitoring data. The discussions at the group meetings have provided unvaluable help and guidance in his current dissertation reseach on statistical methods for monitoring network design. He attended the Multi-scale modeling workshop and he presented a poster at the workshop. He also gave a 1 and 1/2 talk about his research to the network design group.
- **K. Culligan** - student of Bill Gray at Notre Dame
- **C. Abhishek** - student of Casey Miller at UNC-ESE
- **H. Li** - student of Casey Miller at UNC-ESE
- **C. Rupert** - student of Casey Miller at UNC-ESE
- **D. Johnson** - student of Casey Miller at UNC-ESE
- **J. Reese** - student of Tim Kelley at NCSU Math
- **K. Kavanagh** - student of Tim Kelley at NCSU Math
- **D. Sassan -** student in Computer Science at UNC, supported by Miller

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 Inverse Problem Methodology In Complex Stochastic Models

1.0 High-Level Summary: Estimating System Parameters

In many contexts it is necessary to infer parameters of some system from observation of the system itself, and often those parameters are random. For example, one of the most unique properties of HIV infection is its dramatically different courses in different individuals: an HIV-infected person has parameters, which vary over the population, that govern the progression of the disease. These parameters likewise impact the efficacy of treatments, and so design of effective treatments for HIV requires estimation of them. Complexities include measurement error, incomplete observations and the near-certainty that any mathematical representation of the system is incomplete or even incorrect.

The Inverse Problems program has merged complementary views from applied mathematics and statistics to develop novel techniques to use noisy, incomplete data to estimate parameters of complex systems. For HIV, these techniques yield improved therapies—structured treatment interruptions—that are as effective as other therapies but retard or even prevent development of drug resistance.

1.1 Introduction and Overview

In a diverse range of fields, including engineering, biology, physical sciences, material sciences, and medicine, there is heightening interest in the use of complex dynamical systems to characterize underlying features of physical and biological processes. For example, a critical problem in the study of HIV disease is elucidation of mechanisms governing the evolution of patient resistance to antiretroviral therapy, and there is growing consensus that progress may be made by representing the interaction between virus and host immune system by nonlinear dynamical systems whose parameters describe these mechanisms. Similarly, recovery of molecular information for polymers via light beam interrogation may be characterized by a dynamical system approach. Risk assessment is increasingly focused on deriving insights from physiologically-based pharmacokinetic dynamical systems models describing underlying processes following exposure to potential carcinogens.

In all of these applications, a main objective is to use complex systems to uncover such mechanisms based on experimental findings; that is, in applied mathematical parlance, to solve the relevant inverse problem based on observations of the system (data), or, in statistical terminology, to make inference on underlying parameters and model components that characterize the mechanisms from the data. In many settings, there is a realization that dynamical systems should incorporate stochastic components to take account of heterogeneity in underlying mechanisms, e.g., inter-cell variation in viral production delays in within-host HIV dynamics. In addition, heterogeneity may arise from data structure in which observations are collected from several individuals or samples with the broader focus on understanding not just individual-level dynamics but variation in mechanistic behavior across the population. In both cases, it is natural to treat unknown, unobservable system parameters as random quantities whose distribution is to be estimated. There is a large inverse problem literature for systems without such stochastic components; even here, forward solutions (i.e., solutions of the dynamics when parameters are specified) for complex systems often necessitate sophisticated techniques, so that inverse problem methodologies pose considerable challenges. Similarly, there is a vast statistical literature devoted to estimation and accounting for uncertainty in highly nonlinear models with random components, involving hierarchical specifications and complex computational issues.

Despite the potential overlap and emerging challenges, there has been little interaction between applied mathematicians and statisticians to develop relevant inverse problem/statistical inferential methodologies. When combined, the computational and theoretical hurdles posed by both mathematical and statistical issues are substantial, and their resolution requires an integrated effort. Thus, to advance both applied mathematical/statistical research developments and disciplinary science, there is an urgent need for a synergistic joint treatment of fundamental issues involved in fitting complex dynamical systems with stochastic components to data. The major thrust of the SAMSI Program on Inverse Problem Methodology in Complex Stochastic Models entails facilitating the essential cooperative effort required to catalyze collaborative research in this direction.

A crucial first step, to promote communication between members of these two communities who otherwise work independently and introduce each to the perspective and methodologies of the other, has already taken place in the Program Kick-off Workshop. Following the recommendations of the Workshop, short- and long-term future activities will be devoted to supporting and energizing focused collaboration between applied mathematicians, statisticians, and disciplinary scientists. Goals and anticipated outcomes by which success of the Program will be gauged are the forging of new research syntheses between applied mathematicians and statisticians to address timely disciplinary challenges, training of a "new breed" of mathematical-statistical scientist well-versed in the perspectives of both disciplines and poised to make immediate contributions, and dissemination of the new joint perspective at all levels of mathematical/statistical education. Details on these points are given in the following sections.

1.2 Program Activities and Goals

1.2.1 Program Leadership

The Program Leaders, H.T. Banks (Mathematics and CRSC, NCSU) and Marie Davidian (Statistics, NCSU), in conjunction with the Program Committee, are charged with guiding the conception and implementation of Program activities.

1.2.2 Kick-off workshop

The SAMSI Kick-off Workshop on Inverse Problems in Complex Stochastic Models was held September 21–24, 2002, with the goal of introducing the two communities to the research and perspectives of the other to foster new interdisciplinary research thrusts and collaborations. The program offered invited presentations by leading researchers involved in modeling of complex systems, statistical modeling and computation, and disciplinary science, including modeling of HIV dynamics and toxicokinetics (biology and medicine), dielectric materials (engineering), tomography (physics and medicine), and modeling of polymers (material sciences). Lectures provided overviews of the perspectives of applied mathematicians and statisticians in these contexts to demonstrate the need for an integrated framework. The first day featured introductory tutorials on differential equation modeling and statistical inference and modeling providing requisite background for members of each community to interpret and appreciate the perspectives of the other. The full program is given in Appendix B.

The Workshop attracted 122 participants, with 54 and 51 describing themselves as primarily "applied mathematician" or "statistician," respectively, including many prominent researchers in statistics, mathematics, pharmacokinetics, engineering, and physics. Moreover, 68 of the participants were graduate students, postdocs, or young investigators with less than 5 years after receiving a doctoral degree, demonstrating that the Program theme of crossing traditional math/stat disciplinary boundaries has intrigued the future generation of researchers so that there is potential for considerable long-term impact. The Workshop also attracted a diverse group of participants, including 39 females, 11 Hispanics, and 5 African-Americans. The tutorials were attended by 90 participants; a summary of attendance for each session of the main Workshop program is given in Appendix A.

The Workshop concluded with three Discussion Groups, each comprising participants from varied backgrounds and charged with, on the basis of the exposure provided by the Workshop, identifying important future directions for the Program and key issues not covered that should be included in subsequent Program activities, both of which will be instrumental in guiding continuing Program development (Sections 1.2.3 and 1.2.4). Groups also offered additional ideas related to the Workshop theme. A consensus from these discussions was that the Workshop spurred recognition that inverse problems are indeed problems of statistical inference and that even small amounts of uncertainty in data should be taken into appropriate account. Thus, participants saw inverse problems as being at the *real intersection* between applied mathematics and statistics, and emphasized that true collaboration, an activity that SAMSI could spearhead, must take place; however, such collaboration is hampered by a "language barrier." In fact, participants found this reflected in the presentations, which were either mostly statistical or mostly mathematical in perspective, although focused on reaching the other group, and identified possible overlaps in methods; e.g., with regularization techniques and Bayesian inference, that need further elucidation. Accordingly, participants stressed that future SAMSI activities should focus promoting a "common language" that draws from the perspectives of both sides, to foster identification of common themes for which one community could exploit the perspectives of the other and development of joint approaches that build on the best of both. Such activity would arise naturally from pursuit of specific, scientific problems involving difficult statistical and mathematical challenges, whose resolution could not be obtained from a purely applied math or statistical perspective alone but requires a true collaborative effort. The Program should seek to identify such scientific problems to drive methodological research and bring together individuals to work on them; ideally, such teams should include not only mathematical and statistical scientists, but disciplinary scientists and experimentalists. This activity could be followed by further workshops reporting on progress, featuring teams of collaborating mathematical/statistical researchers providing concrete demonstration of how individuals from each community gained new insights from working with those from the other and posing further questions requiring joint resolution. Ecology, finance, and meteorology were identified as further disciplinary areas, not covered in the Workshop, whose challenges could drive methodological collaboration. Detailed summaries of the discussions of each Group, compiled by three SAMSI Postdoctoral Fellows are presented in Appendix D.

Overall, the Workshop was viewed by participants and organizers as an extraordinarily successful first step that has raised awareness and provided a foundation for collaborative efforts.

1.2.3 Research activities

The program had a number of important short-term (1 week to 1 month) visitors including Robert Anderssen (CSIRO, Canberra, Australia; 3 weeks), Karl Kunisch (Institute of Math, Graz, Austria; 2 1/2 weeks), and Fumio Kojima (System Function Science, Kobe University, Kobe, Japan; 1 month), who each played an important role in the Kick-off Workshop (reports from Anderssen and Kojima are included in Appendix D). Following up on the Workshop, we hosted a number of short term visitors during the remainder of the Program, a number of whom requested return visits after participating in the 4 day Workshop. These include (approximate 1 week visits were hosted):

1. Susie Bayarri, Statistician, University Valencia, Spain (Sept.1 - Nov. 21, 2002). Dr. Bayarri is an expert on statistical methods for estimation problems. She participated actively in both the Inverse Problems and Stochastic Computation programs at SAMSI. She also actively participated in weekly Inverse Problem meetings as well the special course taught by Banks and Davidian.

2. Sarah Holte, Fred Hutchinson Cancer Research Center (Nov. 4 - 8, 2002). Dr. Holte is a biostatistician who is collaborating with both applied mathematicians and statisticians on HIV models and methodology development.

3,4. Peter Mueller and Gary Rosner, MD Anderson Cancer Center (Feb. 23 - 26, 2003). Drs. Mueller and Rosner are statisticians, and we coordinated their visits with that of Ackleh to facilitate collaborations on HIV modeling and data analysis.

5. Azmy Ackleh, Mathematics, University of Louisiana (Feb. 23 - 28, 2003). Dr. Ackleh expressed a strong desire to move his research more toward problems involving variability in data and uncertainty in models. He is collaborateing with mathematicians and statisticians in problem-specific investigations entailing use of MCMC methods for parameter estimation as well as combining deterministic with statistical approaches.

A number of other visitors (who were not able to attend the Kick-off Workshop) who are interested in SAMSI-facilitated collaborations, were hosted. These included:

6. Max Gunzburger, Mathematics, Florida State University (Nov. 11 - 12, 2002). Dr. Gunzburger is an expert in reduced order modeling techniques, including Proper Orthogonal Decomposition (POD) and Centroidal Voronoi Tessellations (CVT). He visited to collaborate with both applied mathematicians and statisticians on development

and use of these techniques in inverse problems, 3-D electromagnetic simulations and MCMC simulation methodologies.

7. M. Bendsoe, Mathematics, Technical University of Denmark (November 17 - 21, 2002). Dr. Bendsoe is an expert on shape optimization and design and visited to initiate long term collaboration on sensitivity analysis in complex systems depending on probability measures. Such sensitivity methodology will involve differentiation of solutions of nonlinear ordinary and partial differential equations with respect to probability distributions.

8. W. Ring, Institute of Math, Graz, Austria (Nov. 11 - 29, 2002). Dr. Ring is a young (pre-Habilitation) mathematician from Karl Kunisch's group in Graz who is an expert on level set methods in inverse problems. These methods are currently being developed and used with little or no acknowledgment or treatment of uncertainty in modeling or data. He initiated long term collaborations with both applied mathematicians and statisticians on the development of new methods to account for the the presence of variability in data in the context of inverse and imaging problems.

These visitors had an immediate impact on Program postdocs at SAMSI as well as the SAMSI Graduate Fellows assigned to the Inverse Problem Program.

In addition to the visitor activity, Banks and Davidian ran a seminar series followed by brown bag lunch research discussion session each Tuesday from 10:00 AM to 1:00 PM. Participants include local postdocs, students, and scientists from local labs and industry (Environmental Protection Agency, National Institute of Environmental Health Sciences, CIIT Centers for Health Research, GlaxoSmithKline, etc.).

Many of the non-students in these seminar/discussion sessions also sat in on the special topics course MA/ST 810Q, offered through NCSU with reciprocal enrollment through UNC-CH and Duke. This course, taught by Banks and Davidian on Wednesdays from 4:30-7:00 PM at SAMSI, targeted both applied mathematics and statistics graduate students. 23 students enrolled for credit, with an additional 15–17 post-Ph.D. non-students attending. The course description is given in Appendix E. An important component of the course that reflects the Program philosophy of collaboration was the assignment of students to interdisciplinary teams to complete course work that involves the need for integration of statistical and mathematical concepts.

A Closing Workshop, to be held at SAMSI, summarizing and disseminating the results of these short-term activities, was originally planned for early January 2003. In the interim, IPAM at UCLA announced plans for an Inverse Problem Program (with a different theme from the SAMSI focus on applied mathematics/statistics collaboration), to be held in Fall 2003, with a kick-off workshop planned for May 18–23, 2003, in Lake Arrowhead, California. To promote synergism between the SAMSI and IPAM programs and to encourage and facilitate participation in both workshops by European and other attendees, the Program Closing Workshop has thus been moved to May 14–16, 2003. Invited speakers include L. Sheiner (UCSF), C. Wikle (U. Missouri, Columbia), H. Engle (U. Linz, Austria), E. Somersala (Tech. Univ., Helsinke).

1.2.4 Long-term activities

Several activities planned in the Program were specifically oriented toward long-term results. These include the following.

1. On October 3, we organized an "Electromagnetics Day" at SAMSI. Attendees included Kojima, Banks, Ito (Professor, NCSU), Bardsley (SAMSI Inverse Program Postdoctoral Fellow), Gibson (NCSU graduate student), Joyner (Assistant Professor, West Georgia State University), Pinter (Assistant Professor, University of Wisconsin, Milwaukee), Kepler (SAMSI Program Postdoctoral Associate), Benedict (SAMSI Graduate Associate), and Bill Winfree and Buzz Wincheski (Nondestructive Evaluation Branch, NASA Langley Research Center). Presentations and discussions focused on modeling and data-driven computational methods for nondestructive electromagnetic interrogation of materials using signals in the microwave and terahertz frequency ranges. Long term plans for collaborations include use of model reduction techniques (POD, CVT) in 3-D imaging as well as inclusion of statistical methods to treat measurement and modeling uncertainties. Study of the use of MCMC/Bayesian methods in estimation problems is also planned. Participants have subsequently requested that this become an annual event.

2. The HIV collaborations begun this semester are part of a long term (> 4 years) effort on data-driven modeling and analysis of biological systems using statistical and applied mathematical methods led by Banks and Davidian. In addition to the HIV-specific research, special attention will be given to inverse problems arising in pharmacokinetic, toxicokinetic, and ecological problems. (Specific research teams have already been formed.) These investigations will involve biologists, applied mathematicians, and statisticians collaborating in a true example of the SAMSI philosophy of the coming together of domain scientists, applied mathematicians and statisticians to solve difficult data-driven problems for complex nondeterministic systems.

1.3 Anticipated Outcomes and Measures of Success

1.3.1 Program impact on research

The broad goal of the Program, and hence a key measure of its success, is the formation of real collaborations between applied mathematical and statistical scientists that have the potential to alter fundamentally the thinking in both disciplines and crystallize the realization that both communities share common objectives and interests that can best achieved by a joint effort. In particular, we envision that publications, conferences, and presentations that result from Program-driven collaborations will foster a broad awareness among applied mathematicians of the critical need to consider variation, uncertainty, and the connection between loss functions and statistical models of variable phenomena in the context of inverse problems. Likewise, we see statisticians learning to think outside the constraints of empirically-based models for biological and physical phenomena that focus on "what can be observed" and prediction rather than on uncovering underlying mechanisms that can be successfully described via mathematical constructs. A measure of how well these "culture shifts" have taken hold as a result of the Program will in be gauged initially by the appearance of joint-authored publications in mainline applied mathematical and statistical research journals that demonstrate not only advances in methodology that are possible from joint efforts but that highlight the potential for resulting greater advances in disciplinary science.

Indeed, ultimately, an important indicator of success of the Program will be with respect to its impact on the disciplinary science problems driving the joint mathematical/statistical modeling and inferential efforts. The broad goal is that Program-initiated collaborations will result in the publication and dissemination not only of innovative methodological advances but of new subject-matter insights, e.g., deeper understanding of the mechanisms underlying development of HIV drug resistance, that in turn suggest new hypotheses. Impact on advances in disciplinary science is of course difficult to quantify, but a general short-term measure will be the number of journal publications and conference presentations resulting from collaborative efforts spearheaded by the program; a long-term measure will be the extent to which these efforts are cited and built upon by the disciplinary community itself, reflecting an understanding of the benefits of mathematical and statistical approaches. Already, several Kick-off Workshop participants were disciplinary scientists (e.g., Lewis Sheiner, a physician and leading force in research in human pharmacokinetic analysis; Fumio Kojima, a leader in use of electromagnetic techniques in nondestructive evaluation of materials); thus, there is nascent appreciation of the potential for scientific advance that may be energized by the Program.

With respect to applied mathematical and statistical methodological research, a major objective is math/stat technology transfer, i.e., for the approaches of each community to be exploited by the other to advance both, jointly and separately. Two specific examples with rich potential are the Proper Orthogonal Decomposition dimension-reduction inverse problem computational techniques used by mathematicians and Markov chain Monte Carlo simulation methods used by statisticians. The immediate goal is to achieve this on a small scale through Program-sponsored activities, but with participants "spreading the word" through publications and presentations to effect long-term awareness. The extent to which such transfer occurs will be gauged by the number and significance of such endeavors.

The planned Closing Workshop will disseminate progress made during the Program and crystallize future research directions and is thus expected to have a strong impact on further research in this area. The potential synergism with the IPAM Inverse Problem Program, discussed in Section 1.2.3, will serve to carry forward outcomes of the SAMSI Program to the inverse problem research community.

1.3.2 Program impact on human resources

There will be the obvious (hopefully career-changing) impact on the young investigators (postdocs and students) working in the interdisciplinary environment of the SAMSI Inverse Problem Program. We anticipate that the training will result in young mathematical scientists who are predisposed to work readily on scientific problems in collaborations involving the frontiers of applied mathematics and statistics. For the more senior researchers involved, there is early indication, evidenced in part by the desire of Kick-off Workshop participants to return to SAMSI to begin collaborations, that involvement in SAMSI-initiated research collaborations will have a lasting effect on their research endeavors. This will be reflected in the content of future research publications and presentations of Program participants.

We anticipate that Program postdocs and graduate students immersed in the philosophy of collaboration at the boundary of applied mathematics and statistics will be attractive candidates for positions in academia, government, and industry. The extent to which these participants attain such positions and are guided by the cross-disciplinary philosophy in their work will be important success indicators.

1.3.3 Program impact on education and outreach

The Program is having an immediate impact on the training of applied mathematicians and statisticians through the course MA/ST 810Q. The course emphasizes integration of applied mathematical and statistical perspectives through collaboration and serves as what we hope is a prototype for a future, regular course at NCSU that carries forward the Program philosophy to general graduate education in applied mathematics and statistics. Banks and Davidian plan to compile the course material into a monograph that will serve as inspiration for similar courses at other institutions. Experimentation with and adoption of this new educational model by colleagues in applied mathematics and statistics at other institutions will be an important gauge of the influence of the Program on education.

Outreach activities are already planned. Members of the Inverse Problem Program will participate fully in the June, 2003 Workshop for Undergraduates by presenting a 4 day summary of the Program activities. The Workshop is designed to recruit students to graduate studies and careers in the fields of mathematics and statistics. Follow-up of participants will measure the impact of the Program.

1.4 Additional Material

Papers and technical reports produced under the program are listed in Section I.G.

Workshop participant lists are given in Appendix A.

Workshop programs and abstracts are given in Appendix B.

Workshop evaluations are given in Appendix C.

A Course on Inverse Problem Methodology was taught in Fall, 2002, by Marie Davidian and H.Thomas Banks. A full description of the course is given in Appendix D.

2 Challenges in Stochastic Computation Program

2.0 High-Level Summary: Computing Averages

Across a range of scientific problems from genomics to finance to protecting medical data, a shared characteristic is that key quantities can be expressed as averages. Once a curiosity, this characteristic has catalyzed the entirely new field of stochastic computation, in which the quantities of interest are computed by averaging the results of many simple computations rather than by one complex, lengthy computation. For instance, to compute the area of a complex shape, one may select points "at random" in a larger set, and use the fraction of them falling in the shape (an easy computation) as the estimate of its area. In some cases, stochastic computation is merely more efficient than alternatives; in others, there are no alternatives.

The Stochastic Computation program has addressed randomness-based computations in a wide variety of settings. The implications include improved analyses of genetic data, new models for volatility in financial markets and, although more technical, novel techniques for data analysis that "average models."

2.1 Research Activities

StoCom activities include regular weekly meetings of the four topic research groups, and continual research interactions throughout each week. The program also includes two major workshops (Sept/Oct 2002; June 2003) and a series of four "StoCom Days", i.e., a mini-workshop on each of the four topic areas, in late January and mid-late February 2003. Details of workshops and activities are available at the StoCom website, `www.samsi.info/sto.html`

2.1.1 Model and Variable Selection

The focus of this program is a synthesis of computational methods for model choice, with exploration of best current practice among existing methods, identification of their limitations and development of new methodology. Computation in model choice (selection of a single model or choosing to "average" over many models) involves calculation of marginal likelihoods (integration) for each model, and, in large problems where enumeration is prohibitive, search algorithms to identify a subset of models in order to narrow down comparisons to a manageable set. Research has focused on the following areas:

- The comparison of current and newly proposed stochastic search algorithms for variable selection in linear regression models, including Gibbs samplers, Metropolis-Hastings algorithms with informative proposal distributions, deterministic search based on branch and bound algorithms, stochastic search using sampling without replacement, and perfect sampling. A primary focus in on approaches that scale up with the dimension of the problem.

- The choice of prior distributions. Current default prior distributions for linear models, such as $g-$priors, lead to tractable computations, but not necessarily consistent model selection. We have explored scale mixtures of $g-$priors to create new classes of objective prior distributions with both desirable theoretical properties and computational tractability. We have developed improved algorithms for implementing model selection with Cauchy distributions, one of the recommended scale mixtures that has not been widely utilized in practice. Software will be made available in the near future.

- Methods for computing intractable marginal likelihoods and model probabilities. Comparison of existing methodology such as reversible jump algorithms (that do not require enumeration of models) to approaches that have typically required enumeration of all models. The latter includes Laplace approximations, importance sampling, ratio importance sampling, computation of marginal likelihoods using MCMC samples from each model or from the full model (in the case of variable selection). Future work will continue in developing innovative algorithms that combine the novel search strategies proposed in linear models with the above marginal likelihoods computations to extend these approaches to model spaces where enumeration of models is not feasible.

- Software the can be implemented from R for implementing model selection/model averaging in linear models and extensions will be available from the StoCom web-site.

2.1.2 Contingency Tables

The goals were to bring new computational methods to solve old and new problems of statistical inference for contingency tables. The group's expertise is in three areas: Markov Monte Carlo methods, "perfect" sampling methods, and sequential importance sampling (SIS). These three areas have seen theoretical breakthroughs in recent years.

The first research project was on sampling from tables of genotype data for testing Hardy-Weinberg equilibrium, as in the well-known paper of Guo and Thompson. We were interested in solving two problems with existing approaches: the complexity problem when the total number of individuals in the table is on the order of 10^4, and the "irreducibility" problem when certain allele combinations are lethal and force zero entries in the table. We were able to solve both of these problems in our paper "Monte Carlo Algorithms for Hardy-Weinberg proportions."

The second ongoing project is on contingency tables with ordered covariates. A special case would be logistic regression. Our motivating data set is the well-known Ille-et-Verlaine oesophageal cancer data stratified by age in Breslow and Day (1980), where the data at each age level is quite sparse, and has two ordered covariates, tobacco level and alcohol level. Generalized linear models can be used to fit models of odds-ratios, but the sparsity of these tables makes these approximations unreliable. The problem for us is to develop sampling methods for tables with fixed sufficient statistics. Work has been done on Markov chains for this problem by Diaconis, Graham, and Sturmfels that involves the number theoretic concept of "homogeneous partition identity." But this work is not practical because the number of moves in the Markov chain grows exponentially and is not computable for large tables. A current goal is to find more practical Markov chains or other perfect sampling methods.

2.1.3 Graphical Models

This component has been involved in a range of activities in the exploration, evaluation, synthesis and further development of effective computational methods for fitting graphical models, especially with a view to scaling up to large-scale models. Research continues and will evolve into late summer 2003. Among the research foci are the following areas, and substantial progress has been made across these areas.

- The synthesis and evaluation of current stochastic simulation methods for decomposable graphical models, including variants of Gibbs and Metropolis-Hastings methods. Various modifications of MCMC methods have been developed and evaluated.
- The extension of such methods to nondecomposable models, including implementation of approaches using importance sampling in graphical model space, and the development of approaches linked to complete-conditional regressions, hierarchical/triangular regressions and directed graphical models. New approaches developed this way are now being further investigated, especially in higher-dimensions.
- The exploration and development of annealing and optimisation ideas to generate hybrid methods, especially methods for more rapidly generating large numbers of high probability candidate graphs. New approaches have been defined and found very effective in moderate dimensional problems.
- Theoretical and modelling questions that are critical to applications, including the specification of priors over graph space, tuning parameter evaluation, and measures of complexity.
- The exploration and development of distributed processing for implementation of stochastic computation algorithms for large-scale graphical models, utilising computer (beowulf) clusters. Serial and cluster-relevant code developed for a range of graphical modelling computations will be a deliverable of the program.
- Studies with real and simulated data, including a range of gene expression data sets. A key focus is on scale-up of models and computations to higher dimensions. Initial results in these studies suggest major utility of some of the new approaches developed under this program, and will be a core focus during the next few months of the StoCom program.

2.1.4 Financial Mathematics

The focus in this group is Monte Carlo methods in the context of stochastic volatility (SV) models. Different backgrounds of the participants (probability, statistics and applied math) stimulate active interactions among different approaches and disciplines. Specific activities that have defined the research and will continue into summer are as follows.

- Understanding of pros and cons of various algorithms. Many MCMC/SIS algorithms have been reviewed by our group, especially particle filters and some multi-move

Metropolis methods. They are considered efficient when fitting historical SV models by only using underlying assets (no derivatives involved). We apply these methods to a new proposal — multi-factor SV models with several well separated time scales, e.g. scales corresponding to fast (weekly) and slow (yearly) mean-reversion cycles of volatility. Extensive simulations have been done and a paper in preparation should be ready for submission early Summer 2003.

- Improvement of calibration of SV models based on both stock and option data. For the task of option pricing in the context of SV models, it is inevitable to fit related risk-neutral dynamics by using option data in addition to stock prices. Much more intensive MCMC/SIS computation is required to perform numerical integration for an option price expressed as a conditional expectation under a risk-neutral measure. A key parameter — the volatility risk premium — has to be estimated as part of the Monte Carlo algorithm. A paper in preparation will be submitted during Summer 2003.

2.2 Industrial and Governmental Participation

StoCom had some, albeit limited participation of government and industry-based SAMSI advisers in program formulation. The opening workshop did attract a number of participants from governmental labs and agencies.

One spin-off is a collaborative research project that will involve SAMSI/StoCom postdoc Beatrix Jones working on statistical research in collaboration with the Dr Scott Nichols and colleagues at the National Marine Fisheries Service (Pascagoula – Mississippi Laboratory) once she completes her year with StoCom. Jones will move to Duke to continue her postdoc, continuing in research on graphical models with applications in genomics as her primary activity, but will engage in this spin-off research collaboration for a component of her time.

2.3 Roles of Core Participants (Faculty, Visitors, Postdocs)

2.3.1 Model and Variable Selection

- Merlise Clyde (Senior faculty - Duke, statistics)
  Clyde is co-director of the StoCom program and leader of the component on model selection. She heads up and directs the research on this component, and is involved in administration of the workshops and visitor activities. She will have the lead role in developing manuscripts and web based records of this component.

- Jim Berger (SAMSI Director)
  Berger is involved in research in objective Bayes methodology in variable selection. Berger plays a pivotal role in mentoring post-docs and graduate students.

- Joe Ibrahim (Senior faculty - UNC, biostatistics)
  Ibrahim's research is in calculation of marginal likelihoods based on Ratio Importance Sampling in generalized linear models and prior choice.

- Bani Mallick (Senior faculty - Texas A& M)
  Mallick'a (short term visitor) contributed to the discussion and development of research topics, including perfect sampling in non-parametric high-dimensional wavelet regression models.

- Mark Huber (Junior faculty - Duke, mathematics and statistics)
  While primarily involved in the Contingency Table working group, Huber has contributed new ideas for improving perfect sampling in orthogonal regression, such as wavelet regression and implementation of perfect sampling methods for general variable selection problems.

- Feng Liang (Junior faculty - Duke, statistics)
  Liang's research has centered on objective prior specification for variable selection.

- Helen Zhao (Junior faculty - NCSU, statistics)
  Zhao's research in non-parameteric regression methods bridges classical and Bayesian approaches.

- Rui Paulo (75% SAMSI postdoc)
  Paulo is involved in research on objective Bayes prior specifications and has been instrumental in implementing methods for calculating marginal likelihoods.

2.3.2 Contingency Tables

- Ian Dinwoodie (Senior faculty - Tulane, mathematics – SAMSI Fellow)
  Dinwoodie is co-director of the contingency tables research component, covering research direction and administration. He generates much of the research emphasis at the interface between mathematics and statistics. He will have a lead role in developing manuscripts and web based records of this component.

- Mark Huber (Junior faculty - Duke, mathematics and statistics)
  Huber is co-director of the contingency tables research component, covering research direction and administration. He generates much of the research emphasis at the interface between probability (perfect sampling and related topics) and statistics. He will have a lead role in developing manuscripts and web based records of this component.

- Yuguo Chen (Junior faculty - Duke, statistics)
  Chen is involved in research in designing and implementing efficient sequential importance sampling methods for zero-one and contingency tables. He is actively collaborating with other group members in development of MCMC and perfect sampling strategies. He is also involved in presentations at workshops.

- Adrian Dobra (15% SAMSI postdoc)
  Dobra is involved in research in modelling and computation in contingency tables, especially the development and implementation of MCMC and sequential importance sampling methods for higher-dimensional tables subject to observations on only selected margins. He is also involved in presentations at workshops and undergraduate outreach meetings.

2.3.3 Graphical Models

- Mike West (Senior faculty - Duke, statistics)
  West is director of the StoCom program and leader of the graphical models component. He provides overall research direction, as well as administration and mentoring for junior participants. He will have a lead role in developing manuscripts and web based records of this component. West is also involved in the research, with Dobra, on the contingency tables component.

- Chris Carter (Senior faculty - Hong Kong, statistics – SAMSI Fellow)
  Carter is a visitor (year long) involved in the StoCom research on graphical models. His contributions build on his expertise in the area and he plays a role in mentoring junior participants.

- Beatrix Jones (100% SAMSI postdoc)
  Jones is involved in research in modelling and computation, especially in connection with MCMC methods in decomposable and nondecomposable Gaussian graphical models, and the interfaces with applications in gene expression analysis. Jones also plays a collaborative role in mentoring graduate student participants, and in presentations at workshops and undergraduate outreach meetings.

- Adrian Dobra (15% SAMSI postdoc)
  Dobra is involved in research in modelling and computation in graphical models, particularly in connection with manipulation of large-scale graphs, annealing methods and novel approaches to stochastic search in graphical models, the interfaces with approaches based on regression models, and cluster-based distributed computation. Dobra is involved in presentations at workshops and undergraduate outreach meetings.

2.3.4 Financial Mathematics

- Jean-Pierre Fouque (Senior faculty - NCSU, applied mathematics)
  Fouque directs the financial mathematics component, covering research direction and administration. He generates much of the research emphasis at the interface between mathematics and statistics, guides the research and mentors junior participants. He will have a lead role in developing manuscripts and web based records of this component.

- Chuanshu Ji (Senior faculty - UNC, statistics)
  Ji is involved in collaborative StoCom research concerned with models of interest rate term structures, and option prices. He is mentoring junior researchers, including Beom Lee, and developing theoretical and computational aspects of MCMC for option price problems.

- Yuguo Chen (Junior faculty - Duke, statistics)
  Chen is involved in research in designing and implementing efficient sequential importance sampling methods for filtering and smoothing problems in stochastic volatility

models. He is actively collaborating with other group members in development of MCMC strategies for other more complicated finance models.

- Several other junior faculty have been, and are, participants in meetings and research discussions of the StoCom financial models group, though not core participants. These include Tao Pang (Junior faculty - NCSU, applied mathematics), Sujit Ghosh (Senior faculty - NCSU, statistics) and Ana Valeva (Junior visiting faculty - Duke).

2.4 Additional Material

Papers and technical reports produced under the program are listed in Section I.G.

Workshop participant lists are given in Appendix A.

Workshop programs and abstracts are given in Appendix B.

Workshop evaluations are given in Appendix C.

3 Large-scale Computer Models for Environmental Systems

3.0 High Level Summary: Laboratory-in-a-Computer

Environmental science captures more public attention than any other science except medicine. The environment is tangible, although many long-term effects are not, and there is a sense that scientific tools exist to attack many problems. At the same time, physical experimentation in environmental science is being supplemented or even supplanted by computer models. The reasons are multiple, but often physical experiments are too dangerous, too time-consuming or too expensive, if not outright impossible. (Even if experiments with climate were ethical, they would not be possible.)

The program in Large-Scale Models for Environmental Systems is helping define the paradigm for environmental experiments of the future, which will be conducted using computer models of, for instance, the atmosphere or underground flow of water. Good models (the good laboratories) require collaboration, as has happened extensively in this program, of applied mathematicians—who produce the mathematical representations, computational scientists–who ensure that the model-laboratory actually works, statistical scientists–who deal with uncertainties in data, observations and conclusions, and domain scientists–who ensure that the tools are relevant to the scientific issues.

3.1 Scientific Achievements of the Program

The program on “Large-scale Computer Models for Environmental Systems” has had the following activities up to the time of writing:

1. Two workshops, on “Multiscale Modeling” (February 2–7) and “Optimization” (April 28–30). The third workshop, on “Spatial-temporal Statistics” will take place at NCAR, jointly sponsored by SAMSI and by the Geophysical Statistics Project at NCAR, June 1–6;

2. A one-day workshop on “Physical-Statistical Modeling and Data Assimilation” was SAMSI on March 25, and a second on “Closure Relations for Evolving Models of Multiphase Flow” will take place on May 16;

3. Two graduate-level courses, “Multiphase Transport Phenomena” (Instructors: William G. Gray and Cass T. Miller) and “Environmental Statistics” (Instructor: Richard L. Smith) have been taught at SAMSI;

4. Numerous visitors have attended the program for periods ranging from one week to the whole length of the program;

5. A series of working groups was set up after the first workshop, whose activities are ongoing.

In the remaining parts of this discussion, we outline specific research advcances that have been made (or are underway), dividing the discussion into those focussed on atmospheric science applications and those concerned with porous media.

3.1.1 Atmospheric Science Working Groups

Spatial-Temporal Statistics

This group has been led by Jim Zidek (University of British Columbia, and a long-term visitor to SAMSI). The overall aims of the group are to develop spatial-temporal statistical models for environmental data and to use these models explore issues such as validation of numerical models for atmospheric phenomena, data assimilation, and the design of monitoring networks. Specific new activities that have been made possible by SAMSI include:

1. A collaboration has been established involving Zidek and his co-workers at UBC, several of the Triangle statisticians, and Dr. Prasad Kasibhatla (Duke School of the Environment) on the analysis of a large data set created by Dr. Kasibathla, consisting of a four-month model run of ozone readings (hourly ozone readings on a 36-km. grid covering much of the United States) and associated real observational data from the EPA's AIRS network. This very rich data set should provide the focus for several projects over the next year or so, including developing a spatial-temporal model for the model data and testing it against the real data; developing a strategy for ozone forecasting and high ozone alerts; and improving the design of the real observational network.

2. The presentations by Mark Berliner and Chris Wikle at the February workshop led to numerous discussions on the subject of "physical-statistical modeling" (spatial-temporal modeling in which the form of the temporal dynamics is influenced by physics) and the connections between this work and the applied mathematics activity in data assimilation. To pursue these topics, a one-day workshop was organized at SAMSI on March 25, with Berliner, Wikle and Chris Jones (UNC, Applied Mathematics) as the three presenters. Approximately 30 people attended this. As one example of likely spinoffs, long-term visitor Bent Natvig (University of Oslo) and his student Ingunn Tvete are using these methods in an analysis of earthquake data.

3. David Holland is organizing a subgroup looking at network design. Those involved with this include Holland, Shelly Eberly and Bill Cox from the EPA, Zidek, and local statisticians Fuentes, Smith, Grady (NISS) and Zhu (UNC Statistics). Specific research foci of this group include extensions of the maximum entropy approach to network design to fully Bayesian hierarchical models, and designing networks to monitor extreme values specifically. Activities to date include an invited papers session at the March 2003 ENAR meeting in Tampa, which included papers by Zidek and Holland, and a discussion by Smith. Future work is likely to lead to several research papers as new design objectives are explored.

4. Harry Hurd (adjunct professor in Statistics, UNC) has explored aspects of cyclostationarity in Kasibhatla's ozone data, with promising initial results.

Source Apportionment

This group has been led by Richard Smith (UNC Statistics) and Prasad Kasibhatla (Duke School of the Environment) and focusses on a problem that has attracted a wide literature among atmospheric chemists but not so far among statisticians: the problem of reconciling a set of conjectured sources (or sinks) of a pollutant with monitor readings at locations different from the conjectured sources. An atmospheric transport model relates the sources to the monitor sites, but calculation of source amounts involves solving an inverse problem which is typically ill-conditioned. A talk was given about this problem as it relates to carbon dioxide sources and sinks by Tapio Schneider (Caltech) at the February workshop; Kasibhatla's group is looking as the same problem as it relates to carbon monoxide (CO); at the same time, a local EPA groups including Robin Dennis and Alice Gilliland is using the same technique to solve a problem of attribution of ammonium sources.

In current work, Smith and Kasibhatla are working with two of their PhD students on extending an earlier analysis by Kasibhatla et al to incorporate spatial and temporal correlations among the monitors; this is apparently the first time this has been done in the present context. Future plans include expanding the statistical model into a Bayesian hierarchical framework which will allow many additional sources of uncertainty to be considered (for example, uncertainty in the specification of the atmospheric transport model) and also potentially the inclusion of more data (for example, using satellite data as well as ground-level monitors). This work is likely to continue well beyond the lifetime of the SAMSI project and is expected to provide material for one or more PhD theses.

Montse Fuentes was also active in this working group and has started collaborations with Kasibhatla. Indeed, she is going to take a leave during Fall, 2003 to work with Kasibhatla at Duke and develop interdisciplinary proposals in the area.

Intermittency and Sub-Grid-Scale Processes

This topic involves a collaboration among applied mathematicians (Rich McLaughlin and Roberto Camassa, UNC Applied Math), statisticians (Steve Marron and Richard Smith, UNC Statistics, and Robert Wolpert, Duke Statistics) and subject-matter scientists (including Alberto Scotti, UNC Marine Sciences, and Jason Ching of the EPA) on the interplay between fundamental mathematical analysis, experimentation and statistics relating to intermittent behavior in fluid flow systems and the problem of parameterizing sub-grid-scale phenomena.

One part of this work involves interaction between applied mathematicians,

statisticians and experimental scientists in characterizing intermittency. Intermittency may be defined as the appearance of long-tailed distributions (as opposed to Gaussian distributions) among certain measured variables in fluid flow systems. It has been observed experimentally and in simple cases explained theoretically. One part of the current effort focusses on statistical techniques for determining whether a system is intermittent, making a connection with the statistics of extreme values and processes with long-range correlations.

Another aspect of this work relates to the parametrization of sub-grid scale processes using effective parameters, or effective mixing coefficients. The idea is to replace the complicated physics occuring on unresolvable scales by some much simpler bulk mixing, with coefficients tuned to reflect what is going on on sub-grid scales. The twin goals of this group are to build upon existing mathematical theories, to interact with laboratory experimenters working on mixing and entrainment in stratified fluids; and to interact with EPA scientists working on modeling of boundary layers, and plume models for transport of atmospheric pollutants.

Combining Networks

This group was set up at the instigation of Dr. Peter Finkelstein (EPA) with further participation from EPA scientists David Holland and Joe Sickles, Montserrat Fuentes (NCSU), Marc Serre, Zhengyuan Zhu and Richard Smith (UNC), and several NCSU and UNC PhD students. The EPA often faces the problem of combining data from different networks taking account of spatial dependence and measurement error in each of the networks. An extension that could be viewed as part of the same problem is when one of the "networks" is in fact the output of a computer model, thus creating the problem of optimally combining data from a computer model and a monitoring network but in a different context from other treatments of data assimilation.

As a specific focus for research, the group has chosen to examine the problem of combining aerosol nitrate (ANO3) data from two EPA networks, Castnet and the Speciated Trends Network (STN). These two networks were originally designed for different purposes, Castnet focussing on long-term trends in rural areas and STN primarily for monitoring health effects in cities; close interaction with scientists is needed because of the different designs of the networks and the different measurement techniques used for each. Preliminary discussions have led to several statistical approaches being proposed, that will be developed in more detail over the next several months. If these initial approaches are successful they will likely lead to several publications and long-term collaborations between the statisticians and EPA scientists.

Human Exposure and Health Effects

Given that the primary purpose of EPA air pollution regulations is to protect human health, a vast amount of research takes place on the relationship between air pollution and human health. The role of computer models in assessing human health effects is generally much less than in other areas of research covered by this program, but nevertheless there is a growing field of literature related to personal exposure, which does use comupter models, and in which several SAMSI personnel have been involved. Specific activities are as follows.

1. One project is model validation in the context of a WWW platform known as pCNEM, which has been developed by SAMSI visitor Jim Zidek and his co-investigators to enable remote users to build their own air pollution predictive exposure models online. These distributions, for forecasting exposures to specified pollutants, take account of the complex relationships individuals have with their environments. The research effort has been continued at SAMSI by Zidek, Sandra McBride and Robert Wolpert (Duke), and several EPA personnal, with a focus on validating these models using real data from personal exposure surveys. One piece of research completed at SAMSI (paper shortly to be submitted) uses U.K. data from selected sub-populations in London. Future plans include similar validation studies using similar U.S. data bases. A paper based on this validation work is being written up for the J of Air, Water and Waste Management, and will also be presented as an invited poster at the ISEA conference in Italy, 2003.

2. A second project is to develop a Bayesian hierarchical model for human exposure. Using USEPA panel study data, the project will quantify the contribution of ambient concentrations to personal exposure, taking into account

    - A spatial model of ambient concentration data from local outdoor and ambient monitoring data;
    - Modelling the latent components of personal exposure: activity patterns, proximity to sources and home air exchange rates;
    - Use of individual-level exposure measurements as well as explicitly accounting for bias in measurements and differing uncertainties among different monitors.

    Plans include a presentation at the JSM meeting in August 2003 and subsequent publications in both the statistics and exposure-related literature.

3. The most ambitious use of these studies would be to relate their results directly to large-scale epidemiological studies of air pollution and human health, and specifically to assess how such studies might be biased due to the discrepancy between personal and ambient air pollution. A project just beginning is to use joint distributions of ambient and personal air pollution measures, based on the pCNEM platform and model validation studies just described, in conjunction with large data bases of air pollution and human health such as the NMMAPS study from Johns Hopkins, the raw data from which are now being made available online.

Other Work In Progress

Two other working groups have been formed but have not so far made sufficient progress to report in detail. These are on *Extreme values* (chaired by Amy Grady, postdoctoral researcher at NISS) and on *Uncertainty and Ensemble Forecasting* (chaired by Jerry Davis of the Department of Marine, Earth and Atmospheric Sciences, NCSU).

3.1.2 Porous Media Science Working Groups

Model Formulation and Closure

This activity involves theoretical aspects of porous medium model formulation and closure, with a focus on fundamental aspects related to the translation of operative physical, chemical, and biological processes into well-posed mathematical models that respond to deficiencies existing in current, common model formulations. More detailed objectives are:

- to educate researchers and visitors in advanced methods for model formulation to describe flow and transport in porous medium systems;
- to catalyze an interdisciplinary group working to develop closure relations for evolving models; and

- to evolve improved pore-scale modeling and analysis methods to aid the development of closure relations.

Progress on these research objectives includes the following. An advanced-level course in multiphase porous medium model formulation was taught at SAMSI, by Miller and Gray, during the spring semester. A series of technical reports are being produced on various aspects of multiphase model formulation and closure. An interdisciplinary NSF proposal is being developed concerning closure relations for an evolving class of models. Specific topics that are being devloped into research papers include lattice-Boltzmann modeling at the pore scale; the use of level-set methods for estimation of quantities such as interfacial areas and curvatures needed in evolving closure relations; a more rigorous form of a Darcy's law like expression and highlighting

conditions under which the common form is subject to significant error; and developing a first principles model for multiphase, multispecies transport. Finally, C. Miller and W.G. Gray are working on a first draft of a monograph on evolving multiphase models for multiphase systems.

Personnel involved in this activity include local faculty: D. Adalsteinsson (UNC), R. Camassa (UNC), M.G. Forest (UNC), J. Huang (UNC), C.T. Kelley (NCSU), R.M. McLaughlin (UNC), C.T. Miller (UNC) and M. Minion (UNC); long-term visitors: W.G. Gray (SAMSI Fellow), K. Culligan (WGG student from Notre Dame); and postdocs and graduate students: C. Abhishek, M.W. Farthing, D. Johnson, J.F. Kanney, C.E. Kees, H. Li, C. Pan, J.A. Pedit, J. Reese, C. Rupert, D. Sassan.

Model Solution Approaches

This activity is focussed on the improved solution methods for models of porous medium systems. Such models are typically comprised of systems of partial differential algebraic equations (PDAE's) that are solved using numerical approximation approaches. Because the formulations are complex and varied they lead to a wide range of mathematical forms, the solution of which may be advanced by evolution of improved numerical algorithms, approximation methods, and simulation environments for various classes of problems. Objectives in this case include

- to study algorithms for the solution of PDAE's that arise in models of porous medium systems;
- to investigate evolving methods for spatial discretization of PDAE's;
- to explore the use of methods that adapt aspects of spatial and temporal discretization as a solution evolves; and
- to explore the development of problem solving environments for porous medium models.

Research conducted during SAMSI has resulted or will result in manuscripts on the following topics: comparing methods to solve the steady-state form of Richards' equation; investigating a spatially and temporally adaptive method to solve Richards' equation; exploring the joint use of higher order discontinuous Galerkin methods in space and stiff DAE methods in time for the solution of Richards' equation; and investigating the need for, architectural aspects of, and prototype components of a problem solving environment for environmental modeling.

Participants in this activity are local faculty: C.T. Kelley (NCSU), C.T. Miller (UNC); long-term visitors: C. Dawson (Texas); W.G. Gray (SAMSI Fellow); T.F. Russell (UC-Denver); postdocs and graduate students: C. Abhishek, M.W. Farthing, J.F. Kanney, C.E. Kees, H. Li, J. Reese, C. Rupert, D. Sassan

Optimal Design

This working group is centered on the evolution of improved optimization methods for the design of water supply and contaminant restoration systems. Previous work by Mayer, Kelley and Miller has produced a suite of "benchmark problems" for optimal design in hydrology, with the intention of enabling both the optimization and simulation communities to improve design and evaluations methods for the very difficult problems in this field. These community problems have been used as a focal point for further activities including

- implementing solutions to the community problems in portable software, using several different simulators;
- comparing solutions to the community problems derived using a range of optimization methods, including some that are novel to the community; and
- making both the problems and solutions available to the international community,

The group has already made progress toward these goals, In Kavanagh *et al.* (2003) we report on the implementation, solution, and distribution of two of the community problems.

Tom Russell, a long-term visitor, has been a major contributor to the USGS family of software for modeling of subsurface flow and transport. He has helped the group extend our use of the USGS codes, which were very useful in the previous work. Dawson and Mayer also assisted in this part of the work.

The optimal design problems in this field are not simple. The objective functions and constraints are not smooth, can be discontinuous or even random, and are rarely given by simple formulae. Two of the long-term visitors, John E. Dennis Jr. and Alex Mayer have significant experience in such problems. Dennis is an expert in the analysis and design of algorithms for noisy problems, and Mayer has experience in the application of genetic algorithms to problems in hydrology. These two have assisted us in including more optimization methods in the distribution.

Personnel involved in this section are local faculty: C.T. Kelley (NCSU), C.T. Miller (UNC); long-term visitors: W.G. Gray (SAMSI Fellow), C. Dawson (Texas), J.E. Dennis Jr. (Rice), A.S. Mayer (Michigan Tech), T.F. Russell (UC-Denver); postdocs and graduate students: M.W. Farthing, K.R. Kavanagh, C.E. Kees, J. Reese.

3.1.3 Impact of Scientists from the EPA

The EPA, located in Research Triangle Park, has been heavily involved in the Environmental program. Here is a listing of the key involvements.

- *David Holland* - attended the spatial-temporal working group meetings and has coordinated a subgroup on network design.
- *Peter Finkelstein* - coordinated the combining networks working group
- *Shelley Eberly* - attended meetings of network design group
- *William Cox* - attended meetings of network design group
- *Robin Dennis* - attended meetings of the spatial-temporal statistics and source apportionment groups
- *Alice Gilliand* - attended meetings and gave a presentation at the source apportionment group
- *Jason Ching* - attended meetings and gave a presentation at the working group on intermittency and sub-grip-scale processes
- *Joseph Sickles* - assisted with combining networks working group
- *Janet Burke* - assisted with human exposure working group
- *Ron Williams* - assisted with human exposure working group
- *Ellen Cooter* - attended several group meetings and the workshop on multi-scale modeling, and audited Richard Smith's course in Environmental Statistics
- *Joseph Pinto* - attended several group meetings and the workshop on multi-scale modeling
- *Brian Eder* - attended several group meetings and gave one of the 15-minute talks at the start of the program
- *Tim Watkins*, associate lab director for air toxics - member of the design group and he gave a 1 hour presentation about design issues associated with mercury monitoring.
- *Mike Rizzo* from EPA Region III in Chicago joined the design group during his visit to EPA.

3.1.4 Impact of External Core Participants

James Zidek (Statistics, University of British Columbia) visited SAMSI for five months and was an extremely valuable part of the program. He coordinated the spatial-temporal statistics working group (which was really three separate activities) and co-organized the March 25 one-day workshop. He also participated very actively in two other working groups, those on "source apportionment" and on "human exposure and health effects". He initiated the contact with Prasad Kasibhatla (Duke School of the Environment), whose data and problems motivated both the ozone modeling subgroup (part of the spatial-temporal statistics group) and the source apportionment group. His contacts within the EPA, in previous work on the "pCNEM" WWW platform and continued and extended (jointly with Sandra McBride) during the SAMSI program, made a very valuable contribution to the human exposure activity. Early in the program he gave a seminar on this work which was attended by several EPA researchers as well as local faculty and students. He also continued his previous work on design of monitoring networks, focussing particularly on designing networks to monitor extreme values, on which subject he gave a talk (with acknowledgement to SAMSI) at the ENAR meeting in Tampa. He has written several tech reports on these and related issues, and also gave an invited talk at the Optimization workshop.

Hans Kuensch (Statistics, ETH Zurich) visited for three weeks in February and attended the multi-scale modeling workshop as well as two weeks at SAMSI. He attended several group meetings and gave a seminar on the incorporation of a determinstic PDE into the development of a statistical state-space model, with application to space-time models for lake contamination.

Bent Natvig (Mathematics and Statistics, University of Oslo) visited with his student (Ingunn Tvete) for the first three months of the program. They attended Richard Smith's course on environmental statistics and Natvig made valuable suggestions about it (e.g. reviewing the course notes). They attended the first few meetings of the spatial-temporal working group and Natvig organized a series of discussions on the Berliner-Wikle approach to spatial-temporal modeling, the outcome of which was the March 25 meeting in which both Berliner and Wikle participated. Their own research on earthquake modeling resulted in a program seminar, and a forthcoming paper.

Henri Laurie (Mathematics, University of Capetown) visited for three weeks early in the program and interacted both with local faculty and several other visitors at SAMSI, and with Alan Gelfand at Duke. He gave a seminar on "a comparison problem for cellular automota" which presented a novel approach to spatial statistics using the dynamics of cellular automota to contruct Bayesian hierarchical models, with applications to modeling species diversity in ecology.

David Higdon (Statistics, Los Alamos National Laboratory) visited for the multi-scale modeling workshop (where he gave one of the tutorial lectures) and again for a two-week period in March/April, where he interacted with researchers at both SAMSI and NISS as well as visiting Duke. He gave a talk about Bayesian approaches to model calibration, a subject central to much of the interaction between mathematical modeling and statistics.

Doug Nychka (Geostatistics, National Center for Atmospheric Research) visited for the multiscale modeling workshop (where he gave one of the plenary lectures) and returned for a week in mid-April, where he gave a talk about data assimilation. He also spent a day visiting NC State and met with EPA researchers. He is organizing the NCAR end of the spatial-temporal statistics workshop.

Michael Stein (Statistics, University of Chicago) visited for a one week, giving a talk about spatial-temporal models to SAMSI, and also spending a day visiting NC State and another day visiting the EPA. Discussions among Prof. Stein, Richard Smith and Petrutza Caragea were also most valuable and shedding light on approximation methods in spatial statistics and his critique of Richard Smith's lecture notes will be valuable for their future development.

Sujit Sahu (Mathematics and Statistics, University of Southampton) is visiting for four months as a joint guest of the SAMSI environmental program, the SAMSI stochastic computation program, and Duke University. he has been working mainly with Alan Gelfand on Bayesian approaches to spatial-temporal statistics.

Jared C. Bronski (Mathematics, UIUC) visited samsi and lectured about the rigorous stretched exponential distributions arising in the evolution of a passive scalar diffusing in the presence of a linear, rapidly fluctuating (gaussian white) shear layer (the majda model). Extensions to fractional brownian motion were presented, and new tight asymptotic bounds for the associated small ball problem were presented. Bronski and McLaughlin have begun exploring non-sheared flows, from the point of view of solving the moment equations. the work is ongoing.

Andrew Poje (Mathematics, CUNY) visited and lectured about the use of melnikov dynamical system tools to analyze real ocean tracer fields, and illustrated the success of these strategies in identifying large transport basins. Drew interracted with McLaughlin, Camassa, and briefly with Marron in material related to McLaughlin's cmg proposal (currently in review at NSF). Specifically, research has begun to identify the physical mechanisms at play in the observation of heavy tailed probability distribution functions frequently observed in numerous tracer signals in the environment. The work is ongoing, and we have identified questions of statistical significance in data trends which we are presently discussing with Steve Marron.

Kayo Ide (Atmospheric Sciences, UCLA) visited SAMSI and lectured on lagrangian data assimilation and their study focussing upon the assimilation of a finite dimensional point vortex system through observations of tracer locations. The visit additionally followed the preparation of a CMG NSF proposal between UCLA and UNC focussing upon heavy tailed distributions, their origin, explanation, and ultimate assimilation into predictive systems. Ensuing discussions focussed upon the assimilation aspects of this effort.

Diego del Castillo Negrette (Oak Ridge National Lab) visited and lectured at SAMSI about fractional diffusion equations, and associated stochastic processes of relavence to environmental problems, as well as plasma physics. Diego interracted directly with McLaughlin, Camassa, and Poje in discussions regarding chaos induced heavy tailed tracer distributions. the work is ongoing.

Dan Anderson (Mathematical Sciences, George Mason University) visited samsi and lectured about the motion of Gravity currents in heterogenous porous media. through the visit, simulations were developed and performed for a fully two dimensional, heterogeneous problem. the job has completed, and shows outstanding agreement to the homogenization theory developed for this moving nonlinear elliptic problem. Further, issues regarding the non-uniform asymptotic expansion near the contact line were systematically explored in a simplified geometry. Work along these lines continues.

Anne Bourlioux Mathematics, Univ. of Montreal) lectured at the first workshop on fat tailed probability distributions, giving rigorous mathematical results showing how a gaussian random coefficient in a partial differential equation leads to a non-Gaussian pdf for the evolving quantity. Working with McLaughlin and Camassa, we have extended the result to a new regime, and shown different more intermittent pdf's for this problem under a quasi steady approximation, and what appears to be an algebraic tail.

Igor Mezic (Mechanical Engineering, UC Santa Barbara) visited SAMSI, and lectured in our seminar series on new dynamical systems methods for filtering periodic or qperiodic signals. Interractions through the visit were with Chris Jones, and Leonid Kuznetsov, and with McLaughlin in discussing applying ergodic methods to temporally varying flows to deduce mixing properties of passive scalars.

Jonathan Mattingly (Mathematics, Institute for Advanced Study) visited and lectured at SAMSI regarding the long time simulations of stochastic differential equations, discussing amongst other things, just how invariant measures may assist in developing stable numerical schemes. McLaughlin and Mattingly are discussing using known asymptotic results from homogenization theory as a means for accelerating the convergence of sde schemes.

Tom Russell (Mathematics, University of Colorado) visited for about a month and gave a SAMSI seminar and lecture at a workshop. Joint work was performed with Casey Miller, Matthew Farthing, and Chris Kees on local adjoint methods for a nonlinear test problem. It is envisioned that this will lead to a publication with perhaps some follow up work. Tom also had discussions concerning the use of his transport code for use on the community problems with Kelley and Miller and their groups.

Clint Dawson (Mathematics, University of Texas) visited for about two weeks and gave a samsi seminar and lecture at a workshop. Clint worked with Miller, Matthew Farthing, and Huina Li on discontinuous Galerkin methods for Richards' equations. So far, we haven't had luck with this class of method, which has some theoretical promise compared to existing methods. If we can resolve the problems, we will have a manuscript worth submitting. This is far from certain at this point.

Alex Mayer (Mathematics, Michigan Technological University) visited for about two weeks and gave a samsi seminar and a lecture at a workshop. Alex is working with Kelley and Miller's groups on optimization and the community problems. Kelley and Miller have papers in the works in this area.

3.2 Additional Material

Papers and technical reports produced under the program are listed in Section I.G.

Workshop participant lists are given in Appendix and **programs and abstracts** are given in Appendix B.

Workshop evaluations are given in Appendix C.

Courses that were taught under the program are described in Appendix D.

4 Education and Outreach Program

The SAMSI Education and Outreach program for 2003-2004 includes the following activities:

Outreach days for undergraduates were held on Nov 9, 2002 and Feb 1, 2003. Undergraduates and faculty mentors were invited to SAMSI for a day long program during which the SAMSI Directorate gave an overview of SAMSI and the opportunities for participation were presented. Program leaders for each of the 2002-2003 programs made hour long presentations about their programs.

Undergraduate Interdisciplinary Workshop: A one week workshop for approximately 30 undergraduates is planned for June 9-13, 2003. During this week, the students will be exposed to an intensive program involving formulation of inverse problems, hands-on collection of experimental data, and mathematical and statistical analysis of data. The problem chosen (vibrations of a cantilevered beam with sensors and actuators) is a paradigm for the theme of SAMSI (the interdisciplinary approach involving applied mathematics, statistics, and applications in domain science to solve complex practical problems) and embodies the principles to which we are trying to attract young mathematicians/statisticians.

Industrial Mathematical and Statistical Modeling Workshop: This ten day workshop for Graduate Students is planned for July 21-30, 2003. This workshop will involve approximately 35-40 graduate students selected from a competitive national pool and along with representatives from six industrial/govenment labs to work in teams on problems posed by the industrial representatives.

Diversity: See Section I.H for discussion of the efforts to achieve diversity.

Courses: See the program reviews in Section I.E and Appendix E for discussion of the SAMSI courses.

5 Planning Workshops

5.1 For the Data Mining and Machine Learning Program

On November 11, 2002, a planning workshop was held for the second year program on Data Mining and Machine Learning. In attendance were the program Co-Chairs David L. Banks (FDA) and Alan F. Karr (NISS/SAMSI), James Berger (Duke/SAMSI), Mary Ellen Bock (Purdue, statistics and NAC), William DuMouchel (AT&T), Marc Genton (NCSU, statistics), Jacqueline Hughes-Oliver (NCSU, statistics), David Madigan (Rutgers, statistics), J. S. Marron (UNC/SAMSI), Gary McDonald (NISS), Warren Sarle (SAS Institute), Scott Schmidler (Duke, statistics), Sam Uthurusamy (GM) and Stanley Young (NISS). The workshop was very fruitful, and lead to identification of possible research foci discussed in Section II.A.

5.2 For the Multiscale Model Development and Control Design Program

On December 17, 2002, a planning workshop was held for the second year program on Multiscale Model Development and Control Design. In attendance were the program Co-Chairs Alan Gelfand (Duke, statistics) and Ralph Smith (NCSU, mathematics), Alan Karr (NISS/SAMSI), J. Stephen Marron (UNC/SAMSI), H. Thomas Banks (NCSU/SAMSI), Gregory Forest (UNC, mathematics), and David Schaeffer (Duke, mathematics). In addition, program leaders Chris Wikle (University of Missouri, statistics) and Murti Salapaka (Iowa State University, mathematics) joined by conference call. The workshop was very fruitful, and lead to identification of possible research foci discussed in Section II.A.

5.3 For the Internet Modelling Program

On March 28, 2003, a NISS Affiliates Technology Day was held on Internet Tomography. This was in response to the enthusiasm among the NISS Affiliates shown for this hot topic, that has fairly recently been included in the planned SAMSI program. This workshop was organized by J. S. Marron, Robert Nowak and Walter Willinger, and featured Mark Crovella, Mark Coates, Rene Cruz, Bin Yu, Matt Roughan and Nick Duffield as invited speakers, with Walter Willinger, Don Towlsey, Robert Nowak, Yehuda Vardi, Nina Taft and Eric Kolaczyk as designated discussants. This established important personal ties between SAMSI principals and major players in the internet tomography research community. Informal discussion in the context of this workshop also led to the important idea of adding a Sensor Network component to the SAMSI program.

6 Distinguished Lecture Series

Held on the first Tuesday of every month, this lecture series brought some of the worlds most prominent statistical and mathematical scientists to SAMSI. In addition to their very widely attended lectures, the distinguished visitors held highly useful discussions with SAMSI researchers. The list of distinguished lecturers, and the titles of their talks, follow:

- Simon Tavaré (Professor of Biological Sciences, Mathematics and Preventive Medicine, University of Southern California) lectured on October 1, 2002, on
  Inference From the Fossil Record: When Was the Last Common Ancestor of Extant Primates?

- Lawrence Brown (Miers Busch Professor of Statistics at the Wharton School, University of Pennsylvania) lectured on November 5, 2002, on
  Statistical analysis of a Telephone Call Center: A Queueing Science Perspective.

- Andrew J. Majda (Morse Professor of Arts and Sciences, Courant Institute of Mathematical Sciences and Center for Atmosphere and Ocean Science, New York University) lectured on February 4, 2003, on
  Mathematical Strategies for Stochastic Modeling in Climate.

- William G. Gray (Henry Massman Professor of Civil Engineering, University of Notre Dame) lectured on March 4, 2003, on
  Multiphase porous media flow physics: Integration of information obtained at different scales.

- Grace Wahba (Bascom Professor of Statistics, Biostatistics and Medical Informatics, University of Wisconsin) lectured on April 1, 2003, on
  Statistical Model Building and Model Selection for Smoothing Spline ANOVA Models.

- Stephen Fienberg (Maurice Falk University Professor of Statistics and Social Science, Carnegie Mellon University) lectured on May 6, 2003, on
  Log-linear Models and Computational Algebra: Old Wine in New Bottles?

F. Industrial and Governmental Participation

Government and industry participation in SAMSI programs and activities reflects broad interest in the SAMSI vision. The following summarizes participation during 2002-03.

Environment Program:. Participants came from the Environmental Protection Agency, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, National Center for Atmospheric Research, RTI International and Sandia National Laboratory. Participants from the EPA were particularly active, in some cases even leading working groups; see Section I.E.3.1.3.

Inverse Program: Government and industry participants in this program were drawn from the Air Force Research Laboratory at Brooks AFB, Hutchinson Cancer Research Center, MD Anderson Cancer Center, Analytical Sciences, Inc., Entelos, Inc., GlaxoSmithKline, the NASA Langley Research Center, the National Institute of Environmental Health Sciences, and SAS Institute. Two of the key testbed problems arose at the Hutchinson Cancer Research Center and the NASA Langley Research Center, respectively.

Stochastic Computation Program:. Participants came from the Environmental Protection Agency, GlaxoSmithKline, National Center for Health Statistics, Pacific Northwest National Laboratory, and the RAND Corporation.

One of the strengths of the SAMSI location was realized in these programs, with heavy participation by the Research Triangle Park-based Environmental Protection Agency and National Institute of Environmental Health Sciences.

There was also strong industry/government/national laboratory participation on the National Advisory Council, with members from Microsoft and Los Alamos, and in planning for the following future SAMSI programs.

Data Mining and Machine Learning Program: The Scientific Committee contains one government (FDA) two industrial (General Motors and SAS Institute) members.

Internet Program: The March 28 Internet Tomography Technology Day (see Section I.5.2) also involved significant industrial participation. The Scientific Committee includes 2 industrial members (AT&T and Avaya). Funding has been requested from the National Security Agency for the companion workshop on Sensor Networks. Potential contacts are established, or else are being explored with, both Research Triangle, and nation-wide corporations including AT&T, Avaya, Cisco, Ericsson, IBM, Lucent Technologies, MCI, Microelectronics Center of North Carolina, Motorola, Nortel Networks and Sprint.

Multiscale Program: Significant effort has been focused on the identification of problems involving advanced material architectures with scientists from aerospace (NASA Langley), aeronautic (Boeing), DoD (Kirkland Air Force Base), and industry (Sandia National Laboratories). Some or all of these identified problems will become test-beds for the program.

G. Publications and Technical Reports

Because SAMSI programs have only been running for 6 months, we also indicate below technical reports that are under preparation.

I. INVERSE PROBLEM METHODOLOGY IN COMPLEX STOCHASTIC MODELS

Publications and Technical Reports

- Banks, H.T., D.M. Bortz, G.A. Pinter, and L.K. Potter
  "Modeling and Imaging Techniques with Potential for Application in Bioterrorism," to appear as a Chapter in "Biomathematical Modeling Applications for Homeland Security" (H.T.Banks and C.C.Chavez, eds.), SIAM Frontiers in Applied Mathematics.

- Ma, Y. and M.G. Genton
  "A Semiparametric Class of Generalized Skew-Elliptical Distributions" Submitted to Scandanavian Journal of Statistics.

- Ma, Y., M.G. Genton, and M. Davidian
  "Linear Mixed Effect Models With Semiparametric Generalized Skew-Elliptical Random Effects," To be submitted to Biometrics.

Reports in Preparation (as of April, 2003)

- Ackleh, A.S., H.T. Banks, K. Deng, and S. Hu
  "Parameter Estimation in a Coupled System of Size-Structured Populations"

- Banks, H.T. and J. Bardsley
  "Electromagnetic Parameter Identification for a Debye Polarization Model"

- Banks, H. T. and J. Bardsley
  "Wellposedness for a Second Order System Arising in a Time Domain Electromagnetic Scattering Problem Using Perfectly Matched Layers"

- Banks, H.T., Y. Ma, and L. K. Potter
  "A Simulation Based Comparison Between Parametric and Semiparametric Methods in a PBPK Model"

- Ma, Y., A.A. Tsiatis, and M.G. Genton
  "Semiparametric Efficiency Bounds for Selection Models"

II. CHALLENGES IN STOCHASTIC COMPUTATION

Publications and Technical Reports

- Chen, Y. and D. Small (2003)
  "Testing the Rasch Model via Sequential Importance Sampling"
  Submitted for publication, to Psychometrika.

- Dinwoodie, I.H. (2003)
  "Estimation of Parameters in a Network Reliability Model with Spatial Dependence" Submitted to: Mathematics of Operations Research.

- Dobra, A., C. Tebaldi and M. West (2003)
  "Bayesian Inference for Incomplete Multi-way Tables"
  Submitted for publication, to Biometrika.

- Wong, K. and C. Carter (2003)
  "An Efficient Sampler for Decomposable Covariance Selection Models"
  Submitted for publication.

Reports in Preparation (as of April, 2003)

- Carter, C., F. Wong and R. Kohn (2003)
  "On Estimating Normalizing Constants for Parameter Selection Models"

- Dobra, A., B. Jones, C. Carter, C. Hans, C. Carvalho and M. West (2003)
  "Stochastic Computation in Gaussian Graphical Models"

- Dobra, A., B. Jones, C. Hans, J.R. Nevins and M. West (2003)
  "Sparse Graphical Models for Exploring Gene Expression Data"

- Fouque, J.P., C.H. Han and G. Molina (2003)
  "MCMC Estimation of Stochastic Volatility Models with Multiple Volatility Scales"

- Huber, M., Y. Chen, A. Dobra, M. Nicholas and I.H. Dinwoodie (2003)
  "Monte Carlo Algorithms for Hardy-Weinberg Proportions"

- Ji, C. and B. Lee (2003)
  "Central Limit Theorems in MCMC Computation of Option Prices with Stochastic Volatility Models"

- Liang, F., R. Paulo, M. Clyde, J. Berger (2003)
  "Hyper-Gaussian Distributions for Bayesian Model Choice"

- Paulo, R., G. Molina, C. Kohnen, M. Clyde, J. Berger (2003)
  "Stochastic Computation in Bayesian Variable Selection"

- Wong, F., C. Carter and R. Kohn (2003)
  "Testing for Structure in the Inverse Covariance Matrix"

III. LARGE-SCALE COMPUTER MODELS FOR ENVIRONMENTAL SYSTEMS

Publications and Technical Reports

- Anderson, D.M., R.M. McLaughlin, and C.T. Miller (2003)
  "*The Averaging of Gravity Currents in Porous Media"*
  In review: Physics of Fluids

- Caragea, P. (2003),
- "Approximate likelihoods for spatial processes"
  PhD dissertation, Department of Statistics, University of North Carolina-CH

- Farthing, M.W., C.E. Kees, T.S. Coffey, C.T. Kelley, and C.T. Miller (2003)
  "Efficient Steady-State Solution Techniques for Variably Saturated Groundwater Flow," In press: Advances in Water Resources

- Fuentes, M. (2003)
  "Testing for separability of spatial-temporal covariance functions"
  Submitted for publication

- Kavanagh, K.R., C.T. Kelley, C.T. Miller, M.S.C. Reed, C.E. Kees, and R.W. Darwin (2003)
  "Solution of a Well-Field Problem with Implicit Filtering"
  In review: Optimization and Engineering

- Kavanagh, K. R., C. T. Kelley, C. T. Miller, Mark S. C. Reed, C. E. Kees and Robert W. Darwin (2003)
  "Solution of a Well-Field Design Problem with Implicit Filtering"
  CRSC-TR03-08

- Mclaughlin, Richard, David Adalsteinnson, Nicole Abaid, and Akua Aguapong
  "An Internal Splash: Levitation of Falling Sphere's in Stratified Fluids"
  Submitted for publication

- Pan, C., M. Hilpert, and C.T. Miller (2003)
  "Lattice-Boltzmann Simulation of Two-Phase Flow in Porous Media"
  In review: Water Resources Research

- Pan, C., J.F. Prins, and C.T. Miller (2003)
  "A High-Performance Lattice Boltzmann Implementation to Model Flow in Porous Media," In review: Computer Physics Communication

- Serre, M.L., G. Christakos, H. Li, and C.T. Miller (2003)
"A BME Solution of the Inverse Problem for Saturated Groundwater Flow"
In press: Stochastic Environmental Research and Risk Assessment

- Zidek, J., J. Meloche, G. Shaddick, C. Chatfield and R. White (2003)
"A Computational Model for Estimating Personal Exposure to Air Pollutants with Application to London's PM_{10} in 1997" SAMSI Technical Report 2003-3

Reports in Preparation (as of April, 2003)

- Camassa, Roberto, Ken McLaughlin, Rich McLauglin, and James Bonn
"Effective mixing coefficients for passive transport"

- Chang, H, Fu, H, Le, ND, Zidek, J.V.
"Perspectives on Designing Environmental Monitoring Networks for Measuring Extremes"

- Fu, A, Le, N.D., Zidek, J.V.
"A Statistical Characterization of a Simulated Maximum Annual Rainfall Field Over Canada"

- Kavanagh, K. R., C. T. Kelley, C. T. Miller, M. Reed, C. Kees and R. Darwin
"A Comparision of Optimization Methods for Problems Involving Flow and Transport Phenomena in Saturated Subsurface Systems"

- Kavanagh, K. R. and C. T. Kelley
"Temporal Error Control and Pseudo-transient Continuation for Nonsmooth Dynamics"

- Le, N.D., Sun, L, Zidek, J.V.
"Designing Networks for Monitoring Multivariate Environmental Fields Using Data With Monotone Pattern"

- McLaughlin, Rich, Dan Anderson, and Casey Miller
"Gravity currents in heterogeneous porous media systems"

- Rich McLaughlin, Anne Bourlioux, and Roberto Camassa
"Non-Gaussian PDFs in time varying shear layers"

- Natvig, Bent and Tvete, Ingunn Fride (2003)
"Bayesian hierarchical modelling of spatial and temporal dependencies between earthquakes"

- Zheng, Xiaoyu and M. Gregory Forest
"On the strength of monodomain attractors for sheared nematic polymers"

H. Efforts to Achieve Diversity

From the beginning, SAMSI has focused on achieving diversity. This began with selection of advisory committees. On the National Advisory Committee, 3 of 10 were women, including the Co-Chair, Margaret Wright. On the Local Advisory Committee, 3 of 9 were women and one was an African-American. On the Education and Outreach Committee, there were two African-Americans and one woman on the 7 member committee.

In terms of program leadership, the Inverse Program and the Stochastic Computation Program had Co-Chairs who were women, and the Environmental Modeling Program had three women on the scientific advisory committee. Specific efforts, and successes, of each of the programs towards achieving diversity are indicated below.

Inverse Program: Program leaders actively recruited at HBCU's (Clark Atlanta University, Benedict College, Spellman, Elizabeth City State University, NC A&T) as well as at universities with known programs for under-represented groups (e.g., the program in applied mathematics at Cornell). This resulted in significant participation by under-represented groups including young investigators (defined here as students or professionals less than 5 years after graduate degree). The summary of this participation includes:

- For the Opening Workshop, out of 122 attendees, 90 attended the tutorials for young investigators, 68 were young investigators, 39 were female, 11 were Hispanic, and 5 were African-American.
- From the National and International Core Participants, 5 out of 16 participants were from underrepresented groups.
- From RTP Core Participants, 8 out of 12 were female, 3 out of 12 were African-American.

Stochastic Computation Program: Through repeat advertising and professional networks, extensive efforts were made to bring the StoCom program and its workshops to the attention of women researchers and members of underrepresented groups. The program involves a substantial number of women, especially among the new researchers and students both in the core program and in the workshops. The summary of this participation includes:

- For the Opening Workshop, out of 129 attendees, 74 were young investigators, 40 were female, 8 were Hispanic, and 2 were African-American.
- From the National and International Core Participants, 3 out of 19 participants were from underrepresented groups.
- From RTP Core Participants, 3 out of 19 were from underrepresented groups.

Environmental Modeling Program: The program involves a substantial number of women, especially among the new researchers and students both in the core program and in the workshops. The summary of this participation includes:

- For the Opening Workshop, out of 84 attendees, 31 were young investigators, 26 were female, 4 were Hispanic, and 2 were African-American.
- From the National and International Core Participants, 5 out of 26 participants were from underrepresented groups.
- For the Optimization and Simulation Workshop, out of 49 attendees, 23 were young investigators, 18 were female, and 3 were African-American.

Education and Outreach Program: SAMSI leaders actively recruited at HBCU's (Clark Atlanta University, Benedict College, Spellman, Elizabeth City State University, NC A&T) which resulted in major participation by under-represented groups. For the one day undergraduate workshops in November and February, out of 58 participants, 35 were minorities.

For the Undergraduate Workshop scheduled for June 9-13, 2003, 50% of registrants to date are from HBCU's. We anticipate similar representation at the CRSC/SAMSI Industrial Mathematical and Statistical Modeling Workshop scheduled for July, 2003.

I. External Support

Additional Program Funding

- Kenan Foundation: $50,000
- Army Research Office: $10,000 in support of the Optimization and Simulation Workshop

Affiliated Organizations

Because NISS already has a strong Affiliates program, SAMSI does not intend to have a separate Affiliates program. Instead, in several senses, the affiliate program is a joint program. First, terms of the NISS Affiliates Program allow reimbursement of expenses to attend SAMSI workshops as well as NISS events. To date, this reimbursement has totaled $10,000, and should significantly increase.

Second, reflecting the strong attraction of SAMSI to university departments of mathematical sciences as well as statistical sciences, the university component of the NISS Affiliates Program was recast in January 2003 as the NISS/SAMSI University Affiliates Program. An Assistant Director of NISS has responsibility for operation of this program.

Finally, the NISS Industrial Affiliates and NISS/SAMSI University Affiliates inform the development of SAMSI programs and contribute personnel to the ongoing programs, as detailed in the discussions of the particular programs.

NISS AFFILIATES

Corporations

Amgen, Thousand Oaks, CA
Avaya Labs, Basking Ridge, NJ
General Motors, Detroit, MI
GlaxoSmithKline, Research Triangle Park, NC and Collegeville, PA
Merck & Company, West Point, PA
MetaMetrics, Durham, NC
Pfizer Inc., Groton, CT
RTI International, Research Triangle Park, NC
SAS Institute, Cary, NC
SPSS, Chicago, IL
Telcordia Technologies, Morristown, NJ

Government Agencies and National Laboratories

Bureau of Labor Statistics, Washington, DC
Bureau of Transportation Statistics, Washington, DC

US Census Bureau, Washington, DC
US Environmental Protection Agency, Research Triangle Park, NC
Los Alamos National Laboratory
National Agricultural Statistics Service, Fairfax, VA
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
Pacific Northwest National Laboratory, Richland, WA

NISS/SAMSI University Affiliates

American University, Department of Mathematics and Statistics
UCLA, Department of Statistics and Statistical Consulting Center
Carnegie Mellon University, Department of Statistics
Duke University, Institute of Statistics and Decision Sciences and Dept.of Mathematics
Emory University, Department of Biostatistics
Florida State University, Department of Statistics
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 Mathematical Sciences
University of Maryland, Department of Mathematics
University of Maryland Baltimore County, Department of 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
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
University of Pennsylvania, Department of Statistics
Purdue University, Department of Statistics
Rice University, Department of Statistics
Rutgers University, Department of Statistics
Southern Methodist University, Statistical Science Department
Stanford University, Department of Statistics
Texas A&M University, Department of Statistics

J. Advisory Committees

The four advisory/oversight committees of SAMSI are as follows:

- The Governing Board (GB), which oversees SAMSI's administration, finances, and partner organization relationships. The GB met with the Directorate on September 3, 2002, to review the start-up activities of SAMSI. 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 October 12, 2002, to review the progress in the first-year 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 has met with the Directorate on September 13 and 16, 2002, and on January 6, 2003.

- The Chairs Committee, which 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

  Note that the Chairs are also ex officio members of the LDC. Meetings with the Chairs were held preceding (and during) the LDC meetings mentioned above.

The membership of each of these committees during the past year is given in the table on the following page.

Committee	Name	Affiliation	Field
Governing Board	John Harer	Duke, As.Provost	Mathematics
	Douglas Kelly	UNC, Dean	Statistics
	Jon Kettenring	Telcordia	Math Sciences
	Daniel Solomon	NCSU, Dean	Statistics
National Advisory Committee	Mary Ellen Bock	Purdue	Statistics
	Peter Bickel (Co-Chair)	UC Berkeley	Statistics
	Lawrence Brown	Pennsylvania	Statistics
	David Heckerman	Microsoft	CS&Statistics
	John Lehoczky	Carnegie-Mellon	Probability
	Sallie-Keller McNulty	LANL	Statistics
	George Papanicolaou	Stanford	Mathematics
	Adrian Raftery	Washington	Statistics
	Shmuel Winograd	IBM	Mathematics
	Margaret Wright (Co-Chair)	NYU	CS
Local Development Committee	Andrea Bertozzi	Duke	Mathematics
	Gregory Forest	UNC	Mathematics
	Jean-Pierre Fouque	NCSU	Mathematics
	Alan Gelfand	Duke	Statistics
	Mark Genton	NCSU	Statistics
	Jacqueline Hughes-Oliver	NCSU	Statistics
	Joseph Ibrahim	UNC	Biostatistics
	Christopher Jones	UNC	Mathematics
	Thomas Kepler	Duke	Bioinformatics
	Andrew Nobel	UNC	Statistics
	David Schaeffer	Duke	Mathematics
	Scott Schmidler	Duke	Bioinformatics
	Ralph Smith	NCSU	Mathematics
Chairs Committee	Edward Carlstein	UNC	Statistics
	Clarence E. Davis	UNC	BIostatistics
	Jean-Pierre Fouque	NCSU	Mathematics
	David Morrison	Duke	Mathematics
	Sastry Pantula	NCSU	Statistics
	Dalene Stangl	Duke	Statistics
	William Wilkinson	Duke	Biostatistics
	Warren Wogen	UNC	Mathematics

II. Special Report: Program Plan

A. Plans for 2003-2004

I. DATA MINING AND MACHINE LEARNING PROGRAM (full year)

Data mining and machine learning – the discovery of patterns, information and knowledge in what are almost always large, complex (and, often, unstructured) data sets – have seen a proliferation of techniques over the past several years. Yet, there remains incomplete understanding of fundamental statistical and computational issues in data mining, machine learning and large (sample size or dimension) data sets.

The goals of the SAMSI DM&ML Program are to advance this understanding significantly, to articulate future research needs for DM&ML, especially from the perspective of the statistical sciences, and to catalyze the formation of collaborations among statistical, mathematical and computer scientists to pursue the research agenda.

The program will operate similarly to the 2002-03 Stochastic Computation synthesis program, with several working groups focusing on specific issues of theory and methodology framed by particular testbed data sets. The precise foci of the Working Groups will be identified in conjunction with the September, 2003 Opening Workshop. Based on discussions of the scientific committee and with program participants, current candidates for Working Group foci are:

- 'Large p, small n inference;'
- Bioinformatics;
- Support vector machines;
- Computational experiments that relate performance of DM&ML techniques to problem characteristics;
- Text mining, and other novel, complex forms of data;
- Statistical questions such as sampling, model selection and search, robustness, data quality and multiplicity inherent in many DM&ML techniques.

Computational scalability will be a pervasive theme.

Testbed data sets are currently being identified. Possibilities include bioinformatics (compound activity), product performance (warranty), software engineering (change histories) and spontaneous reporting systems (example: adverse events arising from pharmaceuticals).

Workshops: Two major workshops are planned:

1. *Tutorials and Opening Workshop*, September 6-10, 2003. Confirmed invited speakers include Di Cook (Statistics, Iowa State), Michael Jordan (Computer Science, UC Berkeley), J. S. Marron (Statistics, North Carolina), David Madigan (Statistics, Rutgers) and Robert McCulloch (Statistics, Chicago). The format will be highly participatory, with Birds-of-a-Feather Sessions, a poster session and associated session of two-minute Poster

Sales Talks, a Second Chance Seminar at which anyone can talk, a Young Researchers Session, and Working Group Meetings at SAMSI.

2. *Closing Workshop* on May 20-22, 2004, at which findings of the program will be presented to the community.

We also expect to hold several mini-workshops (1 day) that explore ramifications of DM&ML for other issues, or treat related problems such as text mining, as well as a 'mid-term' workshop in January, 2004, at which program participants will assess progress and set the course of the remainder of the program.

Other Activities: The working groups that form to investigate particular aspects of DM&ML will hold regular weekly meetings, often with attached seminars.

A SAMSI course, *Data Mining and Machine Learning*, will be taught by Professors David Banks and Feng Liang on Wednesdays in the Fall Semester, 2003, from 4:30-7:00 pm.

Program Leaders: David L. Banks (FDA; moving to Duke Statistics in Fall, 2003) and Alan F. Karr are Program co-chairs. The *scientific committee* consists, in addition, of Mary Ellen Bock (Purdue and NAC), Jerome Friedman (Statistics, Stanford), David Madigan (Statistics, Rutgers), William DuMouchel (AT&T) and Warren Sarle (SAS Institute).

Participants: Committed national and international core participants to date are the *SAMSI University Fellow*, Bertand Clarke (Statistics, British Columbia), together with Hugh Chipman (Statistics, Waterloo), Di Cook (Statistics, Iowa State), Prem Goal (Statistics, Ohio State), David Madigan (Statistics, Rutgers), Marianthi Markatou (Statistics, Columbia), Vijay Nair (Statistics, Michigan), Grzegorz Rempala (Biostatistics, Louisville), Michael Tarter (Biostatistics, University of California Berkeley), and Henry Wynn (EURANDOM).

Postdoctoral Fellows for the program are Ernest Fokoue (Ph.D., Statistics, Glasgow, 2001), Assistant Professor at Ohio State, who will take leave to come to SAMSI; and Xiaodong Lin (Ph.D., Statistics, Purdue, 2003), who will come to SAMSI for a year en route to a faculty appointment at the University of Cincinnati. (This will be a shared appointment with NISS, reflecting Lin's interests in privacy-preserving data mining.) Two other NISS postdocs have expressed strong interest in the program, and are likely to participate very meaningfully.

Faculty Release Fellows from the partner universities include David Banks (Statistics, Duke), Marc Genton (Statistics, North Carolina State), Jacqueline Hughes-Oliver (Statistics, North Carolina State), Feng Liang (Statistics, Duke), and Young Truong (Biostatistics, UNC).

Graduate Fellows so far selected are Jen-Hwa Chu (Statistics, Duke) and Fei Liu (Statistics, Duke). Two additional graduate fellows will be selected from UNC and NCSU.

Other local participants that will be extensively involved in the program are Gary McDonald (NISS), Ashish Sanil (NISS), Warren Sarle (SAS Institute), and Stanley Young (NISS). We expect that additional participants, international, national and local, will come forward and be incorporated into the program on an ongoing basis.

II. NETWORK MODELING FOR THE INTERNET (Fall 2003)

Because of the size and complexity of the internet, and the nature of the protocols, internet traffic has proven to be very challenging to model effectively. Yet modeling is critical to improving Quality of Service and efficiency. The main research goal of this program is to address these issues by bringing together researchers from three communities:

- Applied probabilists studying heavy traffic, queueing theory and fluid flow models
- Mainstream Internet traffic measurers/modelers and hardware/software architects
- Statisticians

The timing is right for simultaneous interaction among all three communities, because of the current trend away from dealing with Quality of Service issues through over-provisioning of equipment. This trend suggests that heavy traffic models will be ideally situated to play a leading role in future modeling of Internet traffic, and in attaining deeper understanding of the complex drivers behind Quality of Service. This SAMSI program will catalyze this process through building strong bridges among the three communities.

An additional goal is to enhance contact between these research communities and potential industrial partners. The location of SAMSI is ideal for this purpose, because the Research Triangle is becoming a world center for the networking industry.

Program activities will be centered around three main topics (with a possible 4th topic, depending on pending funding). Main events for each of these will be initiated by a short workshop:

1. *Measurement and Modelling*, September 18-19.
2. *Congestion Control and Heavy Traffic Modeling*, September 19-20.
3. *Internet Tomography*, September 13-14.
4. *Sensor Networks*, September 14-15.

These workshops are intended as a main drawing card for participants. Because of the deeper collaborative goals of SAMSI, our workshops will have a somewhat different format from the usual. First back to back workshops are deliberately paired, with the goal of promoting collaboration between related groups. Another mechanism for encouraging serious engagement, and for encouraging further participation (either through a longer stay at SAMSI, or via return visits for follow up purposes), is the idea of "theme problems".

These theme problems are intended both to engage a large number of researchers, and also to feed into subsequent SAMSI program activities. This subsequent research will be carried out by SAMSI Graduate students, Post-Docs, Triangle University faculty participants, long term visitors, and by return visits from workshop visitors that choose to commit themselves to the ongoing research.

The SAMSI Workshop Theme Problems will be:

1. Measurement and Modelling: "What causes burstiness?" There has been some controversy over this. One important viewpoint is the elegant mathematical theory of heavy tail durations leading to long range dependence. But others have suggested TCP effects as another potential cause. There may be other factors as well. A common statistical principal is that "causation" can only be established via a designed experiment. For this purpose, SAMSI has obtained some time in the UNC Distributed and Real Time Systems Network Lab. The theme problem of the workshop will be to decide: (1) What are the set of problems to be tackled by a series of lab experiments? (2) How should the experiments be designed?

2. Congestion Control and Heavy Traffic Modeling: . Here the theme problem will be "Queueing Delays". Do queueing delays help stabilize TCP or not, and how do they influence burstiness? In the future, queueing delays may become smaller, as a consequence of increasing capacity and of developments such as Explicit Congestion Notification, where a bit in the IP header rather than a dropped packet is used to signal congestion. Can we understand better the inter-relationship of congestion control and queueing delays? Again lab experiments will be used to understand relationships and to test the models. The workshop is intended to focus on a particular set of questions and on the experimental designs needed to answer them.

3. Internet Tomography: The theme problem in this area will be "validation". Our hope is to use the UNC Network as a test bed. Negotiations are currently under way for Avaya to deploy their network analysis tool, which is expected to give a very clear picture of important aspects of the network. The idea is to try out a range of tomography tools, and then compare them with the right answer. Of course the scale is not that of the internet, but it is much larger than the usual lab test beds.

4. Sensor Networks: NSA funding is pending on this topic. We see a large value added to bringing together these people and the Internet Tomography folks. Our partner in this effort is: CENS, http://www.cens.ucla.edu/. Here the theme

problem will be the identification of “focus areas” for future collaboration. Some initial ideas include localization, acoustic (and seismic) tracking, and which data to store and which to send along.

Other key workshop ideas are organizational. Here are some of them:

a. To keep people interested and focused, the time for each workshop will be short: 1.5 days. For students, post-doctoral fellows and participants that are not completely familiar with either area, a half day of tutorials will be offered the day before each pair of workshops.

b. In such a short time, a large number of conventional long talks about current research are impractical. However, since everyone would like to hear about new work, there will be a series of 5 minute presentations near the beginning of the workshop, featuring sound bite versions of current work, and a table for papers to be put out, for those who want to follow up more. Also it is clear that time for personal discussion needs to be allowed in that process.

c. There will be a few talks, by carefully chosen speakers, with the goal of bringing up some of the important issues around the theme problems.

d. There will be small group discussions where the real work will be done on the theme problems. This gives every participant a chance to be an "owner" of the process. To keep things on track, each group will be responsible for reporting back to the main group.

e. With a view towards the future, we are planning a "what I wish I could do" session. Here again presentations will be very short (either 5 or 10 minutes, depending on the number of presenters). This is NOT about current work, or even what is being currently proposed in grant applications. Instead we are looking for ideas that are "farther out there". Ideal for this would be a discussion of problems that are perceived, but where potential solutions are very unclear.

Other Activities: Following each workshop, working groups will be formed to follow up on the theme problems. These will hold regular weekly meetings, often with attached seminars.

Two SAMSI courses, will be taught, one on Tuesdays, the other on Thursdays, both meeting from 4:30 – 7:30. The Tuesday course will be *Statistical Analysis and Modelling of Internet Traffic Data,* taught by J. S. Marron. The Thursday course will be *Long-range Dependence and Heavy Tails,* taught by Murad Taqqu.

Program Leaders: J. S. Marron (North Carolina/SAMSI) and Ruth Williams (California San Diego, are Program co-chairs. The *scientific committee* consists, in addition, of Kevin Jeffay (North Carolina), James Landwehr (Avaya Laboratories), John P. Lehoczky (Carnegie Mellon) , Donald Towsley (Massachusetts) and Walter Willinger (AT&T Research).

Participants: Anticipated core participants include the *SAMSI University Fellow*, Murad Taqqu (Boston), and program-long participation by George Michailidis (Michigan) and Ian Dinwoodie (Louisiana State). Shorter visits are expected from a number of key participants, including Maury Bramson (Minnesota), W. S. Cleveland (Bell Labs), k c claffy (UCSD), Rene Cruz (UCSD), David Culler (Berkeley), Jim Dai (Georgia Tech.), Mark Crovella (Boston), Lorraine Denby (Avaya), Paul Dupuis (Brown), Deborah Estrin (UCLA), Mike Harrison (Stanford), Frank Kelly (Cambridge), Thomas Mikosch (Copenhagen), Robert Nowak (Rice), Vern Paxson (ICIR), Sidney Resnick (Cornell), Rolf Riedi (Rice), Nina Taft (Sprint), Yehuda Vardi (Rutgers), Bin Yu (Berkeley).

Postdoctoral Fellows for the program are Cheolwoo Park (PhD in Statistics from the Seoul National University, currently a Postdoctoral Fellow at UNC), and (anticipated) David Rolls (Mathematics, Queens University, currently a Visiting Assistant Professor at UNC Wilmington).

Faculty Release Fellows from the partner universities are Robert Buche (Mathematics, NCSU), Vladas Pipiras (Statistics, UNC), and F. D. Smith (Computer Science, UNC).

Graduate Fellows so far selected are Juhyun Park and Arka Ghosh (Statistics, UNC), and a graduate student from Computer Science, UNC (who will be responsible for running laboratory experiments).

Other local participants that will be extensively involved in the program are Felix Hernandez Campos (Computer Science, UNC), Kevin Jeffay (Computer Science, UNC), Vidyadhar Kulkarni (Operations Research, UNC), Arne Nilsson (Electrical Engineering, NCSU), Andrew Nobel (Statistics, UNC), Richard Smith (Statistics, UNC), Michele Trovero (SAS and UNC), Amin Vahdat (Computer Science, Duke), Robert Wolpert (Statistics, Duke) and Zhengyuan Zhu (Statistics, UNC). We expect that additional participants, international, national and local, will come forward and be incorporated into the program on an ongoing basis.

III. Multiscale Model Development and Control Design (Spring, 2004)

One of the areas identified by the Panel on Future Directions in Control, Dynamics, and Systems, as a crucial growth area for the next decade, is model development and control design for systems utilizing novel material architectures. To achieve these goals for applications including quantum computing, nanopositioning, artificial muscle design, flow control, liquid crystal polymers, and actuator implants to stimulate tissue and bone growth, it is necessary to develop multiscale modeling hierarchies ranging from quantum to system levels for time scales ranging from nanoseconds to hours. Control techniques must be designed in concert with the models to guarantee the symbiosis required to achieve the novel design specifications. A crucial component in multiscale analysis is the development of homogenization techniques to bridge disparate temporal and spatial scales, and statistical homogenization appears to provide one of the most promising techniques for developing tractable models. *The goal of this SAMSI program is thus to provide a forum to develop nano- and microscale modeling methodologies, statistical* *homogenization techniques, and commensurate robust control designs for advanced material architectures.*

To illustrate two diverse fields in which multiscale modeling is of paramount importance, we summarize briefly perspectives on nanopositioning and electron spin detection, and optimal design of polymeric elastomers. While these are highly specific applications, they illustrate modeling and control requirements for a broad spectrum of current and proposed applications in material science and smart materials, biology and biomedical systems, chemical, electrical, mechanical and computer engineering, quantum physics, applied mathematics and statistics.

All current nanopositioners employ piezoceramic actuators as transducers due to their ultrahigh set point accuracy and stability. Similar devices are being considered for initial experiments to detect electron spins. However, these actuators exhibit inherent hysteresis, constitutive nonlinearities, and thermal noise which must be accommodated to achieve the required angstrom-level accuracy. At low drive frequencies, this can be accomplished through feedback laws but at the higher frequencies required for product diagnostics and real-time tracking of biological processes, the noise to data ratios increase to the point where feedback mechanisms are amplifying noise rather than attenuating unmodeled dynamics. To achieve the desired performance, it is necessary to develop multiscale models which employ quantum and microscale energy relations to predict effective parameters in system models used to characterize the devices. Furthermore, it will be necessary to develop robust control designs based on these multiscale models to achieve the required angstrom-level resolution at high scan rates.

A prototype for the second application is a unimorph comprised of an elastomer bonded to a secondary polymer. Unimorph constructions are currently being considered for use as remote lens cleaners for aerospace missions, acoustic pressure sensors, flow control actuators and artificial jets for aircraft, artificial muscles for robotic components, and actuator implants to stimulate tissue and bone growth. To optimize the performance of the macroscopic devices, it is necessary to optimize the stoichiometry and constitutive behavior of the grafted crystalline groups at the microscopic level. Stochastic homogenization will then be employed to develop a hierarchy of models in which macroscopic measurements are employed to update microscopic models to achieve system level control criteria.

In the context of the proposed program, multiscale models are categorized as having the following scales: quantum, microscopic, mesoscopic, macroscopic, system, full-order numerical and reduced-order numerical. We consider mesoscopic to apply at the lattice or domain level. We note that fundamental energy relations are typically derived at the quantum, microscopic or mesoscopic levels whereas bulk measurements occur primarily at the macroscopic level, and design and control implementation occur for systems which are often constructed from multiple constituent materials. Combined statistical and deterministic modeling methodologies are required at each of the levels as well as to provide statistical homogenization techniques to bridge the levels.

Quantum models are too complex for use in control design at the system level; however, they can be used to predict certain facets of material behavior and to ascertain properties of effective parameters in microscopic or mesoscopic energy models. Models at this level must inherently combine statistical and deterministic properties due to the probabilistic nature of the quantum models. State estimation and compensators used for feedback control must retain this probabilistic nature, and robust control design for systems utilizing quantum devices must incorporate the uncertainties inherent to these devices.

Classical energy relations apply at the microscopic and mesoscopic levels but a number of issues require symbiosis between deterministic and stochastic models. Thermal noise can be significant at this level which necessitates the use of Boltzmann probabilities to determine the likelihood of achieving certain energy levels. Boltzmann statistics provides one of the fundamental tools used to develop hysteresis models for high performance ferroelectric, ferromagnetic and ferroelastic materials, including the piezoceramic employed in the motivating nanopositioner. The complexity of dynamics at the microscopic and mesoscopic levels also necessitates the development of statistical techniques to produce tractable models.

Homogenization of some form is required to extend energy relations developed at the quantum, microscopic or mesoscale levels to obtain macroscale or system models having states commensurate with sensors and control algorithms. A variety of homogenization techniques exist including mathematical homogenization theory based on periodic microstructures and techniques based on dispersion models. The first technique is based on rigorous mathematical analysis but is too computationally intensive to permit direct real-time control implementation or optimal design of a microscopic structure given macroscopic measurements. An attractive alternative is to develop stochastic homogenization techniques which treat coefficients or dynamics of the microscopic models as random variables with unknown probability distributions which must be estimated through inverse problem techniques at the macroscopic level. This yields macroscopic models which incorporate the underlying energy characteristics but have a small number of effective parameters which renders them efficient for system design and control implementation.

From a stochastic perspective, a dominant theme in multiscale modeling focuses on providing stochastic specifications for the analysis of spatial data. In simplest terms, spatial pattern and spatial structure may look very different as we work at different scales of resolution (the so-called ecological fallacy). One inferential problem involves imputing spatial structure at one scale given that spatial data has been collected at a different scale. This problem is referred to as the change of support problem. Ad hoc procedures are apparent but stochastic modeling enables full inference including the opportunity to capture the uncertainty in the modeling. A related problem is to link spatial data layers collected at different scales. In particular, if one is viewed as providing the response variable and the others as providing explanatory variables, then a regression of the response on the others would be sought. Again ad hoc procedures can be envisaged but again stochastic modeling enables formal inference in order to assess adequacy of models, to compare models and to appreciate the variability associated with a particular model.

The goals this Program are both fundamental and broad. From a fundamental perspective, the program will focus on the development of multiscale modeling techniques which span disparate spatial and temporal scales and hence are appropriate for material characterization, material design, and control implementation. The models inherently combine deterministic and stochastic features which must be addressed in a cohesive manner to achieve model and control requirements. From a broad perspective, we will develop the methodologies in a manner which is appropriate for a wide range of materials and high performance applications.

A substantial component of the program will focus on the development of appropriate approximation techniques to obtain convergent full-order numerical models for simulation purposes and highly efficient reduced-order models for real-time experimental implementation of the control designs. One of the reduced-order techniques to be considered is based on the Karhunen-Loeve or Proper Orthogonal Decomposition (POD) theory for obtaining optimal cohesive structures from numerically or experimentally generated data sets. A second aspect of reduced-order model and control design will focus on the development of theory pertaining to LQG balancing for infinite-dimensional systems. Finally, a crucial component of the SAMSI program is the experimental validation of the developed theory to ascertain its subsequent applicability in high performance systems.

The progress made to date in multiscale development has come primarily from within the material science, physics and applied mathematics communities while the robust control design utilizing such models has been developed mostly within the mathematics and electrical engineering communities. The statistical methods currently being employed are, in general, based on principles from statistical physics with few recent contributions from the statistics community. The lack of collaboration between applied mathematicians, material scientists, physicists and statisticians in this area is due to historical reasons with each group tending to focus on that niche of the problem falling within their discipline. The goal of the SAMSI program is to provide an environment which fosters collaboration between applied mathematicians, material scientists, engineers, physicists and statisticians to investigate fundamental and unified techniques for multiscale modeling and control design for advanced materials. To provide motivating applications and to indicate robust control criteria to be developed, certain projects investigated both by CRSC scientists and the general materials community will be employed as catalysts; however, the primary contributions from the program will reside in the collaborative theory and designs fostered between the previously mentioned groups of scientists rather than intermediate results obtained for motivating applications. It is this unique synergistic feature of the SAMSI program which provides the greatest potential for fundamentally contributing to the development of multiscale modeling and control techniques for advanced materials.

Workshops: Three major workshops are planned:

1. *Opening Workshop and Tutorials*, January 17-21, 2004, will focus on mathematical and statistical challenges in the field of multiscale analysis for advanced material architectures.

2. *Workshop on Multiscale Phenomena in Soft-Matter, Nanomaterials*, will be held from February 15-17, 2004.

3. *Workshop on Fluctuations and Continuum Equations in Granular Flow*, will be held from April 15-17, 2004.

Additional mini-workshops, to be scheduled, will address specific issues associated with multiscale modeling, homogenization, large scale simulation, reduced-order system models, and robust control design and implementation.

Other Activities: The working groups that form to investigate particular aspects of the program will hold regular weekly meetings, often with attached seminars.

A SAMSI course, *Multiscale Modeling*, will be taught by Professors Alan Gelfand and Ralph Smith in the Spring Semester, 2004.

We have solicited two projects related to advanced materials for the Industrial Mathematical and Statistical Modeling Workshop for Graduate Students to be held from July 21 - July 29, 2003, through SAMSI sponsorship. The SAMSI postdoc and graduate students for the Multiscale Program will participate in the workshop to help investigate the projects and facilitate their transition into the Multiscale Program. The industrial presenters for these projects and their affiliations are as follows:

- David Dausch, MCNC - Research and Development Institute, Research Triangle Park, NC: Model development for PZT polymers or flexible film actuators.
- Chad Bouton, Battelle Memorial Institute, Columbu, OH: Model development for advanced actuator systems.

Program Leaders: Greg Forest (UNC), Doina Cioranescu (Universite Pierre et Marie Curie), Alan Gelfand (Duke; Co-Chair), David Schaeffer (Duke), Murti Salapaka (ISU), Ralph Smith (NCSU; Chair), and Christopher Wikle (University of Missouri).

Participants: Committed and tentatively committed national and international core participants to date are the *SAMSI University Fellow*, Arthur J. Krener (UC Davis), tentative *SAMSI University* Fellow Andrew Newell (UCSB), Tad Calkins (Boeing), Lawrence Robertson (Kirtland AFB), James Redmond (Sandia National Laboratories), Karla Mossi (Virginia Commonwealth University), Zoubeida Ounaies (Virginia Commonwealth University), Murti Salapaka (Iowa State University), David Jiles (Iowa State University), Marcelo Dapino (Ohio State University), Ioannis Kevrekidis (Princeton University), Don Leo (Virginia Tech), Doug Lindner (Virginia Tech), Tom Bewley (UC San Diego), Miroslav Krstic (UC San Diego), Bassam Bamieh (UC Santa Barbara), Igor Mezic (UC Santa Barbara), and Petar Kokotovic (UC Santa Barbara).

Postdoctoral Fellows for the program are to be determined.

Faculty Release Fellows from the partner universities include Alan Gelfand (Statistics, Duke) and Ralph Smith (Mathematics, North Carolina State).

Graduate Fellows include Jon Ernstberger (Mathematics, NCSU), Jimena Davis (Mathematics, NCSU), Sarah Grove (Mathematics, NCSU), and Eric Vance (Statistics, Duke).

Industrial/Government Laboratory Participants: Significant effort has been focused on the identification of problems involving advanced material architectures in aerospace, aeronautic, DoD, automotive and industrial applications. During the Spring, 2003 semester, we have identified and talked with the following industrial and government laboratory contacts about state-of-the-art problems to be considered in the Multiscale Program.

- Tad Calkins, Boeing: 2-D and 3-D shape memory alloy (SMA) chevrons for noise attenuation in jet engines.
- Lawrence Robertson, Kirtland AFB: Vibration reduction and adaptive optics for membrane mirrors.
- James Redmond, Sandia National Laboratories: Thin film SMA MEMs for flow control.
- Joycelyn Harrison, NASA Langley Research Center: Polymeric elastomers for synthetic jets and flow control.

IV. Planning Workshops (Fall, 2004)

Two planning workshops are being planned for Fall, 2004:

1. Leaders and key participants of the likely Year 3 Program *Computational Biology of Infectious Diseases* will meet to determine the key foci of the program. The date for this meeting has not yet been set.

2. Leaders and potential participants of the likely Year 4 Program *High-Dimensional Inference and Random Matrices* will meet to determine the key foci, both theoretical and applied, of the program. The planning for this program is particularly difficult because of the diversity of possible goals and application areas of the program. We hope to hold this meeting at the AIM Research Conference Center (ARCC) in Stanford for two reasons:

- Two of the proposed leaders (Iain Johnstone and Craig Tracey) are from the area, as are numerous potential participants.
- ARCC has very interesting techniques to focus workshops towards a needed goal, and we hope to use (and learn) this expertise to plan the program.

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.

I. Computational Biology of Infectious Disease (tentative for Fall, 2004)

Within the wider context of computational biology, this program will focus on infectious disease for strategic, scientific and humanitarian reasons. Strategically, studies of infectious disease have had a very long history of successful engagement with mathematics, from Daniel Bernoulli's foundational paper on the epidemiology of vaccination against smallpox, to the adoption of biostatistics as a cornerstone of clinical trials, to the enormous contributions of statistical genetics to understanding of the determinants of susceptibility, and many other contributions.

These successes have not been lost on those concerned over recent biodefense challenges. The NIH and other federal agencies have made decisions to invest substantially in mathematical approaches to biodefense and infectious disease.

The scientific incentive has largely to do with tremendous increases in the availability of relevant data, most obviously from the human and microbial genome projects and their related technologies. The primary process by which the information in the genome becomes manifest in the structure and function of the cell and organism is through gene expression and its regulation. Thus, while focusing on the specific topic of infectious disease, we are given the opportunity to explore the full range of topics of interest in computational biology generally.

The proposed program thus has two major aims.

- To identify those areas where mathematical innovation may have the greatest impact on the basic science and medicine of infectious disease.
- To move toward resolution of the cultural divisions between the biomedical and mathematical sciences and toward the education of a new generation of biologists and mathematicians capable of working in tightly integrated collaborative teams.

The program is currently envisioned as consisting of two intertwining strands: population dynamics, and genomic dynamics. The former category is represented by epidemiology, social network theory, microbial ecology, host-pathogen co-evolution, drug resistance evolution, immune-response modeling, and so on. The latter is the domain of comparative genomics; molecular evolution; gene expression modeling and microarray data analysis; proteomics, drug target identification and vaccine design, and others.

It is hard to imagine a topic that is more timely. The emergence of a novel virus (the agent of SARS), the growing alarm over drug-resistant bacteria, and the concerns over long-term biodefense, coupled with the explosion of genomic data from microbes as well as host organisms, make the areas proposed here some of the most fertile in the applied mathematical sciences today.

Activities of the program will include opening and closing workshops, regular working group meetings, and two courses, ‘Computational Genomics for Infectious Disease’ and ‘Population Dynamics for Infectious Disease.’

There are a wide variety of potential industrial and governmental organizations that are likely to seek involvement in the program. In addition, the NIH, National Institute of Allergy and Infectious Disease, Burroughs-Wellcome Foundation, Ellis Foundation, and the Gates Foundation are potentially interested in the program and are potential sources of additional funding.

We expect this program to have an impact on the practice of computational biology, mathematics and statistics, by making biologists aware of the value to be gained by involving mathematically sophisticated personnel in their research teams, and by making statisticians and mathematicians aware of the most outstanding problems in biology.

Potential Program Leaders: Thomas B Kepler (Biostatistics and Bioinformatics, Duke) and Denise Kirschner (Microbiology and Immunology, Michigan).

Potential Scientific Committee:

Local

Name	**Discipline**	**Affiliation**
Atchley, William	Genetics	NCSU
Cowell, Lindsay	Immunology	Duke
Elston, Tim	Applied Mathematics	UNC
Nobel, Andrew	Statistics	UNC
Schmidler, Scott	Statistics and Decision Science	Duke
Weir, Bruce	Statistical Genetics	NCSU

Global

Name	**Discipline**	**Affiliation**
Anderson, Roy	Theoretical Epidemiology	Imperial College, London
Antia, Rustom	Biology	Emory
Bergstrom, Carl	Biology	University of Washington
Casadevall, Arturo	Microbiology and Immunology	Albert Einstein College of Medicine
Castillo-Chavez, Carlos	Biological Statistics and Computational Biology	Cornell
Gupta, Sunetra	Mathematical Epidemiology	Oxford
Perelson, Alan	Theoretical Biology and Biophysics	Los Alamos
Tan, Man-Wah	Immunology	Stanford
Wong, Wing	Statistics	Harvard

II. Social Sciences (tentative for 2004-2005)

The planned Focused Study Program on the Social Sciences is in the process of being formulated. Currently, we anticipate that it will be a year-long series on linked "sub-programs" that address topics such as:

- Agent-based simulation
- Social networks
- Genetic and environmental influences on behavior
- Longitudinal analysis
- Causality.

Potential program leaders, all of whom are participating in planning discussions, include: Kenneth Bollen (Sociology, North Carolina), Siddartha Dalal (Xerox), Stephen Fienberg (Statistics, Carnegie Mellon), Susan Murphy (Statistics, Michigan), Edward Scheinerman (Mathematical Sciences, Johns Hopkins) and Stanley Wasserman (Psychology, Illinois). Other potential participants are Thomas DiPrete (Sociology, Duke), Andrew Gelman (Statistics, Columbia), Robert Groves (Sociology, Maryland), Kenneth Land (Sociology, Duke), Scott Long (Sociology, Indiana), Adrian Raftery (Washington) and Burt Singer (Demography and Public Affairs, Princeton).

III. Data Assimilation (tentative for Spring, 2005)

With the burgeoning quantities of available data, their role in model formation has become a major topic of research. The assimilation of data into models is at its most advanced level in numerical weather prediction. The most established method is variational in which analysis of model output in light of data is made by a “best-fit” strategy. Sequential data assimilation is a more dynamically based approach in which the Kalman filter is used to accurately estimate the state variables using the expected error statistics.

The issue of assimilating data into models arises in all areas which enjoy a profusion of data. In its broadest sense, it is the subject that arises at the meeting point of data and models. Technology has driven the advances on both sides of the equation: new techniques of measurement have led to an enormous surge in the amount of available data and ever faster computers have given us the capability of new levels of computational modeling. The development of effective methods of data assimilation must now be viewed as one of the fundamental challenges in scientific prediction. Nevertheless, the part of the scientific community interested in these issues has been limited. With the growing need for good methods and the advances in computational and observational capabilities, we now have a tremendous opportunity to bring the relevant scientific areas together in a focused effort aimed at developing new approaches, understanding the underlying issues and testing implementations of new schemes.

Many areas of science, applied mathematics and statistics can be brought to bear on the question of data assimilation. It has been largely developed within the confines of atmospheric sciences, but mathematical and statistical techniques developed for other purposes could offer significant advances, for example dynamical systems theory casts light on the evolution of error covariance, which is the main element of the theory of the Kalman filter used in sequential data assimilation. The statistical approach could be critical in assessing the most useful data for assimilation purposes. Combining these elements in a large dimensional system will require encoding in efficient algorithms, which will bring in computational scientists.

The proposed program will be aimed at fostering these potential interactions and attracting researchers in these areas to the problems of data assimilation. These are self evidently critical problems of our time and we anticipate being able to interest a variety of experts. The methods of data assimilation promise applications in a wide range of areas which offer testing grounds for new ideas and techniques, such as:

- atmospheric science and numerical weather prediction,
- space weather,
- ocean prediction,
- coupled atmosphere-ocean systems,
- biology and tomography,
- shock waves in nuclear explosions,
- hydrology, and
- marine biology.

New and improved techniques in data assimilation could significantly enhance our capabilities of prediction. As such, the social impact could be great in the prediction of hurricane paths, disease development and water control during droughts, to mention just a few examples.

Some of the guiding problems that motivate the research in these areas are:

1. how to compute covariances,
2. how to manage large data-sets,
3. improving models, i.e. can models learn?
4. assimilation of non-prognostic variables,
5. optimal design of experiments based on the results of data assimilation,
6. assimilation of data corresponding to subsidiary variables in coupled models, such as atmospheric data in coupled atmosphere-ocean models, or Lagrangian data in ocean models, etc.

This is an ideal area for a SAMSI program as it is timely, of broad scientific interest and placed right at a point where statistics, applied mathematics and computations meet. We would design workshops and visitor programs to bring together scientists with interests in data assimilation from the viewpoint of specific applications. We would also target researchers in applied mathematics and statistics who have relevant expertise but who have not worked directly in this area.

Possible leaders of program: Kayo Ide (Atmospheric Sciences, UCLA), Christopher Jones (Mathematics, UNC-CH), Douglas Nychka (Statistics, NCAR), Susan Lozier (Earth and Ocean Sciences, Duke University) and Francisco Werner (Marine Sciences, UNC-CH).

IV. High Dimensional Inference and Random Matrices (tentative for Fall, 2005)

The recent growth of interest in ‘random matrix theory’ (RMT) in many areas of mathematics suggests possible developments in statistics, and the idea of SAMSI holding a semester program. As one example, RMT seems to offer tools that allow one to revisit classical multivariate statistics from a ‘large n, large p’ perspective, and perhaps derive approximations and insights that were not available due to the complexity of the distribution theory for fixed n,p or fixed p, large n.

If one takes real statistical contexts and abstracts mathematical/probabilistic questions, one is very likely to come up with issues that have not yet been addressed by the RMT community, and yet are accessible with current tools. For example, the distribution of extreme Wishart eigenvalues for non-identity covariance matrix is a rather natural statistics question that had not been looked at in RMT, and yet results are now beginning to emerge.

Large scale (large n, large p and often both) uses of PCA (aka KLT), CCA and related multivariate techniques occur in many application domains. For example,

- ‘Empirical orthogonal function’ analysis has been a staple of meteorology/climatology since the 1960’s.
- Versions of large scale PCA are involved in contemporary Information Retrieval algorithms, for example as used by search engines such as Google.
- Other examples occur in signal processing, astronomy, ...

Within statistics itself, RMT could contribute to a wide range of areas, including:

- Extreme Sample Eigenvalues in large n, p setting.
- Sample Eigenvectors (associated with extreme eigenvalues)
- Empirical distributions of eigenvalues
- Estimation of large covariance matrices
- Classification & clustering

Possible leaders of program: Iain Johnstone (Statistics, Stanford), Craig Tracey (Mathematics, UC Davis), Ken McLaughlin (Mathematics, UNC).

APPENDIX A – Workshop Participants

The key to **Status** entry is as follows: NRG - New Researcher or Graduate Student
FP - Faculty/Professional

Inverse Problems Methodology in Complex Stochastic Problems
Opening Workshop Participants
September 21-24, 2002

Name	Gender	Affiliation	Department	Status
Ackleh, Azmy	M	U of Louisiana	Mathematics	NRG
Adams, Brian	M	NC State U	Mathematics	NRG
Anderssen, Robert	M	CSIRO		FP
Arellano, Avelino	M	Duke U	Environment	NRG
Ball, Brian	M	NC State U	Mathematics	NRG
Banks, H.T.	M	NC State U & SAMSI	Mathematics	FP
Banks, John	M	U of Washington	Zoology	NRG
Bardsley, Johnathan	M	NC State U & SAMSI	Mathematics	NRG
Bayarri, M.J.	F	U of Valencia	Statistics	FP
Begashaw, Negash	M	Benedict College	Mathematics	NRG
Benedict, Brandy	F	NC State U		NRG
Berger, James	M	Duke U & SAMSI	Statistics	FP
Beun, Stacy	F	NC State U	Mathematics	NRG
Bortz, David	M	U of Michigan	Mathematics	NRG
Brazier, Richard	M	Penn State U	Geosciences	NRG
Camacho, Erika	F	Cornell U	Applied Mathematics	NRG

Cartwright, Natalie	F	U of Vermont	Electrical & Computer Engineering	NRG
Chen, Yuguo	M	Duke U	Statistics	NRG
Chiswell, Karen	F	NC State U	Statistics	NRG
Cintron-Arias, Ariel	M	Cornell U	Applied Mathematics	NRG
Clark, Leona Harris	F	EPA		NRG
Cole, Cammey	F	Meredith College	Mathematics & Computer Science	NRG
Coles, Crista	F	Elon U	Mathematics	NRG
Crank, Keith	M	NSF		FP
Davidian, Marie	F	NC State U	Statistics	FP
Davis, John	M	Baylor U	Mathematics	NRG
Davis, J. Wade	M	U of Missouri	Statistics	NRG
Dediu, Sava	M	Rensselaer Polytechnic U	Mathematics	NRG
DeFacio, Brian	M	U of Missouri	Physics	FP
DeGruttola, Victor	M	Harvard Pub Health	Biostatistics	FP
Draper, David	M	U of CA, Santa Cruz	Applied Math & Statistics	FP
Easterling, Michael	M	Analytical Sciences Inc.		FP
Errnstberger, Jon	M	Murray State U		NRG
Faulkner, Gary	M	NC State U	Mathematics	FP
Flesia, Ana	F	Stanford U	Statistics	NRG
Fouque, Jean-Pierre	M	NC State U	Mathematics	FP
Genton, Marc	M	NC State U	Statistics	FP

Genton, Yanyuan Ma	F	NC State U		NRG
Ghosh, Malay	M	U of Florida	Statistics	FP
Ghosh, Sujit	M	NC State U	Statistics	NRG
Gibson, Nathan	M	NC State U	Mathematics	NRG
Gomatam, Shanti	F	NISS & U of South Florida	Mathematics	NRG
Grunbaum, Alberto	M	U of CA, Berkeley	Mathematics	FP
Haider, Mansoor	M	NC State U	Mathematics	FP
Holte, Sarah	F	Fred Hutchinson Cancer Res Cen	Public Health Sciences	FP
Hood, Jeffrey	M	NC State U	Mathematics	NRG
Horn, Mary Ann	F	Vanderbilt U	Mathematics	NRG
Huang, Yangxin	M	Harvard U	Frontier Science & Technology	NRG
Hurtado, Gerardo	M	SAS Institute		FP
Ito, Kazufumi	M	NC State U	Mathematics	FP
Javier, Walfredo	M	Southern U	Mathematics	NRG
Jeglova, Polina	F	Rensselaer Polytechnic U	Mathematics	NRG
Ji, Lin	M	Rensselaer Polytechnic U	Mathematics	NRG
Ji, Ming	M	San Diego State U	Epidemiology & Biostatistics	NRG
Karr, Alan	M	NISS & SAMSI		FP
Kees, Christopher	M	NC State U	Mathematics	NRG
Kelley, C. Tim	M	NC State U	Mathematics	FP
Kepler, Grace	F	NC State U	Mathematics	NRG

Khan, Taufiquar	M	Clemson U	Mathematics	FP
King, Aaron	M	U of Tennesse	Electrical & Computer Engineering	NRG
King, Belinda	F	AFRL/AFOSR		FP
Kojima, Fumio	M	Kobe U	Computer Systems & Engineering	FP
Koken, Petra	F	NIEHS		NRG
Kunisch, Karl	M	U of Graz	Mathematics	FP
Leon, Selene	F	NC State U	Statistics	NRG
LePage, Raoul	M	Michigan State U	Statistics	FP
Lewis, Brian	M	NC State U	Mathematics	NRG
Li, Haihong	M	Harvard U	Frontier Science & Technology	NRG
Liang, Feng	F	Duke U	Statistics	NRG
Longden-Chapman, Elaine	F	Institute of Physics Publications		FP
Longini, Ira	M	Emory School of Public Health	Biostatistics	FP
Losee, Robert	M	U of North Carolina	Information & Library Sciences	FP
Lubkin, Sharon	F	NC State U	Mathematics	NRG
Luke, Nicholas	M	NC State U	Mathematics	NRG
Madsen, Kristen	F	NC State U	Statistics	NRG
Mallick, Bani	M	Texas A&M U	Statistics	FP
Marron, J.S.	M	U of North Carolina	Statistics	FP
McBride, Sandra	F	Duke U	Statistics	NRG
McCarthy, Maeve	F	Murray State U	Mathematics & Statistics	FP

McLaughlin, Joyce	F	Rensselaer Polytechnic		FP
Medhin, Negash	M	NC State U	Mathematics	FP
Mei, Yajun	M	Cal Tech	Mathematics	NRG
Moodie, Zoe	F	Fred Hutchinson Cancer Research Center	Public Health Sciences	NRG
Mueller, Peter	M	M.D. Anderson Cancer Center	Biostatistics	FP
Musante, C.J.	F	Entelos Inc.		NRG
Nichols, Scott	M	NOAA		NRG
Ogle, Kiona	F	Duke U		NRG
Oughstun, Kurt	M	U of Vermont	Electrical & Computer Engineering	FP
Paciorek, Christopher	M	Carnegie Mellon U	Statistics	NRG
Pantula, Sastry	M	NC State U	Statistics	FP
Park, Jeong-gun	M	Harvard U	Frontier Science & Technology	NRG
Picka, Jeffrey	M	U of Maryland	Mathematics	NRG
Pinter, Gabriella	F	U of Wisconsin	Engineering & Mathematical Sciences	NRG
Potter, Laura	F	GlaxoSmithKline		NRG
Raye, Julie K.	F	Virginia Commonwealth	Mathematics	NRG
Rosner, Gary	M	M.D. Anderson Cancer Center	Biostatistics	FP
Rundell, William	M	NSF		FP
Sacks, Jerome	M	Duke U	Statistics	FP
Santosa, Fadil	M	U of Minnesota	Mathematics	FP
Schafer, Chad	M	U of CA, Berkeley	Statistics	NRG

Sheiner, Lewis	M	U of CA, San Francisco	Laboratory Medicine	FP
Sinha, Debajyoti	M	Medical U of South Carolina	Biometry & Epidemiology	FP
Smith, Cassandra	F	NC State U	Mathematics	NRG
Smith, Marjo	F	Analytical Sciences Inc.		FP
Smith, Ralph	M	NC State U	Mathematics	FP
Taylor, Michael	M	NC State U		NRG
ter horst, Enrique	M	Duke U	Statistics	NRG
Tovbis, Alexander	M	U of Central Florida	Mathematics	FP
Toyoshiba, Hiroyoshi	M	NIEHS		FP
Tsiatis, Anastasios	M	NC State U	Statistics	FP
Vogel, Curtis	M	Montana State U	Mathematics	FP
Walsh, Daniel	M	SAMSI		NRG
Weems, Kimberly	F	NC State U	Statistics	NRG
Whitaker, Shree	F	NIEHS		NRG
Wolpert, Robert	M	Duke U	Statistics	FP
Wu, Hulin	M	Harvard U	Biostatistics	FP
Yamanaka, Takeharu	M	NIEHS		NRG
Yokley, Karen	F	NC State U	Mathematics	NRG
Yoon, Jeong-Rock	M	Rensselaer Polytechnic U	Mathematics	NRG
Zager, Michael	M	NC State U	Mathematics	NRG
Zhang, Daowen	M	NC State U	Statistics	NRG

Zietsman, Lizette	F	Virginia Tech U	Applied Mathematics	FP

Stochastic Computation
Opening Workshop Participants
September 25-October 4, 2002

Name	**Gender**	**Affiliation**	**Department**	**Status**
Andrieu, Christophe	M	U of Bristol	Mathematics	NRG
Banks, H. T.	M	NC State U & SAMSI	Mathematics	FP
Barber, Jarrett	M	Duke U	Statistics	NRG
Bardsley, Johnathan	M	NC State U & SAMSI	Mathematics	NRG
Bayarri, Susie	F	U of Valencia	Statistics	FP
Berger, Jim	M	Duke U & SAMSI	Statistics	FP
Blanchard, Suzette	F	Harvard U		FP
Blattenberger, Gail	F	U of Utah	Economics	FP
Brown, Beth	F	Duke U	Statistics	NRG
Browning, Sharon	F	GlaxoSmithKline		FP
Calabrese, Peter	M	U of Southern CA	Computational & Molecular Biology	NRG
Calder, Catherine	F	Duke U	Statistics	NRG
Carlin, Brad	M	U of Minnesota	Biostatistics	FP
Carter, Chris	M	Duke U & SAMSI	Statistics	FP
Carvalho, Carlos	M	Duke U	Statistics	NRG
Chamu, Francisco	M	U of North Carolina	Statistics	NRG
Chen, Li	F	North Carolina State U	Statistics	NRG

Chen, Tom	M	Colorado State U	Engineering	NRG
Chen, Yuguo	M	Duke U	Statistics	NRG
Chen, Zhen	M	NIEHS		NRG
Chiswell, Karen	F	North Carolina State U	Statistics	NRG
Choi, Seo-Eun	F	Florida State U	Statistics	NRG
Cloete, Nicoleen	F	U of Auckland	Mathematics	NRG
Clyde, Merlise	F	Duke U	Statistics	FP
Crank, Keith	M	National Science Found	Mathematical Sciences	FP
Cripps, Ed	M	U of New South Wales	Statistics	NRG
Cuellar, Milena	F	London School of Economics	Statistics	NRG
Davidian, Marie	F	North Carolina State U	Statistics	FP
Davis, Jerry	M	North Carolina State U		FP
Dinwoodie, Ian	M	Duke U & SAMSI	Statistics	FP
Dobra, Adrian	M	Duke U & SAMSI	Statistics	NRG
Doucet, Arnaud	M	Cambridge U	Engineering	FP
Draper, David	M	U of CA, Santa Cruz	App Mathematical Sci	FP
Dunson, David	M	NIEHS		FP
Engel, Dave	M	Battelle-Pacific National Laboratory	Statistics & Quantitative Sciences	FP
Fill, Jim	M	Johns Hopkins U	Mathematical Sciences	FP
Fokoue, Ernest	M	Ohio State U	Statistics	NRG
Fouque, Jean Pierre	M	North Carolina State U	Mathematics	FP

Fuentes, Montserrat	F	North Carolina State U	Statistics	NRG
Gamundi, Emily	F	Tulane U	Mathematics	NRG
Gelfand, Alan	M	Duke U	Statistics	FP
Genton, Yanyuan Ma	F	NC State U & SAMSI	Mathematics	NRG
Ghosh, Sujit	M	North Carolina State U	Statistics	FP
Godsill, Simon	M	Cambridge U	Engineering	FP
Golinelli, Daniela	F	U Southern CA & RAND	Computational & Molecular Biology	NRG
Gomatam, Shanti	F	NISS & U of South Florida		NRG
Green, Peter	M	U of Bristol	Statistics	FP
Han, Chuan Hsiang	M	North Carolina State U	Mathematics	NRG
Hans, Chris	M	Duke U & SAMSI	Statistics	NRG
Haran, Murali	M	U of Minnesota	Statistics	NRG
Clark Harris, Leona	F	U.S. EPA		FP
Hartemink, Alex	M	Duke U	Computer Science	NRG
Hobert, Jim	M	U of Florida	Statistics	FP
Holland, Dave	M	U.S. EPA		FP
Holloman, Chris	M	Duke U	Statistics	NRG
Holte, Sarah	F	Fred Hutchinson Cancer Res Cen	Public Health Sciences	FP
Hoogerheide, Lennart	M	Erasmus U	Economics	NRG
House, Leanna	F	Duke U	Statistics	FP
Huber, Mark	M	Duke U	Mathematics & Statistics	NRG

Ibrahim, Joe	M	U of North Carolina	Biostatistics	FP
Iversen, Ed	M	Duke U	Statistics	FP
Ji, Chuanshu	M	U of North Carolina	Statistics	FP
Johnson, Andrea	F	North Carolina State U	Statistical Genetics	NRG
Jones, Beatrix	F	SAMSI		NRG
Karr, Alan	M	NISS		FP
Khan, Taufiquar	M	Clemson U	Mathematics	NRG
Kohnen, Christine	F	Duke U	Statistics	NRG
Kunas, Myra	F	Tulane U	Mathematics	NRG
Lee, Beom	M	U of North Carolina	Statistics	NRG
Leonard, Michael	M	SAS Institute		FP
Li, Huina	M	U of North Carolina	Environmental Sciences & Engineering	NRG
Li, Li	F	North Carolina State U		NRG
Liang, Feng	F	Duke U	Statistics	NRG
Liao, Ming	M	Duke U	Statistics	NRG
Lin, Hsin-I	F	Tulane U	Mathematics	NRG
Liu, Jane	F	UBS Warburg		NRG
Liu, Jun	M	Harvard U	Statistics	FP
Liu, Kejun	M	North Carolina State U		NRG
Liu, Shufeng	F	U of CA, Santa Cruz	App Mathematical Sci	NRG
Lopes, Hedibert	M	Fed U of Rio de Janeiro	Statistics	NRG

Losee, Robert	M	U of North Carolina	Information & Library Sciences	FP
Madsen, Kristen	F	North Carolina State U	Mathematics	NRG
Mallick, Bani	M	Texas A&M U	Statistics	FP
Marchev, Dobrin	M	U of Florida	Statistics	NRG
McBride, Sandra	F	Duke U	Statistics	NRG
McCulloch, Rob	M	U of Chicago	Grad School of Business	FP
McKeague, Ian	M	Florida State U	Statistics	FP
Mitha, Faheem	M	U of North Carolina	Statistics	NRG
Molina, German	M	Duke U	Statistics	NRG
Monahan, John	M	North Carolina State U	Statistics	FP
Murdoch, Duncan	M	U of Western Ontario	Statistics	FP
Nicholas, Mike	M	Duke U	Mathematics	NRG
Nichols, Scott	M	NOAA		FP
O'Brien, Sean	M	NIEHS		NRG
Ogle, Kiona	F	Duke U	Biology	NRG
Pan, Doris	F	U of North Carolina	Statistics	NRG
Paulo, Rui	M	NISS & SAMSI		NRG
Pawel, David	M	U.S. EPA		FP
Pittman, Jennifer	F	Duke U	Statistics	NRG
Raggi, Davide	M	Duke U		NRG
Ramakrishnan, Dinesh	M	Duke U	Electrical Engineering	NRG

Ray, Surajit	M	Penn State U	Statistics	NRG
Rigat, Fabio	M	Duke U	Statistics	NRG
Rodrigues-Yam, Gabriel	M	Colorado State U	Statistics	NRG
Rosenberger, Jim	M	Penn State U	Statistics	FP
Sanil, Ashish	M	NISS		NRG
Scott, Kevin	M	NC State U & SAS		NRG
Setzer, R. Woodrow	M	U.S. EPA		FP
Talih, Makram	M	Yale U	Statistics	NRG
Tavare, Simon	M	U of Southern CA	Computational & Molecular Biology	FP
Thomas, Andrew	M	Imperial College London	Primary Care & Population Health Sci	FP
Thomas, Len	M	U of St. Andrews	Mathematics & Statistics	FP
Tvete, Ingunn Fride	F	U of Oslo	Mathematics	NRG
Valeva, Anna	F	Duke U	Statistics	NRG
van Dijk, Herman	M	Erasmus U	Economics	FP
van Oest, Rutger	M	Erasmus U	Economics	NRG
Varadarajan, Vijay	M	Duke U	Electrical Engineering	NRG
Vasudevan, Sathyanarayanan	M	Duke U	Electrical Engineering	NRG
Verloo, Didier	M	Institute of Tropical Medicine		FP
Vogel, Curtis	M	Montana State U	Mathematics	FP
Wang, Yazhen	M	U of Connecticut	Statistics	NRG
Watanabe, Karen	F	Tulane U	Mathematics	FP

Wei, Rong	M	National Center for Health Statistics		FP
West, Mike	M	Duke U	Statistics	FP
Wolfe, Patrick	M	Cambridge U	Engineering	NRG
Wolpert, Robert	M	Duke U	Statistics	FP
Yokley, Karen	F	North Carolina State U	Mathematics	NRG
Zhang, Daowen	M	North Carolina State U	Statistics	FP
Zhang, Hao (Helen)	F	North Carolina State U	Statistics	NRG

Large-Scale Computer Models for Environmental Systems
Workshop on Multi-Scale Modeling
Workshop Participants

February 2-7, 2003

Name	Gender	Affiliation	Department	Status
Allen, Michael	M	Tennessee Tech U	Mathematics	FP
Bailey, Barbara	F	U of Illinois	Statistics	FP
Banard, Dawn	F	Duke U	Statistics	NRG
Barber, Jarrett	M	Duke U	Statistics	NRG
Beale, J. Thomas	M	Duke U	Mathematics	FP
Behringer, Bob	M	Duke U	Physics	FP
Bennethum, Lynn	F	U of Colorado	Mathematics	FP
Berger, James	M	SAMSI		FP
Berliner, Mark	M	Ohio State U	Statistics	FP
Bourlioux, Anne	F	U of Montreal	Mathematics & Statistics	FP

Bronski, Jared	M	U of Illinois	Mathematics	FP
Celia, Michael	M	Princeton U	Civil & Environmental Engineering	FP
Chamu, Francisco	M	U of North Carolina	Statistics	NRG
Chen, Li	F	North Carolina State U	Statistics	NRG
Christensen, Britt	F	Technical U of Denmark	Environment & Resources	NRG
Cooter, Ellen	F	U.S. EPA	NOAA	FP
Culligan, Katherine	F	U of Notre Dame	Civil Eng & Geological Sciences	NRG
Davis, Jerry	M	North Carolina State U		FP
Fadimba, Koffi	M	U of South Carolina - Aiken	Mathematics	FP
Farthing, Matthew	M	U of North Carolina	Environmental Sciences & Engineering	NRG
Ferreira, Marco	M	Federal U of Rio de Janeiro	Mathematics	NRG
Finkelstein, Peter	M	U.S. EPA		FP
Fleishman, Lauren	F	U of North Carolina	Environmental Sciences & Engineering	NRG
Frierson, Dargan	M	Princeton U	Applied & Computational Math	NRG
Fuentes, Montserrat	F	North Carolina State U	Statistics	FP
Gelfand, Alan	M	Duke U	Statistics	FP
Genton, Marc	M	North Carolina State U	Statistics	FP
Gerber, Edwin	M	Princeton U	Applied & Computational Math	NRG
Grady, Amy	F	NISS		NRG
Gray, William	M	U of Notre Dame	Civil Eng & Geological Sciences	FP
Hack, James	M	NCAR		FP

Hall, Timothy	M	NASA		FP
Han, Chuan Hsiang	M	North Carolina State U	Mathematics	NRG
Heo, Tae-Young	M	North Carolina State U	Statistics	NRG
Higdon, Dave	M	Los Alamos National Laboratory	Statistical Sciences Group	FP
Johnson, Deona	F	U of North Carolina	Environmental Sciences & Engineering	NRG
Kalnay, Eugenia	F	U of Maryland	Meterology	FP
Kanney, Joseph	M	U of North Carolina	Environmental Sciences & Engineering	NRG
Karr, Alan	M	NISS		FP
Katsoulakis, Markos	M	U of Massachusetts	Mathematics	FP
Kelley, C. Tim	M	North Carolina State U	Mathematics	FP
Kerr, Robert	M	U of Warwick	Mathematics	FP
Konate, Dialla	M	Virginia Tech	Mathematics	FP
Kramer, Peter	M	Rensselaer Polytechnic Institute	Mathematical Sciences	FP
Krim, Hamid	M	North Carolina State U	Electrical & Computer Engineering	FP
Kuensch, Hans	M	ETH Zentrum	Statistics	FP
Lee, MiHyun	F	North Carolina State U	Statistics	NRG
Li, Huina	M	U of North Carolina	Environmental Sciences & Engineering	NRG
Ma, Liyun	F	North Carolina State U	Statistics	NRG
Madsen, Kristen	F	North Carolina State U	Mathematics	NRG
Majda, Andrew	M	Courant Institute		FP
McBride, Sandra	F	Duke U	Statistics	NRG

McLaughlin, Richard	M	U of North Carolina	Applied Mathematics	FP
Miller, Casey	M	U of North Carolina	Environmental Sciences & Engineering	FP
Miller, Robert	M	Oregon State U	Oceanic & Atmospheric Sciences	FP
Mukherjee, Bhramar	F	U of Florida	Statistics	NRG
Nail, Amy	F	North Carolina State U	Statistics	NRG
Natvig, Bent	M	U of Oslo	Mathematics	FP
Nychka, Doug	M	NCAR		FP
Pan, Doris	F	U of North Carolina	Statistics	NRG
Park, Juhyun	F	U of North Carolina	Statistics	NRG
Parlange, Marc	M	Johns Hopkins U	Geography & Env Engineering	FP
Pedit, Joseph	M	U of North Carolina	Environmental Sciences & Engineering	NRG
Pinto, Joseph	M	U.S. EPA		FP
Reese, Jill	F	North Carolina State U	Mathematics	NRG
Restrepo, Juan	M	U of Arizona	Mathematics	FP
Rupert, Carl	M	U of North Carolina	Environmental Sciences & Engineering	NRG
Sampson, Paul	M	U of Washington	Statistics	FP
Schneider, Tapio	M	California Inst of Technology	Geological & Planetary Sciences	FP
Smith, Richard	M	U of North Carolina	Statistics	FP
Speed, Michael	M	Texas A&M U	Statistics	FP
Stevens, Bjorn	M	U of CA - Los Angeles	Atmospheric Sciences	FP
Stuart, Andrew	M	Warwick U	Mathematics	FP

Tebaldi, Claudia	F	NCAR		FP
Tribbia, Joe	M	NCAR		FP
Tvete, Ingunn Fride	F	U of Oslo	Mathematics	NRG
Vance, Eric	M	Duke U	Statistics	NRG
Vanden Eijnden, Eric	M	Courant Institute		FP
Welty, Claire	F	Drexel U	Civil & Architectural Engineering	FP
Wikle, Christopher	M	U of Missouri	Statistics	FP
Yortsos, Yannis	M	U of South Carolina	Engineering	FP
Zheng, Xiaoyu	F	U of North Carolina	Mathematics	NRG
Zhu, Zhengyuan	M	U of North Carolina	Statistics	FP
Zidek, Jim	M	U of British Columbia	Statistics	FP

Large-Scale Computer Models for Environmental Systems
Workshop on Simulation & Optimization
Workshop Participants

April 28-30, 2003

Name	**Gender**	**Affiliation**	**Department**	**Status**
Adams, Brian	M	North Carolina State U	Mathematics	NRG
Bartelt-Hunt, Shannon	F	U of Virginia	Civil Engineering	NRG
Begashaw, Negash	M	Benedict College	Mathematics	FP
Carle, Steven	M	Lawrence Livermore National Laboratory	Geoscience & Environmental Tech	FP
Chiswell, Karen	F	North Carolina State U	Statistics	NRG
Crank, Keith	M	National Science Foundation	Mathematical Sciences	FP

Culligan, Katherine	F	U of Notre Dame	Civil Engineering	NRG
Darwin, Robert	M	North Carolina State U	Mathematics	NRG
Davis, Stephen	M	U.S. Army Research Office	Computational Mathematics Program	FP
Dawson, Clint	M	U of Texas at Austin	Aerospace Eng & Engineering Mechanics	FP
Dennis, John	M	Rice U	Computational & Applied Mathematics	FP
Dentcheva, Darinka	F	Stevens Inst of Technology	Mathematical Sciences	FP
Farthing, Matthew	M	U of North Carolina	Environmental Sciences	NRG
Finkel, Daniel	M	North Carolina State U	Operations Research	NRG
Gibson, Nathan	M	North Carolina State U	Mathematics	NRG
Gray, Genetha	F	Sandia National Laboratories	Computational Sci & Mathematics Research	NRG
Gray, William	M	U of Notre Dame	Civil Engineering & Geological Sciences	FP
Harmon, Russell	M	U.S. Army Research Office	Mechanical & Env Sciences Division	FP
Hatch, Andrew	M	North Carolina State U	Applied Mathematics	NRG
Howington, Stacy	M	US Engineer Research & Development Center	Waterways Experiment Station	FP
Jenkins, Lea	F	Clemson U	Mathematical Sciences	FP
Kavanagh, Kathleen	F	North Carolina State U	Mathematics	NRG
Kees, Chris	M	North Carolina State U	Mathematics	NRG
Kelley, C.T.	M	North Carolina State U	Mathematics	FP
Lasater, Matthew	M	North Carolina State U	Mathematics	NRG
Levy, Rachel	F	North Carolina State U	Applied Mathematics	NRG
Luke, Nicholas	M	North Carolina State U	Applied Mathematics	NRG

Mayer, Alex	M	Michigan Tech U	Geology	FP
McLaughlin, Dennis	M	Massachusetts Inst of Technology	Civil & Environmental Engineering	FP
Medhin, Negash	M	North Carolina State U	Mathematics	FP
Miller, Casey	M	U of North Carolina	Environmental Sciences & Engineering	FP
Neuman, Shlomo	M	U of Arizona	Hydrology & Water Resources	FP
Pan, Doris	F	U of North Carolina	Environmental Sciences & Engineering	NRG
Petzold, Linda	F	U of CA, Santa Barbara	Mechanical & Env Engineering	FP
Reese, Jill	F	North Carolina State U	Applied Mathematics	NRG
Russell, Thomas	M	U of Colorado at Denver	Mathematics	FP
Serre, Marc	M	U of North Carolina	Environmental Sciences & Engineering	NRG
Shoemaker, Christine	F	Cornell U	Civil & Environmental Engineering	FP
Silantyev, Valentin	M	Northeastern U	Mathematics	NRG
Smith, Richard	M	U of North Carolina	Statistics	FP
Wheeler, Mary	F	U of Texas at Austin	Computational & Applied Mathematics	FP
Williams, Pamela	F	Sandia National Laboratories	Computational Sci & Mathematics Research	FP
Woodward, Carol	F	Lawrence Livermore National Laboratory	Center for Applied Scientific Computing	FP
Yeh, William	M	U of CA, Los Angeles	Civil & Environmental Engineering	FP
Yokley, Karen	F	North Carolina State U	Mathematics	NRG
Yu, Jie	F	SAMSI		NRG
Zheng, Xiaoyu	F	U of North Carolina	Mathematics	NRG
Zhu, Zhengyuan	M	U of North Carolina	Statistics	NRG

Zidek, Jim	M	U of British Columbia	Statistics	FP

Education & Outreach
One-Day Workshop Participants
November 9, 2002

Name	**Gender**	**Affiliation**	**Major**	**Status**
Adams, Shauntay	F	Spelman College		S
Almond, Natalie	F	UNC-Wilmington	Applied Mathematics	S
Bakewell, Edward	M	UNC-Wilmington	Mathematics	S
Beauford, Angela	F	Spelman College		F
Billing, Emily	F	UNC-Wilmington	Mathematics & Physics	S
Blavo, Selasi	M	Clark Atlanta University	Chemical Engineering	S
Bolarinwa-Keita, Lola	F	Spelman College	Mathematics	S
Chung, Gary	M	Clark Atlanta University		F
Cork, Marlon	M	Benedict College	Computer Science & Mathematics	S
Davis, Angela	F	Benedict College	Mathematics	S
Ellison, Charles	M	NC State University	Computer Science & Mathematics	S
Fabiano, Richard	M	UNC-Greensboro		F
Ferrier, Edward	M	UNC-Greensboro	Philosophy of Science	S
Futrell, Robert	M	NC State University	Computer Science & Applied Mathematics	S
Gillens, Kayla	F	Clark Atlanta University	Electrical Engineering	S
Joshua, Jami	F	UNC-Greensboro	Pure Mathematics	S
Kelly, Ishika	F	Benedict College	Computer Science	S

Lane, Curtis	M	Benedict College	Physics	S
Lawallen, Nathan	M	NC State University	Applied Mathematics & Statistics	S
Lugo, Gabriel	M	UNC-Wilmington		F
Medhim, Negash	M	NC State University		F
Newkirk, Paula	F	Clark Atlanta University	Mathematics	S
Prince, Shauntee	F	Benedict College	Mathematics	S
Robinson, Kito	M	Clark Atlanta University		S
Saunders, Tanisha	F	Spelman College	Mathematics	S
Sharpe, Marcietta	F	NC State University	Applied Mathematics	S
Thatcher, Aaron	M	UNC-Wilmington	Applied Mathematics	S
Tucker, Dorielle	F	Clark Atlanta University		S
Windham, Kevin	M	NC State University	Applied Mathematics	S
Xu, Yao-Huan	M	Benedict College		F

Education & Outreach
One-Day Workshop Participants
February 1, 2003

Name	**Gender**	**Affiliation**	**Major**	**Status**
Adeniyi, Ifedapo	M	Clark Atlanta University	Engineering	S
Bayless, Jonathan	M	Clemson University	Mathematical Sciences	S
Begashaw, Negash	M	Benedict College		F
Chung, Gary	M	Clark Atlanta University		F
Coulson-Clark, Margery	F	Elizabeth City State University		F

Cox, Michelle	F	Elizabeth City State University	Mathematics & Applied Mathematics	S
Davis, Jimena	F	Clemson University	Mathematical Sciences	S
Gant, Raymond	M	Benedict College	Pure Mathematics	S
Griffin, Javon	M	Elizabeth City State University	Mathematics	S
Harrell, Shonte	F	Spelman College	Mathematics	S
Holness, Jabari	M	Benedict College	Computer Science	S
Houston, Johnny	M	Elizabeth City State University		F
Iyangar, Balaji	M	Benedict College		F
Khan, Nashini	F	Benedict College	Mathematics	S
Khan, Taufiquar	M	Clemson University		F
Lee, Eric	M	Clark Atlanta University		S
Lugo, Gabriel	M	UNC-Wilmington		F
Mathias, Jimela	F	Spelman College	Mathematics	S
Maxwell, Sherian	F	Benedict College	Computer Science	S
Mynnet, Julia	F	UNC-Wilmington	Mathematics	S
Parker, Sonia	F	UNC-Wilmington	Mathematics	S
Shah, Nagambal	F	Spelman College		F
Simmons, Susan	F	UNC-Wilmington		F
Stanley, Erica	F	Clark Atlanta University	Computer Science	S
Swannack, Charles	M	Clemson University	Computer Engineering	S
Townsend, Howard	M	Benedict College	Computer Science	S

Uter, Simone	F	Elizabeth City State University	Computer Science	S
Walters, Daron	M	Clark Atlanta University	Computer Science	S
Wu, Meng	M	UNC-Wilmington	Mathematics	S
Ziemiecki, Ryan	M	UNC-Wilmington	Statistics	S

APPENDIX B – Workshop Programs and Abstracts

I. Inverse Problems Methodology in Complex Stochastic Problems

A. Opening Workshop Program
September 21-24, 2002

- **Saturday, September 21 -- North Carolina Biotechnology Center**

12:45-1:30: Registration

1:30-3:30: Tutorial on Differential Equation Modeling and Inverse Problems

H. T. Banks

This tutorial will focus on modeling and inverse problem concepts for complex dynamical systems. Topics include ordinary and partial differential equation modeling; time, space, and state dependent coefficient estimation; nonlinear mechanisms; identifiability, ill-posedness, stability, and regularization in inverse problems. Computational methods to be discussed are finite differences, finite elements and modal approximations; reduced order modeling including Proper Orthogonal Decomposition (POD) / Principal Component Analysis (PCA) based techniques. A review of related available software will be given.

3:30-4:00: Coffee Break

4:00-6:00: Tutorial on Statistical Modeling and Inference
M. Davidian

This tutorial will provide an overview of considerations underlying statistical approaches to taking appropriate account of uncertainty in fitting nonlinear mathematical models to data and to making formal statistical inference. Sources of variation in different data structure will be discussed, including situations where repeated observations on a system are available for a single sample or individual and on multiple such samples or individuals. Consequences of failure to acknowledge variation appropriately will be elucidated. Incorporation of deterministic mathematical models within appropriate statistical frameworks that explicitly represent relevant sources of variation, including nonlinear fixed and random effects models, will be outlined. Associated methods for parameter estimation and inference both by frequentist and Bayesian approaches will be discussed. Available software will be reviewed and demonstrated, and its limitations for incorporation of complex dynamical systems highlighted.

6:00-8:00: Welcome Reception at NISS/SAMSI Building

- **Sunday, September 22 -- Radisson Governor's Inn**

8:15 – 9:00 Registration

Morning: HIV Dynamics
9:00 - 9:45

"Mathematical Modeling for the HIV Response to Treatment: Some Past Achievements and Future Direction"
Victor De Gruttola, Harvard School of Public Health

9:45 - 10:30
"Mathematical Models of HIV Infection: Past, Present, and Future"
Sarah Holte, Fred Hutchinson Cancer Research Center

10:30-11:00 Coffee Break

11:00 -11:45
"Using Mathematical Models to Evaluate HIV Vaccination Strategies"
Ira Longini, Emory School of Public Health

12:00-1:30 Box Lunch at Radisson

Afternoon: Electromagnetics
1:30 - 2:15
"A Brief Introduction to Electromagnetism"
Brian DeFacio, University of Missouri

2:15 - 3:00
"Dispersive Pulse Dynamics and Precursor Fields in Dielectric and Conductive Materials"
Kurt Oughstun, University of Vermont

3:00-3:30 Coffee Break

3:30 - 4:15
"New Directions/Methodologies for E&M Modeling and Interrogation"
Richard Albanese, AFRL, Brooks AFB

- **Monday, September 23**

8:30 – 9:00 Registration

Morning: Polymers
9:00 - 9:45
"Recovering Molecular Information from the Mixing of Wheat-Flour Dough"
Robert Anderssen, CSIRO

9:45 - 10:30
"A Nonparametric Bayesian Approach to Inverse Problems"
Robert Wolpert, Duke University

10:30-11:00 Coffee Break

Mid-Morning: Tomography
11:00 - 11:45
"Network Tomography: A Nonlinear Inverse Problem for a Stochastic Process"
Alberto Grunbaum, University of California Berkeley

<u>**Late Morning**: Physiologically Based Pharmacokinetics</u>
11:45 - 12:30
"Incorporating Variability and Uncertainty into Human Health Risk Assessment"
Laura Potter, EPA

12:30-1:30 Box Lunch at Radisson

<u>**Afternoon**: Parameter Estimation and Inverse Problems Statistics Perspective</u>
1:30 - 2:20
"Statistical Methodology for Inverse Problems"
David Draper, University of California Santa Cruz

2:30 - 3:20
"Mixture Modeling"
Peter Mueller, M. D. Anderson Medical Center

3:30 - 4:00 Coffee Break

4:00 - 4:50
"Statistical Modeling and Inverse Problems"
Gary Rosner, M. D. Anderson Medical Center

- **Tuesday, September 24**

<u>**Morning**: Parameter Estimation and Inverse Problems Applied Mathematics Perspective</u>
9:00 - 9:50
"An Applied Mathematician's Prospective on Regularization Methods"
Curt Vogel. Montana State University

10:00 - 10:50
"Numerical Techniques for Least Squares Problems with Non-Quadratic Penalty Terms"
Karl Kunisch, Technische Universitat Graz

10:50-11:20 Coffee Break

11:20 - 12:10
"Regularization of Ill posed Inverse Problems by Priori Information"
Fadil Santosa, University of Minnesota

12:10-1:30 Box Lunch at Radisson

<u>**Afternoon**</u>
1:30 - 3:00
Group Discussions: Challenges and Directions

3:15 - 4:30
Summary Group Reports

B. Opening Workshop Speaker Abstracts

Richard Albanese
AFRL, Brooks AFB

TITLE: "New Directions/Methodologies for E&M Modeling and Interrogation"

ABSTRACT:
The primary research and development interest I will address in this presentation is the imaging of objects obscured by covering layers or "clutter". Examples are detection of tumors within the human body, and detection, from the air, of military weapons hidden under trees.

Electromagnetic pulses called "precursors" have the ability to deeply penetrate materials. The use of precursors for imaging is being explored in a formal structured manner. A probabilistic interpretation of precursor propagation will be described and further research along this or analogous lines will be suggested. The use of deterministic precursors to interrogate material properties will be described as will difficult statistical questions arising from their use.

The reality of electromagnetic loss in materials requires one to consider Maxwell's equations coupled to the heat conduction equation. Challenges associated with this system of equations are outlined. In a similar manner, the fact that materials are electrostrictive (respond to an electromagnetic field with movement) leads to a coupling of the Maxwell and acoustic equations. Again, challenges will be outlined. Imaging using a combination of electromagnetic and acoustic wave fronts will be discussed.

Finally, an inverse problem involving electromagnetic fields, but arising in neurological research, will be addressed. Microelectrode monitoring of nervous system activity is an important current research effort with a mathematical core. The statistical aspects of neural network structure determination using a small set of measuring points will be illustrated.

Robert Anderssen
CSIRO

TITLE: "Recovering Molecular Information from the Mixing of Wheat-Flour Dough"

ABSTRACT:
The study of the mixing of wheat-flour dough on recording mixers, such as the Farinograph and Mixograph, is central to issues related to the milling of hard and soft wheats, the mixing of the resulting flours with water and other ingredients and the baking of the resulting doughs, as well as the design of scientific and industrial mixers. It plays an even greater and more fundamental role in the breeding of new wheat varieties. It is this latter aspect that will be the focus of the talk, though related aspects of bread, cake, pasta and pastry making will not be ignored. It is guaranteeing their quality that is the control back to which plant breeding must respond.

The overall goal of the talk is an examination of some of the recent statistics and mathematics involved with extracting molecular information from various types of cereal chemistry experiments.

Brian DeFacio

University of Missouri

TITLE: “Mathematical Modeling of Toxicology Endpoints”

ABSTRACT:
The Maxwell equations are one of the most beautiful structures in Physics, in large part due to their remarkable agreement with experiments - at present this agreement is to within an accuracy of 11 to 12 decades and extends over 31 decades in frequency.

This talk will start with Maxwell's equations for the electromagnetic fields (E;B) as 3-D (space) vectors possibly depending on time living in a vacuum, their interpretation, the boundary conditions they satisfy and their "wave behavior"(which includes the conditions for solutions to exist)..The Minkowski E&M fields (E,D;B,H) and the sources (rho,J)in 3-D space will be cast into an equivalent Maxwell system. SI (International Scientific) units will be used throughout(with comment).Next, the constitutive relations, the Lorentz force law and the combined first and second laws of equilibrium thermodynamics will be presented.

Then the (local) balance equation for the E&M sources (rho,J) and the global electric charge Q conservation will be shown. The Poynting theorem for the flow of E&M power flow will be discussed. The class of LIHS (Linear, Isotropic, Homogeneous, Stationary) media, the polarization of E&M waves, chiral media and their relation to polarization and the Fresnel equations for reflection /refraction of a homogenous E&M wave by a smooth interface.

We will conclude with a couple of recent examples of remarkable behavior of matter in interaction with E&M fields and comments on direct and inverse problems for electromagnetism are presented.

Victor DeGruttola

Harvard School of Public Health

TITLE: “Mathematical Modeling for the HIV Response to Treatment: Some Past Achievements and Future Direction”

ABSTRACT:
Mathematical modeling of the dynamics of the Human Immunodeficiency Virus Type I (HIV-1) has led to important insights into the biology of HIV-1 infection. The widespread belief that HIV-1 and the CD4+ T-lymphocytes it infects are cleared rapidly (half-lives on the order of hours or days) resulted from such modeling. These models may be fit separately to information on each study subject or observational unit, but nonlinear mixed effects models have been applied to data on all patients simultaneously. The latter approach can be more efficient in some settings (especially settings with sparse data) and is more appropriate for modeling within- and between-subject variation. Furthermore, such methods are better able to detect model misspecification. In addition to providing information about mechanisms of infection, viral dynamic models can also provide useful early information about drug efficacy. Such efficacy may wane, however, as a result of the development of resistant virus within treated individuals. Virological rebound may occur for this or for other reasons at any time initiation of therapy. Nonlinear mixed effects models have been used in this setting as well, to show an association between lower nadir of viral burden following treatment and longer time to rebound. One reason for this association might be that rates of mutation in the HIV-1 genome are related to the amount of ongoing replication.

Modeling the process of development of resistance at a mechanistic level would be an important achievement, comparable to the development of dynamic models. Many mutations that confer resistance to available classes of drugs have been identified, but their exact role, rate of appearance, and interactions are still under investigation. Understanding the relationship between viral genotype and drug efficacy is important for: 1) selection of appropriate treatment regimens for individual patients, 2) sequencing treatments, 3) drug development. Ultimately it would be valuable to develop models relating the presence of resistance-inducing mutations at onset of treatment, or their appearance during treatment, to viral dynamic parameters. The high dimensionality of viral genetics data, however, poses challenges for this research

David Draper

University of California-Santa Cruz

TITLE: "Statistical Methodology for Inverse Problems"

ABSTRACT:
The following class of scientific problems at the interface between applied mathematics and statistics has arisen with greatly increased frequency in recent years: (1) the real-world process at issue is modeled as the solution to a system of partial differential equations (PDEs), which can typically only be solved numerically by a complicated (deterministic) computer program; (2) the output of this program is typically a curve or surface unfolding in time and space; (3) the inputs to this program (typically, the boundary conditions for the PDEs and the values of physical constants arising in the equations) are not known with certainty; (4) this uncertainty is typically quantified by means of (prior) probability distributions over the unknown quantities in (3), obtained by expert elicitation; and (5) Monte Carlo methods are employed to explore the mapping from inputs to outputs, by making repeated random draws of vectors of inputs and passing them through the PDE-solver, resulting in a set of random curves or surfaces. In this talk I will describe methods for analyzing data arising in this way, using as a case study a problem in nuclear waste disposal risk assessment.

Alberto Grunbaum

University of California-Berkeley

TITLE: "Network Tomography: A Nonlinear Inverse Problem for a Stochastic Process"

ABSTRACT:
The usual tomographic model which is appropriate in the case of X-rays ignores scattering and aims at imaging the attenuation coefficient as a function of location. The use of low energy sources like an infrared laser requires a more sophisticated model based either on the full wave equation or on some kind of transport theory or stochastic model.

I will discuss one such model in detail and connect this physically motivated inverse problem with a more general problem for networks. The basic problem is to infer characteristics of the network from appropriate boundary measurements. Some preliminary results are very encouraging but the field is quite open and there are many more questions than answers.

Sarah Holte

Fred Hutchinson Cancer Research Center

TITLE: "Mathematical Models of HIV Infection: Past, Present, and Future"

ABSTRACT:
In this talk, I will describe the use of mathematical models in the area of modeling infection dynamics in HIV infected individuals. Early work in this area by Ho, Wei, and Perelson resulted in dramatic new insights into the behavior of the virus during the long latent phase of infection that most individuals experience. Since that time, numerous researchers have applied more or less complex models, both with and without the use of experimental data with varying degrees of success.

I will outline the salient features of some of these models and offer suggestions about what needs to done in order to move forward to successfully contribute to research in HIV infection dynamics. Specifically, I will outline the problems associated with formal statistical inference in these models that must be addressed.

Karl Kunisch
University of Graz, Austria

TITLE: "Numerical Techniques for Least Squares Problems with Non-Quadratic Penalty Terms

ABSTRACT:
Quadratic penalty terms are widely used for stabilization of ill-posed problems. They provide the advantage of allowing a straight-forward statistical interpretation, can be linked to maximum likelihood formulations, and are simple to differentiate. For linear inverse problems the resulting optimality system are the normal equations for which numerical solution techniques are readily available.

The use of non-quadratic penalty terms, on the other hand, is well-motivated by their ability to preserve edges and corners in the objects that must be identified by means of the inverse problem formulation and allows reduction of over smoothing that is typical for quadratic penalty terms for such features. Functionals which are most frequently used in this context are of BV (bounded variation) - type. For one-dimensional problems this formulation is related to the taut-string algorithm, which is well analyzed in the statistical literature. A common feature to these approaches is the appearance of non-differentiable cost-functionals. Their numerical treatment is not standard and will be the focus of this lecture.

Related difficulties arise if the fit-to-data criterion is not of quadratic type but it rather of contains L-1 terms, as motivated by robust statistics considerations. Again this results in a non-differentiable optimization problem the structure of which must be utilized for the development of efficient numerical algorithms.

Ira M. Longini, Jr
Emory School of Public Health

TITLE: "Using Mathematical Models to Evaluate HIV Vaccination Strategies"

ABSTRACT:
In anticipation of an effective HIV vaccine, it is important that vaccination policies be developed to assure that vaccine resources are used to optimal effect. Because the worldwide demand for

vaccine can be anticipated to exceed initial production capacities, it is essential that countries develop strategies for achieving effective HIV control with limited vaccine supplies. The objective of this study was to evaluate HIV vaccine distribution strategies for several HIV epidemic scenarios. A deterministic mathematical model consisting of a system of nonlinear differential equations was developed. The model was fit to data from a number of developing country settings. Vaccination strategies achieving optimal control of HIV transmission with available vaccine supplies were investigated. The findings demonstrate that the population level impact of a vaccination program can vary dramatically as a result of the HIV vaccination policies enacted. Furthermore, some strategies (e.g., targeting vaccination to small high-risk groups, such as injection drug users) may be considered optimal when evaluated by their cost-effectiveness, but would not be optimal if evaluated by the criterion of achieving the greatest absolute reduction in HIV transmission with a given amount of vaccine. The best HIV vaccination strategies will be dependent on the specific HIV vaccine effects and the transmission dynamics of the HIV epidemics in the scenario countries.

Peter Mueller

M.D. Anderson Cancer Center

TITLE: "Mixture Modeling"

ABSTRACT:
We consider a class of models and related inference schemes that arise when collecting data from a number of individuals, or experimental units. We assume that for each individual an inverse problem is solved by means of probabilistic inference, i.e., using a formulation of the inverse problem as a statistical inference question. A typical example is the analysis of population pharmacokinetic (PK) data. To understand patterns of variation and mechanistic behavior across the population we embed inference for each individual in a larger, encompassing model.

We argue for mixture models as prior probability models in this modeling context and discuss modeling and posterior inference in such models. We will review appropriate Markov chain Monte Carlo (MCMC) simulation methods, including simulation across variable dimension models, model averaging, convergence diagnostics, etc. The discussion will be focused on mixture models. But many of the methods and approaches discussed are valid in general, for any posterior inference in complex probability models.

Kurt E. Oughstun

University of Vermont

TITLE: "Dispersive Pulse Dynamics and Precursor Fields in Dielectric and Conductive Materials"

ABSTRACT:
A general description of the precursor fields associated with the propagation of a rapid rise-time signal with fixed carrier frequency in a general homogeneous, isotropic, locally linear, causally dispersive, nonmagnetic medium is presented. The description is based upon the modern asymptotic theory, which relies upon the dynamical evolution of the saddle points of the complex phase function that appears in the exact Fourier-Laplace integral representation of the propagated field. These saddle point dynamics primarily depend upon the high and low frequency structure of the dispersion relation for the dispersive medium which is obtained here directly from the

general Kramers-Kronig relations for the complex dielectric permittivity. Material dispersion dependent conditions for the appearance of the Sommerfeld and Brillouin precursors are obtained. These general conditions are exemplified by both the Debye and Rocard-Powles models of rotational polarization phenomena and the Lorentz model of resonance polarization phenomena in lossy dielctric media. The persistence of the Brillouin precursor in the presence of conductivity is also presented.

Laura K. Potter

Glaxo Smith Kline

TITLE: "Incorporating Variability and Uncertainty into Human Health Risk Assessment"

ABSTRACT:
The potential health risks associated with exposure to toxic environmental contaminants often vary widely across the human population. Factors that contribute to such inter-individual variability include age and sex, as well as differences in metabolic rates, circulation, fat content and disease state. Environmental regulators take this variability and general uncertainty into account when determining levels of safe exposures for chemicals, often with the use of safety factors that divide the dose level by a standard amount. A challenge for environmental health scientists is to develop more precise, scientifically based methods to better quantify uncertainty and population variability in both the toxicokinetic profiles and the dose-response behaviors of toxic compounds. In this presentation an overview of the risk assessment process will be given, with a focus on the current modeling techniques that are utilized. Probabilistic methods for incorporating variability and uncertainty into these models will be discussed.

Gary Rosner

M.D. Anderson Cancer Center

TITLE: "Statistical Modeling and Inverse Problems"

ABSTRACT:
In this talk, I will use a case study to present some statistical methodology aimed at making inference in inverse problems. Although ostensively a biomedical problem, the methodology has broader application, extending to situations in which multiple sources of variation may need parsing to improve inference. The motivating example is the measurement of the concentration of a pharmaceutical in patients over time, where we want to allow for between-individual variation, as well as within individual variation. I will also discuss the problem of finding optimal study designs when the data consist of measurements that can be characterized by a complex dynamic nonlinear model with subject-specific parameters.

Fadil Santosa

University of Minnesota

TITLE: "Regularization of Ill posed Inverse Problems by Priori Information"

ABSTRACT:
Many inverse problems arising in applications are illposed by nature. Regularization procedure, which are often employed to stabilize the problem, can come in many different forms. The use of

a priori information is of paramount importance in devising regularization. In many problems, we often know what the solution should look like. The challenge is how to build in our preference into the solution procedure and thereby stabilizing the problem.

In this presentation, I will describe two simple model inverse problems and proceed by reviewing deterministic approaches of using a priori information in regularizing them. A goal of this talk is to engage statisticians in the audience to develop a unified regularization approach where a priori statistical information is also incorporated.

Curt Vogel
Montana State University

TITLE: "An Applied Mathematician's Prospective on Regularization Methods"

ABSTRACT:
We will be reviewing abstract concepts like compactness and well-(and ill-) posedness for an abstract parameter identification problem

$$d = Aq.$$

Practical computations require a discretization of this abstract problem. In addition, measured data is discrete and contaminated with error. Hence we must consider a model like

$$\mathrm{d} = A\mathrm{q}_{\mathrm{true}} + \eta \qquad (1)$$

Our goal is then to accurately estimate the true (discretized) parameter $\mathrm{q}_{\mathrm{true}}$ from d contaminated by noise .
We will review the concept of regularization and present some concrete examples. One Such example is Tikhonov regularization (also know as MAP estimation), where we minimize the functional

$$T(\mathrm{q}) = \ell(A\mathrm{q};\mathrm{d}) + \alpha P(\mathrm{q}). \qquad (2)$$

The key to effective regularization is the incorporation of prior information .The first term on the right hand-side of (2) quantifies the degree to which $A\mathrm{q}$ fits the data d . If statistical information about the noise η in (1) is available, this term can be taken to be the negative of the log likelihood functional corresponding to the noise distribution.

The second term on the right-hand-side of (2) quantifies prior information about the parameter. In certain statistical models provide a natural prior. More often, an ad hoc selection is made. In this case, one must also select the so-called regularization parameter in (2). We will close with a discussion of statistically based regularization parameter selection procedures.

Robert Wolpert
Duke University

TITLE: "A Nonparametric Bayesian Approach to Inverse Problems"

ABSTRACT:
We propose a new method for making inference about the solution to a Fredholm integral equation of the first kind, i.e., of estimating an unknown measure H(ds) upon observing some

values of the integral G(t) of a known kernel k(x,s) with respect to H(s), using L'evy processes as Bayesian prior distributions for modeling uncertainty about H(ds).

Inference is based on simulation-based MCMC methods. The method is illustrated with a problem in rheology, a branch of polymer chemistry.

C. *Program for the Inverse Closing Workshop*
May 14-15, 2003, NISS-SAMSI Building

- **Wednesday, May 14, 2003**

9:15	Opening Remarks
9:30-10:30	"Statistics for Science: Perspectives from Pharmacokinetics/Pharmacodynamics" Lewis Sheiner, University of California, San Francisco
10:30-11:30	"A Simulation Based Comparison Between Parametric and Semiparametric Methods in a PBPK Model" Yanyuan Ma, SAMSI
11:30-11:45	Discussions
11:45-1:00	Lunch
1:00-2:00	"Non-Stationary Inverse Problems and Dynamical Prior Models" Erkki Somersalo, Helsinki University of Technology
2:15-3:00	"Inverse Problems in Complex Model Validation" Danny Walsh, SAMSI
3:00-3:30	Discussions

- **Thursday, May 15, 2003**

9:30-10:30	"A Hierarchical Bayesian Spatio-Temporal Model for Predicting the Spread of Invasive Species Given Uncertain Observations" Christopher K. Wikle, University of Missouri-Columbia
10:45-11:30	"2D Electromagnetic Parameter Identification for a Debye Polarization Model" Johnathan Bardsley, SAMSI
11:30-11:45	Discussions
11:45-1:00	Lunch
1:00-2:00	"Iterative Regularization of Nonlinear Inverse Problems: Deterministic Convergence Theory, Ideas on Incorporating Uncertainty" Heinz Engl, Johannes Kepler Universitat

2:00-3:30 Closing Discussions and Remarks

II. CHALLENGES IN STOCHASTIC COMPUTATION

A. *Opening Workshop Program*
September 25-October 1, 2002, Radisson Governor's Inn

- **Wednesday, September 25**

12:45-1:30 Registration

1:30-3:00 **Tutorial 1**: Introduction to Stochastic Computation
John Monahan, North Carolina State University

3:00-3:30 Coffee Break

3:30-5:00 **Tutorial 2** (Part 1): Introduction to Gibbs Sampling and Markov Chain Monte Carlo Methods
Alan Gelfand, Duke University

- **Thursday, September 26**

8:00-8:30 Registration

8:30-10:00 **Tutorial 2** (Part 2): MCMC with Applications to Hierarchical Models
Brad Carlin, University of Minnesota

10:00-10:30 Coffee Break

10:30-12:30 **Tutorial 3**: Overview of Perfect Simulation Methods
Duncan Murdoch, University of Western Ontario

12:30-2:30 Lunch

2:30-3:30 **Tutorial 4** (Part 1): From Chain Polymers To Nonlinear Dynamic Systems: An introduction to Sequential Monte Carlo
Jun Liu, Harvard University

3:30-4:00 Coffee Break

4:00-5:30 **Tutorial 4** (Part 2): Sequential Sampling Techniques
Simon Godsill, Cambridge University

- **Friday, September 27**

8:45-9:30 Registration

9:30-11:30 **Tutorial 5**: Aspects of Computation for Genomic Variation Data
Simon Tavare, University of Southern California

11:30-1:30 Lunch

1:30-3:00 **Tutorial 6** (Part 1): Advanced Monte Carlo Methods
Peter Green, University of Bristol

3:00-3:30 Coffee Break

3:30-4:30 **Tutorial 6** (Part 2):Some Advanced MCMC Techniques
Jun Liu, Harvard University

- **Saturday, September 28**

9:15-10:00 Registration

10:00-10:20 Opening Remarks by *Jim Berger & Mike West*

Session 1: MCMC Theory and Methods
10:20-12:00 Chair: Merlise Clyde

"Sequential Importance Sampling with Pilot Exploration"

Jun Liu

"Adaptive Polar Sampling: a Flexible and Robust Monte Carlo Method"

Herman van Dijk

12:30-1:30 Lunch

Session 2: More MCMC Theory and Methods
1:30-3:10 Chair: John Monahan

"A Strategy for MCMC Acceleration"

David Draper

"Honest MCMC via Drift and Minorization"

Jim Hobert

3:10-3:40 Coffee Break

Session 3: And more MCMC
3:40-5:20 Chair: Feng Liang

"An Efficient Sampler for Decomposable Covariance Selection Models"

Chris Carter

"Analysis of van Dyk and Meng's Data Augmentation Algorithm for the Multivariate t Model"

Dobrin Marchev

8:00-11:00 **RECEPTION & POSTER SESSION** at NISS/SAMSI Building. Transportation provided.

- **Sunday, September 29**

Session 4: Particle Filtering and Sequential Methods
9:00-10:40 Chair: Patrick Wolfe

"Parameter Estimation in General State-Space Models using Particle Methods"

Arnaud Doucet

"Smoothing and Filtering with Particle Approximations"

Simon Godsill

10:40-11:10 Coffee Break

Session 5: Stochastic Computation in Contingency Tables and Discrete Distributions
11:10-12:50 Chair: Christine Kohnen

"Posterior Distributions over Spaces of Contingency Tables"

Adrian Dobra

"Transform Methods for the Hypergeometric Distribution"

Ian Dinwoodie

12:50-2:00 Lunch

Session 6: Stochastic Computation in Spatial Statistics
2:00-3:40 Chair: Sandra McBride

"Hidden Markov Models and Disease Mapping"

Peter Green

"Stochastic Computation Strategies for Fitting Spatial Data Models"

Alan Gelfand

3:40-4:10 Coffee Break

Session 7: Stochastic Computation in Nonparametric Regression and Curve Fitting
4:10-5:50 Chair: Susie Bayarri

"Stochastic Simulation Based Methods for Bayesian Curve and Surface fitting"

Bani Mallick

"Estimation and Variable Selection in Nonparametric Heteroscedastic Regression"

Robert Kohn

- **Monday, September 30**

Session 8: Stochastic Computation in Population Genetics
9:00-10:40 Chair: Ed Iversen

"Approximate Bayesian Computation in Population Genetics"

Simon Tavaré

"Sequential Importance Sampling with Resampling in Molecular Population Genetics"

Yuguo Chen

10:40-11:10 Coffee Break

Session 9: More Stochastic Computation in Genetics
11:10-12:50 Chair: Mike West

"Pedigree Data Analysis with Crossover Interference"
Sharon Browning

"MCMC in Parentage Analysis"
Beatrix Jones

12:50-2:00 Lunch

Session 10: Perfect Sampling and Novel Methods of Stochastic Computation
2:00-4:30 Chair: Robert Wolpert

"Perfect Sampling for Some Mixtures of Distributions"

Mark Huber

"Perfect Sampling for Non-Markovian Queues"

Duncan Murdoch

"On the Possibility, and Impossibility, of Interruptible Perfect Sampling"

Jim Fill

- *Tuesday, October 1*

Session 12: Stochastic Computation in Stochastic Volatility
9:00-10:40 Chair: German Molina

"An MCMC Approach for Multivariate Dynamic Models with Stochastic Volatility"

Jane Liu

"Multivariate Stochastic Volatility Models: Portfolio Allocation, Financial Contagion and Regime Switching"
Hedibert Lopes

10:40-11:10 Coffee Break

Session 13: Stochastic Computation in Hierarchical Models and Applications
11:10-12:50 Chair: Sujit Ghosh

"Bayesian Hierarchical Quantile Regression"

David Dunson

"Signature Recognition via MCMC"

IAN MCKEAGUE

12:50-2:00 Lunch

Session 14: Stochastic Computation in Financial Modelling
2:00-3:30 Chair: Chuanshu Ji

"Monte Carlo Simulations in Finance"

Jean-Pierre Fouque

"Variance Reduction Methods in Monte Carlo Simulations for Stochastic Differential Equations"

Yazhen Wang

3:30-4:15 Coffee Break

4:15-5:30 **SAMSI Distinguished Lecture**
"Inference from the Fossil Record: When was the Last Common Ancestor of Extant Primates"
Simon Tavaré

B. Opening Workshop Poster Session

Saturday, September 28, 2002
8:00 – 11:00 p.m., NISS / SAMSI Building

- "Price Elasticity of the Demand for Water: A Meta-Analysis"
  Gail Blattenberger (University of Utah)

- "Simulation Based Sequential Analysis of Markov Switching Stochastic Volatility Models"
  Carlos Carvalho (Duke University)

- "A Bayesian Approach for Assessing Heterogeneity in Generalized Linear Models"
  Zhen Chen (National Institute of Environmental Health Sciences)

- "MCMC for a distribution over random graphs: fitting a selection model to allele data"
  Nicoleen Cloete (University of Auckland)

- "Space-time Modelling of Sydney Harbour Winds"
  Ed Cripps (University of New South Wales)

- "Performance of MCMC techniques for parameter estimation of nonlinear stochastic dynamical systems"
  Milena Cuellar (London School of Economics)

- "A Comprehensive Stochastic Risk Model for the Hanford Nuclear Site"
  Dave Engel (Pacific Northwest Laboratories)

- "MCMC sampling for Factor Analytic models"
  Ernest Fokoue (Ohio State University)

- "Eigenvalue Bounds for Markov Chains on Nonnegative Lattice Points"
  Emily Gamundi (Tulane University)

- "Bayesian inference in a hidden stochastic two-compartment model for feline hematopoesis"
  Daniela Golinelli (RAND Statistics Group)

- "Towards independent sampling for Bayesian disease mapping"
  Murali Haran (University of Minnesota)

- "Efficient Sampling from nonstandard distributions using neural network approximations"
  Lennart Hoogerheide (Erasmus University Rotterdam)

- "Empirical Bayes Analysis of cDNA Microarray Data"
  Leanna House (Duke University)

- "Scaling Up Bayesian Model Averaging"
  Christine Kohnen (Duke University)

- "Sequential Monte Carlo calibration of stochastic volatility models for interest rate term structure"
  Beom Lee (University of North Carolina-Chapel Hill)

- "Resampling in pairwise relationship inference"
  Li Li (North Carolina State University)

- "A Bayesian Population Approach to Modelling Bioaccumulation of Polycyclic Aromatic Hydrocarbon in an Aquatic Food Web"
  Hsin-I Lin (Tulane University)

- "BMM: A Simple, General Algorithm for Drawing from Densities"
  Rob McCulloch (University of Chicago)

- "Some perfect sampling algorithms on continuous state spaces"
  Faheem Mitha (University of North Carolina, Chapel Hill)

- "MCMC Sampling under Multiple Constraints"
  German Molina (Duke University)

- "Small Sets and Domination in Perfect Simulation"
  Giovanni Montana (University of Chicago)

- "McMC Analysis of Gaussian Spatial Processes Having Product Correlation Form"
  Rui Paulo (Duke University)

- "Distance-based Model Selection with Application to the Analysis of Gene Expression Data"
  Surajit Ray (Pennsylvania State University)

- "Efficient Gibbs Sampler implementations to the Constrained Normal Linear Regression Model"
  Gabriel Rodrigues-Yam (Colorado State University)

- "Application of Particle Filtering to State-space Models of Wildlife Population Dynamics"
  Len Thomas (University of St Andrews)

- "Some Empirical Results using Adaptive Polar Sampling"
  Rutger Van Oest (Erasmus University Rotterdam)

- "MCMC Techniques in Veterinary Diagnostic Test Validation"
  Didier Verloo (Institute of Tropical Medicine)

- "Efficient Stochastic Computation for Audio Signal Processing"
  Patrick Wolfe (University of Cambridge)

C. Opening Workshop Speaker Abstracts

Christophe Andrieu

University of Bristol

TITLE: "Controlled MCMC for Automatic Sampler Calibration" (joint work with C.P. Robert and E. Moulines)

ABSTRACT:
In this talk we present an original and general framework for automatically optimising the statistical properties of Markov chain Monte Carlo (MCMC) samplers. Classical MCMC samplers usually depend on parameters, say \theta, that need to be tuned in order to lead to efficient algorithms. It is well known for example that the performance of a Metropolis-Hastings algorithm will heavily depend on the choice of a 'good' proposal distribution, which might depend on some parameters. A natural goal is therefore to optimise the set of parameters \theta, on which the sampler depends, in order to satisfy some statistical criteria. The methodology we propose allows for the self-tuning of the Markov chain process in the light of its history: the sampler therefore learns "on the fly" the optimal set of parameters. The method is supported by theoretical results which prove the convergence of the method under fairly general and applicable conditions. A particular emphasis is given on non-asymptotic and explicit bounds on the convergence of ergodic averages. These bounds have clear practical interpretations in terms of basic properties of the transition probability of the Markov chain and give some insight into the potential benefits of such adaptation schemes. We present several detailed examples of applications and numerical experiments. In particular we show how optimal blocking can be chosen in the context of stochastic volatility models (cf Shephard & Pitt, 1997).

Gail Blattenberger
University of Utah

TITLE: "Price Elasticity of The Demand for Water: A Meta-Analysis"

ABSTRACT:
Severe drought in Utah has focused attention on current outmoded pricing strategies. A critical question is the impact of price on water demand. Data unavailability limits the feasibility of a local study, but the abundance of empirical work on water demand elsewhere renders meta-analysis an appropriate alternative. A meta-analysis is conducted using a hierarchical linear model estimated with maximum likelihood and MCMC methods. The focus of this analysis is the implication for forecasting price impacts in the local situation.

Sharon Browning
North Carolina State University

TITLE: "Pedigree data analysis with crossover interference"

ABSTRACT:
We propose a new method for calculating likelihoods on pedigree genetic data that incorporates crossover interference using the chi-square models. The method is based on importance sampling of unobserved inheritance patterns conditional on the observed genotype data, and takes advantage of fast algorithms for no-interference models while using reweighting to allow for interference. We show that the method is effective for arbitrarily many genetic markers with small pedigrees.

Chris Carter
Duke University

TITLE: "An Efficient Sampler for Decomposable Covariance Selection Models"

ABSTRACT:
This paper proposes an efficient sampler for the Bayesian estimation of covariance matrices using Markov chain Monte Carlo (MCMC). The methods apply to decomposable covariance selection models with a hyper inverse Wishart (HIW) prior for the covariance matrix. The conjugate properties of the HIW distribution are used to generate from reduced conditional distributions. In particular, the covariance matrix is integrated out of all conditional distributions and is not generated in the MCMC. The resulting sampler is shown to have a much faster convergence rate than existing methods. The computational complexity of one iteration of the MCMC is shown to be similar to existing methods, so the gain in convergence rate is significant. An efficient mixture estimate of the posterior mean of the inverse covariance matrix is given.

Yuguo Chen
Duke University

TITLE: "Sequential Importance Sampling with Resampling in Molecular Population Genetics"

ABSTRACT:
We review the sequential importance sampling (SIS) algorithms developed in population genetics based on the coalescent model. Different approaches to formulating the proposal distribution for these SIS methods are explained and compared. We also review the technique of resampling, which often substantially increases the efficiency of sequential importance sampling algorithms. Although the usual approach to resampling does not work so well for statistical inference problems in population genetics, we develop a new resampling schedule which can improve efficiency significantly. We give insights into our resampling schedule, and discuss some possible generalizations of the idea.

Zhen Chen
National Institute of Environmental Health Sciences

TITLE: "A Bayesian Approach for Assessing Heterogeneity in Generalized Linear Models"

ABSTRACT:
Generalized linear mixed models (GLMMs) are used routinely for analyzing clustered data arising in a broad variety of applications. In Bayesian analyses, inverse Wishart or inverse gamma priors are almost always used for the covariance of the random effects, for computational convenience and to enforce the positive definite constraint on the covariance matrix. In this article, we propose a new class of prior distributions based on a Gaussian structure for variance component parameters underlying the random effects covariance. The proposed prior assigns positive probability not only to the full model but also to reduced models that exclude one or more of the random effects. This structure facilitates Bayesian inferences about the covariance structure, while also accounting for uncertainty in the random effects model in estimating the population parameters. A Markov chain Monte Carlo algorithm is proposed for posterior computation, and the approach is illustrated using data on prenatal exposure to PCBs and psychomotor development.

Seo-eun Choi
Florida State University

TITLE: "A Statistical Approach to the Ocean Circulation Inverse Problem" (joint work with Ian McKeague and Kevin Speer)

ABSTRACT:
Scientific progress in oceanography is often based on explaining physical relationships among quantities such as pressure, salinity, temperature and current velocity. For example, such relationships are used to estimate current velocity from pressure when pressure is only determined up to a constant of integration from available data. Our object is to produce a map of current velocity in the North Atlantic based on sparse measurements along ship tracks. Oceanographers have formulated this as a (deterministic) inverse problem. We propose an alternative statistical approach based on Bayesian methods. We use the equations of motion describing the ocean (Navier-Stokes fluid equations) to specify the prior. Our preliminary work concentrates on the Stommel Gulf Stream model and the development of a Gibbs sampler to extract features from the posterior velocity field.

Nicoleen Cloete
University of Auckland

TITLE: "MCMC for a distribution over random graphs: fitting a selection model to allele data"

ABSTRACT:
In the absence of selection effects, the stochastic development of a genealogy with a population of fixed size is often modeled using the Kingman coalescent process. This process determines a probability distribution over rooted binary trees of fixed dimension. Recently Neuhauser and Krone gave a stochastic model generalizing the Kingman coalescent in a natural way to include the effects of selection. The new model determines a distribution over a class of graphs of randomly variable dimension. Our aim is to carry out Bayesian inference for the selection parameter of the model of Neuhauser and Krone, from allelic data, using Markov chain Monte Carlo. We describe an algorithm for this purpose, focusing on efficiency considerations.

Edward Cripps
University of New South Wales

TITLE: "Space-time Modelling of Sydney Harbour Winds" (joint work with William Dunsmuir, David Nott and Chris Wikle)

ABSTRACT:
In this paper we develop a space-time statistical model for local forecasting of surface level wind fields in a coastal region with complex topography. Our statistical model makes use of output from deterministic numerical weather prediction (NWP) models, which are able to produce forecasts of surface wind fields on a spatial grid. When predicting surface winds at observing stations errors can arise due to sub-grid scale processes not adequately captured by the NWP model, and our statistical model attempts to correct for these influences. In particular, we use information from observing stations within the study region as well as topographic information to account for local bias. We use Bayesian methods for inference in our model, with computations

carried out using Markov chain Monte Carlo (MCMC) algorithms. Emprirical performance of our model is described.

Milena C. Cuellar
London School of Economics

TITLE: "Performance of Markov Chain Monte Carlo (MCMC) techniques for parameter estimation of chaotic dynamical systems"

ABSTRACT:
We study the performance of MCMC techniques in the estimation of parameters of well known chaotic systems. We focus our study on the sensitivity of the results to the choice of noninformative priors in particular models. We discuss the implications of the implementation of MCMC techniques--based on pure stochastic modeling--for the characterization of deterministic chaotic systems. Our study is a controlled numerical experiment that takes as input the time series generated by the Logistic, Tinkerbell, Henon, and Henon-like maps, with dynamical and measurement noise.

I. H. Dinwoodie
Tulane University

TITLE: "Transform methods for the hypergeometric distribution" (joint work with Laura Matusevich and Ed Mosteig)

ABSTRACT:
Two new methods for computing with hypergeometric distributions on multi-dimensional lattice points are presented. One uses Fourier analysis, and the other uses Groebner bases in the Weyl algebra. Both are very general and apply to log-linear models that are graphical or non-graphical. The main application is the exact conditional test of parameter significance.

Adrian Dobra
Duke University

TITLE: "Posterior Distributions over Spaces of Contingency Tables"

ABSTRACT:
In this talk we present two Bayesian approaches for sampling from the posterior distribution of a contingency table having a fixed set of marginal totals. Both of these methods are based on data augmentation and model averaging. However, they are essentially different because one of them makes use of Markov bases (Diaconis and Sturmfels, 1998) to generate a feasible table consistent with the fixed marginals, while the other employs a novel sequential cell sampling algorithm described in Dobra, Tebaldi and West (2002). We discuss advantages/disadvantages of using these two methods and present examples.

Arnaud Doucet
University of Cambridge

TITLE: "Particle Methods for Control of Nonlinear Non-Gaussian State-Space Models" (joint work with Vladislav Tadic & Sumeet S. Singh)

ABSTRACT:
Particle methods have been widely used to solve optimal filtering problems in general state-space models. We consider here the more challenging control problem; i.e. how to determine a control policy maximizing an expected average or discounted reward. This problem arises in statistics, experimental design, finance, signal processing, communications etc. Except for a few cases (linear Gaussian models and a quadratic cost), it is well-known that the Bellman equation cannot be solved analytically. We propose here a simple simulation-based method combining particle filtering and stochastic approximation which allows to solve efficiently this problem.

David Draper
University of California, Santa Cruz

TITLE: "A strategy for MCMC acceleration" (joint work Shufeng Liu)

ABSTRACT:
In this work we explore the usefulness of a simple idea for reducing the autocorrelation time of MCMC output, in settings where the posterior distribution is unimodal and some rough estimates of the location of the mode and the scale of the posterior (e.g., from maximum likelihood considerations or preliminary MCMC output) are available.

Dave Engel
Battelle-Pacific Northwest National Laboratory

TITLE: "A Comprehensive Stochastic Risk Model for the Hanford Nuclear Site"

ABSTRACT:
From its creation in 1943 until recently, the facilities on the 560 square-mile Hanford Site in Washington State, USA, were dedicated primarily to the production of weapons-grade plutonium for national defense. Decades of production of nuclear materials using up to 10 reactors have left nuclear and chemical contamination on the site and in the adjoining Columbia River. A computer system, SAC (Systems Assessment Capability), was created to examine the current and future risks on the Hanford Site and in the Columbia River from all wastes on the site. The latest SAC analysis incorporates release information from
720 waste sites and generates stochastic estimates of several risk metrics over a 1,100-year modeling period using a Monte-Carlo approach. This presentation will describe the approach and the computer platform used for the release, transport, and risk computations (distributing the analysis over a 128 CPU Linux cluster) along with an uncertainty analysis performed on the assessment. Example results will be provided for a 25-realization analysis based on the Hanford-baseline cleanup plan. The work illustrates the incorporation of uncertainty concepts into modeling projections that can assist in setting environmental cleanup objectives.

Jim Fill
Johns Hopkins University

TITLE: "On the Possibility, and Impossibility, of Interruptible Perfect Sampling"

ABSTRACT:
I will discuss various aspects of perfect sampling. A particular focus will be recent joint work with Keith Crank in which we establish, for various scenarios, whether or not interruptible perfect sampling from the stationary distribution is possible when an ergodic finite-state Markov chain can only be viewed passively. In particular, we prove that such sampling is not possible using a single copy of the chain. Such sampling is possible when enough copies of the chain are available, and we provide an algorithm that terminates with probability one.

Ernest Parfait Fokoue
Ohio State University

TITLE: "MCMC sampling for Factor Analytic models"

ABSTRACT:
In the study of the factor analysis model, two of the most common goals are the determination of the number of factors and the derivation of a "simple" structure. The classical approach deals with these two tasks separately, and often resorts to ad-hoc methods. This paper adopts a Bayesian approach to those problems, and adapts ideas from stochastic geometry and Bayesian finite mixture modelling to construct an ergodic Markov chain having the posterior distribution of the complete collection of parameters (including the number of factors) as its equilibrium distribution. The algorithm proposed combines a Gibbs sampler updating scheme with the discrete simulation of a continuous-time birth-and-death point process to produce a sampling scheme that efficiently explores the posterior distribution of interest. The MCMC sample path obtained from the simulated posterior then provides a flexible ingredient for most of the inferential tasks of
interest. Illustrations on both artificial and real tasks are provided,while challenges and ideas for future improvements are also presented.

Simon Godsill
University of Cambridge

TITLE: "Smoothing and filtering with particle approximations" (joint work with William Fong)

ABSTRACT:
We describe recent developments in Monte Carlo filtering and smoothing for state space models in which parameters exhibit slow and smooth evolution with time. In such cases particle methods are well known to become degenerate. We discuss methods based upon a multiscale representation of the system dynamics
in which certain critical parameters are constrained to vary slowly and smoothly with time. We also discuss how to make the methods adaptive to the local smoothness of parameter variation. Examples are given in time-varying autoregressive models for speech.

Peter Green
University of Bristol

TITLE: "Hidden Markov models and disease mapping"

ABSTRACT:

We present new methodology to extend Hidden Markov models to the spatial domain, and use this class of models to analyse spatial heterogeneity of count data on a rare phenomenon. This situation occurs commonly in many domains of application, particularly in disease mapping. We assume that the counts follow a Poisson model at the lowest level of the hierarchy, and introduce a finite mixture model for the Poisson rates at the next level. The novelty lies in the model for allocation to the mixture components, which follows a spatially correlated process, the Potts model, and in treating the number of components of the spatial mixture as unknown. Inference is performed in a Bayesian framework using reversible jump MCMC. The model introduced can be viewed as a Bayesian semiparametric approach to specifying flexible spatial distribution in hierarchical models.

Murali Haran
University of Minnesota

TITLE: "Towards independent sampling for Bayesian disease mapping"

ABSTRACT:
Spatial Poisson models using conditional autoregressions are commonly used in Bayesian modeling of areal data. Inference for such models is generally carried out via Markov chain Monte Carlo methods. In this paper, we propose a method for producing independent samples from the posterior distribution. We provide a systematic method for producing heavy tailed proposal distributions that can be used in a perfect sampling scheme for these models. This scheme is along the lines of the perfect sampling algorithm described in Moller and Nicholls (1999). We describe the application of our methods to three Minnesota cancer data sets.

Jim Hobert
University of Florida

TITLE: "Honest MCMC via Drift and Minorization" (joint work with Galin Jones, Brett Presnell and Jeffrey Rosenthal)

ABSTRACT:
I call an MCMC algorithm "honest" if it is possible to calculate standard errors using a valid central limit theorem along with a consistent estimate of the (unknown) asymptotic variance. In this talk, I will explain how drift and minorization conditions on the underlying Markov chain make honest MCMC possible.

Mark Huber
Duke University

TITLE: "Perfect Sampling for Some Mixtures of Distributions"

ABSTRACT:
Coupling from the past (CFTP) has been utilized to generate perfect samples from a posterior distribution for mixtures by Hobert, Robert, and Titterington. I'll discuss their approach and methods for speeding up their algorithm, as well as an alternate method based on the Randomness Recycler (RR) protocol. RR can be faster than CFTP in certain problem instances, as well as having several technical advantages.

Beatrix Jones
Statistical and Applied Mathematical Sciences Institute

TITLE: "MCMC in Parentage Analysis"

ABSTRACT:
Parentage Analysis studies typically are interested in characterizing mating and perhaps offspring dispersal patterns for a natural (wild) population. Typically parentage, particularly paternity, cannot be directly observed. Instead, genetic information is collected on offspring and potential parents. This information rarely unequivocally determines parentage, but suggests likely candidates. An appealing approach to analysis of such data is to sample from the joint posterior of the true parentage assignments and the parameters governing mating and dispersal. (Typically the latter are of primary interest.) The presentation is illustrated with examples from plant and insect populations.

Robert Kohn
University of New South Wales

TITLE: "Estimation and Variable Selection in Nonparametric Heteroscedastic Regression" (joint work with Paul Yau)

ABSTRACT:
The article considers a Gaussian model with the mean and the variance modeled flexibly as functions of the independent variables. The estimation is carried out using a Bayesian approach that allows the identification of significant variables in the variance function, as well as averaging over all possible models in both the mean and the variance functions. The computation is carried out by a simulation method that is carefully constructed to ensure that it converges quickly and produces iterates from the posterior distribution that have low correlation. Real and simulated examples demonstrate that the proposed method works well. The method in this paper is important because (a) it produces more realistic prediction intervals than nonparametric regression estimators that assume a constant variance; (b) variable selection identifies the variables in the variance function that are important; (c) variable selection and model averaging produce more efficient prediction intervals than those obtained by regular nonparametric regression.

Beom Lee
University of North Carolina

TITLE: "Sequential Monte Carlo calibration of stochastic volatility models for interest rate term structure"

ABSTRACT:
Two MCMC-based methods, Metropolized independence sampling and sequential importance sampling, are applied to the calibration of stochastic volatility models for interest rate with term structure. Different versions of stochastic volatity models, such as the Log-AR(1) model and the CIR model, are studied. Both simulated data and real data are used in the model fitting.

Li Li

NCSU Bioinformatics Research Center

TITLE: "Resampling in pairwise relationship inference" (joint work with Sharon Browning)

ABSTRACT:
Traditional likelihood calculations for pedigree data analysis are based on algorithms developed by Elston and Stewart, and Lander and Green, which are infeasible when both pedigree size and number of markers are large. More recently, Sequential Importance Sampling (SIS) has been applied to pedigree data analysis making calculations for large pedigrees with a large numbers of markers possible. However, when the system gets large, the variance of the importance sampling weights increases while efficiency and accuracy of the method decrease. In our pairwise relationship inference study, we incorporate a resampling strategy into SIS to solve the problem. Instead of sampling each sampled path (inheritance pattern) along the chromosome independently, we sample N paths in parallel and after a certain number of marker loci, either deterministically or dynamically, draw a new sample set from the current sample set according to their importance sampling weights. Simulation studies are performed to test the efficiency and accuracy of our method comparing to SIS. Effective Sample Size is 4 to 100 times greater with resampling than with SIS only, with greatest gains for densely spaced markers.

Hsin-I Lin
Tulane University

TITLE: "A Bayesian population Approach to Modelling Bioaccumulation of Polycyclic Aromatic Hydrocarbon in an Aquatic Food Web"

ABSTRACT:
Food-web bioaccumulation models are useful in predicting pollutant concentrations in food web organisms and quantifying the ecological and human health risks. The existing natural variation between organisms in the environment is often ignored by using averaged values in models. Incorporating the inherent uncertainty of parameters into models can increase the reliability of model predictions. Bayesian statistical analysis, which treats model parameters as random variables, provides a mechanism to quantify the variability in model predictions based upon the variability in model parameters. In this study, we developed an aquatic food-web bioaccumulation model to account for the inter-individual variability of physical parameters (e.g. body weight, lipid fraction) in the predator and its prey. A set of differential equations was used to predict the concentration in food web organisms based upon the polycyclic aromatic hydrocarbon (PAH) concentrations in water, sediment, and rate constants of uptake, elimination, and metabolism. Sediment, water, and biota were collected from control and contaminated sites in the LaBranche Wetlands, Louisiana and analyzed for PAH concentration using gas chromatograph/mass spectrometry. A Markov chain Monte Carlo (MCMC) technique was used to derive the posterior distributions of the model parameters. The results were used to predict the distribution of contaminant concentrations in the population. In this study, the MCMC method was applied to a food web bioaccumulation model, which provides a quantitative expression of the variability in predicted concentrations

Jane Liu
UBS Warburg

TITLE: "An MCMC Approach for Multivariate Dynamic Models with Stochastic Volatility"

ABSTRACT:
We discuss Monte Carlo methods for Bayesian dynamic linear models (DLM). General Bayesian DLMs with unknown time varying factor sensitivities and variances do not have closed form sequential analysis. The learning of the time varying variances is commonly based on variance discounting. Variance discounting models do not specify the Markov evolution process for the variances, which constitutes a major challenge in the research of simulation based techniques for DLMs. Previous research has explored using matrix-variate Beta and Wishart distributions to provide such an evolution equation. This, however, turns out to be very restrictive.

We provide a general evolution process for variance-covariance matrices using the Bartlett decomposition of Wishart variables. Sequential updating and backward smoothing equations are derived for the dynamic variance model. They're generalized to deal with singular variance-covariance matrices. The efficient forward filtering backward sampling technique is used in the proposed Gibbs sampler for general DLMs to sample both the conditional variances and the conditional factor sensitivities. We apply this approach to a US equity sector model, and compare our result with that of some alternative approaches.

Jun Liu
Harvard University

TITLE: "Sequential Importance Sampling with Pilot Exploration" (joint work with Junni L. Zhang)

ABSTRACT:
The sequential importance sampling method and its various modifications have been developed intensively and used effectively in diverse research areas ranging from polymer simulation to signal processing and statistical inference. We propose a new variant of the method, sequential importance sampling with pilot-exploration resampling (SISPER), and demonstrate its successful application in folding polypeptide chains described by a two-dimensional hydrophobic-hydrophilic (HP) lattice model. In this method, a small random sample of pilot exploration paths are sent out to gather future information, and this pilot information is used in weighting the partial chains for enrichment and pruning. Numerical results showed that SISPER out performed several existing methods, e.g., the genetic algorithm, Prune-enrichment Rosenbluth method, and the evolutionary Monte Carlo, in finding the ground folding states of 2D HP sequences. Our recent application of SISPER to variable selection problem also showed promising results. It is likely that the main ideas underlying SISPER can be applied more broadly to optimization and integration problems in other research areas, such as statistical computing, signal processing, and bioinformatics.

Hedibert Freitas Lopes
Federal University of Rio de Janeiro

TITLE: "Multivariate Stochastic Volatility Models: Portfolio Allocation, Financial Contagion and Regime Switching"

ABSTRACT:
Some of the most recent developments in simulation-based sequential analysis are implemented to multivariate stochastic volatility problems. We improve upon existing methods by allowing the model parameters to be estimated sequentially either through traditional polinomial dynamic trends or based on latent variables describing the several changes in the parameters. We also

extend the traditional Markov switching stochastic volatility model by allowing sequential forecasts, which are crucial for comparing competing models in financial econometric applications. We present two macroeconometric applications. In the first one we mainly investigate the sequential portfolio allocation problem amongst exchange rates for several countries. In the second one we study forms of financial contagion among emergent economies.

Bani Mallick
Texas A&M University

TITLE: "Stochastic simulation based methods for Bayesian curve and surface fitting" (joint work with Chris Holmes)

ABSTRACT:
This talk demonstrates how to apply the stochastic computation methods for curves and surface fitting problems. First we will exploit a Bayesian method of estimating a variety of curves by a sequence of piecewise polynomials. A joint distribution is set up over both the number and position of the knots and the coefficients of the polynomials. We use reversible jump Markov Chain Monte Carlo methods to compute the posterior distributions of the parameters in this varying dimension problem. The methodology has been successful in giving good estimates for "smooth" functions as well as functions which are not differentiable, and perhaps not even continuous at a finite number of points. The method will be extended for surface fitting problems with several variables. Next we will show how to apply the method of coupling from the past (CFTP) to this curve and surface fitting problem with orthogonal predictors. Specifically we will use wavelets and Demmler-Reinsch smoothing splines as our basis functions. Using CFTP we construct a coupled Gibbs sampler that automatically determines its own run time to convergence so removes the need to rely on approximate convergence diagnostics as in usual MCMC. A striking results in all of our examples is the rapid convergence of the Gibbs sampler. Possessing this knowledge allows us to repeatedly regenerate the sampler following known convergence and hence generate independent and identically distributed samples from the posterior modal space.

Dobrin Marchev
University of Florida

TITLE: "Analysis of van Dyk and Meng's Data Augmentation Algorithm for the Multivariate t Model"

ABSTRACT:
Let *pi* denote the posterior distribution that results when data from a multivariate location-scale Student's *t* distribution is combined with the standard non-informative prior. van Dyk and Meng (2001, JCGS) developed an efficient MCMC algorithm for sampling from *pi* and provided empirical evidence suggesting that their algorithm converges to stationarity much faster than the standard Gibbs sampler. In this paper, we formally analyze the Markov chain underlying Meng and van Dyk's algorithm. Specifically, we derive drift and minorization conditions and show how these conditions can be used to get rigorous bounds on the total variation distance to stationarity and to do regenerative simulation.

Robert McCulloch
University of Chicago

TITLE: "BMM: A Simple, General Algorithm for Drawing from Densities" (joint work with John Barnard
and Xiao-Li Meng)

ABSTRACT:
In this paper we propose a simple and general approach for drawing from a probability distribution represented by a density. We focus on drawing from a bivariate density, but in principle, the method could be used in any dimension. All that the method assumes is that we are able to evaluate the density. The method is based on the Metropolis algorithm. The Metropolis algorithm works by accepting candidate draws
in such a way that the stationary distribution of the resulting sequence of draws is the one we want. The key is our approach for generating candidate draws. We divide the support of the distribution into disjoint regions. Within each region, we evaluate the density at strategically chosen locations. One of the regions is stochastically chosen in such a way that a region where the density evaluations are larger is more likely to be chosen. We then iterate by partitioning the chosen region and repeating the process. We continue partitioning and choosing sub-regions until only a small region is left and then draw from this region (e.g uniformly). This draw is then used as the candidate in the Metropolis algorithm. The examples illustrate the use of the method to draw from bivariate densities. The examples show that the method is very robust. We draw from a highly correlated bivariate normal and a bimodal distribution with little dependence in the draws. KEY WORDS: random variate generation, simulation, Monte-Carlo, Metropolis algorithm, binary search, bivariate distribution

Ian McKeague
Florida State University

TITLE: "Signature recognition via MCMC"

ABSTRACT:
A Bayesian model for off-line signature recognition involving the representation of a signature through its curvature is developed. The prior model makes use of a spatial point process for specifying the knots in a spline approximation restricted to a buffer region close to a template curvature, along with an independent time warping mechanism. In this way, prior shape information about the form of the signature can be built into the analysis. The data model is based on additive Gaussian noise superimposed on the underlying curvature. The approach is implemented using MCMC and applied to a collection of documented instances of Shakespeare's signature.

Faheem Mitha
University of North Carolina

TITLE: "Some perfect sampling algorithms on continuous state spaces"

ABSTRACT:
The research on perfect sampling has been growing rapidly since the seminal work of Propp and Wilson. Our work focuses on applying various perfect sampling methods to the case of continuous distributions, often required in Bayesian inference problems. Specifically, we will study the randomness recycler originally proposed by Jim Fill and Mark Huber, and the multishift coupler proposed by David Wilson, along with Duncan Murdock's method of mixing with an independence sampler.

German Molina
Duke University

TITLE: "MCMC sampling under multiple constraints, with application to analysis of a traffic network"
(joint work with M.J. Bayarri and James O. Berger)

ABSTRACT:
Our application deals with the assessment and propagation of uncertainty in a complex computer traffic model. We approach the estimation/propagation of uncertainty in the data through a probabilistic network. We augment the parameter set with latent counts representing number of cars at each location in the system. Sampling the 200-dimensional space requires dealing with a posterior subject to 37 linear restrictions. These restrictions come from direct observation of some car counts and the physical structure of the network. We find a direct way to reparameterize the posterior in one simple step. The reparameterized posterior has an additional problem of iteration-dependent support for most full conditionals. We show a way to exactly compute the support in one single computation per iteration and full conditional. A third problem is to find non-negative, integer-valued starting values for 127 latent counts subject to the 37 restrictions. We identify sufficient conditions to avoid dealing with the Diophantine problem of finding starting values, providing a direct way to generate starting values. The output of our MCMC is used as input to the complex computer model. We provide a comparison of results before and after uncertainty propagation.

D.J. Murdoch
University of Western Ontario

TITLE: "Perfect Sampling for Non-Markovian Queues" (joint work with G. Takahara)

ABSTRACT:
We show how variations on Propp and Wilson's (1996) coupling from the past algorithm may be applied to give perfect draws from steady state distributions of queues with general arrival and service times. The methods are illustrated on a queue with bursty arrivals.

Duncan Murdoch
University of Western Ontario

TITLE: "On the perfect simulation of stochastic differential equations" (joint work with Xiaoqiang Li)

ABSTRACT:
Propp and Wilson (1996) proposed the idea of coupling from the past (CFTP) for the purpose of simulating exactly from the limiting distribution of a Markov chain. The multishift coupler (Wilson, 2000) is an important tool to apply the idea of CFTP on a continuous state space. In this paper, we describe a coupled Euler scheme which applies the multishift coupler to the usual Euler method of simulation of a stochastic differential equation. We apply our algorithm to achieve a pre-specified total variation distance from the exact distribution of the solution to constant diffusion SDEs. For non-constant diffusion SDEs, we can apply the It\^{o} formula to transform

into the constant diffusion SDEs. In addition, we describe ideas for extending our method to get a perfect sample.

Rui Paulo
National Institute of Statistical Sciences

TITLE: "McMC Analysis of Gaussian Spatial Processes Having Product Correlation Form"

ABSTRACT:
In the context of evaluating complex computer models, the problem of Bayesian analysis involving Gaussian spatial processes, with partially unknown mean and covariance structures, is encountered. To carry out needed McMC posterior computations, a special product form of the covariance function is considered. This, together with use of a design strategy that leads to at least some of the variables being observed in a Cartesian product fashion, allows for use of a Kronecker product formulation of the covariance matrix to considerably speed up the calculations involved in the sampling mechanism. For this problem, we devise and compare two McMC strategies, based on the ability to calculate the maximum likelihood estimates and the availability of a closed form expression for both the integrated likelihood and the Fisher information matrix.

Surajit Ray
Penn State University

TITLE: "Distance-based Model-Selection with application to Analysis of Gene Expression Data"

ABSTRACT:
Multivariate mixture models provide a convenient method of density estimation, model based clustering and explanations for the actual data generation process. But the problem of choosing the number of components (g) in a statistically meaningful way is still a subject of considerable research . Available methods for estimating g include optimizing AIC and BIC, estimating the number through nonparametric maximum likelihood, hypothesis testing and Bayesian approaches with entropy distances. In our current research we present several rules for selecting a finite mixture model, and hence g, based on estimation and inference using a quadratic distance measure. In one methodology the goal is to find the minimal number of components that are needed to adequately describe the true distribution based on a non parametric confidence set for the true distribution. We also present results for selecting g based on a risk analysis that includes a penalty for overfitting. Another less formal methodology is based on a measure analogous to R^2 in regression. Finally we fine tune our methods to analyze gene-expression data from micro-arrays, and compare them with other competitive methods.

Gabriel Rodriguez-Yam
Colorado State University

TITLE: "Efficient Gibbs Sampler implementations to the Constrained Normal Linear Regression Model"

ABSTRACT:
In this work we will consider various efficient implementations of the Gibbs Sampler to the problem of linear regression with normal errors and linear constraints on the regression

parameters. The frequentist approach to this problem has been widely studied (Judge and Takayama, 1966; Lovell and Prescott 1970; Liew 1976; Gallant and Gerig 1980; Gouriéroux et. al. 1982), but distribution theory on the estimates is difficult to handle analytically and only approximations are available. Related work from a Bayesian perspective is given in Geweke (1986), where expectations of functions relative to the posterior distribution are computed via Monte Carlo integration. Gelfand, et. al. (1992) suggest an approach to routinely analyze problems with constrained parameters using the Gibbs Sampler. In order to implement this procedure, it is necessary to draw a sample from a multivariate normal random vector X restricted to a subset R defined by a set of linear constraints. When the "unconstrained" probability of R is small, the direct naïve procedure of drawing from the un-constrained normal distribution until a value in R is obtained is impractical. The rejection method, also known as the accept-reject procedure, is a powerful technique for simulating data in the univariate case (Devroye, 1986; Robert and Casella, 1999), but in the general multivariate case the design of an efficient rejection method is more difficult (Devroye 1986, Johnson 1987). Alternatives of using the Gibbs Sampler to sample from a truncated multivariate normal distribution have been widely considered (Geweeke, 1991; Robert, 1994; Hahivassiliou, et. al., 1996; and Hahivassiliou and McFadden, 1997), but the problem of slow convergence and poor mixing may appear when some off-diagonal elements of the "unconstrained" correlation matrix of X are high (O'Hagan, 1994; Gilks and Roberts, 1995; Roberts, 1995). Thus, in order to implement the Gibbs Sampler to the constrained regression problem successfully, two problems need to be addressed. First, we need an efficient simulation from a constrained multivariate normal distribution, and second, the mixing of the resulting Markov chain used in the Gibbs sampler must be fast. This research addresses both issues.

Makram Talih
Yale University

TITLE: "Portfolio selection as covariance selection: time-varying graphical models"

ABSTRACT:
We interpret the classical portfolio selection problem as a covariance selection problem. Individual stock returns, or, by aggregation, industry portfolio returns can thus be jointly modeled as graphical Gaussian models. To account for the changing covariance structure in such data, we allow the edges of the underlying graphs to vary, albeit slowly, over time. Using well-established algorithmic criteria for edge deletion and addition in decomposable graphs (Giudici and Green, 1999), a carefully designed proposal strategy allows for a relatively expeditious exploration of the state space.

Simon Tavare
University of Southern California

TITLE: "Approximate Bayesian computation in population genetics"

ABSTRACT:
I will review an interesting and practical approach to approximate Bayesian computation using rejection methods, and illustrate how the method can be used for the analysis of single nucleotide polymorphism data and restriction fragment length polymorphism data.

Len Thomas

University of St Andrews

TITLE: "Application of Particle Filtering to State-space Models of Wildlife Population Dynamics"
(joint work with Stephen T. Buckland, Ken B. Newman, and John Harwood)

ABSTRACT:
Wildlife populations are highly structured stochastic systems, about which there is usually incomplete information. Nevertheless, it is often important to make inferences about the system - for instance when the population is of economic or conservation concern. We propose a unified framework for defining and fitting stochastic discrete time, discrete stage models of wildlife population dynamics. The biological system is described by a state-space model, where the true but unknown state of the population is modelled in a state process, and this is linked to survey data by an observation process. All sources of uncertainty in the inputs, including uncertainty about model specification, are readily incorporated. By dividing the state process into sub-processes, complex models can be easily constructed from manageable building blocks. We illustrate the approach with a model of the British grey seal metapopulation. We use the auxiliary particle filter algorithm of Lui and West to fit the model and calculate smoothed estimates of the states and model parameters. Several technical issues remain, including setting appropriate priors on states and parameters, specifying an appropriate model for the observation error, local kernel smoothing of parameter estimates and methods for further reducing particle depletion.

Didier Verloo
Co-ordination Centre for Veterinary Diagnostics

TITLE: "MCMC techniques in veterinary diagnostic test validation"

ABSTRACT:
Reliable estimates of the diagnostic sensitivity and specificity of tests are needed in many settings. These estimates are needed by veterinary practicioners to update clinical inferences, planners of surveillance programs for sample-size calculations, epidemiologists to adjust prevalence estimates, odds ratios and other parameters for misclassification and by risk analysts for complete specification of probabilities in their scenario pathways. Although modern biotechnology has allowed major breakthroughs, however, diagnostic test validation has been frequently hindered by the lack of a gold standard, i.e. an exact knowledge of the true status of the tested individuals. In that way, new tests have to be compared with imperfect existing ones, and the performance of a new test can seriously be under- or overestimated in this way leading to wrong decisions in many settings. Therefore, epidemiologists and biostatisticians have shown an increasing interest
in Latent Class Analysis (LCA) which models associations between observed categorical variables by assuming that a non-observed (latent) variable is determining these associations. In diagnostic test validation, the true disease status of an individual can be considered as a dichotomous latent variable with two categories, diseased and not diseased. The deterministic approach to solve those models are based on maximum likelihood estimations of the conditional probabilities but suffers from local maxima/minima, sparse data or unidentifiabelity of some models (lack of degrees of freedom df). Also the modelling of conditional dependence between tests using covariates (fixed effects) or random effects consumes df and increases the mathematical difficulty. Bayesian inference using MCMC techniques can bring solutions where the deterministic methods fail or become to complicated. This will be demonstrated with different validation projects for diagnostics of veterinary diseases. Next to this these MCMC techniques

provide full probability distributions of the posteriors which makes it straightforward to implement for further quantitative risk analysis based on Monte Carlo simulation.

Yazhen Wang
University of Connecticut

TITLE: "Variance Reduction Methods in Monte Carlo Simulations for Stochastic Differential Equation"

ABSTRACT:
Continuous-time models play a central role in the modern finance theory. These models are usually defined through stochastic processes governed by stochastic differential or partial differential equations. Derivative pricing based on the models is usually very complicated and lacks of explicit formulas. Numerical methods such as Monte Carlo simulations are resorted to evaluate the pricing. This talk will present some results in using wavelet like bases to reduce variances in Monte Carlo and Quasi Monte Carlo based evaluation of the derivative pricing for stochastic differential equation models.

Patrick Wolfe
University of Cambridge

TITLE: "Efficient Stochastic Computation for Audio Signal Processing"

ABSTRACT:
Bayesian hierarchical models provide a natural and effective means of exploiting prior knowledge concerning the time-frequency structure of sound signals such as speech and music---something which has often been overlooked in traditional approaches to audio applications such as speech enhancement, signal compression, and music modelling. Having constructed a Bayesian model and prior distributions capable of taking into account the time-frequency characteristics of typical audio waveforms, we focus here on efficient methods of obtaining samples from, and/or point estimates of, the resultant posterior.

Using straightforward Gibbs sampling schemes, we have thus far been able to formulate noise removal methods capable of superior performance, both objectively and subjectively, in comparison with standard techniques. Given the potential of such methods in critical applications such as mobile communications and hearing aid design, we aim next to develop algorithms operating in near real-time, using sequential block-based processing with low latency. However, the bandwidth even of low-fidelity signals necessitates a sampling rate of several kilohertz; thus, the size of typical audio datasets renders standard MCMC methods infeasible in real-world applications, and presents a unique challenge in stochastic computation.

D. *Program & Guest Speaker Abstracts for Mid-Term Workday for Contingency Tables*
January 23, 2003, NISS-SAMSI Building

10:00-10:45 "Monte Carlo Algorithms for Approximating Exact Conditional Probabilities"
Brian Caffo, Johns Hopkins University

11:00-11:30 "Bayesian Inference in Incomplete Multi-way Tables"

	Adrian Dobra, Duke University
11:40-12:10	"Exact sampling for Hardy-Weinberg Equilibrium" Mark Huber, Duke University
12:10-1:30	Lunch
1:30-2:00	"Markov chains for Hardy-Weinberg Equilibrium" Ian Dinwoodie, Duke University
2:10-2:40	"Sequential Importance Sampling for Hardy-Weinberg Equilibrium" Yuguo Chen, Duke University
3:00-3:50	"Latte: Lattice point enumeration and applications to multiway contingency tables" Ruriko Yoshida, University of California at Davis
4:00-4:15	Further Problems Yuguo Chen, Ian Dinwoodie, Adrian Dobra, Mark Huber, Michael Nicholas

Brian Caffo
Johns Hopkins University

TITLE: "Monte Carlo Algorithms For Approximating Exact Conditional Probabilities"

ABSTRACT:
Conditional inference eliminates nuisance parameters by conditioning on their sufficient statistics. For contingency tables conditional inference entails enumerating all tables with the same sufficient statistics as the observed data. For moderately sized tables and/or complex models the computing time to enumerate these tables is often prohibitive. Monte Carlo approximations offer a viable alternative provided it is possible to obtain samples from the correct conditional distribution. This talk presents a Markov Chain Monte Carlo solution algorithm based on the Metropolis/Hatings/Green algorithm with a sequentially generated rounded normal candidate. The talk concludes with a discussion of general software implementation of these methods.

Rudy Yoshida
University of California, Davis

TITLE: "LattE: Lattice point enumeration and applications to multiway contingency tables"

ABSTRACT:
We are presenting polyhedral-algebraic algorithms to count the number of contingency tables with fixed given marginals. We will describe our program LattE, the first implementation which applies Barvinok's cone decomposition and the symbolic algebra of power series to the algorithms. Barvinok's cone decomposition is the algorithm that can decompose any cone into unimodular cones. If we fix the dimension, Barvinok's cone decomposition runs in polynomial time. After we

decompose each vertex cone of a given polytope into unimodular cones, we can write the formal power series and count the number of lattice points inside the given polytope.

This decomposition can be used to write Hilbert-function like generating functions counting lattice points at dilations of the polytope (Ehrhart polynomials).

E. *Program for Mid-Term Workday for Model Selection*
January 30, 2003, NISS-SAMSI Building

10:00-10:50 "Empirical Bayes versus Fully Bayes Variable Selection"
Ed George, University of Pennsylvania

10:50-11:15 Discussion, followed by Coffee Break and Informal Discussion

11:15-11:45 "Zellner-Siow Priors and Variable Selection"
Rui Paulo, NISS and Duke University

11:45-12:15 "Hypergeometric G-priors for Variable Selection"
Feng Liang, Duke University

12:15-12:30 Discussion

12:30-1:30 Lunch

1:30-2:00 "Novel Search Strategies for Variable Selection"
Merlise Clyde, Duke University

2:00-2:15 Discussion

2:15-2:45 "Perfect Sampling in Model Selection"
Mark Huber, Duke University

2:45-3:10 Discussion and Break

3:10-4:00 "Bayesian Computation: From Marginal Posterior Densities to Marginal Likelihoods"
Ming-Hui Chen, University of Connecticut

4:00-4:15 Discussion

4:15-4:30 Future Directions

F. *Program for Mid-term Workday for Graphical Models*
February 13, 2003, NISS-SAMSI Building

9:00-9:15 Arrival and introductions

9:15-9:30 Introductory remarks

Mike West, Duke & SAMSI

9:30-10:30 "Gaussian decomposable models in statistics: Background and modelling aspects"
Chris Carter, SAMSI-Duke University Fellow

10:30-11:30 "MCMC approaches and model specification"
Beatrix Jones, SAMSI Research Fellow

11:30-12:00 Coffee and conversation

12:00-1:00 "Decoding gene expression control using generalized gamma networks"
Paola Sebastiani, University of Massachusetts

1:00-2:15 Lunch and informal discussions

2:15-3:15 "Stochastic search and optimization approaches"
Adrian Dobra, Duke & SAMSI Research Fellow

3:15-4:15 "Approaches via dependency networks"
Chris Hans, SAMSI-Duke Graduate Research Assistant

4:15-5:00 Tea and informal discussions

G. *Program and Speaker Abstracts for Mid-Term Workday for Financial Modeling*
February 20, 2003, NISS-SAMSI Building

10:00-11:00 "MCMC methods for financial econometrics"
Nick Polson, University of Chicago

11:00-11:30 "Sequential Monte Carlo methods in stochastic volatility models"
Beom Lee, University of North Carolina

11:30-12:00 Discussion

12:00-1:30 Lunch

1:30-2:00 "Short time-scale in stochastic volatility and option pricing"
Sean Han, North Carolina State University

2:00-2:30 "Bayesian estimation of Stochastic volatility models with long-term and short-term volatility components"
German Molina, Duke University

2:30-3:00 Discussion

3:00-3:15 Break

3:15-4:15 "Particle Filtering and Applications"
Rene Carmona, Princeton University

4:15 Discussion

Rene Carmona
Princeton University

TITLE: "Particle Filtering and Applications"

ABSTRACT:
The purpose of the talk is to present new theoretical results on the asymptotic properties of the optimal filter of a Hidden Markov Model, and to report on numerical experiments with the particle filters. We shall consider engineering applications such as in-car intelligent navigation systems, and financial applications such as volatility tracking and fixed income affine model fitting.

Sean Han
North Carolina State University

TITLE: "Short time-scale in stochastic volatility and option pricing"

ABSTRACT:
The presence of a short time-scale in S&P 500 can be identified by use of the empirical structure function, or variogram, of the high-frequency log absolute returns. We show that a well-separated longer time-scale can be ignored under the pricing measure but not for short time-scale. Option pricing problem under fast-scale volatility is dealt by asymptotic analysis such that the approximated option price takes into account the skew of implied volatility.

Beom Lee
University of North Carolina

TITLE: "Sequential Monte Carlo methods in stochastic volatility models"

ABSTRACT:
This talk provides an introduction to calibration of stochastic volatility models via MCMC methods. Sequential importance sampling is applied to modify the traditional MCMC for generating latent volatility series. Two simulation dynamics are involved: the asset price (or return) process governed by a real-world probability measure, and the option data related to a risk-neutral probability measure.

German Molina
Duke University

TITLE: "Bayesian estimation of Stochastic volatility models with long-term and short-term volatility components"

ABSTRACT:
In this talk we propose a multifactor model as an approach to the estimation of different time scales. We describe the multiple move algorithm (Carter & Kohn, 1994, Shephard, 1994) that will allow us to sample the T by K matrix of volatilities at once, where K is the (fixed) number of

factors and T the number of univariate observations. The existence of fast and slow time scales imposes additional restrictions that will be added as constraints in our MCMC. We compare the results we obtain when estimating 2-factor simulated data with a 1-factor model versus a 2-factor model, when those factors are associated with different time scales. We will discuss several problems that arise when estimating this model and outline possible future areas of research.

Nick Polson
University of Chicago

TITLE: "MCMC methods for financial econometrics"

ABSTRACT:
This chapter discusses MCMC based methods for estimating continuous-time asset pricing models. We describe the Bayesian approach to empirical asset pricing, the mechanics of MCMC algorithms and the strong theoretical underpinnings of MCMC algorithms. We provide a tutorial on building MCMC algorithms and show how to estimate equity price models with factors such as stochastic expected returns, stochastic volatility and jumps, multi-factor term-structure models with stochastic volatility, time-varying central tendency or jumps and regime switching models.

H. Tentative Program for Closing Workshop
June 26-28, 2003

- **Thursday, June 26**: Radisson Governor's Inn

8:00 Continental Breakfast

8:15-8:45 Registration Check-In

8:45-9:00 Welcome and Introduction

Session 1 - Graphical Models A
9:00-10:30 Chair: Chris Hans, Duke University & SAMSI

"MCMC for Bayesian data analysis"
Paulo Giudici, University of Pavia

"Graphical Gaussian model selection in a Bayesian framework"
Helene Massam, York University

10:30-11:00 Coffee break

Session 2 - Graphical Models B
11:00-12:30 Chair: Carlos Carvalho, Duke University

"MCMC and stochastic search approaches in Gaussian graphical models"
Beatrix Jones, SAMSI and Duke University

"Compositional regressions, DAGs and fitting high-dimensional Gaussian graphical models"
Adrian Dobra, Duke University and SAMSI

12:30-1:30 Lunch

<u>Session 3</u>
1:30-3:00 Chair: Chris Carter, Duke University & SAMSI

"MCMC exact p-values and some applications"
Julian Besag, University of Washington

"Stochastic simulation methods for the number of components in a mixture"
Sujit Sahu, University of Southhamption

3:00-3:30 Tea Break

<u>**Session 4 -** Model Selection A</u>
3:30-5:00 Chair: Joe Ibrahim, University of North Carolina

"Mixtures of g-priors and variable selection"
Feng Liang, Duke University

"Marginal likelihoods and Bayes factors"
Rui Paulo, SAMSI

- **Friday, June 27**: Radisson Governor's Inn

8:00 Continental Breakfast

8:30-9:00 Registration Check-In

<u>**Session 5 -** Model Selection B</u>
9:00-10:30 Chair: Susie Bayarri, University of Valencia

"Nonparametric regression variable selection"
Helen Zhang, North Carolina State University

"Model search"
Merlise Clyde, Duke University

10:30-11:00 Coffee break

<u>**Session 6 -** Contingency Tables & Related Topics</u>
11:00-12:30 Chair: Ian Dinwoodie, Tulane University & Duke SAMSI Fellow

"Markov chain moves for generating contingency tables with fixed weighted row sums"
Mark Huber, Duke University

"Sequential importance sampling for generating contingency tables with fixed weighted row sums"
Yuguo Chen, Duke University

12:30-1:30 Lunch

<u>**Session 7 -** Contingency Tables & Related Topics A</u>

1:30-3:00 Chair: Mark Huber, Duke University

"Making inferences from arbitrary sets of conditionals and marginals for contingency tables"
Aleksandra Slavkovic, Carnegie Mellon University

"Algebraic geometry of Bayesian networks with hidden variables"
Luis David Garcia, Virginia Tech

3:00-3:30 Tea Break

Session 8 - Contingency Tables & Related Topics B
3:30-5:00 Chair: Michael Nicholas, Duke University

"Variations on Barvinok's counting algorithm and applications to multiway contingency tables"
Ruriko (Rudi) Yoshida, University of California, Davis

"Binary graph models"
Seth Sullivant, University of California, Berkeley

- **Saturday, June 28**: SAMSI-NISS Building

8:00 Continental Breakfast

Session 9 - Financial Models A
9:00-10:30 Chair: Yuguo Chen, Duke University

title tba
Chuansu Ji, University of North Carolina

"Multi-scale stochastic volatility"
Jean-Pierre Fouque, North Carolina State University

10:30-11:00 Coffee break

Session 10 - Financial Models B
11:00-12:30 Chair: Chuansu Ji, University of North Carolina

"MCMC to estimate multi-scale stochastic volatility models"
German Molina, Duke University

"Pricing Asian options and variance swaps with volatility scales"
Sean Han, North Carolina State University

12:30-1:30 Lunch

Session 11 - Financial Models and Time Series
1:30-3:00 Chair: Jean-Pierre Fouque, North Carolina State University

title tba
Tao Pang, North Carolina State University

"Bayesian analysis of random coefficient autoregressive models"
Sujit Ghosh, North Carolina State University

3:00-3:30 Tea Break & Close

III. LARGE-SCALE COMPUTER MODELS FOR ENVIRONMENTAL SYSTEMS

A. Program for the Workshop on Multi-Scale Modeling
February 2-7, 2003, Radisson Governor's Inn

- **Sunday, February 2**

8:00-8:30 Registration

8:30-10:00 "An Introduction to Porous Medium Systems"

Cass T. Miller, University of North Carolina

10:00-10:30 Coffee Break

10:30-12:00 "Physical Parameterization Techniques in Large-Scale Atmospheric Models"
James Hack, National Center for Atmospheric Research

12:00-1:30 Lunch

1:30-3:00 "An Introduction to Methods of Homogenized Averaging for Problems with Multiple Scales"
Richard McLaughlin, University of North Carolina

3:00-3:30 Coffee Break

3:30-5:00 "Some Multiscale Approaches in Statistical Computation and Modeling"
David Higdon, Los Alamos National Laboratory

- **Monday, February 3**

8:00-8:30 Registration

8:30-8:45 Opening Remarks
James Berger, Director of SAMSI
Richard Smith--Program Leader, University of North Carolina

SESSION I
8:45-10:15

"Breeding, Data Assimilation and Probability of Atmospheric-Ocean Models"

Eugenia Kalnay, University of Maryland

10:15-10:45 Coffee Break

SESSION II
10:45-12:15

"Inference for Misaligned Spatial Data Layers: The "Change of Support" and "Modifiable Areal Unit" Problem"
Alan Gelfand, Duke University

"Anthropogenic Carbon in the Ocean Inferred from Tracer Measurements"
Timothy Hall, National Aeronautics and Space Administration

12:15-1:30 Lunch

SESSION III
1:30-3:00

"Practical Considerations in the Use of Averaging Procedures in Porous Media Flow Studies"
William Gray, University of Notre Dame

3:00-3:30 Coffee Break

SESSION IV
3:30-5:00

"Density-Coupled Transport in Heterogeneous Aquifers"
Claire Welty, Drexel University

"Boundary Conditions Between Porous Media and Bulk Fluid"
Lynn Bennethum, University of Colorado-Denver

8:00-10:00 **Poster Session and Reception**
NISS-SAMSI Building
Transportation from Radisson Governor's Inn will be provided by Carolina Livery

- **Tuesday, February 4**

8:00-8:30 Registration

SESSION I
8:30-10:00

"Statistical Models for Models"
Doug Nychka, National Center for Atmospheric Research

10:00-10:30 Coffee Break

<u>SESSION II</u>
10:30-12:00

"Physical-Statistical Modeling"
Mark Berliner, Ohio State University

"Hierarchical Bayesian Modeling of Multiscale Spatio-Temporal Processes"
Chris Wikle, University of Missouri

12:00-1:15 Lunch

<u>SESSION III</u>
1:15-2:45

"Effective Transport by a Spatiotemporal Mean Flow with Small-Scale Periodic Fluctuations"
Peter R. Kramer, Rensselaer Polytechnic Institute

"Inertial Particles in a Random Field"
Andrew Stuart, University of Warwick

2:45-3:15 Coffee Break

3:15-4:15 Breakout Discussion Groups

4:15-4:30 Break

4:30 **SAMSI Distinguished Lecture:** Andrew Majda, New York University
"Mathematical Strategies for Stochastic Modeling in Climate"

- **Wednesday, February 5**

8:00-8:30 Registration

<u>SESSION I</u>
8:30-10:00

"Aggregating Clouds"
Bjorn Stevens, University of California-Los Angeles

"Determining Sources and Sinks of Carbon Dioxide from Atmospheric Concentration Measurements: A Multiscale Inverse Problem"
Tapio Schneider, California Institute of Technology

10:00-10:30 Coffee Break

<u>SESSION II</u>
10:30-12:00

"Coarse-grained Stochastic Processes and Monte Carlo Simulations in Lattice Systems"

Markos Katsoulakis, University of Massachusetts

"Multi-scale Dynamical Systems with Stochastic Effects: Mathematical Aspects and Numerical Techniques"
Eric Vanden Eijnden, New York University

12:00-1:30 Lunch

SESSION III
1:30-3:00

"Statistical Assessment of Air Quality Numerical Models"
Montserrat Fuentes, North Carolina State University

"Nonstationary Spatial Covariance Modeling Via Spatial Deformation and Extensions to Covariates and Models for Dynamic Atmospheric Processes"
Paul D. Sampson, University of Washington

- **Thursday, February 6**

SESSION I
8:30-10:00

"Turbulent Scalar Statistics and Application to Reactive Scalar Modeling"
Anne Bourlioux, University of Montreal

"Convection of Miscible Binary Mixtures in Porous Media"
R.P. Behringer, Duke University

10:00-10:30 Coffee Break

SESSION II
10:30-12:00

"Field Experiments and Observations of Subgrid-Scale Processes in the Atmospheric Surface Layer"
Marc Parlange, Johns Hopkins University

"Pushing the State of the Art in Ocean Modeling and Data Assimilation: Intrinsic Variability of Ocean Currents"
Robert Miller, Centre National de Recherches Meteorologiques and Oregon State University

12:00-1:30 Lunch

SESSION III
1:30-3:00

"Upscaling and Hysteresis in Models of Soil Moisture, Evaporation and Transportation"
Mike Celia, Princeton University

"Non-Equilibrium Effects in the Capillary Pressure-Saturation Relationship for Two-Phase Flow in Porous Media"
Majid Hassanizadeh, Delft University of Technology

3:00-3:30 Coffee Break

<u>SESSION IV</u>
3:30-5:00

"Long Tails in Probability Distributions in Passive Scalar Transport"
Jared Bronski, University of Illinois at Urbana-Champaign

"Interaction of Waves and Currents"
Juan Restrepo, University of Arizona

- **Friday, February 7**

<u>SESSION I</u>
8:30-10:00

"Challenges in Parameterizations for Global Circulation Models of the Atmosphere"
Joe Tribbia, University Corporation for Atmospheric Research

"A Bayesian Analysis of Regional Climate Change Projections, Based on a Multi-Model Ensemble"
Claudia Tebaldi, University Corporation for Atmospheric Research

10:00-10:30 Coffee Break

<u>SESSION II</u>
10:30-12:00

"Forecasting and Hindcasting in Multiscale, Nonlinear Systems"
Gregory Eyink, Johns Hopkins University

"Flow, Transport and Reaction in Porous Media: Upscaling from Pore-Network Models"
Yannis Yortsos, University of Southern California

B. Workshop on Multi-Scale Modeling Poster Session
February 3, 2003, NISS-SAMSI Building

Prasheen Kumar Agarwal, North Carolina State University
"Bootstrapping of Spatially Correlated Data"

Li Chen, North Carolina State University
"Bayesian Hierarchical Spatio-Temporal Models for Wind Prediction"

Britt S. B. Christensen, Technical University of Denmark
"Fluid Interfaces at the Pore-Scale: Trying to Make Observations and Lattice Boltzmann Simulations Meet"

Katherine A. Culligan, Notre Dame University
"Interfacial Area Measurements for Unsaturated Flow Through a Porous Medium"

Marco A. R. Ferreira, Universidade Federal do Rio de Janeiro
"A New Class of Multi-scale Random Fields with Applications to the Estimation of Permeability Fields"

Peter L. Finkelstein, ***US Environmental Protection Agency***
"Observations of the Integral Form of the Similarity Theory Stability Correction Terms for Momentum and Temperature (Qsub_m and Q sub_T) over Agricultural Fields and Forests"

Dargan Frierson, ***Princeton University***
"A Simple Model of Tropical Precipitation Response to Sea Surface Temperature"

Edwin Gerber, Princeton University
"A Mechanism and Simple Model of the North Atlantic Oscillation and Annular Modes"

Leonid Kuznetsov, University of North Carolina
"A Method for Assimilation of Lagrangian Data"

Richard McLaughlin, University of North Carolina
"An Internal Splash: Levitation of Falling Particles in Stratified Fluids"

Doris Pan, ***University of North Carolina***
"Lattice-Boltzmann Simulation of Immiscible Two-fluid Flow in Porous Media"

Joseph A. Pedit, University of North Carolina
"Modeling Sorption of Diuron and Lindane by an Aquifer Sand"

C. Workshop on Multi-Scale Modeling Speaker Abstracts

R. P. Behringer
Duke University

TITLE: "Convection of Miscible Binary Mixtures in Porous Media"

ABSTRACT:
We have carried out extensive studies of convection and enhanced diffusion of miscible binary mixtures that are confined to a thin porous layer. These studies have been carried out using two novel visualization techniques, one involving Magnetic Resonance Imaging (MRI) and the other involving novel constructions of porous media that allow conventional shadowgraph visualization. Throughout these measurements, we use mixtures of water and ethanol. Key control parameters for binary-fluid porous media convection (BF-PMC) include the Rayleigh number, R, and the separation ratio. Linear theory predicts that for negative separation ratios, there will be a Hopf bifurcation to oscillatory convection rolls with increasing R. However, the experiments show a very different behavior. Rather, there is a hysteretic transition at Rayleigh numbers well above the expected Hopf value to a state of steady convection rolls. One possible explanation may be that once convection begins, there is an effective stirring of the fluid due to enhanced diffusion. To investigate this possibility, we have carried out theoretical studies and

also direct measurements of mixing using NMR techniques. Concerning the former, we have constructed an eight-mode Lorenz-like truncation of the full equations (assuming Darcy's law) of motion that also includes a velocity-dependent diffusive mixing term. Using NMR, we have measured the amplitude of this term directly for one of the media used in the convection experiments. From the model, we find that the experiments can indeed be explained if the mixing amplitude is large enough. However, the strength of the mixing amplitude that is needed is over two orders of magnitude too large compared to the measured value. Although a mixing effect may still play a key role in BF-PMC, a simple enhanced diffusion model does not correctly describe the experiments.

Lynn S. Bennethum,
University of Colorado at Denver

TITLE: "Boundary Conditions between Porous Media and Bulk Fluid"

ABSTRACT:
The boundary conditions between two fluids or two phases at the microscale are well understood and commonly accepted. However at the macroscale, the boundary conditions are varied and often mis-applied. In this talk we discuss what has been learned about boundary conditions from hybrid mixture theory, a rigorous, methodical, upscaling approach.

Hybrid mixture theory, pioneered by W. Gray and M. Hassanizadeh, consists of classical mixture theory applied to a multi-phase system with volume averaged field equations. Averaging each field equation (e.g. conservation of mass, momentum balance, energy balance, and entropy inequality) produces macroscopic variables which are precisely defined in terms of their microscopic counterparts. Constitutive restrictions are obtained by exploiting the entropy inequality using the Coleman and Noll method. Treating the bulk phase fluid at the same spatial scale as the homogenized porous medium produces the boundary conditions to be discussed.

Mark Berliner
Ohio State University

TITLE: "Physical-Statistical Modeling"

ABSTRACT:
I discuss the hierarchical Bayesian framework for analyzing physical problems. This paradigm provides opportunities for the combination of physical reasoning and observational data in a coherent analysis framework, but in a fashion which manages the uncertainties in both information sources. A key to the hierarchical viewpoint is that separate statistical models are developed for the process variables studied and the observations conditional on those variables. Modeling process variables in this way enables incorporation of physics across a spectrum of levels of intensity ranging from qualitative use of physical reasoning to strong reliance on numerical models. Selected examples from this spectrum are reviewed. A simple example involving modeling at different spatial scales is discussed.

Anne Bourlioux
Universite de Montreal

TITLE: "Turbulent Scalar Statistics and Application to Reactive Scalar Modeling"

ABSTRACT:
The PDF and various conditional statistics of a passive scalar in a turbulent flow are interesting for their own sake and also play a key role in modeling reactive scalars (for example, in models for combustion, pollutant transport, bio-organisms population dynamics, etc). Many models will make assumptions about the behavior of those relevant statistics. In this talk, I will present results obtained with an idealized set-up that allows for very clean numerical or analytical processing of the data to be used for model validation and development. Despite its simplicity, the set-up still leads to a non-trivial intermittent dynamics, very similar to what is observed in some experiments. I will contrast the reliability and efficiency of this approach with the more traditional numerical route to validating the models for the statistics of the scalar, that consists in generating huge databases of numerical solutions for the appropriate partial differential equations and then process those discrete solutions to extract the relevant statistics for the scalar, for comparison with the presumed model.

Jared Bronski
University of Illinois

TITLE: "Long Tails in Probability Distributions in Passive Scalar Transport"

ABSTRACT:
We present some work (joint with Rich McLaughlin) on rigorous estimates for the probability distributions in a model of passive scalar transport originally due to Majda. We will also discuss some work on a generalization of this model by Vanden Eijnden. We also attempt to put these mathematical results in a physical context.

Michael A. Celia
Princeton University

TITLE: "Upscaling and Hysteresis in Models of Soil Moisture, Evaporation, and Transpiration"

ABSTRACT:
Soil moisture dynamics are central to vegetation growth, ground-water recharge, and land surface-atmosphere interactions. While the basic equations of two-phase (air-water) flow may be applied to this system, the appropriate spatial scale over which to define averaged quantities is not always obvious. In some cases, detailed simulations using multi-dimensional flow equations are used, while in other cases simplified, spatially averaged models are used. A computational study involving upscaling from the highly spatially-resolved scale to a spatially averaged scale covering the entire root zone provides insights
into how upscaled models relate to highly-resolved models. Analysis of computational results indicates that dimensionless groups can provide guidelines for conditions under which certain upscaled models may be appropriate. In addition, computational results indicate that upscaled evaporation and transpiration functions exhibit hysteresis, despite having no hysteresis at the small scale. This observation leads to the conjecture that hysteresis is caused by upscaling.

Greg Eyink
John Hopkins University

TITLE: “Forecasting and Hindcasting in Multiscale, Nonlinear Systems”

ABSTRACT:
The utility of large-scale models of the atmosphere, ocean and groundwater flow are considerably enhanced by combining them with observational datasets, using so-called "data assimilation" methods. The most widely accepted mathematical framework for doing so is the optimal estimation theory based upon calculation of conditional statistics of model states or parameters given information from the measurements. Such statistical methods allow one both to assess model uncertainty and to validate models. Exact, rigorous methods have been known since the 1960's for calculating such conditional statistics for nonlinear stochastic dynamics. For Markov processes, recursive-in-time estimation algorithms consist of two steps: evolution of statistics between measurements by the dynamical equations, and updating of distributions at measurement times by Bayes rule. However, these methods are not applicable without approximation to geophysical systems with a wide range of excited scales and with irreducibly many degrees-of-freedom. We'll discuss here some practical approximation schemes to determine the conditional statistics in multiscale systems. In particular, we'll consider two approximate methods to evolve statistics between measurements, moment-closure and ensemble (or "particle") methods. We'll also consider two approximations of the Bayes rule update, a least-squares linear interpolation or Kalman scheme and a mean-field conditional analysis. The methods shall be illustrated and compared in some simple models of climate transitions and of contaminant transport by random advection.

Montserrat Fuentes
North Carolina State University

TITLE: “Statistical Assessment of Air Quality Numerical Models”

ABSTRACT:
Evaluation of physical models for air quality applications is crucial to assist in control strategy selection. The high risk of getting the wrong control strategy has costly economic and social disbenefits, which will be increasingly important for the proposed new multi-pollutant standards and combined controls strategies. Here we develop a formal method for evaluation of numerical models. We specify a simple model for both, the output of the air quality models and the observations in the ground, in terms of the unobserved ground truth, and estimate the model in a Bayesian way. This yields solutions to the model evaluation and bias removal problems simultaneously. It provides improved evaluation of air quality numerical models via the posterior predictive distribution of the ground observations, and enables us to remove the bias in the numerical models output by estimating additive and multiplicative bias parameters in the model.

In our approach we take into account the lack of stationarity of the air pollution process. We present a new model for non-stationary spatial processes that is based on the concept of spatial spectra, this means spectral functions which are space dependent. We propose a nonstationary nonparametric and various parametric approaches to estimate the spectral density of a nonstationary spatial process. We

apply our methods to data on SO_2 concentrations.

Alan E. Gelfand
Duke University

TITLE: "Inference for Misaligned Spatial Data Layers: The "Change of Support" and
"Modifiable Areal Unit" Problems"

ABSTRACT:
An established problem in working with spatial data is the matter of having acquired spatial data at one scale of resolution and seeking to infer about it a different scale. The so-called change of support problem is usually concerned with data obtained at point-referenced level with interest in inference at block/areal unit level or perhaps vice versa. The modifiable areal unit problem is concerned with data obtained for one set of areal units with interest in inference about what would be expected for an alternative set of units. With increased collection of spatial data, a related problem arises. One routinely finds spatial data layers which are obtained at different scales of resolution with interest in building regression models to enable the use of one layer to explain another. A variety of ad hoc algorithmic approaches can be proposed to address these two problems. However, they may not be satisfying from an inferential perspective. In this presentation we will review some of these ad hoc procedures but the focus will be on fully model-based methods. Through a range of environmental and ecological examples we will describe how stochastic modeling can be brought to these problems, how we can fit these models and what sorts of inference can be extracted.

William G. Gray

University of Notre Dame

TITLE: "Practical Considerations in the Use of Averaging Procedures in Porous Media Flow Studies"

ABSTRACT:
Averaging procedures are commonly applied to conservation equations to change the scale of analysis from the microscale, or pore scale, to a larger macroscale representative of a section of the porous medium. The procedure leading to the macroscale equations involves the following steps: i) Develop microscale mass, momentum, energy, and entropy equations for phases, interfaces between phases, and common lines where the interfaces come together; ii) Average the equations to obtain their macroscale counterparts, including terms for which constitutive forms are needed; iii) Constrain the entropy equation with the conservation equations and the macroscale thermodynamic relations; iv) Obtain constitutive expressions as needed to close the system of equations.
Although these steps have been applied with sufficient frequency that their elements are well understood, in fact each step presents a continuing set of challenges for those trying to ensure that mathematical elegance strongly intersects with the physical configuration of a system under study. In this presentation, the mechanics of averaging theory will be explored along with some

of the specific conceptualizations and aspects of multiphase flow in porous media that require continuing creative investigation.

James Hack
National Center for Atmospheric Research

TITLE: "Physical Parameterization Techniques in Large-Scale Atmospheric Models"

No Abstract Available

Timothy M. Hall
National Aeronautics and Space Administration

TITLE: "Anthropogenic Carbon in the Ocean Inferred from Tracer Measurements"

ABSTRACT:
The ocean is a major buffer for increases in atmospheric CO_2 caused by anthropogenic activity, but there remains large uncertainty on the rate of ocean uptake and the amount of anthropogenic carbon (?C) presently stored. Ocean ?C cannot be directly measured, but considerable effort has gone into indirect estimates. In one technique tracers with little or no natural background are used as proxies for ?C. Unfortunately, no tracer is a perfect proxy, and it has not been clear how best to apply the information from ocean tracers to ?C.

Here, we present and apply a new technique to estimate ?C concentrations from tracer measurements. Tracers are used to constrain Green's functions from the ocean surface to interior points, and the Green's functions are applied to the history of surface-water ?C. Compared to previous techniques the Green's function technique makes no assumptions on the role of mixing versus bulk advection, and as a result is less biased. The technique is physically compelling because the Green's function can be interpreted as a distribution of transit times ("ages") since interior water masses made last surface contact, thereby generalizing the widespread use of "age" in ocean transport studies. Using CFC-12 as a tracer, we estimate that the Indian Ocean contains 8 to 11 Gt of ?C, compared to a previous estimate of 14 Gt.

Majid Hassanizadeh
Delft University of Technology

In collaboration with Michael A. Celia, Princeton University

TITLE: "Non-Equilibrium Effects in the Capillary Pressure-Saturation Relationship for Two-Phase Flow in Porous Media"

ABSTRACT:
Capillarity plays a central role in the description of multiphase and unsaturated flows in porous media. In quantitative modeling of multiphase flow, a relationship is needed to describe capillary pressure as a function of other medium properties. Although the underlying processes that

determine the distribution of fluid phases in porous media are extremely complicated, the main theoretical and practical tool currently used to quantify the capillary pressure function is an empirical relationship between capillary pressure and saturation expressed as:

$$P^n - P^w = P^c = f(S) \qquad (1)$$

where P^n and P^w are the average pressures of non-wetting and wetting phases, respectively, P^c is capillary pressure, and S is the wetting phase saturation.

This simple model is implicitly assumed to account for all microscale effects and processes that influence the equilibrium distribution of fluids, such as surface tension, presence of fluid-fluid interfaces, wettability of solid surfaces, grain size distribution, and microscale heterogeneities. All of these effects are essentially lumped into the $P^c - S$ relationship. Moreover, this relationship is obtained experimentally under equilibrium conditions. Thus, to obtain a drainage (or imbibition) curve, one starts with a wet (or dry) soil sample, then the capillary pressure is increased (or decreased) incrementally and at each step the water content is measured after equilibrium is reached. The time to equilibrium after each imposed pressure increment ranges from a few hours to many days, depending on the soil type and saturation degree. The typical time needed to construct a complete capillary pressure-saturation curve is in order of weeks or longer. Now, the question arises whether such curves adequately describe the relationship between $P^n - P^w$ and S in drainage or imbibition events with a time scale in the order of hours. In fact, there is ample theoretical and experimental evidence that this simple relationship is not unique but it depends on the flow dynamics; it depends on both the history and the rate of change of saturation. The dependence of capillary pressure-saturation curves on the history of flow is known as capillary pressure hysteresis; this is a well-known effect and has been the subject of extensive investigations. The dependence of capillary curves on the rate of change of saturation is due to non-equilibrium effects; it is much less known and is not quantified properly. The latter effect, also called dynamic effect in the literature, is the subject of this presentation.

First, we present a few examples from the unsaturated flow literature where it is shown that capillary pressure-saturation curves obtained dynamically under flow conditions differ significantly from the curves obtained under static conditions. An list of possible microscale effects causing these deviations is presented and discussed.

Next, we present a new theory of capillarity obtained in the framework of a macroscopic theory of multiphase flow in porous media. In this theory, equation (1) is replaced with:

$$P^n - P^w = P^c - \tau \frac{\partial S}{\partial t} = f(S) - \tau \frac{\partial S}{\partial t} \qquad (2)$$

where τ [$M\tau^1 T^{-1}$] is a material coefficient that is a measure of the non-equilibrium effect. It appears that this new term works as a damping effect.

We can show that the abnormalities in capillary pressure curves reported in the soil physics literature can, in general, be explained by the above equation. The magnitude of the new dynamic capillary pressure coefficient is determined from available data. Its value ranges from 10^4 to 10^8 kg/m.s.

Almost all experiments reported in the literature relate to unsaturated flow. Recently, we have performed an extensive set of laboratory experiments involving two immiscible phases, namely water and TCE. We have carried out both drainage and imbibition experiments under static as well dynamic conditions. The observations are in line with the unsaturated flow experiments mentioned above. The dynamic coefficient is determined for water-TCE system.

To examine whether these dynamic effects may be significant in practical situations, a continuum-scale simulator has been constructed in which equation (2) is included. This simulator is then run to determine the range of coefficients for which discernable effects occur. Results from such simulations indicate that for large values of the non-equilibrium coefficient (yet within the values determined from experiments), the non-equilibrium term has a significant effect on the simulation results. This indicates that non-equilibrium effects may be important and numerical simulators for two-phase flow should generally include the corresponding additional term in the capillary pressure-saturation relationship.

Dave Higdon

Los Alamos National Lab

TITLE: **"**Some Multiscale Approaches in Statistical Computation and Modeling"

ABSTRACT:
Multiscale concepts can play a key role in specifying spatial models and in speeding up computationally demanding computer simulations. Such approaches are often crucial for dealing with statistical inference involving large systems. This tutorial will consider some approaches for developing parsimonious and tractable representations of spatial, and space-time fields that model different scales separately. In addition I'll also go over a multiscale version of Markov chain Monte Carlo which utilizes computer simulations run at different grid resolutions. Such an approach can be very effective in certain inverse problems.

A summary of the topics is then:

Specifying Multiscale Models
Statistical approaches to representing "upscaled" fields.
Multiscale Markov chain Monte Carlo

Eugenia Kalnay
University of Maryland

TITLE: **"**Predictability, "Errors of the Day" and Breeding"

ABSTRACT:
In large chaotic systems like the atmosphere, predictability varies from place to place and day to day, depending on the stability of the underlying flow. We will describe how to use the simple method known as breeding to estimate the "errors of the day". We will show applications ranging from the Lorenz (1963) 3-variable model to realistic models used for numerical weather prediction with millions of variables.

Markos A. Katsoulakis

University of Massachusetts

TITLE: "Coarse-Grained Stochastic Processes and Monte Carlo Simulations in Lattice Systems"

ABSTRACT:
Diverse scientific disciplines ranging from materials science to Catalysis to biomolecular dynamics to climate modelling involve nonlinear Interactions across a large range of physically

significant length scales. In this talk we present a new class of coarse-grained stochastic processes and corresponding Monte Carlo simulation methods describing computationally feasible mesoscopic length scales, derived directly from microscopic lattice systems. Our main paradigm is a microscopic spin flip model for the adsorption and desorption of molecules between a surface and the overlying gas phase, while such types of microscopic, spin flip processes have also been proposed recently as providing prototype models for unresolved features of moist atmospheric convection.

We demonstrate analytically and numerically that the new coarse-grained stochastic models can capture large scale structures while retaining significant microscopic information. The requirement of detailed balance is utilized as a systematic design principle to guarantee correct noise fluctuations for the coarse-grained model. The coarse-grained stochastic algorithms provide large computational savings without increasing programming complexity or CPU time per executed event compared to microscopic Monte Carlo simulations. This is a joint work with A. J. Majda (Courant Institute) and D. G. Vlachos (Chemical Engineering, University of Delaware).

Peter Kramer

Rensselaer Polytechnic Institute

TITLE: "Effective Transport by a Spatiotemporal Mean Flow with Small-Scale Periodic Fluctuations"

ABSTRACT:

The transport of material and heat in the ocean or atmosphere is influenced strongly by both the prevailing large-scale mean flow structure and disordered turbulent motion prevalent on smaller scales. To obtain some insight into the effects of turbulent mixing, various authors over the last decade have studied the transport of passive material in model flows which have a periodic structure. For such flows, one can develop a rigorous homogenization theory to describe the effective mixing on large scales and long times. We will present an extension of these homogenization studies to a class of model flows which consist of a superposition of a large-scale mean flow with a small-scale periodic structure (both of which can depend on space and time). Using the two-scale convergence method, we can rigorously derive homogenized equations for these models, in which the mean flow and periodic structure are shown to interact nonlinearly. Different kinds of homogenized equations can emerge, depending on the spatio-temporal relationships between the mean flow and periodic fluctuations and the relative magnitude of the molecular diffusivity. It is shown that the small-scale structure is responsible for an enhancement in the diffusivity as well as for the presence of an additional effective drift, both of which are functions of space and time, with values depending upon the local properties of the mean flow. The physical manifestations of the interaction between the mean flow and the periodic fluctuations will be illustrated through some simple model examples.

Richard M. McLaughlin

University of North Carolina

TITLE: "An introduction to Methods of Homogenized Averaging for Problems with Multiple Scales"

ABSTRACT:
The purpose of this lecture is to present an introduction for scientists, statisticians, mathematicians, postdocs, and graduate students, the ethod and uses of multi-scale homogenized averaging in several environmentally relevant examples arising in the turbulent mixing of transported quantities, and effective behavior in heterogeneous porous media. Important problems in the environment concern the parametrization of unresolvable phenomena by effective closures. This tutorial lecture will demonstrate idealized scenarios in which such closures may be obtained with mathematical rigor, and will further identify the limitations of such averaging, and the need for improved mathematical methods to handle more complex problems. The methods will be presented to a broad audience, and explicit, simplified examples will be utilized to exhibit the averaging in a concrete fashion.

Casey Miller

University of North Carolina

TITLE: "An Introduction to Porous Medium Systems"

ABSTRACT:
The purpose of this lecture is to provide an introduction to porous medium systems for mathematicians, statisticians, and scientists who are not familiar with such systems. This lecture is especially appropriate for graduate students and post docs who wish to learn about the specific interdisciplinary challenges arising from the study of porous medium systems. Specific topics covered will include the traditional continuum modeling approach; model closure approximations; model formulation for problems involving single-phase flow, multiphase flow, and species transport and reactions; origin and importance of multiscale problems; and model solution approaches. We will also motivate the need for evolving theories of multiphase flow and transport, which will be covered in detail by the plenary and invited speakers.

Robert Miller

Oregon State University

TITLE: **"**Pushing the State of the Art in Ocean Modeling and Data Assimilation:
Intrinsic Variability of Ocean Currents"

ABSTRACT:
The variability of ocean currents, especially western boundary currents is sometimes attributed to variability in the winds. The winds are certainly variable, but wind variability alone cannot account for observed and modeled current variability.

This intrinsic variability cannot be understood in the context of the linear and linearized theories of ocean circulation developed in the 50's and 60's. The Kuroshio, the dynamical analogue of the Gulf Stream in the Atlantic, provides an unambiguous example on which theories can be tested and general strategies developed. The Kuroshio is observed to take distinct alternate paths off the coast of Japan. Each persists for years, and the transition takes place in months. Limited area

models and basin scale models of varying complexity exhibit this behavior, apparently for different reasons.

Understanding of the Kuroshio and the general circulation of the North Pacific will require application of tools not widely used in the physical oceanography community. Ultimately conflicts among model results must be resolved by appeal to data, and thus results of data assimilation experiments, which themselves must account for nonlinearity as well as noise.

Marc Parlange

Johns Hopkins University

TITLE: "Field Experiments and Observations of Subgrid-Scale Processes in the Atmospheric Surface Layer"

ABSTRACT:
We will discuss some fundamental properties of atmospheric turbulence based on field observations and how these properties relate to parameterizations used in Large Eddy Simulation (LES.) The atmospheric surface layer temperature and velocity data have been obtained in a series of field experiments using various arrays of three - dimensional sonic anemometers. The data are analyzed to develop an improved understanding of the statitical properties of the subgrid-scale (SGS) heat flux and shear stress near the land surface. The data are used to evaluate the ability of various SGS models to reproduce statistical and geometric properties of the heat flux and shear stress and test possible improvements based on atmospheric stability. The important role of coherent structures near the land surface on SGS physics is highlighted.

Doug Nychka

National Center for Atmospheric Research

TITLE: "Statistical Models for Models"

ABSTRACT:
Numerical models for environmental and geophysical processes produce output that can summarized and probed using statistical models. This talk will review some recent areas where statistics can support numerical modeling. These include: building stochastic parameterizations, detecting distinct dynamical regimes, quantifying extreme behavior and spatial heterogeneity and detecting coherent structures. The areas of application, while focused on the atmosphere, range across many scales and highlight some basic statistical tools such as nonparametric regression, spatial statistics, wavelets and clustering.

Juan M. Restrepo
University of Arizona

TITLE: "Interaction of Waves and Currents"

ABSTRACT:

There are a variety of oceanic problems in which the interaction of waves and currents, at the space/time scales typical of the currents, are crucial. Examples are pollution dynamics, thermal transport, sediment dynamics, the dynamics of biological populations. The spatial scales can be as large as the continental shelf; the times scales as long as seasons or years. These are scales many times larger than those of the waves, thus, in most instances, computing wave/current interactions can be very impractical or impossible if the waves need to be resolved.

A recently developed theoretical framework will be presented in which waves and currents can be modeled at these larger scales. The model not only circumvents the impracticality of resolving motions at wave scales, it also describes precisely how the waves and currents interact. It is thus expected that such model be of great practical utility as well as a robust theoretical tool for the investigation of a variety of oceanic physical phenomena.

Paul D. Sampson
University of Washington

TITLE: "Nonstationary Spatial Covariance Modeling Via Spatial Deformation and Extensions to Covariates and Models for Dynamic Atmospheric Processes"

ABSTRACT:
Nonstationary, anisotropic correlation structure is now widely recognized to underly most atmospheric space-time processes, depending the temporal and patial scales of measurement and analysis. We will begin with a review and demonstration of current Bayesian methods and software for fitting the Sampson-Guttorp spatial deformation model to space-time air quality data. This methodology, while computationally intensive, lends insight into processes underlying observed monitoring data. We will comment on issues of spatial and temporal scale and the relevance of this modeling to the evaluation of air quality models. We will then discuss two proposals for future development. The first proposal addresses the incorporation of fixed or dynamic covariates into a space-time model by deforming and embedding the original coordinate space into a higher dimensional space in order to incorporate covariates or dynamics without abandoning the convenience of a stationary correlation model. The second is a proposal, due to Gneiting, for space-time covariance models for dynamic error structures in the context of data assimilation. The aim is to account explicitly for dynamic atmospheric processes such as wind fields.

Tapio Schneider
California Institute of Technology

TITLE: "Determining Sources and Sinks of Carbon Dioxide from Atmospheric Concentration Measurements: A Multiscale Inverse Problem."

ABSTRACT:
Measurements of carbon dioxide concentrations in the atmosphere indicate that about half of the carbon dioxide emitted by human activities have been absorbed by the biosphere and the oceans. However, where and by what processes the carbon dioxide has been absorbed is uncertain. The uncertainties are due in part to insufficient spatial resolution of the current network of measurement stations, and in part to uncertainties about the transport of carbon dioxide and its interaction with biological processes.

Bjorn Stevens
University of California-Los Angeles

TITLE: "Aggregating Clouds"

ABSTRACT:
Rationalizations of the Earth's climate depend critically on the representation of cloud processes. In this talk we review more than a decade of concerted effort to understand one particular cloud/climate interaction, that associated with shallow stratiform clouds in the marine boundary layer. Particular emphasis is placed on highlighting the interplay of large-scale (climate) models, simple theory, numerical analysis, large-eddy simulations, and (most recently) novel observations to improve our understanding of this one particular type of cloud climate interaction. Key physical issues which are identified and focused-on include the representation of entrainment at a very stratified interface, and the interplay of processes, particularly those involving precipitation, across a range of sub-grid scales. Although recent work is providing some closure on the former issue, the latter remains a great challenge.

Andrew Stuart
Warwick University

TITLE: "Inertial Particles in a Random Field"

ABSTRACT:
When particle and fluid time constants are commensurate, then preferential concentration, whereby the configurations of inertial particles are highly correlated through a turbulent velocity field, can occur. The purpose of this work is to propose a straightforward mathematical model for this phenomenon, and then to use the model to study various scaling limits of interest and to study numerically the effect of inter-particle collisions. The model we propose comprises Stokes law for particle motions, with a Gaussian random field for the velocity field. The velocity field is found from a streamfunction which satisfies a linear stochastic PDE of OU type. Thus the complete model is in the form of a stochastic PDE coupled to an ODE.

The scaling limits corroborate experimental evidence about the lack of preferential concentration for large and small Stokes number. They also make new predictions about the possibility of preferential concentration at large times, and lead to stochastic differential equations (SDEs) governing this phenomenon, using techniques developed by Majda, Timofeyev, Vanden Eijnden and Kramer. Straightforward analysis of the coupled
stochastic PDE-ODE system leads to a proof that the predicted limiting SDEs are correct, using strong convergence methods.

A fast algorithm which incorporates inter-particle elastic collisions is described. The effect of collisions is shown to be negligible for the most part, although in some cases they are shown to have an interesting anti-diffusive effect.

Claudia Tebaldi
National Center for Atmospheric Research

TITLE: "A Bayesian Analysis of Regional Climate Change Projections, Based on a Multi-model Ensemble" (joint work with Richard Smith)

ABSTRACT:

Different General Circulation Models (GCMs) produce different climate change projections, especially when evaluated at subcontinental (regional) scales. The problem of combining their sometimes discordant responses into a summary measure, and relative uncertainty bounds, becomes especially relevant when analyzing these smaller scales, since they are more likely to become inputs for policy makers and the impacts research community than global scale measures of climate change.

There are heuristic criteria in the interpretation of GCMs output generally accepted, that awaited formalization: it makes sense to give larger weight to the output of those GCMs that show better performance in reproducing present day climate (i.e. have smaller bias) and that agree with the consensus (i.e. do not appear like outliers).

We present a suite of Bayesian models that combine climate reproductions and projections from 9 different GCMs over regions discretizing the land masses of the globe, and formalize the former weighing criteria of "discounting bias" and "rewarding convergence", in order to provide a final distributional estimate of regional climate change. We start from possibly the simplest univariate (one region at a time) model, show how to improve it by making it "robust" and accounting for correlation between present and future climate response within a single GCM, and finally present a multivariate version, accounting for correlation between regional averages of a GCM output. Posterior distributions of a climate change measure and other parameters of interest are derived, and possible extensions suggested.

Joe Tribbia
National Center for Atmospheric Research

TITLE: "Challenges in Parameterizations for Global Circulation Models of the Atmosphere"

ABSTRACT:
The problem of parameterization of physical process in large scale numerical models of the atmosphere will be developed first from a formal perspective of specifying sub-grid effects as distributions conditional upon resolved scales of motions. Historically, focus has been concentrated upon models of the 'average' or mean tendencies and the models have been of two classes: sub-grid models which depend upon scale similarity those which are developed from idealized process models. I will describe some recent efforts aimed at moving beyond the historical focus on the mean and argue that neither the pure scaling approach nor the process model approach to parameterization is likely to suffice due to the spatio-temporal intermittency of small scale atmospheric motions.

Eric Vanden Eijnden
Courant Institute of Mathematical Sciences

TITLE: "Multi-scale Dynamical Systems with Stochastic Effects: Mathematical Aspects and Numerical Techniques"

ABSTRACT:

Asymptotic techniques leading to efficient numerical schemes for multi-scale dynamical systems will be presented and illustrated on several examples motivated by atmosphere-ocean sciences.

Claire Welty
Drexel University

TITLE: "Density-Coupled Transport in Heterogeneous Aquifers"

ABSTRACT:
Application of spectral perturbation techniques to miscible density-coupled transport in heterogeneous media predicts an effective macrodispersion tensor that is dependent upon fluid as well as porous medium properties. Theoretical results also show that the dispersion tensor is asymmetric and dependent upon mean displacement distance of a solute body. Stochastic mean simulations that incorporate the theoretical results for a dense solute plume sinking in an aquifer will be presented to illustrate the theory. Comparison of the theoretical results to large-scale laboratory observations and fine-grid numerical simulations will also be presented.

Christopher K. Wikle
University of Missouri

TITLE: "Hierarchical Bayesian Modeling of Multiscale Spatio-Temporal Processes"

ABSTRACT:
Spatio-temporal processes are ubiquitous in the environmental and physical sciences. This is certainly true of atmospheric and ecological processes, which typically exhibit many different scales of spatial and temporal variability. The complexity of these processes and large number of observation/prediction locations precludes the use of traditional covariance-based space-time statistical methods. Alternatively, we consider conditionally-specified (i.e., hierarchical) spatio-temporal models. Physical and dynamical constraints are easily incorporated into the conditional formulation, so that the series of relatively simple, yet physically realistic, conditional models leads to a much more complicated joint space-time covariance structure than can be specified directly. Furthermore, multiresolution spectral formulations, which naturally allow spatial nonstationarities, reduce dimensionality, and introduce sparseness, are easily included in this conditional framework. Time permitting, examples will be presented that feature atmospheric and ecological processes.

Yannis C. Yortsos
University of Southern California

TITLE: "Flow, Transport And Reaction In Porous Media: Upscaling From Pore-Network Models"

ABSTRACT:
Flow and transport in porous media finds a number of applications, from the recovery of subsurface fluids to processes in packed beds and porous columns. Typically, the evolution in space and time of the macroscopic variables of the displacement is sought. Uncovering this dependence requires the coupling of mechanisms at the pore-scale with the heterogeneity of the porous medium, particularly at the pore-network scale, and the delineation of displacement and flow patterns. In this talk, we describe methodologies for implementing such upscaling processes,

using statistical physics tools. In the case of connected immiscible fluids, the dominance of capillary forces at the small scale leads to Invasion Percolation in a Gradient. A diagram delineating the regimes of fully developed drainage is developed. When the displaced phase is disconnected, the mobilization of trapped ganglia is necessary. We describe a method based on Darcian dynamics to simulate this process. Issues associated with miscible displacement are also discussed. Finally, we address problems of drying, dissolution and combustion, where mass transfer and reactions are of importance and where the upscaling from the pore-network scale presents interesting new issues.

SAMSI Distinguished Lecture

Andrew J. Majda

Courant Institute of Mathematical Sciences

TITLE: "Mathematical Strategies for Stochastic Modeling in Climate"

ABSTRACT:
Modeling and predicting the features of short-term climate presents a major scientific challenge due to the enormous range of interacting physical phenomena on many different spatio-temporal scales. This lecture will introduce the audience to two important contemporary examples involving multiple scale interaction for climate involving the tropical Western Pacific and low frequency variability for the midlatitude atmosphere. Then emerging mathematical strategies for stochastic modeling of the unresolved degrees of freedom will be described for these two contrasting physical situations. In the case of the tropical Western Pacific, novel parameterization models utilizing stochastic Ising models and associated systematic coarse-grained birth-death processes will be discussed. A systematic mathematical strategy for reduced stochastic modeling for the mid-latitude atmosphere will be discussed through idealized examples which emphasize the new phenomena which emerge through such an approach; these phenomena will be discussed in the context of the regression-fitting strategies utilized in contemporary climate modeling.

D. Program and Abstracts for One-Day Workshop
March 25, 2003, NISS-SAMSI Building

9:30-10:15 "Essentials of Physical-Statistical Modeling"
Mark Berliner, Ohio State University

10:15-11:00 "Hierarchical Modeling of Advection Diffusion Processes"
Chris Wikle, University of Missouri

11:00-11:30 Coffee break

11:30-12:30 "Bayesian Hierarchical Modeling of Air-Sea Interaction"
Mark Berliner, Ohio State University

12:30-1:15 Lunch

1:15-2:15 "Hierarchical Bayesian Boundary Value Problems"

	Chris Wikle, University of Missouri
2:15-2:30	Coffee break
2:30-4:00	"Ocean Modeling and Lagrangian Analysis" Christopher Jones, University of North Carolina

Mark Berliner
Ohio State University

TITLE: "Essentials of Physical-Statistical Modeling"

ABSTRACT:
I will review selected aspects of physical modeling and the transition to stochastic modeling in the presence of uncertainty. The notions presented are linked to the hierarchical statistical modeling with emphasis on Bayesian approaches. An brief example of Bayesian analysis for stochastic differential equation process models relying on the Fokker-Planck equation is presented. Further motivation of physical statistical modeling is derived from use of combinations of simplified physical models and numerical techniques in the presence of observational data.

TITLE: "Bayesian Hierarchical Modeling of Air-Sea Interaction"

ABSTRACT:
We develop a Bayesian hierarchical model for aspects of vigorous air-sea interactions on a basin scale. The approach relies on both physical reasoning and statistical techniques for data processing and uncertainty management. The crucial component of the modeling involves development of a stochastic model for the ocean conditional and atmospheric behavior based on quasi-geostrophic modeling and a stochastic model for the atmospheric elements. To demonstrate the strategy, we apply it in the context of an Observing System Simulation Experiment. An ocean "truth" simulation is driven by idealized surface winds in a test bed domain abstracted from the Labrador Sea. This truth simulation is not based on quasi-geostrophic modeling, but rather uses a primitive equation-shallow water equation model. Artificial observations analogous to scatterometer and altimeter data are incorporated and comparisons made with the evolution of the "truth" simulation over a ten day experiment. The presentation reviews Berliner, Milliff, and Wikle (2003, In Press), Journal of Geophysical Research-Oceans.

Christopher Jones
University of North Carolina

TITLE: "Ocean Modeling and Lagrangian Analysis"

ABSTRACT:
Much data in the ocean is Lagrangian in that it is derived from subsurface floats or surface drifters following the flow. At the same time, many questions about the ocean involve transport and mixing, and therefore raise Lagrangian issues. Advances in the use of dynamical systems ideas have afforded a deeper understanding of Lagrangian transport. An application that elucidates chlorophyll

dispersion in the Gulf of Mexico will be presented. A strategy for model analysis, based on this theory, at both model input and output, will be described. This will include Lagrangian data assimilation, optimal float placement design and a new approach to model evaluation.

Chris Wikle
University of Missouri

TITLE: "Hierarchical Modeling of Advection Diffusion Processes"

ABSTRACT:
The hierarchical Bayesian approach to physical-statistical modeling will be illustrated with a simple advection-diffusion equation. The methodology will be illustrated from both a physical-space and spectral-space perspective. An ecological example showing the spread of an invasive species will be presented.

TITLE: "Hierarchical Bayesian Boundary Value Problems"

ABSTRACT:
Boundary value problems are ubiquitous in the atmospheric and ocean sciences. Typical settings include bounded, partially bounded, global and limited area domains, discretized for applications of numerical models of the relevant fluid equations. Often, limited area models are constructed to interpret intensive datasets collected over a specific region, from a variety of observational platforms. These data are noisy and they typically do not span the domain of interest uniformly in space and time. Traditional numerical procedures cannot easily account for these uncertainties. A hierarchical Bayesian modeling framework is developed for solving boundary value problems in such settings. By allowing the boundary process to be stochastic, and conditioning the interior process on this boundary, one can account for the uncertainties in the boundary process in a reasonable fashion. In the presence of data and all its uncertainties, this idea can be related through Bayes' Theorem to produce distributions of the interior process given the observational data. The method is illustrated with an example of obtaining atmospheric stream function fields in the Labrador Sea region, given scatterometer-derived observations of the surface wind field.

E. Program for the Workshop on Simulation and Optimization
April 28-30, 2003, Radisson Governor's Inn

- **Monday, April 28, 2003**

<u>Optimization of Remediation Systems</u>

8:30-9:00 Registration

9:00-9:45 "Groundwater Optimization: Advantages and Disadvantages of Derivative-Based, Heuristic and Surface Approximation Methods"
Christine Shoemaker, Cornell University

9:45-10:30 "The Inverse Problem Of Parameter Structure Identification"
William Yeh, University of California, Los Angeles

10:30-11:00 Coffee Break

11:00-11:45 "Optimization of Engineering Design of Subsurface Environmental Remediation Systems- Development and Testing of Community Benchmark Problems"
Alex Mayer, Michigan Technological University

11:45-12:30 "Optimal Design for Groundwater Flow and Remediation Problems"
Kathleen Kavanagh, North Carolina State University

12:30-1:30 Lunch

Future of Simulation: Emerging Models

1:30-2:15 **"A Highly Efficient Conditional Moment Algorithm for Transient Flow in Random Porous Media"**
Shlomo Neuman, University of Arizona

2:15-3:00 "Future of Simulation: Emerging Models"
Dennis McLaughlin, Massachusetts Institute of Technology

3:00-3:30 Coffee Break

3:30-4:15 "On the Governing Equations of Flow in Porous Media"
William G. Gray, Notre Dame University

4:15-4:45 Open Discussion

6:30-8:00 **Poster Session and Reception**
NISS-SAMSI Building
Transportation from Radisson Governors Inn will be provided by Carolina Livery

- **Tuesday, April 29, 2003**

Stochastic and Deterministic Simulation

8:30-9:00 Registration

9:00-9:45 "The Department of Defense's ADaptive Hydraulics/Hydrology Model (ADH)"
Stacy Howington, Engineer Research and Development Center

9:45-10:30 "Advanced Modeling Techniques for Numerical Simulation of Complex Subsurface Hydrosystems"
Mary F. Wheeler, Texas Institute for Computational and Applied Mathematics

10:30-11:00 Coffee Break

11:00-11:45 "Annoying Issues In Numerical Simulation Of Subsurface Transport"
Thomas F. Russell, University of Colorado at Denver

11:45-12:30 "Nonlinear Solution and Sensitivity Methods for Variably Saturated Flow"
Carol S. Woodward, Lawrence Livermore National Laboratory

12:30-1:30 Lunch

1:30-2:15 "Discontinuous Galerkin Methods for Convection-Diffusion Problems"
Clint Dawson, University of Texas at Austin

2:15-3:00 "Non-Newtonian Fluid Flow Through an Extrusion Filter"
Lea Jenkins, Clemson University

3:00-3:30 Coffee Break

3:30-4:15 "Pore-Scale Modeling for Closure of Multiphase Models Derived Using Thermodynamically Constrained Averaging Theory Approaches"
Cass T. Miller, University of North Carolina

4:15-4:45 Open Discussion

- **Wednesday, April 30, 2003**

8:30-9:00 Registration

9:00-9:45 "Constrained Optimization of Expensive Functions Using Surrogates"
John Dennis, Rice University

9:45-10:30 "Stochastic Programming Approach to Uncertainty and Risk"
Darinka Dentcheva, Stevens Institute of Technology

10:30-11:00 Coffee Break

11:00-11:45 "Adaptive Numerical Methods for Sensitivity Analysis of Differential-Algebraic Equations and Partial Differential Equations"
Linda Petzold, University of California Santa Barbara

11:45-12:30 "Using Nonlinear Interior Point Methods for Optimal Groundwater Contamination Containment"
Pamela J. Williams, Sandia National Laboratories

12:30-1:30 Lunch

1:30-2:15 "Optimal Designs for Monitoring the Extremes of Environmental Processes"
Jim Zidek, University of British Columbia

2:15-3:00 "Environmental Applications of the Stochastic Engine"

Steven F. Carle, Lawrence Livermore National Laboratory

3:00-3:30 Coffee Break

3:30-4:15 "Implicit Filtering"
C. T. Kelley, North Carolina State University

4:15-4:45 Open Discussion

F. Simulation and Optimization Poster Session
April 28, 2003, NISS-SAMSI Building

Shannon L. Bartelt-Hunt
University of Virginia

TITLE: **"Optimal Design of a Waste-Disposal Liner Containing Sorptive Amendments"**

ABSTRACT:
Clay and geomembrane composite liner systems are commonly used to protect groundwater supplies from contamination by municipal and industrial waste disposal facilities. The primary design characteristic for these earthen liners is maintenance of a low hydraulic conductivity to minimize contaminant advection. Under this condition, the primary pollutant transport mechanism is diffusion, which may contribute significantly to the mass flux of organic contaminants out of the liner. The magnitude of the diffusive flux is controlled by two factors: weak contaminant sorption and the short diffusion length. The addition of a geomembrane layer to the design of earthen liners can further serve to minimize pollutant advection, but diffusion through this layer still occurs and the geomembrane may fail due to rupture along a seam or if punctured.

Another method of enhancing pollutant sorption and thus minimizing pollutant flux out of liners is to amend conventional liners with materials capable of strongly sorbing organic pollutants, such as organophillic bentonite, activated carbon, shale and fly ash.

A genetic algorithm optimization model linked to a 1-D aqueous transport model will be used to determine the optimal liner design that minimizes the overall costs of the system while still meeting specified aqueous transport goals at the base of the liner. Decision variables will include the type of amendment incorporated in each layer of the liner, the weight percentage of the amendment in each layer, and the total number of layers in the overall liner design. Laboratory data on the sorption, permeability, and mechanical properties of the proposed amendments will be collected and used as input data for the model.

Katherine A. Culligan
University of Notre Dame

TITLE: "On the Measurement of Interfacial Areas in Multiphase Porous Media Flow"

ABSTRACT*:*
Multiphase flow and contaminant transport in porous media are strongly influenced by the presence of fluid-fluid interfaces. Recent theoretical work based on conservation laws and the second law of thermodynamics has demonstrated the need for quantitative interfacial area information to be incorporated into multiphase flow models. We have used the GSECARS microtomography beam-line at the Advanced Photon Source to produce three-dimensional pore-scale images of multiphase flow through a porous medium. Both air/water and oil/water systems have been investigated. Analysis of the high-resolution images (17 micron pixels) allows for computation of interfacial areas and saturation. Corresponding pressure measurements were made during the course of the experiments. Representative Elementary Volume analyses were done on quantities of interest. Also presented here are plots of capillary pressure as a function of experimentally measured interfacial area and saturation. Interfacial areas were estimated using isosurfaces and two-point correlation functions.

Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38.

Matthew W. Farthing

University of North Carolina-Chapel Hill

TITLE: "Efficient Steady-State Solution Techniques for Variably Saturated Groundwater Flow"

ABSTRACT:
Most discretization approaches for Richards' equation lead to nonlinear systems that are large and difficult to solve. The solution of nonlinear systems for steady-state problems can be particularly challenging, since a good initial guess for the steady-state solution is often hard to obtain, and the resulting linear systems may be poorly scaled. Common approaches like Picard iteration or variations of Newton's method have their advantages but perform poorly with standard globalization techniques under certain conditions. Pseudo-transient continuation has been used in computational fluid dynamics for some time to obtain steady-state solutions for problems in which Newton's method with standard line-search strategies fails. Here, we examine the use of pseudo-transient continuation as well as Newton's method combined with standard globalization techniques for steady-state problems in heterogeneous domains. We investigate the methods' performance with direct and preconditioned Krylov iterative linear solvers. We then make recommendations for robust and efficient approaches to obtain steady-state solutions for RE under a range of conditions.

Matthew W. Farthing

University of North Carolina-Chapel Hill

TITLE: "A PSE for Modeling Transport Phenomena in Porous Medium Systems"

ABSTRACT:
Porous medium systems are usually modeled by systems of conservation equations that are closed by constitutive relations and solved approximately using numerical methods. While this approach is well established and routinely used, many aspects of the underlying model formulations and solution approaches change routinely, especially for reactive contaminant

transport and multiphase flow and transport problems. This situation poses a significant computational science challenge, since the models, and resultant codes, are of high complexity and subject to frequent change.

Here, we review approaches used to model porous medium systems and highlight some of their shortcomings. We discuss the development of problem solving environments for a range of target problems and suggest an approach for developing a problem solving environment for transport phenomena in porous medium system. We then present preliminary efforts for modeling batch reaction systems as well as unsaturated flow and reactive transport.

Daniel Finkel
North Carolina State University

TITLE: "The DIRECT Global Optimization Algorithm Applied to Groundwater Remediation Flow Problems"

ABSTRACT:
The DIRECT Optimization Algorithm searches a bound design space for the global minimum by systematically slicing the space into hyper-rectangles and sampling these rectangles in their centers. One of DIRECT's features is that it uses no gradient information in its search. Instead, DIRECT proceeds by using a modified Lipschitz Optimization approach. In particular, DIRECT varies the Lipschitz Constant of the objective function to perform multiple searches throughout the design space.

In this work, we introduce the reader to DIRECT, and briefly describe how it searches for the global minimum. We also present numerical results of DIRECT's application to a well field design problem.

Genetha Gray
Sandia National Labs

TITLE: "Derivative Free Optimization for Transmembrane Protein Structure Determination"

ABSTRACT:
Many of the questions posed in the fields of biology and medicine cannot be answered completely with traditional laboratory experimentation. Hence, numerical algorithms are emerging as a crucial component of biological research. In this poster, we give an example of one such problem: determining the structure of transmembrane proteins. We begin with a basic description of transmembrane proteins and explain why understanding the structure of these proteins is important in the study of numerous diseases and for drug development. We describe the biological characteristics that make it amenable to optimization and examine its mathematical properties. Finally, we compare and contrast some results obtained using different derivative free optimization algorithms.

KATHLEEN KAVANAGH

North Carolina State University

TITLE: “Solution of a Groundwater Water Control Problem with Implicit Filtering”

ABSTRACT:
We describe the application of a parallel implementation of the implicit filtering algorithm to a control problem from hydrology. We seek to control the temperature at a set of drinking water wells by placing barrier wells between the drinking water wells and the well that injects heated water from an industrial site. We seek the optimal pumping rates at the barrier wells. The resulting objective function is nonconvex, nonsmooth, and has several local minima.

Chongxun Pan
University of North Carolina

TITLE: “A High-Performance Lattice Boltzmann Implementation of Multiphase Flow in Porous Media”

ABSTRACT:
The lattice Boltzmann (LB) method is a powerful approach for simulating computational fluid dynamics at a micro-kinetic level. However, computational limitations are of great concern when applying LB simulations for multiphase porous medium systems because the resolution needed to resolve fundamental phenomena at the pore scale leads to very large lattice sizes, hence substantial computational and memory requirements that necessitate the use of massively parallel computing approaches.

Conventional LB implementations for simulating flow in porous media store the full lattice, making parallelization straightforward but wasteful. In this work, we investigate a two-stage implementation consisting of a sparse domain decomposition stage and a simulation stage that avoids the need to store and operate on lattice points located within a solid phase. A set of five domain decomposition approaches are investigated for single and multiphase flow through both homogeneous and heterogeneous porous medium systems on different parallel computing platforms. An orthogonal recursive bisection method is found to yield the best performance of the methods investigated, showing near linear scaling and substantially less storage and computational time compared to regular decomposition approaches.

JILL REESE
North Carolina State University

TITLE: “Optimal Design for Groundwater Flow and Remediation Problems”

ABSTRACT:
Optimal design problems arise naturally in management applications from subsurface hydrology. Implementing and solving such a problem requires developing the appropriate model, simulating the subsurface, developing an objective function and appropriate constraints, and minimizing the objective function via optimization software. For example, applications can include water resource management or the remediation of a contaminated site. In this work we describe how to implement and solve such optimal design problems and compare optimization techniques.

VALENTIN M. SILANTYEV
Northeastern University

TITLE: “Explicitly Solvable Kirchhoff and Riabouchinsky Models with Partially Penetrable Obstacles and Their Application for Estimating the Efficiency of Hydraulic Turbines in Free Flows”

ABSTRACT:
The classical Kirchhoff and Riabouchinsky models consider the streamlining of an impervious obstacle by ideal inviscid incompressible fluid with separation of the flow from the edges of the obstacle, which is introduced in order to avoid d'Alembert paradox (zero thrust on the obstacle in the case of streamlining without separation). The free boundary problems arising from these models can be solved explicitly for some kind of obstacles using the Kirchhoff transform (passing to the hodograph variable that eliminates free boundaries and reduces the problem to constructing the conformal representations of prescribed domains depending on some parameters). In my recent works with A.M.Gorlov, A.N.Gorban', and M.E.Braverman these models are generalized for the case of a partially penetrable obstacle absorbing the energy from the flow passing through the obstacle. The study of these generalized models was motivated by the engineering problem of estimating the efficiency and optimizing the parameters of the brand new free flow hydraulic turbines, which do not require the construction of water dams for their operation. The turbines of these kind can be used for harnessing energy from alternative “clean” resources, like ocean currents or tidal streams, without harming the environment or causing a potential risk for the surrounding area in the case of emergency or deliberate damage. It will be shown that in some situations the generalized models can be also explicitly solvable by using the same technique as in the classical case. The theoretical estimates of the efficiency and other parameters obtained from these models proved to be in good agreement with the results of practical tests performed by engineers.

G. Simulation and Optimization Workshop Speaker Abstracts

Steven F. Carle

Lawrence Livermore National Laboratory

TITLE: “Environmental Applications of the Stochastic Engine”

The Stochastic Engine (SE) is a flexible computational tool for inversion or optimization of large non-linear or underdetermined problems with non-unique solutions. Such difficult problems are common to environmental applications. The SE uses Bayesian inferencing to incorporate prior information into a Metropolis-type search algorithm that efficiently samples a multi-dimensional solution space. For imaging problems in environmental applications, the SE approach enables: (1) incorporation of hard or soft data as prior information, (2) estimation of probability distributions for lithology types at a grid of locations, and (3) preservation of geologic realism in multiple solutions consistent with data. The SE offers flexibility because forward calculations are used to drive the solution search algorithm, and different forward models can be plugged into the SE. The SE uses a staged approach, whereby the SE is run successively using forward models of increasing complexity and! computational intensity (e.g, lithology>flow>transport). The staged approach enables consideration of multiple processes in determination of the solution space, while weeding out unlikely solutions using less computationally intensive forward models at early stages. Environmental applications of the SE thus far have focused on determination of the spatial distribution of lithology types given prior geologic knowledge, electrical resistance tomography data, geophysical logs, and flow measurements. The code "TSIM" is used for the initial lithology generation stage to generate geologically plausible realizations. TSIM has been modified to control stepsize in the Metropolis search by controlling how similar a new realization is to a previous lithologic configuration. In these environmental applications, the SE finds lithologic configurations that are consistent with prior geologic knowledge and available direct and indirect data on lithology.

This work was performed under the auspices of the U.S. Department of Energy (DOE) by the University of California Lawrence Livermore National Laboratory (LLNL) under Contract W-7405-Eng-48.

Clint Dawson
University of Texas at Austin

TITLE: “Discontinuous Galerkin Methods for Convection-Diffusion Problems”

ABSTRACT:
Discontinuous Galerkin (DG) methods have attracted much interest recently as a means for solving a variety of problems, most notably, elliptic flow problems and convection-diffusion systems. These methods have become popular because of their ability to use very general, even nonconforming meshes, general approximating spaces, adaptive grids, and because they satisfy local mass conservation. Several variants of the DG method have been proposed. In this talk, we will focus on DG methods for convection-diffusion problems. Methods based on both primal and mixed formulations will be discussed, and numerical results for geoscience applications, including contaminant transport and shallow water equations, will be given. Issues related to local and global mass conservation, and why it is important, will also be discussed.

JOHN DENNIS
Rice University

TITLE: “Constrained Optimization of Expensive Functions Using Surrogates”

ABSTRACT:
Many engineering design problems are governed by coupled systems of partial differential equations, table lookups, and who knows what. It is common in these problems for the governing simulations to return no value for a high proportion of feasible points. This talk is a hurried introduction to the filtered direct search method incorporated into the current Boeing Design

Explorer release and the public domain NOMAD code. This talk goes quickly over the algorithmic and theoretical details to save time for supporting numerical results. Technical reports giving the omitted details are available from www.caam.rice.edu and www.crpc.rice.edu.

Darinka Dentcheva
Stevens Institute of Technology

TITLE: "Stochastic Programming Approach to Uncertainty and Risk"

ABSTRACT:
Stochastic programming provides a methodology for optimization in the presence of uncertainty and risk. We shall discuss some basic modeling approaches and formulate the most popular models in stochastic programming. Two-stage and multi-stage models, as well as problems involving probability functions will be presented. Brief overview of their main structural properties will be given. Fundamental ideas of numerical methods for solving stochastic optimization problems will be presented. We shall discuss also how limited information about probability distributions can be used in optimization models. At the end of the talk we shall suggest approaches to modeling risk. Several examples will illustrate the concepts and ideas.

William G. Gray
University of Notre Dame

TITLE: "On the Governing Equations of Flow in Porous Media"

•

ABSTRACT:
In the subsurface environment, the presence of a solid phase provides a tortuous network through which a fluid must flow. This network imposes challenges on efforts to model the flow that are different from those encountered in modeling surface or atmospheric processes. If models are to be successful in accounting for heterogeneities, simulating flow of multiple fluid phases, and providing information on contaminant transport, they must be built on sound theoretical principles as well as effective and efficient numerical techniques. The purpose of this presentation is to indicate some of the shortcomings of the theoretical foundation. In particular, it is noted that a hunt for appropriate parameters to complete a model is unsatisfactory if those parameters merely support governing equations that do not properly account for the physical processes of importance. The suggestion is made that improvement is needed in the forms of the governing equations as well as the coefficients in those equations.

Stacy Howington
US Engineer Research and Development Center

TITLE: "The Department of Defense's ADaptive Hydraulics/Hydrology Model (ADH)"

ABSTRACT:
The Department of Defense is responsible for cleaning up groundwater contamination at present and formerly-used defense sites. Common contaminants include solvents, metals, and explosives. The DoD relies on numerical modeling to help assess the extent of contamination, identify the source of contamination,

estimate the future threat, and design remedial measures. Computational domains often are large to include known boundary conditions. However, fine resolution is needed to describe geologic variability, to capture flow fields in the vicinity of wells, drains, trenches, or where groundwater and surface water interact, and to resolve contaminant plumes. The resulting computational meshes typically range from 500,000 to millions of elements. The DoD's Corps of Engineers also has responsibilities in surface water for maintaining navigation, flood control, and environmental quality. These require the use of numerical models to simulate hydrodynamics and sediment and constituent transport.

A new model is nearing completion in response to several needs not met by present simulation tools. Among these needs are dynamic mesh refinement and coarsening, efficient distributed computing, and multi-physics coupling. The ADH (ADaptive Hydraulics/Hydrology) model is a multi-dimensional, continuous finite element code that simulates partially saturated groundwater flow, shallow overland flow using either diffusive wave or full St. Venant equations, hydrostatic or non-hydrostatic Navier-Stokes equations, and constituent and non-cohesive sediment transport. The model is written in the C programming language and permits only simplex elements (tetrahedra, triangles, and lines). The computational mesh adapts during the simulation in response to error indicators. As the mesh adapts, the computational load is redistributed among processors.

A central idea in the development of this code is reuse of investment in parallelism, adaption, solvers, and preconditioners. Thus, the computational engine forms the center of the code to which many equation sets are 'strapped'. At present, the model is being extended to permit coupling among different flow physics and to include a multiple-continuum transport scheme and more efficient methods for simulating flow over initially dry ground. For the future, we want to make use of available computational power to automate model calibration and sensitivity analysis and to optimize remedial design in uncertain geology.

Lea Jenkins
Clemson University

TITLE: "Non-Newtonian Fluid Flow Through an Extrusion Filter"

ABSTRACT:
In this talk, we discuss the flow of a non-Newtonian fluid through an extrusion filter. The problem is important in the fiber industry, where companies are interested in extending filter lives. Debris is deposited as the fluid moves through the filter, which changes the flow velocities. Darcy's equation, modified by a power law, is used to model the fluid velocity. We discuss previous three-dimensional models for the filter medium, and we also discuss our efforts to model the medium using partial differential equations. We present our method of handling the nonlinearities, and our finite element formulation.

Kathleen Kavanagh

North Carolina State University

TITLE: "Optimal Design for Groundwater Flow and Remediation Problems"

ABSTRACT:
Recently a set of optimization problems was proposed in the environmental engineering literature for benchmarking purposes. We present a subset of these problems, describe some implementation approaches, and provide numerical results for a well-field design problem and a hydraulic capture problem. Both of these applications lead to objective functions that are nonsmooth and have local minima, hence making them challenging problems for traditional, gradient based optmization techniques.

C. T. Kelley
North Carolina State University

TITLE: “Implicit Filtering”

ABSTRACT:
Implicit filtering is a sampling method that uses difference gradients in a projected BFGS iterations. The size of the difference increment is reduced as the optimization progresses. In this talk we will discuss implementation decisions, similar decisions that must be made in all sampling methods, and the theoretical support for those decisions.

Alex Mayer
Michigan Technological University

TITLE: “Optimization of Engineering Design of Subsurface Environmental Remediation Systems- Development and Testing of Community Benchmark Problems”

ABSTRACT:
The U.S. EPA has estimated that remediation of contaminated soil and groundwater will cost on the order of tens of billions of dollars. Application of optimization has the potential to provide more efficient engineering solutions for remediation systems. Much work has been accomplished in the area of optimization of subsurface remediation systems in the last two decades. A wide range of optimization techniques- from traditional, gradient-based techniques to evolutionary algorithms- have been applied. However, one of the roadblocks to further success in the field is that there is little uniformity in the composition of the problems that are used by researchers when testing their optimization technique of choice. The consequence is that it is difficult to compare the results of different studies, and ultimately, to recommend the best optimization techniques to be used for a given type of problem.

In an attempt to overcome this difficulty, we have designed a set of systematic test problems to be attacked by the engineering and mathematics community, as a means for benchmarking and comparing optimization approaches. The test problems pose many of the difficulties anticipated in solving real-world problems such as (a) mixed continuous and integer, nonlinear objective functions, (b) the combination of boundary conditions and system parameters gives rise to complex relationships between the objective function, the decision variables, the constraints, and the state

variables, (c) evaluation of the objective function is based on solving model equations that are difficult to solve accurately and quickly; and (d) the number and range of decision variables is potentially enormous. Furthermore, the problem specifications have been designed to encourage, not restrict, innovation in optimal design and subsurface science. This talk will describe the design and execution of the test problems.

Dennis McLaughlin

Massachusetts Institute of Technology

TITLE: "New Models for New Times: Remote Sensing, Data Assimilation, and Coupled Problems"

ABSTRACT:
Environmental simulation models have long been used mostly to investigate the future consequences of alternative design and management options. Such models are generally intended to be faithful reproductions of reality and are based on abstract fundamental principles. Measurements of predicted variables are included in the modeling process primarily as aids to calibration. Recent developments in ubiquitous sensing (both remote and in situ) have opened up new possibilities and challenges for hydrologic modeling. The vast amounts of data becoming available are likely to reveal processes and behavior not previously observed or predicted. This has already occurred in oceanography and other earth sciences. New data sources may eventually make it possible to monitor the planetary environment in real time, at high resolution. Such developments suggest that environmental models can be viewed as data processing tools, useful primarily for extracting information from measurements. This viewpoint has significant implications for model design. In particular, the modeling process needs to be enlarged to consider the nature of the measurements and the scale of the problem as well as the physical/chemical/biological principles thought to govern the observed system. This will be illustrated with a few examples.

Cass T. Miller

University North Carolina

TITLE: "Pore-Scale Modeling for Closure of Multiphase Models Derived Using Thermodynamically Constrained Averaging Theory Approaches"

ABSTRACT:
Traditional models for multiphase flow in porous media are closed using empirical, ad hoc approaches. Evolving approaches are based upon a rigorous set of conservation principles and thermodynamically constrained closure approximations. These new models require new types of closure approximations. We explore the use of Lattice-Boltzmann models as a means of simulating multiphase models. Aspects considered include model forms, pore morphology, pore-scale simulation, pressure-saturation-interfacial area, curvatures, and viscous coupling.

Shlomo P. Neuman
University of Arizona

TITLE: "A Highly Efficient Conditional Moment Algorithm For Transient Flow In Random Porous Media"

ABSTRACT:
We present a highly efficient parallel computational algorithm for transient flow in random porous media of finite extent subject to uncertain forcing terms. The algorithm combines finite elements with numerical Laplace transform inversion to solve recursive approximations of otherwise exact nonlocal equations for the mean and variance-covariance of hydraulic head and flux. The random head and flux are nonstationary in space-time due to arbitrary forcing and conditioning on measured values of hydraulic conductivity at discrete points in space. Recursive approximation is carried to second order in log hydraulic conductivity conditional standard estimation error. The conditional moment solution compares well with "ground truth" Monte Carlo simulations of superimposed mean-uniform and convergent flows in a rectangular domain when log hydraulic conductivity exhibits large random fluctuations. The algorithm requires much less computer time than is needed for Monte Carlo statistics to stabilize, regardless of whether both solutions are computed sequentially or in parallel. The computational advantage of the parallel moment algorithm over parallel Monte Carlo simulations becomes more pronounced as grid size increases. Another computational advantage of the moment algorithm is that it allows analyzing a variety of flow scenarios by modifying forcing term statistics without recomputing Green's functions as long as the boundary configuration remains unaltered.

Linda Petzold
University of California, Santa Barbara

TITLE: "Adaptive Numerical Methods for Sensitivity Analysis of Differential-Algebraic Equations and Partial Differential Equations"

ABSTRACT:
Sensitivity analysis of differential-algebraic equation (DAE) systems generates essential information for design optimization, parameter estimation, optimal control, model reduction, process sensitivity and experimental design. Recent work on methods and software for sensitivity analysis of DAE systems has demonstrated that forward sensitivities can be computed reliably and efficiently. However, for problems which require the sensitivities with respect to a large number of parameters, the forward sensitivity approach is intractable and the adjoint (reverse) method is advantageous. In this talk we give the adjoint system for general DAEs and investigate some of its fundamental analytical and numerical properties. We describe our new adjoint DAE software and outline some issues which are critical to the implementation.

Defining the adjoint sensitivity system and writing the appropriate software to describe it can be a very challenging problem for large-scale engineering systems, particularly when it comes to finding appropriate boundary conditions for the adjoint partial differential equation (PDE) system. Therefore our goal for both DAE and PDE systems has been the development of

methods and software in which generation and solution of the sensitivity system are transparent to the user. This has been largely achieved for DAE systems. We will propose a solution to this problem for PDE systems solved with adaptive mesh refinement.

Thomas F. Russell
University of Colorado, Denver

TITLE: "Annoying Issues in Numerical Simulation of Subsurface Transport"

ABSTRACT:
It is an empirical fact that most practical numerical simulations of subsurface transport processes are carried out with "old" or "crude" discretization techniques. Most prominent among these is first-order upstream finite differences. When such methods are used in an inner simulation loop, their inaccuracy limits the usefulness of outer-loop optimization procedures. The persistence of these old techniques, despite their drawbacks, is largely due to the difficulty of formulating better methods that are robust, fast, and accurate. The presentation will outline the attributes of various old and new methods for transport, and will speculate on how optimization might be used in the design of better inner-loop transport-simulation methods.

Christine Shoemaker
Cornell University

TITLE: "Groundwater Optimization: Advantages and Disadvantages of Derivative-Based, Heuristic and Surface Approximation Methods"

This talk will review applications of a variety of optimization methods to groundwater remediation optimization, including a description of side-by-side comparisons of 7 alternative methods on a bioremediation problem that involves finite difference solution to a system of highly nonlinear partial differential equations.

I will also discuss our current research project on the use of surface approximation methods in combination with other optimization techniques. Such methods can significantly improve computational efficiency of algorithms for the optimization of a continuous function *f(x)* that is costly to evaluate. In groundwater remediation this function *f(x)* is the objective function that requires solution of the partial differential equations describing groundwater hydrology, reactive transport and microbe population growth. A nonlinear surface approximation *R(x)* is a continuous nonlinear multivariate approximation to *f(x)* based on evaluations of *f(x)* at a limited number of values of *x.* I will show that *R(x)* can be used as part of optimization algorithms in order to reduce the number of points at which we compute *f(x),* and thereby reduce computational cost. I will discuss the use of augmented radial basis function surface approximation methods to solve nonconvex deterministic optimization problems. The optimization search done on the response surface to select the next point *x* for evaluation of *f(x)* can then be based on a number of methods, including modern heuristic or gradient-based search procedures. Numerical results will be presented for nonconvex deterministic optimization in serial and parallel algorithms on test functions.

Mary Fanett Wheeler
Texas Institute for Computational and Applied Mathematics

TITLE: "Advanced Modeling Techniques for Numerical Simulation of Complex Subsurface Hydrosystems"

ABSTRACT:
We describe a methodolgy for simulating coupled flow, transport and reaction processes over large space and time scales in subsurface hydrosytems. As the relevant processes strongly differ in the various subdomains of a subsurface hydrosystem, different model concepts must be chosen for the subdomains, and special coupling methods must be applied to account for the interaction processes between these subdomains as well as for moving subdomain boundaries. Important examples include natural attenuation which is used in tens of thousands of contaminated sites in the United States in place of or in conjunction with engineering remediation systems, migration and storage of carbon-dioxide in the subsurface and atomic waste disposal sites.

Computational results demonstrating this methodolgy for multiphase flow and reactive transport are presented. In addition, current efforts in developing a posterioi error bounds are discussed.

Pamela J. Williams
Sandia National Laboratories

TITLE: "Using Nonlinear Interior Point Methods for Optimal Groundwater Contamination Containment"

ABSTRACT:
The design of optimal groundwater contamination containment strategies have received increased interest over the past decade as decision makers consider trade-offs between preventing further contamination through short-term containment strategies or adopting the wait-and-see approach while seeking complex long-term bioremediation strategies. Many researchers have used active-set methods, which are combinatorial in nature, to identify and analyze containment strategies. We propose the use of a nonlinear interior-point method to determine the optimal groundwater contamination containment strategy. An advantage of an interior-point method is that estimation of active inequality constraints is unnecessary. In our implementation, we have developed techniques to speed-up time to solution and to incorporate parameter sensitivity information into the finite-difference gradient calculations.

We will discuss the problem formulation, aspects of the interior-point method, and the interface to the groundwater flow and contaminant transport models. In addition, we will present numerical results and outline directions of future research.

Carol S. Woodward
Lawrence Livermore National Laboratory

TITLE: "Nonlinear Solution and Sensitivity Methods for Variably Saturated Flow"

ABSTRACT:
Simulation of water resource management problems often requires the solution of large problems with many spatial zones. In addition, effective use of simulation solutions requires knowledge of the uncertainty introduced into the solution by variances in problem data. Current techniques for obtaining this information can require many runs of the simulation code and can be very time-consuming, especially for large-scale problems. In this two-part talk, we will look at both solution methods for these problems as well as uncertainty quantification based on sensitivity analysis.
In part 1, we present nonlinear solvers for variably saturated porous media flow problems. In particular, we consider the set of nonlinear equations arising from a finite difference discretization of the time-dependent Richards' equation. The solvers make use of multigrid methods in various forms. We first consider solvers which use a Newton-Krylov approach for handling the nonlinearities, and apply multigrid as the linear system preconditioner. Multigrid is used in an attempt to achieve scalability. We compare the effectiveness of two multigrid preconditioning algorithms in the context of large-scale, three-dimensional domains with heterogeneous, discontinuous permeability fields. We will also present recent work on the development of nonlinear multigrid methods for similar problems.

In part 2, we will look at sensitivity analysis for these problems. Sensitivity analysis techniques give a way to compute solution uncertainties by using information on the sensitivities of the solution to various parameters. These sensitivities are just the solution derivative with respect to the parameter in question, and equations for them can be derived by differentiating the original model problem. The resulting sensitivity equation is linear and can be solved in tandem with the model equations. First order estimates of solution uncertainties can be developed from these sensitivities with a straightforward additional calculation.

This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

William W-G. Yeh
University of California, Los Angeles

TITLE: "The Inverse Problem of Parameter Structure Identification"

ABSTRACT:
This research proposes a combinatorial optimization scheme for solving the inverse problem of parameter structure identification. Identification of distributed parameters invariably involves a large number of unknowns. Because of data limitation, the formulated inverse problem is inherently non-unique and unstable. Parameterization reduces the degrees of freedom of the unknown distributed parameter and defines the spatial attributes of a distributed parameter in terms of parameter dimension, parameter pattern and parameter values. Traditional parameterization methods include zonation, finite element, as well as other interpolation schemes. In this research, we demonstrate the use of Voronoi tessellation for parameterization. Accordingly, the inverse problem seeks to identify the number and locations as well as the values of the basis points associated with the Voronoi tessellation. We first use a genetic algorithm to search for the near-optimal parameter pattern and values. We then use grid search and a quasi-Newton algorithm iteratively to improve the GA solution. We use either the sensitivity equations

or the adjoint state equations to calculate the gradient vector. We also discuss a proposed universal parameterization scheme that unifies various continuous and piece-wise constant (zonation) structures. Numerical experiments are conducted to demonstrate the proposed methodology.

James V. Zidek
University of British Columbia

TITLE: "Optimal Designs for Monitoring the Extremes of Environmental Processes"

ABSTRACT:
The Np hard problem of combinatorial optimization confronting the designer of an environmental monitoring network is very computationally demanding. Add to that the special technical and conceptual problems arising in the of measurement of extremes and one has a very challenging optimization problem indeed.

In this talk, we describe methods for augmenting an environmental monitoring network, in particular, one that uses entropy to sidestep the pervasive problem of ill-defined objectives. We demonstrate that approach by showing how to extend a monitoring network in Vancouver, that measures hourly small particulate airborne concentrations. Our approach uses a hierarchical Bayesian spatial predictive distribution for potential sites, conditional on all available data. We indicate how, in constructing that distribution, we over the challenges presented by high strong autocorrelations and incomplete data, that available having a staircase pattern due to different start times of existing network stations.

How well would that design solution work if site extremes such as the yearly 98th percentile of daily average levels were the metric of interest? Our predictive distribution allows us to explore that issue through simulation of the concentration field and discover difficulties, depending on how the assessment is made. We will propose a design strategy for extremes that gets around some of these difficulties along with an exact branch and bound algorithm for finding the optimal design.

G. *Program for One-Day Workshop on Porous Media*

May 16, 2003, NISS-SAMSI Building

9:00-9:45 *Closure Issues*
Greg Forest and Roberto Camassa

9:45-10:30 *Theory and APS experiments*
Katherine Culligan

10:30-11:00 Break

11:00-11:45 *Image Processing*
Dorthe Wildenschild

11:45-12:30 *Lattice Boltzmann Simulation*
Doris Pan

12:30-1:30 Lunch

1:30-2:15 *2D Micromodels*
Laura Pyrak-Nolte

2:15-3:00 *Pore Drainage*
Markus Hilpert

3:00-3:30 Break

3:30-4:15 *Level Set Methods*
David Adalsteinsson

4:15-5:00 Group - Plans for Future

H. Workshop on Spatial-Temporal Statistics Tentative Schedule
June 1-6, 2003, UCAR-Boulder, Colorado

- **Sunday, June 1**

7:00-9:00 Reception at the Millennium Hotel

- **Monday, June 2**

8:45-10:00 Robin Dennis, Environmental Protection Agency

10:00-10:15 Break

10:15-11:30 Noel Cressie, Ohio State University

11:30-1:00 Lunch and Posters

1:00-2:15 Brad Carlin, University of Minnesota

2:15-3:15 Francesca Dominici, John Hopkins University

3:15-3:30 Break

3:30-4:30 Robert Wolpert, Duke University

4:30-5:30 Richard Smith, University of North Carolina

- **Tuesday, June 3**

8:45-10:00 Mark Berliner, Ohio State University

10:00-10:15 Break

10:15-11:30 Chris Wikle, University of Missouri
Ralph Milliff, Colorado Research Associates

11:30-1:00	Lunch
1:00-2:15	Jeff Anderson, GFDL
2:15-3:15	Jay Breidt and Scott Urquhart, Colorado State University
3:15-3:30	Break
3:30-4:30	Richard Kleeman, Courant Institute of Mathematical Sciences
4:30-5:30	Montserrat Fuentes, North Carolina State University

- **Wednesday, June 4**

8:45-11:30	Gabriela Hegerl, Duke University Ben Santer, Lawrence Livermore National Laboratory Tom M. L. Wigley, National Center for Atmospheric Research
11:30-1:00	Lunch
1:00-2:15	Andrew J. Majda, Courant Institute of Mathematical Sciences
2:15-3:15	Richard McLaughlin, University of North Carolina
3:15	Late afternoon free with local hiking

- **Thursday, June 5**

8:45-10:00	Alan Gelfand, Duke University
10:00-10:15	Break
10:15-11:30	Doug Nychka, National Center for Atmospheric Research
11:30-1:00	Lunch
1:00-2:15	Marc Serre, University of North Carolina George Christakos, University of North Carolina
2:15-3:15	Michael Stein, University of Chicago
3:15-3:30	Break
3:30-4:30	Philippe Naveau, Unviersity of Colorado at Boulder Marc Genton, North Carolina State University

- **Friday, June 6**

8:30-12:00	Working groups

IV. EDUCATION & OUTREACH

A. *Program for One-Day Workshop*
November 9, 2002

8:30-9:00	Continental Breakfast
9:00-10:00	Welcome & SAMSI Overview Dr. Jim Berger, Director of SAMSI
10:00-10:30	Coffee Break
10:30-11:30	Inverse Problems in Science & Engineering Dr. H.T. Banks, N.C. State & SAMSI
11:30-1:00	Lunch and Informal Discussions
1:00-2:00	Stochastic Computation Dr. Mike West, Dr. Merlise Clyde & Dr. Mark Huber, Duke
2:00-2:30	Coffee Break
2:30-3:00	The Statistics of Climate Change Dr. Richard Smith, UNC-Chapel Hill
3:00-3:30	Porous Medium Science: Mathematical and Statistical Challenges Dr. Casey Miller, UNC-Chapel Hill

B. *Program for One-Day Workshop*

February 1, 2003

8:30-9:00	Continental breakfast
9:00-10:00	Welcome and Overview of SAMSI-its purpose, goals, and opportunities Dr. Alan Karr, Associate Director-SAMSI & Director-NISS
10:00-10:30	Break
10:30-11:30	Inverse Problems in Science and Engineering Dr. H.T. Banks, North Carolina State University
11:30-1:00	Lunch and informal discussions between guests, faculty and SAMSI students, postdocs
1:00-2:00	Stochastic Computation Dr. Mike West, Dr. Merlise Clyde, and Dr. Mark Huber, Duke University
2:00-2:30	Break
2:30-3:00	Assessing Uncertainty in Weather Prediction

Dr. Montserrat Fuentes, North Carolina State University

3:00-3:30 Multi Scale Phenomena: Mixing and Entrainment in Stratified Fluids
Dr. Richard M. McLaughlin, University of North Carolina-Chapel Hill

APPENDIX C – Workshop Evaluation Summaries

Workshop participants were given an evaluation questionnaire to complete in each of the SAMSI workshops. A sample questionnaire is given on the following page.

Below are the summaries of the participant evaluations for the four main scientific workshops held to date. The rating scale was 1-5 (lowest to highest). The five questions addressed in the tables were:

1. Scientific Quality
2. Staff Helpfulness
3. Meeting Room/AV Facilities
4. Lodging
5. Local Transportation

Inverse Workshop
H.T. Banks, Leader
Sept. 21-24, 2002
Total Responses: 55
Total Participants: 121

Item	Rating 1	2	3	4	5	N/A
1		1	3	22	27	2
2			1	11	42	1
3			1	20	34	
4			3	12	26	14
5	3	2	7	9	13	21

StoCom Workshop
Mike West, Leader
Sept. 25-Oct. 1, 2002
Total Responses: 38
Total Participants: 134

Item	Rating 1	2	3	4	5	N/A
1			3	14	21	
2			2	7	29	
3		1	7	11	19	
4		2	7	6	12	11
5	1	7	6	8	7	9

ENV Multiscale Workshop
Richard Smith, Leader
Feb 2-7, 2003
Total Responses: 25
Total Participants: 84

Item	Rating 1	2	3	4	5	N/A
1				7	16	2
2				5	20	
3	1	2	3	12	7	
4		1	5	8	5	6
5	1	6	4	2	5	7

ENV Optimization Wrkshp
C.T. Kelley, Leader
Apr 28-30, 2003
Total Responses: 14
Total Participants: 49

Item	Rating 1	2	3	4	5	N/A
1				6	8	
2			1	2	11	
3			1	4	9	
4				3	4	7
5			2	2	3	7

APPENDIX D – Course Descriptions for 2002-2003

I. *Inverse Problem Methodology in Complex Stochastic Models* ***(Fall 2002)***

A special topics course will be offered this fall in connection with the SAMSI Program on "Inverse Problem Methodology in Complex Stochastic Models," described in a previous email. Detailed background about the rationale for this course follows the course description.

MA/ST 810Q-003-- Inverse Problem Methodology in Complex Stochastic Models (3 credit hours. S/U)

Official course description: Overview of mathematical modeling of complex dynamical systems and statistical considerations for fitting these models to data, with emphasis on the interface between applied mathematics and statistics. Ordinary and partial differential equation modeling of biological, physical, and other phenomena; time, space, and state-dependent coefficient estimation; nonlinear mechanisms; identifiability, ill-posedness, stability, and regularization in inverse problems; and computational methods. Fundamentals of statistical modeling and inference; sources of variation in complex data structures; frequentist and Bayesian methods for parameter estimation and assessment of uncertainty. Mathematical and statistical topics will be integrated through discussion of numerous examples from the biological and physical sciences and engineering.

Prerequisites:

For mathematics students: Undergraduate-level differential equations (required), MATLAB or similar computing experience (desirable), Numerical analysis (desirable).

For statistics students: Statistical inference (at the level of Casella and Berger; required), Linear models theory (required), Splus, R, MATLAB, or similar computing experience (desirable).

(Mathematics students need not have statistics student prerequisites, and vice versa.)

The course will meet at a time to be announced in the NISS building in Research Triangle Park.

Please direct questions to Marie Davidian (davidian@stat.ncsu.edu).

Background: A current "hot topic" in biological sciences, physical sciences, engineering, epidemiology, and a host of other areas is the use of complex nonlinear dynamical systems models to characterize phenomena of scientific interest. These models involve (mostly) deterministic systems of ordinary and partial differential equations that are used to characterize interactions among various components of a biological, physical, or other system. For example, there is much current interest in using these models to describe known and hypothesized mechanisms involved in the interplay over time between the

HIV virus and the immune system within organisms (like humans), understanding of which has the potential to suggest new treatment strategies. In general, such models allow a formal description of mechanisms in terms of parameters (some of which may be functions rather than real-valued). In the HIV example, parameters might include viral clearance rates, immune cell infection rates, and new virus production rates.

The solution of the system of differential equations implies an expression (that is likely not available in a closed form) for quantities of interest at any time point. For example, a system to model HIV dynamics determines expressions for the concentration of virus (or "viral load"), concentration of T-cells (an immune system component), etc., present in the organism at any time following administration of antiretroviral therapy. Given the values of the parameters that describe the system and some initial conditions, evaluating such expressions must in general be carried out via complex numerical techniques; this is referred to as "forward solution."

As development of these models and computational strategies for evaluating them have improved, there is great interest in applying them to data. For example, data on "viral load" and measures of immune status may be collected longitudinally on one or more organisms; that is, values of the system at certain time points have been observed (possibly subject to variation due to measurement error and other sources). From such data, it is of interest to learn about the underlying mechanisms (values of parameters). Thus, for application to data, the problem is the reverse of "forward solution:" given (possibly error-prone) values taken on by system components at known time points, can the values of the parameters be recovered, knowledge of which may provide insight into the underlying mechanisms? Recovering the values of the parameters is a form of what is known in this literature as an "inverse problem." Of course, statisticians would call this "parameter estimation," and would be concerned about taking proper account of the various sources of variation, development of valid inferential strategies, and efficiency of estimation. This requires that the complex dynamical systems be embedded in a statistical framework appropriate to the data structure.

The key feature of applying complex dynamical systems to data in this way is that they inherit the computational challenges of both the mathematical and statistical elements. For example, to fit a statistical model in which the mathematical model has been incorporated via maximum likelihood optimization or Bayesian computational techniques, evaluation at any time point requires the forward solution to be computed at the current value of the parameters. How to resolve the joint computational challenge is part of the focus of the SAMSI Inverse Problem program and is an area where new research directions will be established.

This course is meant to give statistics and applied mathematics students the necessary background to appreciate both the "big picture" and some of the specific mathematical and statistical challenges in this endeavor. For statistics students, the course will serve as an introduction to differential equation modeling, assuming no prior exposure, focusing on the aspects that are most important for statisticians to understand. For applied mathematics students, the course will serve as an introduction to statistical inference and considerations involved in statistical modeling (e.g. the need to take appropriate account of multiple sources of variation in complex data structures). These topics will be highlighted by discussion of subject-matter examples from numerous fields of application.

The diversity of the audience will provide a unique forum for students to learn about this area and relevant applications and to understand the perspectives of applied mathematicians and statisticians and the terminology each group uses.

II. Environmental Statistics (Spring 2003)

This is an advanced graduate course that covers many of the statistical methodologies used in the environmental sciences, including spatial statistics, time series analysis and extreme value theory applied in such areas as air pollution modeling, global climate change and the prediction of extreme environmental events such as floods and hurricanes. The course will be largely based on the material in a 10-lecture course given at the University of Washington in June, 2001, under the auspices of the Conference Board of the Mathematical Sciences.

The course during Spring 2003 will be given one lecture each week, on Wednesdays 3:00 pm - 6:00 pm, at the National Institute of Statistical Sciences, 19 Alexander Drive, Research Triangle Park. The course is part of the SAMSI program on Large Scale Computer Models for Environmental Systems.

The course has no specific prerequisites but a graduate-level knowledge of linear models and regression analysis, at the level of Statistics 174, will be assumed. Some familiarity with other methods of applied statistics, such as design of experiments, time series and multivariate analysis, will be valuable but is not assumed. Assessment will depend on the number of participants in the course but is likely to be through either a term paper or a presentation to the class. There will not be regular homeworks or a formal written exam.

III. Multiphase Transport Phenomena (Spring 2003)

A. Summary

This course is focused on the development of models to describe transport phenomena in multiphase porous medium systems. Of special interest is systems that involve more than one fluid phase and a solid phase. Such systems occur routinely in both engineered and natural contexts. Porous medium systems lying below the Earth's surface (e..g., soil, groundwater, petroleum reservoir systems) are of special interest. Conservation of mass, momentum, angular momentum, energy, and entropy are considered for phases and species in volumes, interfaces, common lines, and common points. Multiphase thermodynamics are developed and used to constrain the development of closure relations. Models are developed across a range of scales, and recent computational and experimental approaches to close these models are addressed.

This course will be offered at the Statistical and Applied Mathematical Sciences Institute (SAMSI) in Research Triangle Park. The content is suitable for a wide range of scientists, engineers, mathematicians, and statisticians interested in the formulation of conservation principle-based models to describe transport phenomena in complex systems.

B. General Information

Instructors:	W.G. Gray and C.T. Miller
Office	SAMSI, Research Triangle Park
Office Hours	By Appointment
Office Phone	919-685-9350
E-mail	wggray@nd.edu and casey miller@unc.edu
Course Schedule	Tuesdays 4:30–7:00 pm at SAMSI, RTP
Course Materials	blackboard.unc.edu

C. Grading Bases

Grades will be based upon a course project. The format of this project is described below.

D. Course Objectives

Objectives of the course are:

- To develop an understanding of the assumptions and limitations of standard approaches used to formulate mathematical models to describe the conservation of mass, momentum, and energy in multiphase systems;
- To develop an understanding of evolving methods of continuum mechanics that may be used to formulate conservation-based models of multiphase systems that are more rigorous than standard approaches;
- To explore methods that may be used to guide the development of the general form of constitutive relations needed to close conservation equation formulations of multiphase systems; and
- To investigate evolving computational and experimental approaches to advance specific constitutive relation forms and parameter dependencies needed to close multiphase models.

E. Background Required

The subject matter of this course is broad and quantitative in nature. A good background in mathematics through partial differential equations, vector and tensor calculus, calculus-based physics, and fluid mechanics are needed. A background in statistics, thermodynamics, and physical chemistry would be helpful. Rarely does a student have a complete grounding in each of these areas. Thus, it is usually necessary for all students to do some background reading in areas in which their background is deficient for the material covered in this course.

F. Course Reference Materials

Because the focus of this course is on evolving methods for modeling multiphase flow, a suitable book does not yet exist. Therefore, other forms of reference materials will be relied upon, including periodic course notes, reprints, reports, computer codes, and literature references. These materials will be placed on blackboard to facilitate electronic distribution.

G. Course Project

The course project will consist of original independent work done by the student in an area related to the course. Topics will be arrived at mutually by the student and the instructors. The instructors will supply a list of potential topics, and each student may tailor a topic to an area of special interest. A suitable project will investigate some aspect of evolving models of multiphase systems: model formulation, constitutive equation determination for model closure, model solution methods, or model analysis. The subject material of the course is rich in potential topics, many of which have not yet been investigated or published in the literature. The course project should be typed in a manuscript format, including abstract, introduction, background, formulation, results, discussion, conclusions, and reference sections. Further guidance will be provided.

F. Course Outline

1. Standard Approaches for Modeling Multiphase Systems
2. Fundamental Aspects of Porous Medium Systems
3. Microscale Modeling: I. Conservation Equations for Volumes
4. Microscale Modeling: II. Conservation Equations for Interfaces, Curves, and Points
5. Microscale Modeling: III. Entropy Inequality and Model Closure
6. Macroscale Modeling: I. Conservation Equations for Volumes
7. Macroscale Modeling: II. Conservation Equations for Interfaces, Curves, and Points
8. Macroscale Modeling: III. Thermodynamic Considerations
9. Macroscale Modeling: IV. Geometric Considerations
10. Macroscale Modeling: V. Single-Phase Fluid Flow and Transport
11. Macroscale Modeling: VI. Closure for Two-Phase Fluid Flow
12. Multiphase Modeling: VII. Closure for Two-Phase Fluid Flow and Species Transport
13. Closure Relation Approaches: I. Microscale Modeling
14. Closure Relation Approaches: II. Pore-Scale Modeling and Evolving Experimental Approaches