# SAMSI 2006-07 Program on Development, Assessment and Utilization of Complex Computer Models

# Final Report

M.J. Bayarri
Chair, Program Leaders
On behalf of the Working Groups Leaders

March 2008

## Contents

**1 Introduction** **1**

**2 Program Organization** **2**

2.1 Subprograms . . . 2

2.2 Working Groups . . . 4

**3 Research Goals and Activities** **8**

3.1 Program Level . . . 8

3.2 Air Quality . . . 9

3.2.1 Introduction. Research Goals. . . . 9

3.2.2 Specific activities. . . . 10

3.3 Calibration of Computational Models of Cerebral Blood Flows . . . 11

3.3.1 Introduction. Research Goals. . . . 11

3.3.2 Specific activities. . . . 12

3.4 Climate and Weather . . . 13

3.4.1 Introduction. Research Goals. . . . 13

3.4.2 Specific Activities . . . 13

3.5 Dynamics of Infectious Diseases . . . 15

3.5.1 Introduction. Research Goals. . . . 15

3.5.2 Specific Activities . . . 16

3.6 Engineering Methodology . . . 17

3.6.1 Research Goals. . . . 18

3.6.2 Specific Activities . . . . . . . . . . 18
3.7 Granular Materials - Engineering Applications . . . . . . . . . . 19
3.7.1 Introduction. Research Goals. . . . . . . . . . . 19
3.7.2 Specific Activities . . . . . . . . . . 19
3.8 Inference and Uncertainty Analysis of Hydrological Models . . . . . . . . . . 20
3.8.1 Introduction . . . . . . . . . . 20
3.8.2 Research Goals . . . . . . . . . . 21
3.8.3 Research Activities and Current State of Work . . . . . . . . . . 22
3.9 Methodology . . . . . . . . . . 23
3.9.1 Introduction. Research Progress. . . . . . . . . . . 23
3.9.2 Specific Activities . . . . . . . . . . 25
3.10 Statistical Mechanics of Granular Flow . . . . . . . . . . 27
3.10.1 Introduction. Research Goals. . . . . . . . . . . 27
3.10.2 Specific Activities . . . . . . . . . . 28
3.11 Systems Biology . . . . . . . . . . 28
3.11.1 Introduction. Research Goals. . . . . . . . . . . 28
3.11.2 Specific Activities . . . . . . . . . . 29
3.12 Terrestrial Models . . . . . . . . . . 29
3.12.1 Introduction. Research Goals. . . . . . . . . . . 30
3.12.2 Specific Activities . . . . . . . . . . 30

**4 Publications** **32**
4.1 Special Technometrics Issue . . . . . . . . . . 32
4.2 Published Papers . . . . . . . . . . 32
4.3 Submitted Papers and Technical Reports . . . . . . . . . . 35
4.4 Work in Progress . . . . . . . . . . 36

**5 Workshops** **38**

**6 Other Activities** **42**
6.1 Research sessions, working get-together, brainstorming sessions ... . . . . . . . . . . 42
6.2 Submitted Research Projects . . . . . . . . . . 43
6.3 Other Collaborative Research and tentative Projects . . . . . . . . . . 44

**7 Education and Outreach** **47**
7.1 Credit Courses . . . . . . . . . . 47
7.2 Workshops for Students and New Investigators . . . . . . . . . . 47
7.3 Presentations at Professional Meetings . . . . . . . . . . 48
7.4 Seminars . . . . . . . . . . 53

7.5 Graduate students. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56

# 1 Introduction

Mathematical models intended for computational simulation of complex real-world processes are a crucial ingredient in virtually every field of science, engineering, medicine, and business, and in everyday life as well. Cellular telephones attempt to meet a caller's needs by optimizing a network model that adapts to local data, and people threatened by hurricanes decide whether to stay or flee depending on the predictions of a continuously updated computational model.

Two related but independent phenomena have led to the near-ubiquity of models: the remarkable growth in computing power and the matching gains in algorithmic speed and accuracy. Together, these factors have vastly increased the applicability and reliability of simulation—not only by drastically reducing simulation time, thus permitting solution of larger and larger problems, but also by allowing simulation of previously intractable problems.

The intellectual content of computational modeling comes from a variety of disciplines, including statistics and probability, applied mathematics, operations research, and computer science, and the application areas are also remarkably diverse. Despite this diversity of methodology and application, there are a variety of common challenges - detailed below - in developing, evaluating and using complex computer models of processes, which directly relate to the mission of SAMSI.

# 2 Program Organization

## 2.1 Subprograms

The study of computer models needs to take place in the context of actual computer models. But because of the inherent complexity of computer models, and the very different types of such models, this SAMSI program is articulated in sub-programs, focusing on specific computer modeling applications. This approach allows in-depth exploration of specific types of computer models, while maintaining an overall 'SAMSI umbrella' that allows quick transfer of techniques developed in one sub-program to another. The following subprograms were conducted during the year.

**Environmental/Ecological/Climate Models Subprogram.** The environmental modeling subprogram dealt with problems and research fields at the interface between statistics and environmental modeling. These include Problems of model calibration in the presence of structural model deficits and input uncertainty, problems of decision-oriented model application under high uncertainty about model structure and parameter values, and problems of universality or transferability of environmental models.

This Subprogram was led by Peter Reichert (EAWAG. It had three distinctive working groups: *Air Quality, Climate and Weather*, and *Terrestrial Models*. A fourth working group, *Hydrology*

*Models*, was also started, but completely merged with the *Methodology* working group.

**Subprogram on Uncertainty in Models of Granular Materials: Sources and Consequences.** Granular materials are ubiquitous. This subprogram aimed to develop a better understanding of the variability that appears in - indeed, often dominates - the observed behavior of granular materials during flow and deformation. The field exhibits rich yet poorly understood physics. For example, today there is no first- principles explanation of the creation and breaking of force chains; there is no theory of the propagation of sound in a granular medium. There is a need for a working theory of the behavior of granular materials that can describe practical applications, such as a theory of materials handling, of bin loading, of granular avalanches and pyroclastic flows. The segregation of granular materials, by size, shape, density or composition, is but one example of a problem of fundamental physics with enormous practical application.

There is active research in physics and in several branches of engineering on questions of granular material flow and deformation. Yet in many ways the field is in its infancy, not unlike the theory of fluid dynamics at the time Navier, Boltzman, and Stokes. This Subprogram broadened and deepened the discussion among physicists, statisticians, mathematicians and engineers, in pursuing new ideas to describe granular materials. More specifically, the Subprogram was ideal for interactions of statisticians who engage: experimental physicists and engineers examining the significant role of fluctuations in granular deformation; experimental physicists and mathematicians investigating segregation of granular materials; engineers and mathematicians analyzing and computing macro-scale mathematical models of granular flow to understand uncertainty in those models.

This Subprogram was led by Bruce Pitman (U. Buffalo). It formed two working groups, one on *Engineering Applications of Granular Materials*, and the second on the *Statistical Mechanics and Physics of granular flows.*

**Engineering Subprogram.** The engineering subprogram studied frequently-occurring problem areas in finite-element and other engineering models. Specifically, it focussed on the problems of Validation, Calibration, and Combining Data from physical experiments and computer experiments. The emphasis was on applications where the computer models require substantial running times and the physical models are difficult or expensive, so that, in some cases, physical experiments can be conducted for only subcomponents of the desired system or a physical simulator may only be possible for the desired system. Issues of combining codes from system components to produce valid codes for the entire system can then arise. The design of both the physical and computer experiments was also and issue of special interest.

This Subprogram was led by Tom Santner (Ohio State U.). It had one working group: *Engineering Methodology.*

**Biological Modeling Subprogram.** This Subprogram focused on three types of biological models. The first on models to predict cerebral blood flow. As a first step, a fluid dynamic model for the Circle of Willis was developed; improvements were investigated, and boundary conditions were carefully considered. Model calibration based on partial data was undertaken.

A second focus of the program was on system biological models. Models used in systems biology range from small biochemical networks modeled with a set of coupled ODEs that can be simulated quickly on a standard PC, through to large spatio-temporal models of a whole cell (or cell population) behaviour run on large computing facilities. Some models are deterministic, while others are intrinsically stochastic, giving different output on each run. All typically contain uncertain parameters that must be estimated from sparse, noisy experimental data. Additionally, there is often uncertainty regarding model structure. Particular problems that arise in the context of systems biology models include: estimating large numbers of parameters from sparse data, parameter estimation using complex multivariate data, simultaneous estimation of model parameters and structure, and estimating parameters of complex stochastic models.

Finally, models for the dynamics of infectious diseases were contemplated. In particular, for the impact of drug therapy and resistance on acute viral infections. These models are based on a multi-scale approach, integrating within-host models (i.e. ones that describe infection within a given individual) with between-host (epidemiological) models that describe the spread of infection at the population level. Numerous questions regarding fitting these models to data, validating the models and using them for assessment of the spread of viral infection were contemplated.

This Subprogram was led by Darren Wilkinson (U. of Newcastle). It had three active working groups: *Calibration of Computational Models of Cerebral Blood Flow, System Biology* and *Dynamics of infectious diseases.*

**Methodology Subprogram.** This Subprogram engaged in an in-depth treatment of methodological issues that arise in the design, analysis and utilization of computer models across many fields of application. This Subprogram evolved in close collaboration with the four disciplinary subprograms (Environmental/Ecological Models, Engineering Models, Uncertainty in Models of Granular Materials, and Biological Modeling), engaging them in an overall research umbrella.

In trying to predict reality (with uncertainty bounds), some of the key issues that arose were: use of model approximations (emulators) as surrogates for expensive simulators, for calibration/prediction tasks and in optimization or decision support; dealing with high-dimensional input spaces; validation and utilization of computer models in situations with very little data, and/or functional (possibly multivariate) outputs; non-homogeneity, including jumps and phase changes on the output as we move around the input space; implementation and transference methodology to current practice; efficient MCMC algorithms and prior assessments; optimization and design.

This Subprogram was led by M.J. Bayarri. It had one working group: *Methodology*, with considerable overlap with the rest of working groups. The working group for *Hydrology models* entirely merged with the Methodology working group.

## 2.2 Working Groups

The active working groups are met weekly throughout the year to pursue their particular research topics identified in the kickoff and posterior workshops and/or subsequently chosen by the working group participants. Some few working groups had their activity concentrated in a shorter period of time. The working groups consisted of SAMSI visitors, postdoctoral fellows, graduate students, and local faculty and scientists. A number of working group members did not reside in SAMSI nor in the area, and took active part on the meetings via teleconferencing and Webex access. The working groups had active web pages in which material, notes, agendas and members were posted.

**Air Quality** working group was led by Serge Guillas (Georgia Institute of Technology), and it focused on the study of Air quality computer models. Statisticians and an EPA scientist, who run and studied the Air Quality computer model CMAQ, were actively involved. Permanent participants included Serge Guillas (Georgia Tech), Chungsheng Ma (NCSU), Daiwen Kang (EPA), Fei Liu (Duke U.), Daiwen Kang (EPA).

**Calibration of Computational Models of Cerebral Blood Flows** The working group was led by Pierre Gremaud (NCSU). It focused on the study and calibration of models describing cerebral blood flows. The working group consisted of applied mathematicians, statisticians and physicists; it also worked closely with experimentalists (group of Dr. Novak, Beth Israel Deaconess Medical Center, Harvard University).

The permanent participants were Kristen DeVault (NCSU), Pierre Gremaud (NCSU), Mette Olufsen (NCSU), Guillaume Vernières (SAMSI postdoc/UNC) and Darren Wilkinson (Newcastle University).

**Climate and Weather** working group, led by Montserrat Fuentes (NCSU), Steve Sain (NCAR) and Jonathan Rougier (U. Bristol). This working group focused on problems related to climate and weather physical models, and their use and limitations for short and long term forecast.

The participants were: Jeff Anderson (NCAR), Howard Bondell (NCSU), Tsui-Long Chen (NCSU), Jim Crooks (SAMSI), Mike Dietze (Harvard U.), Dorin Drignei (U. Oakland), Montse Fuentes (NCSU), James Gattiker (U. Southampton, now LANL), Serge Guillas (Georgia Tech, now UCL), Josh Hacker (NCAR), Cari Kaufman (SAMSI and NCAR), H. Lee (USC), Chunsheng Ma (NCSU), Pablo Mininni (NCAR), Nancy Nichols (U. Reading), Astrid Maute (NCAR), Shane Reese (BYU), Brian Reich (NCSU), Arthur Richmond (NCAR), Jonathan Rougier (U. Bristol), Steve Sain (NCAR), Christine Shoemaker (Cornell U.), Justin Shows (NCSU), Randy Sitter (SFU), Elaine Spiller (SAMSI), Leonard Smith (Oxford and LSE), Robert Tardif (NCAR), Guillaume Vernieres (UNC).

**Dynamics of Infectious Diseases.** This working group was led by H.T. Banks (North Carolina State University), Ariel Cintrón-Arias (SAMSI/NCSU), and Alun Lloyd (NCSU). Its focus was on longitudinal data acquisition and analysis, population and within-host models, and statistical and mathematical methodologies.

The participants were Tom Banks (NCSU), Jeff Borggaard (Virginia Tech), John Burns (Virginia Tech), Adam Childers (Virginia Tech), Ariel Cintrón-Arias (SAMSI/NCSU), Gene Cliff (Virginia Tech), Cammey Cole (Meredith College), Jim Crooks (SAMSI), Marie Davidian (NCSU), Jimena Davis (NCSU), Sava Dediu (NCSU), Stacey Ernstberger (NCSU), Sarah Grove (NCSU), Shuhua Hu (NCSU), Abdul Jarrah (Virginia Tech), Sarah Lynn Joyner (NCSU), Grace Kepler (NCSU), Reinhard Launberbacker (Virginia Tech), Alun Lloyd (NCSU), Henning Mortveit (Virginia Tech), Golnar Newbury (Virginia Tech), Betty Paredes-Alvarez (Virginia Tech), Carlos Rautenberg (Virginia Tech), Peter Reichert (SAMSI), Johnny Samuels (NCSU), Daniel Sutton (Virginia Tech), Karyn Sutton (Arizona State), Alan Veliz-Cuba (Virginia Tech), Paola Vera-Licons (Virginia Tech), Lizette Zietsman (Virginia Tech).

**Engineering Methodology** The Engineering Methodology Working Group was led by Thomas Santner (the Ohio State University and SAMSI). This group worked on problems that occur in physics-based computer models. Among the features that make these models challenging are their long running times, the possibility of parallel physical (or near-physical) experiments, the presence of calibration and tuning parameters in the codes, and the nature of their output which can range from (easily-studied) real-valued output to (competing) multivariate ones to functional output.

The participants of this working group were Don Bartel (Cornell University), Dianne Bautista (The Ohio State University), Susie Bayarri (University of Valencia), Thomas Bengtsson (Bells Laboratories), Tiangang Cui (University of Auckland), Ian Dinwoodie (Duke University), Chris Gotwalt (SAS Institute), Genetha Gray (SANDI Laboratories), Eitan Greenshtein (N.Carolina State University/SAMSI), Gang Han (The Ohio State University), Dave Higdon (Los Alamos National Laboratories), Ying Hung (Georgia Tech), Herbie Lee (UC-Santa Cruz), Simon Lunagomez (Duke University/SAMSI), Abhyuday Mandal (University of Georgia), Scott Mitchell (Sandia Laboratories), Max Morris (Iowa State University), Nancy Nichols (Reading University), Abani Patra (University of Buffalo), Angie Patterson (General Electric), Mark Perry (LSE) Bruce Pitman (University of Buffalo), Grant Reinman (Pratt & Whitney), Jerry Sacks (NISS), Thomas Santner (The Ohio State University), Randy Sitter (Simon Fraser University), Curtis Storlie (NC State University), Laura Swiler (Sandia Laboratories), Matt Taddy (UC at Santa Cruz), Shih-Chung Tsai (GM Corporation), Gentry White (NC State University), and Henry Wynn (LSE).

**Granular Materials - Engineering Applications** working group was led by E. Bruce Pitman (University at Buffalo). This group was interested in implementing Bayesian methodology to applications involving granular materials. One of the principal applications is the hazard risk assessment of landslides or granular avalanches. Another application is to the forces and loads on the walls of a bin or hopper that stores granular material.

The group on engineering applications included Susie Bayarri (U. Valencia/SAMSI), J. Berger (SAMSI), Dorin Drignei (Oakland U), Michael Goldstein (Durham U.), Nancy Nichols (Reading U.), Abani Patra (U. Buffalo), Luis Pericchi (U Puerto Rico), Bruce Pitman (Buffalo), Michael Shearer (NCSU), Robert Wolpert (Duke), post-doc Elaine Spiller (Duke) and students Jennifer Joyce (UNC) and Simon Luna-Gomez (Duke); student Dalbey participated from Buffalo, and faculty member Calder in Buffalo contributed regarding the specific application to volcanic eruptions at Montserrat.

**Inference and Uncertainty Analysis of Hydrological Models** working group, led by Peter Reichert (Eawag), focused on calibration and uncertainty analysis of hydrological watershed models. A simple hydrological model was used to be more efficient in developing techniques designed for more complex (and slow) hydrological and, more generally, ecological, models. Two main topics were approached: (a) moving bias description from output to the internal model structure with the aid of time-dependent parameters; (b) emulating dynamic hydrological models.

Due to the large number of working groups and the methodological focus of the work, the group merged with the methodology working group and worked closely together with the engineering methodology working group.

**Methodology** working group led by Susie Bayarri (U Valencia/SAMSI) and Robert Wolpert (Duke U) worked across multiple working groups to identify methodological issues deserving investigation, to work on the calibration and validation of models, as well as to address challenging issues in the implementation of large and complex models with multiple sources of uncertainty.

Participants were Dianne Bautista (OSU), Susie Bayarri (U Valencia/SAMSI), Sunyoung Bu (UNC), Tsuei-Long Chen (NCSU), Ariel Cintron-Arias (SAMSI), James Crooks (SAMSI), Tiangang Cui (Auckland U), Michael Goldstein (Durham U), Genetha Gray (Sandia), Eitan Greeshtein (SAMSI/Duke U), Serge Guillas (Georgia Tech), Gang Han (OSU), Leanna House (Durham U), Mark Huber (Duke U), Ying Hung (Georgia Tech), Cari Kaufman (SAMSI/NCAR), Herbie Lee (UCSC), Fei Liu (Duke U), Tom Loredo (Cornell U), Simon Lunagomez (Duke U), Chunsheng Ma (NCSU), Max Morris (Iowa State U), Nancy Nichols (U Reading), Tony O'Hagan (U Sheffield), Abani Patra (U Buffalo), Rui Paulo (Polytechnic U Lisbon), Luis Pericchi (U Puerto Rico), Bruce Pitman (U Buffalo), Zhiguang Qian (U Wisconsin), Jonathan Rougier (U Bristol), Tom Santner (OSU), Christine Shoemaker (Cornell U), Leonard Smith (Oxford U and LSE), Elaine Spiller (Duke U), David M. Steinberg (Tel Aviv U), Curtis Storlie (NCSC), Matt Taddy (UCSC), Gentry White (NCSU), Darren Wilkinson (Newcastle U), Robert Wolpert (Duke U), Henry Wynn (LSE).

**Statistical Mechanics of Granular Flow** working group was led by Sorin Mitran (UNC). This working group was interested in applying statistical and statistical mechanical approaches to new discoveries in the physics of granular materials. Using simulation techniques to complement experiments, the topics under investigation included stresses on particles in a simple Couette cell assembly, and the creation, propagation, and break-up of void regions in shear cells.

The statistical physics working group included Jennifer Joyce (UNC), Tom Loredo (Cornell U.), Bob Behringer (Duke), Peter Mucha (UNC), Karen Daniels (NCSU), Bruce Pitman (U. Buffalo), Sorin Mitran (UNC).

**Systems Biology** Systems Biology is an exciting new area of scientific research, made possible only in recent years due to parallel developments in computational and experimental disciplines. In particular, advances in molecular biology technology has enabled study of cellular function and dynamics at levels of detail previously impossible. This has given new insight into biochemical network dynamics, and opens up the possibility of developing computational models of key aspects of cellular function. Such models have huge potential for furthering our understanding of biological processes; among other things, functional foods, rational drug design, personalised medicine and combination therapies are all targets for systems biology approaches.

Participants were Darren Wilkinson (Newcastle), Richard Boys (Newcastle), Richard Yamada (Cornell), Jarad Niemi (Duke), Lynn Kuo (U. Conn.), Michael Breen (EPA), Miyuki Breen (NC State), Kevin Brown (Harvard), Samuel Kou (Harvard), Daniel Henderson (Newcastle), Kristen Dang (UNC), Susie Bayarri (U. Valencia).

**Terrestrial Models** working group led by Jim Clark (Duke) concentrated on linking seasonal climate/weather variation to models of forest dynamics, including gaining understanding in how this intermediate scale variation in climate affects biodiversity and ecosystem processes. Activities have focused on parameterization and analysis of a large forest simulator.

Jim Crooks has developed an emulator for stochastic simulators, the specific application being the large forest simulator. The emulation method, based on the Kernel Stick-Breaking Process model of Dunson and Park, is extremely flexible and is able to infer highly non-Gaussian output distributions without a large number of simulator runs. Crooks presented this method at the 9th Case Studies in Bayesian Statistics workshop in October 2007 and at the Environmental Protection Agency - RTP in March 2008. The air quality modeling group at EPA was extremely interested in the emulator in particular and in statistical methods for addressing various other issues in computer experiments. Given funding Crooks will be working with this air quality group full time starting in July.

Cari Kauffman and Jonty Rougier have been building a model to predict the spatial and temporal evolution of soil moisture under varying climatic conditions. Soil moisture is an important factor in tree growth and fecundity, and these predictions are part of the Clark group's larger goal of studying the effect of climate change on forest dynamics. Our approach has been to build a state-space model in which the temporally evolving soil moisture field is the unknown state vector. Our initial predictions used a simple linear model, which could be fit using a Kalman filter, but they did not take advantage of valuable prior knowledge about the evolution of soil moisture over time. In our new model, this process is governed by a simple, catchment-level hydrology model we developed based on the literature, combined with a stochastic model for distributing total water according to local topography. We are now developing an MCMC algorithm for sampling the non-time-varying parameters, with an embedded smoothing algorithm to sample the temporally evolving soil moisture fields conditional on these parameters.

Permanent participants included Jim Clark (Duke), Cari Kaufman (SAMSI), Sean McMahon (Duke), Mike Dietze Harvard), Jim Crooks (SAMSI), Benoit Courbaud (Duke).

# 3 Research Goals and Activities

## 3.1 Program Level

Apart from the goals specified for each of the working groups, unified goals of the overall program were to *i)* promote interaction between the different working groups, *ii)* gain understanding and widen the possible goals by exposing researchers in a very specific area/models to similar problems and goals of related models, *iii)* foster collaboration between modelers, statisticians, applied mathematicians and scientists, *iv)* promote active participation of new researchers and students. These goals were nicely achieved, and were clearly reflected in working groups with large overlapping, denoting the widened interest of researchers, and miscellaneous mix of researchers from the different areas as well as of senior and new researchers.

Most working groups were a good mix of mathematicians, modelers, statisticians, and scientists from specific areas, whether attending in person or remotely. This resulted in a extraordinary enrichment of all involved: learning simultaneously about the characteristic of science behind the computer model, the mathematics/engineering translating science into models, the numerical implementation, and the statistical learning from field data is an extremely instructive exercise, in which the limitations and potentials of each component of the global use of the model is exposed and analyzed, not in isolation, but in connection with all the other components. This is the best (maybe the only) scientific road for assessment of existing models, development of better models, and adequate utilization of them. SAMSI goal of promoting interaction and achieve results not possible otherwise was exceptionally achieved and demonstrated in this Program; SAMSI played an instrumental role in the exciting achievements of the working groups.

Simultaneously, also in virtually all working groups, there was a good mix of senior people, postdocs and students. It is remarkable that this Program attracted exceptionally good participants; all Sub-programs had the participation of indisputable top leaders in their areas. No less important is the representations of extremely good post-docs and graduate students. The mix resulted in very vital and motivating working groups, with the exciting activities detailed bellow.

The participants were also a good mix in geographical terms; most participants were non locals, coming mainly from all over USA, but also with an important representation from Europe. In similar terms, minorities, in particular women were very well represented. The Program Leaders have strived to have women representation in all working groups, workshops and activities. In particular, when invitations were issued (for participation in different activities), adequate representation (senior/junior, geographical and minorities) were actively pursued. Computer Models is an area where extremely few senior, influential women are active at the moment, and most of them were asked to participate (obviously, not all of them could make it, although all showed a lot of interest in the Program). Even with this restriction, the Program achieved a good representation of minorities in all its activities.

We next briefly describe the goals, activities and achievements of the different working groups. More detailed information appear in the respective web pages.

## 3.2 Air Quality

### 3.2.1 Introduction. Research Goals.

The working group identified some important issues for research:

- *Model calibration.* For these models, no systematic studies of the tuning parameters have been done. Some sensitivity analyzes are available, and some statisticians in the group will use them at first. There is a need to run the models under various parameterizations, following a design of experiment, to find good parameterizations. We used the Bayesian calibration in this framework. We considered 30 runs of a pollution model and field data (1000 EPA stations scattered over the US, with a very specific distribution). We obtained preliminary results for calibration over the Southwest USA.

- *Model downscaling.* Can a statistical help with downscaling rather than forecasting a very high resolution? It clearly depends on what resolution is used for the inputs, and the parameterizations done. We obtained results for a regression models using local ancillary variables, in order to predict pollution levels at a sub-grid scale. A paper was published on the topic: Guillas S., Bao J., Choi Y., Wang Y. Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta, *Atmospheric Environment*, **42** (6), 1338-1348, 2008.

As a longer term goal, group participants intend to examine the problem of human exposure. The central question is how to use air quality model outputs for public health research, specifically epidemiological work.

### 3.2.2 Specific activities.

The Air Quality working group held seven meetings over the Fall of 2006. A web page

`http://www.samsi.info/200607/compmod/workinggroup/air/index.html`

describes the topics covered and some presentations. We discussed calibration of an air quality model. S. Guillas and colleagues at Georgia Tech then run this model for designed inputs. Observations from one station in Atlanta was used to calibrate. An important issue was the calibration of functional outputs (specially pursued by Fei Liu) and it is still under development. This research was new since no calibration of an air quality model had been performed in the past.

On the topic of downscaling, contributions focused mainly in two studies, which are still under way:

1. S. Guillas, Chunsheng Ma. Space-time downscaling of regional ozone forecasts with nonseparable covariance models. In this work we introduce a new type of space-time covariances that enable us to better fit the differences between model outputs and observations. The resulting downscaling should improve on the result for time series.

2. S. Guillas, A. Gelfand and S. Sahu, Bayesian downscaling of an air quality model. In this work, uncertainties are naturally assessed by the Bayesian approach. Meteorological variables are used to locally improve the regional forecasts, using strong priors to give preference to the chemistry transport model.

Dana Draghicescu (CUNY-Hunter College) was invited by the working group to give a presentation on modeling and prediction of probability distribution functions and quantiles for space-time environmental processes. A collaboration between S. Guillas and Dana Draghicescu started on the topic of quantile maps for air quality assessment.

## 3.3 Calibration of Computational Models of Cerebral Blood Flows

### 3.3.1 Introduction. Research Goals.

The long term goal of the effort was to explain the autoregulation mechanisms taking place in the vascular system in response to outside stimuli (such as posture changes from sitting to standing for instance). The group focused its attention on one particular subsystem of the cerebral vasculature: the Circle of Willis. This network of about sixteen vessel segments plays a key regulation role. Further, its topology may be different from patient to patient. It is expected that an improved understanding of the mechanical properties, both fluid and elastic, of this subnetwork will also lead to better predictive capabilities regarding risks of strokes.

The working group identified some important issues for research:

- *Model selection.* A lot of recent research in hemodynamics has focused on detailed modeling/calculation of localized features of blood flows such as for instance flows in the vicinity of aneurysms. A more systemic and global approach has to be considered in order to take into account circulation in the entire Circle of Willis. Further, the model has to be simple enough to allow patient dependent topological changes to be easily taken into account, ruling out full three-dimensional approximations.

  A pseudo one-dimensional approach was used in which blood is considered as a non-Newtonian fluid and visco-elastic reactions of the vessels are taken into account.

- *Mathematical and numerical issues.* The mathematical structure of the equations corresponds to systems of nonlinear hyperbolic balance laws which are linked, from vessel to vessel, by boundary conditions. This highly nonstandard type of problems has recently received a lot of attention in other fields (communication networks, traffic flows, pipeline network management, etc...). Pseudo-spectral methods were used as solvers.

- *Model calibration.* One the main goals of the working group was model calibration. A key point that differentiates our work from that of other groups in this field was our direct access to high quality data through collaboration with Dr. Novak's group. Ensemble Kalman filtering techniques were used for calibration purposes as several material properties are not accessible to measurement.

- *Numerical simulation of strokes* The influence of the topology of the cerebral vasculature on stroke likelihood and subsequent survival rate is not very well understood. Further, for obvious reasons, very few data can be collected from stroke survivors. A reliable numerical model could thus potentially play an important predictive role. This research line is still actively pursued.

### 3.3.2 Specific activities.

Some specific activities for the working group included:

– The chosen numerical model was calibrated and successfully validated against clinical data[1] This study is one of the very few where a cerebral blood flow model was calibrated against clinical, as opposed to synthetic, data.

– Post-processing of raw data from Dr. Novak's group. Due to the measurement technique, flow properties (namely velocity) can only be recorded one vessel at a time. The set of available data is thus not synchronized. Various statistical issues linked to this fact were (and continue to be) investigated.

[1] K. DeVault, P.A. Gremaud, V. Novak, M.S. Olufsen, G. Vernieres and P.Zhao, *Blood flow in the Circle of Willis: modeling and calibration*, submitted for publication in Multiscale Modeling and Simulation: a SIAM Interdisciplinary Journal.

- Comparisons of various calibration techniques (Newton based methods vs. Kalman filtering) were conducted. Kalman filtering has proved to be more reliable and less time consuming than the other techniques investigated.
- The conditions at the boundary between the part of the vascular network that is modeled (Circle of Willis) and the rest of the network are delicate. A comparison of two different ad hoc models is under way. The suitability of more complex conditions such as impedance models linked is under study[2],[3].
- For the Special, ongoing area on **Numerical simulation of strokes**, some specific activities include:
  * Based on the chosen model, a detailed study of the effect of stroke (partially or totally occluded vessels) has been initiated. Preliminary results indicate the remarkable potential for the Circle of Willis to maintain acceptable levels of perfusion throughout the brain even under very trying circumstances (such as high asymmetric inflows)[3], [4].
  * The comparison of those preliminary numerical results with clinical data is under way.

## 3.4 Climate and Weather

Numerical models are crucial to simulate geophysical, chemical and ecological processes and to understand the relationship among components in the Earth system. As models have become larger and more complex, their construction, validation and analysis are no longer amenable to simple approaches and statistical summaries. Statistical science in the past 20 years has advanced to handle the interpretation of complex multivariate, spatial and temporal data sets and it is well-suited to tackle the massive outputs from numerical experiments that are now the norm in the geosciences. This SAMSI working group, organised in partnership with the National Center for Atmospheric Research (NCAR), was tasked with applying cutting edge statistical methods to the needs of geophysical model development and to make statisticial scientists aware of the particular scientific issues and research in the geophysical modeling community.

### 3.4.1 Introduction. Research Goals.

An exciting development for this working group was the close collaboration between statisticians, applied mathematicians and NCAR modelers, which resulted in a better understanding and characterization of uncertainty in climate and weather predictions. In particular our work addresses

[2]K. DeVault, P.A. Gremaud, *Analysis and Numerical Analysis of Structured tree boundary conditions for hemodynamics*, in preparation.

[3]K. DeVault, *Numerical study of two problems in fluid flow: cavitation and cerebral circulation*, urlhttp://www.lib.ncsu.edu/theses/available/etd-03072008-112809/

[4]K. DeVault, P.A. Gremaud, V. Novak, M.S. Olufsen and P.Zhao, *Numerical simulation of strokes in cerebral blood flows*, in preparation

the characterization, elicitation, handling, and visualisation of the system uncertainty that arises from model-based predictions, problems of decision-oriented model applications, and improvement of weather and climate forecasts.

The first few meetings of this working group were held jointly with the Climate and Weather working group from the SAMSI Fall Program on High Dimensional Inference and Random Matrices; these were devoted mainly to identifying and working on key references in the literature. Following the Joint SAMSI-NCAR Workshop in November, the activities were more focused on specific goals and models. More specifically, this focused on the following five projects:

1. Stochastic parameterization of WRF-1D
2. Data assimilation for WFF-1D
3. Evaluation of physical models
4. Comparison of regional climate models
5. Turbulance modelling.

#### 3.4.2 Specific Activities

There were five working subgroups doing research on the five scientific problems described below. The groups had weekly meetings, and a joint monthly meeting with all the members of the climate modeling working group.

**Project 1: Stochastic parameterization of WRF-1D.** NCAR scientists: Tardif, Hacker; SAMSI leader: Fuentes; people involved: Shows, Bondell

The objective of this work was to build a stochastic model representing the temporal variability in the characteristics of clouds (cloud height and amount of water in the cloud), based on the statistical properties of observed clouds at the Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) instrumented site in the Southern Great Plains (SGP). The model will subsequently be used to provide input to the WRF-1D atmospheric model to study the response of the modeled atmospheric boundary layer to stochastic cloud radiative forcing.

**Project 2: Data assimilation for WFF-1D.** NCAR scientists: Hacker, Anderson; SAMSI leader: Guillaume Vernieres; people involved: Fuentes, Spiller

The WRF-1d (Weather Research Forecast) is a column model derived from the 3d WRF. One of its purposes is to estimate key parameters than can be used in the 3d WRF for the purpose of nowcasting and forecasting. We aimed to estimate the uncertainties in the parameters and the dynamics of the WRF-1d model using wind speed and temperature observations measured at 10 meters height.

The code for the WRF-1d was migrated to the Topsail cluster at UNC. The ensemble filtering is being developed using MPI and will be design to use the 128 nodes that we have access to, the idea being to use a large ensemble for a better representation of the PDFs associated with the optimum parameters.

**Project 3: Evaluation of physical models.** NCAR people: Sain, Maute, Richmond; SAMSI leader: Guillas; people involved: Rougier, Reese

TIEGCM is a simulator of the processes in the upper-atmosphere. In our experiment, a single TIEGCM evaluation computes the daily response of magnetic perturbations at the ground ($H$,$D$, and $Z$ component), and the up-/poleward ExB drift velocity, at specified sites above the surface of the earth. We have an initial ensemble of 30 evaluations in which three parameters have been varied in a maximin latin hypercube design. These parameters are the amplitude of the migrating tide, the phase of the tide, and the minimum electron density.

Our objective is to calibrate up to seven of the parameters, including the three mentioned above. Our initial ensemble does not span the observational data we have at each site, suggesting that we need to increase the range over which the three parameters are varied, and/or vary the remaining four parameters. To this end, our first task was to emulate the TIEGCM response to the three model parameters. This allows us to judge whether moving these parameters beyond their current limits will improve the calibration, and, in conjunction with the modellers, whether such a move is 'physical'.

Emulating the simulator response poses a number of interesting and unusual challenges. The outputs at a given site are a periodic function of time. Similarly, the emulator's response at different site must conform to the geometry of the surface of the earth. One plan to address this is to use regressors that are a tensor product of Fourier and Spherical Harmonics basis functions, the latter with higher order in the zonal terms to account for the stronger effect of latitude. This is quite a demanding task, particularly if we attempt to impose the same kind of structure on the emulator residual process. However, it has the advantage of combining information from multiple sites (up to twenty five); there is clearly some smoothness in the TIEGCM outputs.

An alternative plan is to emulate the TIEGCM response to the model-parameters at each site, which avoids the need for spherical harmonics, and might be able to take advantage of functional modelling for the time-response. We are pursuing both of these plans.

**Project 4: Comparison of regional climate models.** NCAR: Sain; SAMSI leader: Kaufman; people involved: Reich, Bondell

The NARCCAP project is running a variety of 50km resolution regional climate models (RCMs) for North America, with boundary conditions provided by a variety of global climate models (GCMs) and reanalysis datasets (interpolated data products based on observations). This experimental design lends itself to a functional ANOVA approach, in which the output response can be decomposed into main effects and interactions. This technique allows us to partition variability in the model response as a diagnostic technique for comparing models. The SAMSI group working on this project has obtained data from a similar project giving temperature output over the UK in a 2 RCM x 2 GCM x 2 C02 scenario experiment. Some preliminary analysis, spatially-averaged, shows that the largest source of variability is the C02 scenario, followed by the GCM, followed by the RCM. That is, the two RCMs have a high degree of agreement, given a particular GCM and scenario. This group is extending this analysis to model spatial surfaces rather than overall means. Currently the group is exploring one-way functional ANOVA modelling using Gaussian process priors on the mean functions using simulated data, and studying the connections between this method and existing methods using splines.

**Project 5: Turbulance modelling.** NCAR: Mininni; SAMSI leader: Ma. This group developed log-Gaussian frameworks for turbulence modelling.

### 3.5 Dynamics of Infectious Diseases

#### 3.5.1 Introduction. Research Goals.

The main research goal were

- To address model validation in the context of biological processes, including infectious diseases, viral dynamics, and dynamical networks.
- To learn about data acquisition in public domains.
- To provide tutorial discussions about mathematical modeling of epidemic processes between human hosts, and at the level of population of cells.
- To discuss algebraic modeling approaches for cellular networks.
- To introduce Bayesian methods for inference of biological network dynamics.
- To convey the applicability of statistical model selection.
- To discuss statistical and mathematical methodologies addressing model bias.

#### 3.5.2 Specific Activities

This working group scheduled four monthly meetings:

January 15, 2007 held in the Center for Research in Scientific Computation of NCSU.

February 16, 2007 held in SAMSI.

March 16, 2007 held in SAMSI.

April 13, 2007 held in SAMSI.

On Jan. 15, we had an organization meeting and decided on the following three monthly dates. During the second part of the morning we had a hands-on exploration of several public sources of epidemiological longitudinal data including: Centers for Disease Control and Prevention, European Influenza Surveillance Scheme, Canadian Centre for Infectious Disease Prevention and Control, Sentinelle Network and Sentiweb, and World Health Organization. Various links to access these databases can be found in: http://www4.ncsu.edu/ acintro/samsi_dyn_infc/data_sources.html

In addition, two discussions were held on Jan. 15 conveying population as well as within-host dynamics with the following titles and facilitators:

1. *Introduction to deterministic epidemiological models: linear stability analysis, basic reproductive numbers, and sensitivity analysis* led by Ariel Cintron-Arias.

2. *Introduction to within-host dynamics: estimates of reproductive numbers obtained from patient viral load counts* led by Alun Lloyd.

During the Feb. 16's meeting there were five discussions addressing gene regulatory networks, algebraic models, graph dynamical systems, estimation of influenza reproductive numbers, and demographic stochasticity in epidemic models with the following titles:

1. *Pathosystems biology: an introduction to discrete models of cellular networks* led by Reinhard Laubenbacher.

2. *Polynomial dynamical systems* facilitated by Abdul Salam Jarrah.

3. *SIR Models over graphs, graph dynamical systems and the EpiSims project* led by Henning S. Mortveit.

4. *Estimation of seasonal effective reproductive numbers of influenza A(H3N2)* led by Ariel Cintron-Arias.

5. *Demographic stochasticity in epidemic models* facilitated by Alun Lloyd.

On Mar. 16, we had the following discussion addressing boolean networks, biochemical network dynamics, re-activation of latent viruses, evolutionary computation tools, model selection, with following titles:

1. *On the number of attractors of boolean networks* led by Alan Veliz-Cuba.

2. *Stochastic modelling and inference for biological network dynamics* led by Darren Wilkinson.

3. *A mathematical model for induced reactivation of latent herpes virus* led by Grace Kepler.

4. *Akaike information criterion* led by Shuhua Hu.

5. *Reverse engineering of biological systems: discrete mathematics and evolutionary computation tools* led by Paola Vera-Licona.

The meeting held on Apr. 13, there were two discussions about endocrine networks and model validation with the following titles:

1. *Modeling endocrine networks* led by E. M. Cliff.

2. *Analyzing possible causes of model bias with stochastic, time-dependent parameters* led by Peter Reichert.

## 3.6 Engineering Methodology

As mathematical models of physical phenomenon have become more available and their computer implementations have become increasingly sophisticated, it has become common practice to refer to the codes which represent input/output relationships as "computer models" and the process of running the code at trial inputs as a "computer experiment." Indeed, engineering decisions are frequently made either by a combination of physical and computer experiments, or even based solely on the resulting (complex) computer codes.

### 3.6.1 Research Goals.

The research of this working group evolved mainly around the following issues

- *Screening for high-dimensional inputs.* The research goal was to determine methodology for identifying active inputs in high-input codes. In addition, we seek to determine designs for the computer experiment to facilitate this screening.

- *Models for Dynamical Systems.* Dynamical systems are time-evolving system and thus can be thought of as a type of functional output. The goal was to produce a physics-based emulator of a dynamical system. More generally, how does one perform emulation based on highly correlated real-valued or functional output.

- *Calibration of Computer Codes.* In functional cases one reduces dimensionality by expressing computer model output using a system of basis functions. There are several proposals in the literature for creating such bases but no consensus on how this should be done when the goal is calibration of computer codes.

- *Simultaneous Calibration and Tuning of Computer Codes.* While methodology for using calibration parameters to describe the uncertainty in computer model output, there is no parallel development for setting tuning parameters. Thus the research question is, how should one perform (simultaneous calibration) and setting of tuning parameters? What metrics should be used to compare the outputs of the computer model and physical experiment?

- *Validation of Computer Codes.* Are there better choices of calibration models that make the tuning parameter and computer model bias terms more orthogonal than that proposed by Kennedy and O'Hagan (2001). These quantities are highly correlated and can be difficult for MCMC methods to accurately predict.

### 3.6.2 Specific Activities

The Engineering Methodology working group has focused on developing dynamic model approximations to computer codes of coupled systems. One application of such a model is in coupled differential equation model of the water output from a watershed area. Another potential application area is to describe the deflection of the individual blades a coupled system of coupled rotor fan blades. The use of these models as competing emulators to standard GASP emulators, their calibration, and their accuracy as approximators to the complex computer model are all important points of research.

Another project that was completed during the period of support was to the biomechanical design of acetabular cups that are used in hip prosthesis replacements. The goal was to design these devices to be robust to surgical variability in their insertion and to variability in patient bone quality and to variability in loading the hip.

Other engineering projects have undertaken by members of the hydrology group and individual graduate students associated with this program (see additional information regarding Gang Han, Dianne Bautista, and the Hydrology working group).

Some papers that have appeared or that will appear or are under preparation as a direct consequence of the Program appears listed in the next Section.

## 3.7 Granular Materials - Engineering Applications

### 3.7.1 Introduction. Research Goals.

The working group identified the following main global goals which were actively pursued during the year:

- *Experimental design, calibration and forecasting* of granular avalanches. Preliminary focus was on using the Bayes linear methodology, but other methodologies were explored in collaboration with the Methodology working group. Other Applications, including stress loads and flows in bins and hoppers, were also considered.

- *Extreme events.* Included in this work is a focus on extreme events - those large but rare pyroclastic flows and lahars that cause widespread devastation - as well as developing quantitative statistical tools that inform hazard assessment. We developed statistical models for the inputs to the computer model that are appropriate for extreme events; refinement are being considered.

- *Model comparison and evaluation.* Interfacing with the methodology group, an important goal of this working group was to develop a rational basis for comparing models of avalanches or bin loads and flows that include different physics as well as other phenomenological modeling factors, to yield an integrated modeling effort that offers better predictive capabilities. The issue was considered and discussed at lenght but we still do not have a clear recommendation for evaluation of models with spatially varying outputs. We hope to pursue it in the future, as collaboration is still very active.

There are several measures of success in this working group - successful application of a Bayesian approach to the issues of computational experimental design, calibration and forecasting of avalanches/pyroclastic flows; application of statistical ideas on variation and uncertainty to engineering problems of stresses on and flows inside bins and hoppers. Papers and presentations on these ideas give specific metrics.

### 3.7.2 Specific Activities

During the course of the year, the group gravitated towards the Soufriere Hills volcano on Montserrat as a source for data. This volcano began its current eruption cycle in 1995, and has produced a few very large flows and many - literally hundreds - of smaller flow events. These eruptions have caused dozens of fatalities, and the evacuation of the southern third of the island, including the capital, Plymouth. Eliza Calder, a geologist at Buffalo, collected and analyzed this data, and shared her expertise with us along the way. In January, 2007, a new dome bubble began to grow at the summit of the volcano, and geo-scientists anticipate another large eruption soon. Flow simulations at Montserrat, using the most recent digital elevation data, and with an eye toward developing revised hazard map, are being developed.

The following specific activities, correspoding to some of the goals above, were pursued during the year:

- *Experimental design, calibration and forecasting.* One research goal for the group was in the application of Bayes linear methodology for experiment design, calibration and forecasting of granular avalanches. Earlier work by Patra, Dalbey and Pitman, was extended to develop an adaptive methodology for constructing an emulator. Comparison with Gaussian Process demonstrates the strengths and weaknesses of both approaches.

- *Extreme events.* Another goal was an examination of extreme events to inform hazard assessment. We investigated 'rare events' in the context of mass flows at Montserrat, an active volcano that is erupting regularly. A model for eruptions using the "peaks over threshold" approach to large flows is used to compute the probability of extreme events in a given period of time.
- *Sampling key parameters.* Analytic and computational approaches to sampling some of the principal parameters in mass flow models, in particular, friction angles were developped.
- *Sequentially designed emulators for boundaries.* The input distribution along with the simulators can not directly be used to compute by Monte Carlo the probability of catastrophic events, as these are very rare and the simulator very slow. The group investigated instead in a novel direction in which an emulator is adaptively fitted to characterize a critic boundary defining the region of inputs producing catastrophic events, thus allowing the probability computation directly from the statistical model.

Papers about these four research activities are in various stages of development, as detailed in the Section below

## 3.8 Inference and Uncertainty Analysis of Hydrological Models

### 3.8.1 Introduction

A major conceptual problem in hydrological, and - with increasing temporal resolution of data - also in ecological models, are systematic deviations of the results of predictions of deterministic models from data. This invalidates simple statistical assumptions on the distribution of the error term and makes uncertainty estimates of inferred model parameters and derived model results unreliable. Systematic measurement errors due to miscalibration of river flow gauges are usually not a main contributor to these problems. The major causes are (i) input uncertainty, (ii) model structure deficits, and (iii) the inadequateness of a deterministic description of the hydrological system. A description of these inadequacies by a bias term in model output is not satisfactory, as we are unable to predict the bias when using the model to extrapolate system behaviour into the future or to extended ranges of driving conditions.

The application of Bayesian techniques in hydrology or, more generally, in environmental modelling, is often limited by the computational demand of their implementation. Computational resources may become limiting if computationally demanding models have to be used for computationally demanding statistical inference techniques, such as Markov chain Monte Carlo techniques. Statistical emulators of demanding simulation programs can provide a solution to this problem, but currently used emulators still have deficits. They do not make use of our knowledge about mechanisms described by the model, and they can run into numerical difficulties for densely spaced output dimensions, such as the time dimension for dynamic models.

### 3.8.2 Research Goals

We divided our activities into two research areas.

**(a) Bias reduction with time-dependent parameters.** The introduction of a statistical description model bias by an additive model inadequacy function to model output (Kennedy and O'Hagan, 2001) makes it possible to consider model bias explicitly, to improve the reliability of parameter estimates (as the statistical assumptions are no longer severly violated), and it provides limited support for analysing the cause of the bias. However, as it focuses on model output exclusively and does not make the attempt of describing an intrinsic mechanism of the system described by the model, it is hard to use for a mechanistic identification of the cause of the model structure deficit and for extrapolation. This approach was extended recently for time-dependent models (Tomassini et al., submitted) by attempting to model the bias intrinsically in the model by making selected parameters a time-dependent stochastic process. An analysis of the identified time dependence of the parameter (by trying to relate it to external and internal model variables) can facilitate the analysis of mechanistic causes of model bias. The unexplained contribution to the bias can be considered by incorporating the stochastic process for the parameter that is most likely not to be deterministic in the model formulation. If both of these contributions can correctly be identified, this can be expected to lead to improved model predictions as compared to a bias term in model output.

**(b) Emulating hydrological models.** Statistical emulators have been proposed to make Bayesian analysis possible for very computationally expensive simulation programs (Currin, et al., 1991; Kennedy and O'Hagan, 2001; Santner et al., 2003). However, only few attempts have been made so far to apply this approach to dynamic models for which relatively dense time series can be generated. Another deficit of previously developed emulators is the lack of consideration of (known) mechanisms represented by the model. It can be expected, that consideration of this knowledge could improve the interpolation and extrapolation capability of the emulator. It was the goal of this research field to address these two deficits by developing a mechanism-based emulator for dynamic simulation models.

### 3.8.3 Research Activities and Current State of Work

We divide the progress into the research field introduced above.

**(a) Bias reduction with time-dependent parameters.**

In this research field of the SAMSI program we implemented a simple hydrological model that was used as a research tool to explore the methodology of using time-dependent parameters to support the identification of causes of bias in simulation models. The methodology was successfully applied to this model and proved to be a promising research direction for further developing bias modelling, in particular when having predictions in mind. A paper describing the methodology and its application was submitted for publication.

**(b) Emulating hydrological models.**

In this research field of the SAMSI program, together with the working groups on methodology and engineering methodology, we developed a technique for constructing mechanism-based emulators of dynamic simulation models and applied this concept to the simple hydrological simulation model mentioned above. The idea is to base the emulator on a simple linear, discrete-time state-space model and on a description of its innovation terms by Gaussian Processes as a function of model parameters. This concept proved to be promising. So far, two papers about this emulator were written (one submitted, one to be submitted in May 2008).

Publications within the SAMSI program are listed in the next Section

**Other References**


Kennedy, M.C., O'Hagan, A., Bayesian calibration of computer models, Journal of the Royal Statistical Society, Series B, 63(3), 425-464, 2001.

Kuczera, G., Kavetski, D., Franks, S., Thyer, M., Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterizing model error using storm-dependent parameters, Journal of Hydrology, in press.

O'Hagen, A., Curve fitting and optimal design for prediction, Journal of the Royal Statistical Society, Series B, 40(1), 1-42, 1978.

Santner, T.J., Williams, B.J., Notz, W.I., The Design and Analysis of Computer Experiments, Springer, New York, 2003.

Tomassini, L., Reichert, P., Künsch, H.-R., Buser, Ch., Borsuk, M.E., A smoothing algorithm for estmating stochastic, continuous-time model parameters and an application to a simple climate model, submitted.

Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, Water Resources Research 41, W01017, doi:10.1029/2004WR003059, 2005.


## 3.9 Methodology

The Methodology Working Group addressed a range of theoretical and computational issues that arise in the development, assessment, and utilization of complex computer models— especially the obstacles or opportunities for improvement encountered by more than one of the other working groups. Short presentations from a range of participants and guests were used as catalysts to help focus attention and creative efforts at emerging areas of opportunity for advancing this computational methodology.

### 3.9.1 Introduction. Research Progress.

The Methodology subgroup addressed several goals identified in earlier work by the group:

- *Design* for both computer model runs and proposed aquisition of field data. In low-dimensional spaces we have a clear idea of the risks and benefits of traditional 'space filling' designs (such as Latin hypercubes) or, conversely, of designs favoring more 'extreme values' of model inputs; the latter offer (seemingly) more precise determination of model parameters, while the former offer better opportunities for assessing model fit and accurate reporting of uncertainty. In higher dimensional spaces our intuition and abilities to visualize are weaker, and technical obstacles interfere with our ability to balance the twin desires to spread out data input points and to 'fill in' the model space; there is a need for more experience with different design strategies, particularly in high dimensional spaces. In collaboration with the Granular Material Working Group, we developed a hybrid methodology that begins with a Latin hypercube design and then devotes a disproportionately large computational effort near the 'boundary' region (often far enough out in a distributional tail that conventional sampling would yield little information) where interest lies.

  A second goal has been established of developing sequential designs, both "early bail-out" strategies based on rapid identification of unfeasible regions, and the more challenging goal of generating particularly good sequential designs to guide the choice of subsequent observations in applications where observations (either field or very costly simulators) are expensive. Some progress has been made toward this latter goal, by dynamic selection of design points, revealing an opportunity for further exploration.

- *Emulators.* Since computer models are usually very costly to run, an 'emulator' (fast simulator, surrogator) is required, usually for optimization, design and statistical analyses (calibration, validation, prediction, etc.). Traditionally, GASP (Gaussian separable process) models with simple isotropic covariance structures developed in the geostatistical sciences are the 'default' choice, but serious comparison with alternative processes and study of the limitations of GASPs is still lacking. Another promising possibility is use of 'rougher' or 'simpler' computer models, or a combination of both. Some progress has been made toward reducing the GASP computational burdon in high-dimensional spaces, through modular models which explore parts of the parameter space at a time, and through dynamic allocation of design points (see above).

- *Huge spaces* offer big challenges, for both input and data spaces. This includes functional input/data spaces. One obvious difficulty with 'traditional' GASP is its failure to scale up to even moderate dimensions, so either imaginative numerical methods have to be used (such as dynamic selection of design points), or dimension reduction strategies have to be adopted. Another promising approach for future work is the design and application of kernel-based non-Gaussian emulators; these appear to be slower than GASP for small-to-moderate sized problems, but they scale better to moderate-to-large ones.

- *Bayesian approaches to Calibration, Validation and Prediction* A main use of computer models is to predict reality in the presence of (often great) uncertainties. Predictions should always be made along with 'confidence' bands. Also, calibration and validation of computer models have to be performed simultaneously. All these issues are best addressed through a Bayesian approach with an input-dependent discrepancy or 'bias' function, which naturally incorporates all the uncertainties. This is indeed the prevalent methodology since the pioneering work of Kennedy and O'Hagan (2001). However, the confounding between unknown parameters in model inputs and the bias function has not been openly recognized and addressed till recently. This confounding airs many issues of implementation (appropriate priors), numerical analysis (issues on MCMC mixing and convergence), and interpretation and reporting of the results.

- *Prediction for untried (or altered) scenarios.* Using computer models for prediction in situations in which data are lacking or very scarce lies at the very heart of development of computer simulators. Exactly how to best extrapolate models and bias, with appropriate measures of uncertainty, is an important area of research that is not yet solved satisfactorily. A combination of expert judgement, statistical insights and modelers knowledge will be needed for successful extrapolation.

- *Multivariate outputs.* Computer models output are multivariate in nature. Usually, each output dimension is addressed separately, but sometime simultaneous consideration of multivariate output is required. GASPs and other generalizations, as well as hierarchical Bayes implementations can be utilized and should be investigated.

- *Approximations.* Because of the high-dimensional input spaces, confounding, little external prior information and scarce data, the brute-force Bayes is sometimes infeasible (especially for somewhat 'automatic' use without expert oversight), and suitable approximations need to be investigated. For instance, some of the parameters (specially those involved in the emulator) may have to be replaced by estimates (or even by fixed values). Our Working Group explored some of these options, including the 'Bayes linear' approach initiated by Working Group member Michael Goldstein.

The group considered a few 'hands-on' models to focus efforts on specific issues, including emulation of the TITAN2D simulator of granular flow, emulation of coupled harmonic oscillators intended as a simple model for turbine blades, and simple pedagogic models.

### 3.9.2 Specific Activities

- *Design* Visitor Henry Wynn (Statist., London School of Economics, UK) presented pioneering work developing algebraic methods to support design, and new methods for the rapid identification of feasible regions in design space.

Visitor David Steinberg (Statist. & Oper. Res., Tel Aviv University, Israel), an expert on design, presented a group of lectures on new ways of generating promising designs for computer experiments.

Elaine Spiller (SAMSI post-doc) presented a reanalysis of simulated avalanch and volcanic flow data and designs from the work of Bruce Pitman (Mathematics, University of Buffalo) and colleagues, illustrating how sharp discontinuities in the parameter space may warrant special attention in the design phase.

- *Emulators* Participant Herbie Lee (Applied Math. and Statist., UC-Santa Cruz) presented his work on "treed Gaussian processes" intended to overcome the isotropy limitation of GASPs. Jim Crooks (SAMSI/Duke post-doc) presented motivating problems from the SAMSI Terrestrial Working Group, where GASPs (and other Gaussian methods) seem inappropriate. Robert Wolpert (Statistical Science, Duke Univ.) presented novel non-Gaussian methods that may offer an alternative to GASPs.

- *Huge spaces* Bruce Pitman (Mathematics, Univ. Buffalo) presented complex PDE models for avalanch modeling, along with early efforts to develop simulators (based on the Bayes linear approach of Michael Goldstein (Math. Sci, Univ Durham, UK); Goldstein presented more details and insight about this approach. The Working Group explored different ways of reducing the very high dimensionality of the avalanch problem— such as considering only a small number of locations, or approximating the vast number of uncertain friction coeficients with constants or stationary processes.

  Several subject-area specialists presented their high-dimensional models to help the Working group appreciate the breadth of these problems and look for common themes— these include SAMSI visitor Peter Reichert (Environ. Sci, ETH, Switzerland) speaking on hydrology problems (indluding the need for time-dependent parameters, with the potential of increasing model dimensions dramatically); Ken Reckhow (Environ. Sci, Duke Univ.) speaking on water quality models; Bernt Mueller (Physics, Duke Univ) speaking on models for high-energy nuclear collisions; Jim Clark (Environ. Sci and Biology, Duke Univ) speaking on dynamic terrestrial models. Doug Nychka (head of IMAGe at NCAR) arranged for remote presentations on several aspects of climate modeling by his colleagues at NCAR (Johannes Feddema of Univ Kansas on anthropogenic land cover change experiments; Pablo Mininni of NCAR on statistical properties of turbulent flows; Hanli Liu, Art Richmond and Michael Wiltberger of NCAR, on upper atmosphere modling issues; and Joshua Hacker and Gordon Bonan of NCAR on planetary boundary layer uncertainties).

- *Bayesian approaches to Calibration, Validation and Prediction* Michael Goldstein (Math. Sci, Univ Durham, UK) presented several lectures on the Bayes Linear approach to statistical modeling, and the Working Group sought similarities and contrasts with the GASP model emulator approach to computer model validation and assessment. Tony O'Hagan (Statistics, Univ. Sheffield, UK) offered a series of talks exploring the approach pioneered in Kennedy and O'Hagan (2001).

- *Prediction for untried (or altered) scenarios.* The work by Bruce Pitman and Abani Patra (Mech. & Aero. Eng., Univ. Buffalo) and its reanalysis by Elaine Spiller (SAMSI post-doc) helped illuminate the obstacles one faces in interpolation and extrapolation of emulators and simulators, in a problem with sharp discontinuities. New methods are needed to raise cautionary 'red flags' when such discontinuities are recognized during emulation and simulation.

  Wolpert and Simon Lunagomez (Statistical Science, Duke Univ.) presented models and methods for using volcanic eruption data from Pitman, Patra, and Eliza Calder (Geology, Univ Buffalo) to make long-term predictions of the rate of cataclysmic eruptions based on mid-term data of a wide range of eruption magnitudes.

- *Multivariate outputs* Lenny Smith (Maths, Oxford Univ, UK) presented climate models with coupled outputs, illuminating a range of ways that models may be uncertain in their parametric aspects or predictions, and showing how multidimensional methods may be needed.

- *Approximations.* Cari Kaufman (SAMSI/NCAR postdoc) presented a dimension-reduction technique called 'covariance tapering' in which model Gaussian covariance functions are approximated by much sparser ones, leading to computationally-tractible approximations for high-dimensional problems. A number of possibilities for extension of this work emerged from discussions within the Working Group.

- *Use of derivative data for model calibration and prediction.* New tools and software are being developed by Working Group participants to exploit *derivative* information generated by recent simulators— for example, in suggesting better experimental designs. Early work has been completed on the development of a new approach to emulation employing derivative information to guide the specification, calibration and prediction of bias. Usual GASP emulators are very good for interpolation but not for extrapolation. Emulating the derivative information (as well as the usual state information) captures a bit of the physical dynamics of the process, offering a chance to improve extrapolation. Of course this is challenging for complex high-dimensional simulators, but linear approximations of the problem remain computationaly feasible. Some members of the Working Group (particularly Bayarri, Berger and Paulo) have pursued this approach, describing their work in a Working Group presentation.

- *Multiscale Models.* Jim Zidek (U British Columbia) presented a discussion of the difficulties facing environmental modellers trying to predict human health effects of environmental pollutant levels, with data given at spatial and temporal scales ranging over eight orders of magnitude and more. The stragegy of building "locally stationary" models emerged.

- *Dynamic Emulators.* A joint meeting of the Methodology and Engineering Working Groups considered novel approaches to the emulation of dynamic models with time-dependent parameters. Led by Peter Reichert (Environ. Sci, ETH, Switzerland) and Gentry White (NCSU), the focus was on hydrological models for soil water uptake.

## 3.10 Statistical Mechanics of Granular Flow

The main theme of this working group is to investigate whether it is possible to deduce continuum behavior from statistical mechanics analysis of grain interaction. The distinguishing feature of granular flow by comparison to normal fluid mechanics is that the control volume over which we seek a continuum behavior is a mesoscopic system which contains a much smaller number of particles ($10^3$) than those that typically found in fluid mechancis ($10^{23}$). Such mesoscopic systems are difficult to treat using standard methods of statistical mechanics. However, changing the focus to the statistical mechanics of the possible intergranular interactions allows the investigation of a system with a large number of components ($10^{20}$). To this end a number of research goals which attempt to elucidate the intergranul interaction have been formulated.

### 3.10.1 Introduction. Research Goals.

The following themes were pursued by the group :

- *Numerical simulations of few grain interactions.* All current particle interaction algorithms use a simplified grain interaction model based upon an analytical treatment of the collision of two isolated grains. The idealized situation treatable by analytical methods is known not to hold in dense granular flows. To supplement the isolated two-body interaction, numerical simulations have been carried out of interactions between multiple grains (from 3 to 10 grains in the system) with full explicit solution of the elasticity equations. The objective is to allow direct numerical simulation to inform models on granular physics especially as regards segragation and fluctuations.

- *Segregation and fluctuations.* A new viewpoint on segregation and fluctuation starting from analysis of possible integrain interactions is under investigation. The attractive feature of this approach is the applicability of the standard tools of statisitical mechanics. The relevance of the approach to the problem of deducing continuum granular flow behavior is not yet decidied at this point.

### 3.10.2 Specific Activities

A number of meetings were held early on to allow communication between experimental researchers and theorists. The main result of this communication was a better appreciation of the difficulties inherent in describing the detailed physcis of the grain interaction process. In particular, the best characterized experimental systems are formed by plastic disks that polarize light passing through them and thus allow visualization of the internal stress field (Behringer). The disks exhibit viscoplastic behavior. This has led to the problem of developing a simulation capability for such materials (Mitran) in order to first investigate two-body behavior in detail. As of March, 2007 the basic two-body problem was solved. The next stage was to build a sufficiently extensive database of simulations that can then be compared to experimental results and afterwards inform theoretical treatments of segregation and fluctuation.

## 3.11 Systems Biology

### 3.11.1 Introduction. Research Goals.

There are a number of obstacles in the way of building useful dynamic models of biological processes. Some are experimental, some are computational, but many are mathematical and/or statistical, and it is these mathematical and statistical challenges in particular that were the focus of this WG. Particular issues include choosing an appropriate kind of model - spatial/non-spatial, deterministic/stochastic, discrete/continuous in time/space. Then given a particular choice, how can we choose and validate the processes being modelled, and the form of rate equations to be used? Given these, how can we use very indirect and noisy experimental measurements to estimate key model parameters? How do we reconcile top-down descriptive statistical models fitted to available data with bottom-up mathematical models of the underlying mechanisms? What implications do new experimental technologies have for model development, choice, and assessment? What are the key mathematical and statistical developments required to advance the field of systems biology most rapidly?

The kickoff workshop (described below) identified effective parameter estimation and associated identifiability and model validation issues as being a major obstacle for efficient model-building. This therefore represented the primary focus of the WG. Multi-scale modelling and simulation was also identified as requiring urgent development effort, and this area was also the subject of working group activity.

### 3.11.2 Specific Activities

The WG formed out of the Biosystems Modeling workshop (detailed below), and met weekly at SAMSI, involving SAMSI visitors and locals, as well as other participants via teleconference/internet.

The WG initially focussed on parameter estimation problems encountered by two of the participants. The first was a deterministic biochemical network model arising from work being carried out at the Environmental Protection Agency. Some simple parameter estimation methods had already been applied to the model (and associated data). The working group examined the application of more sophisticated Bayesian analysis techniques to the problem in order to better understand the parameterisation issues. The WG benefitted from the visit to SAMSI of Richard Boys and Daniel Henderson (Newcastle U.), experts in estimation and calibration of both deterministic and stochastic biological models.

The second model was a stochastic model of molecular motor dynamics being developed at Cornell. Parameter estimation for complex stochastic kinetic models is still a cutting-edge research problem, and this therefore presented a more formidable challenge to the working group. This model was the subject of an intensive research session in April, when Richard Yamada (Cornell) visited SAMSI.

Various other models and modelling techniques were considered within the general theme of the working group, and these involved varying subsets of the WG participants. These interactions led to a number of papers, now submitted or in preparation, as outlined in the "Additional information". In particular, a stochastic model of mtDNA deletion accumulation considered during the March intensive research session has led directly to the submission of two papers.

### 3.12 Terrestrial Models

The Terrestrial Models subgroup of the "Development, Assessment and Utilization of Complex Computer Models" program was aimed at understanding vegetation response to climate change, incorporating processes that operate at fine spatial scales. Predictions of biodiversity response primarily emphasize climate envelopes, translating the climate range where a species is found today to maps of 2 x CO2 predictions of future climate. Yet real populations are controlled by processes that operate at spatial scales of meters, being limited by dispersal and soil variation (including hydrology). These fine-scale processes depend on climate, but their impacts depend on heterogeneity. The challenge of understanding climate impacts entails i) downscaling regional climate to realistic variation at relevant scales and ii) parameterizing that variation appropriately for forest response.

#### 3.12.1 Introduction. Research Goals.

This research was intended to develop models that integrate regional climate change with landscape heterogeneity that can be used to explore biodiversity response. We specifically considered intermediate scale variation in climate, between fast weather (minutes to days) and slow climate (years), with biodiversity response. Although this is arguably the most logical scale to consider, as it describes the periodic droughts, warm springs, etc, to which plants respond, all modeling to date has been on year-to-year variation. Moreover, current efforts focus on a "climate envelope" approach, sometimes coupled with species area curves. This approach is problematic for a number of reasons discussed in Ibanez et al. (Ecology 2006).

The landscape heterogeneity is being addressed with models that downscale regional climate to moisture and temperature gradients related to topography. Identified research goals were:

- *Overall model.* A hierarchical inference/forward simulation approach seems to be a useful vehicle for this analysis and can engage a wide range of interests and activities, merging an existing complex inferential model and a forward simulator that uses the estimates from the inferential model.
- *Biodiversity predictions* ffor climate change scenarios is a main goal: it is possible and would have a wide audience and impact.
- *Emulator.* Detailed analysis of the simulation output and construction of the emulator (Gaussian process model or others) based on simulation of detailed models. Can we fill in the parameter space? What are the sensitivities? Which parameters, in what states? Identify data needed to improve prediction.
- *Different amounts and types of data available from different investigators.* Many investigators have different types of information that bear on essentially the same sets of processes. What types can be included in a general scheme? How to best scale inference and simulation to large regions? How should high frequency forcing response be scaled to demographic scales (individual allocation over years)? How to deal with variables that are completely hidden? Investigate connections to regional hydrology.

#### 3.12.2 Specific Activities

During the duration of the Program, we worked toward the goal of biodiversity predictions for climate change scenarios. Based on preliminary results, we expect such predictions to look much different from those currently in the literature. The elements of this analysis may ultimately include compilation of seasonal climate predictions, from our participants affiliated with NCAR, the parameterized model to seasonal climate, which was and still is in development here at Duke, and construction of an emulator. There are three modeling interrelated modeling efforts, summarized here

**Downscaling regional climate** We developed a statistical model to predict soil moisture over space and time under varying climate scenarios, using data from sites in North Carolina. Cari Kauffman and Jonty Rougier are building a model to predict the spatial and temporal evolution of soil moisture under varying climatic conditions. Soil moisture is an important factor in tree growth and fecundity, and these predictions are part of the Clark group's larger goal of studying the effect of climate change on forest dynamics. Our approach has been to build a state-space model in which the temporally evolving soil moisture field is the unknown state vector. Our initial predictions used a simple linear model, which could be fit using a Kalman filter, but they did not take advantage of valuable prior knowledge about the evolution of soil moisture over time. In our new model, this process is governed by a simple, catchment-level hydrology model we developed based on the literature, combined with a stochastic model for distributing total water according to local topography. We developed an MCMC algorithm for sampling the non-time-varying parameters, with an embedded smoothing algorithm to sample the temporally evolving soil moisture fields conditional on these parameters. This allow us to sample from the posterior predictive distribution for soil moisture under varying climate scenarios, and these samples can then be used as input to the Clark group's forest simulator

**Biodiversity simulation** The hierarchical inference/forward simulation approach of Govindarajan et al. (2004, 2007) was used for this analysis. For this project the model was extended to include landscape variation in soil moisture and temperature variation. This work was led by Sean McMahon.

**Model emulation** Jim Crooks developed an emulator for stochastic simulators, the specific application being the large forest simulator. The emulation method, based on the Kernel Stick-Breaking Process model of Dunson and Park, is extremely flexible and is able to infer highly non-Gaussian output distributions without a large number of simulator runs. Crooks presented this method at the 9th Case Studies in Bayesian Statistics workshop in October 2007 and at the Environmental Protection Agency - RTP in March 2008. The air quality modeling group at EPA was extremely interested in the emulator in particular and in statistical methods for addressing various other issues in computer experiments. Given funding Crooks will be working with this air quality group full time starting in July.

**References**


1. S. Govindarajan, M. Dietze, P. K. Agarwal and James Clark (2004). A scalable simulator for forest dynamics. In *Proceedings of the twentieth annual symposium on Computational geometry*, pp 106–115. ACM Press: New York, USA.

2. Govindarajan, S. M. Dietze, P. Agarwal, and J.S. Clark. (2007). A scalable algorithm for dispersing populations. Journal of Intelligent Information Systems, in press.

3. I. Ibañez, J. Clark, M.C. Dietze, K.Feeley, M. Hersh, S. Ladeau, A. McBride, N. Welch and M. Wolosin (2006). Predicting Biodiverisy Change: Outside the Climate Envelope, Beyond the Species-Area Curve. Ecology, 87(8), 2006, pp. 18961906

# 4 Publications

This SAMSI Program was very productive, and was behind directly or indirectly of a large number of publications. Here is a list of some of the publications and technical reports produced under the Program, including those under preparation.

## 4.1 Special Technometrics Issue

David Steinberg is organising a special issue with articles on complex computer models. The planning of this issue arose as a consequence of this SAMSI program, and many program participants have submitted results arising from the program to this special issue.

## 4.2 Published Papers


1. Bai, P., Banks, H. T., Dediu, S., Govan, A. Y., Last, M., Lloyd, A., Nguyen, H. K., Olufsen, M. S., Rempala, G. A., Slenning, B. D. (2007) Stochastic and deterministic models for agricultural production networks. *Mathematical Biosciences and Engineering*, **4**(3):140.

2. Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C.H., and Tu, J. (2007). A Framework for Validation of Computer Models. *Technometrics* **49** # 2 pp.138-154.

3. Bayarri, M.J., Berger, J.O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R.J., Paulo, R., Sacks, J., Walsh, D (2007). Computer Model Validation with Functional Output. *Annals of Statistics* **35** #5, pp 1874-1906.

4. M. J. Bayarri, J. O. Berger and G. Molina (2007). Incorporating Uncertainties into Traffic Simulators. In *Recent Advances in Modeling and Simulation. Tools for Communication Networks and Services* (A. Netjat Ince and Arnold Bragg, eds.). pp. 330-347. Springer.

5. Boys, R. J., Wilkinson, D. J., Kirkwood, T. B. L. (2008) Bayesian inference for a discretely observed stochastic kinetic model, *Statistics and Computing*, available on-line.

6. P. F. Craigmile, N. A. Cressie, T. J. Santner, and Y. Rao (2006) "Bayesian Inferences on Environmental Exceedances and Their Spatial Locations," *Extremes*, **8**(3), 143-159.

7. N. A. Cressie, B. E. Buxton, C. A. Calder, P. F. Craigmile, C. Dong, N. J. McMillan, M. Morara, T. J. Santner, K. Wang, G. Young, and J. Zhang (2007) "From Sources to Biomarkers: A Hierarchical Bayesian Approach for Human Exposure Modeling," *Journal of Statistical Planning and Inference*, **137**, 3361-3379.

8. Foley, M. and Fuentes, M. (2008). A statistical framework to combine multivariate spatial data and physical models for hurricane surface wind prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, in press.

9. Fuentes, M., Reich, B., and Lee, G. (2008). Spatial-temporal mesoscale modelling of rainfall intensity using gage and radar data. *Annals of Applied Statistics*, in press.

10. Hacker, J. P., J. L. Anderson, and M. Pagowski (2007). Improved vertical covariance estimates for ensemble-filter assimilation of near-surface observations. *Mon. Wea. Rev.*, 135, 10211036.

11. Hacker, J. P. and D. Rostkier-Edelstein (2007). PBL state estimation with surface observations, a column model, and an ensemble filter. *Mon. Wea. Rev.*, in press.

12. Serge Guillas, Jinghui Bao, Yunsoo Choi and Yuhang Wang, February 2008. Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta, *Atmospheric Environment* **42**, 6, Pages 1338-1348.

13. Hacker, J. P., J. L. Anderson, and M. Pagowski, 2007: Improved vertical covariance estimates for ensemble-filter assimilation of near-surface observations. *Mon. Wea. Rev.*, 135, 10211036.

14. Hacker, J. P. and D. Rostkier-Edelstein, 2007: PBL state estimation with surface observations, a column model, and an ensemble filter. *Mon. Wea. Rev.*, accepted.

15. F. Liu, M.J. Bayarri, J.O. Berger, R. Paulo, J. Sacks (2008). A Bayesian Analysis of the Thermal Challenge Problem. *Computer Methods in Applied Mechanics and Engineering* (in press), doi: 10.1016/j.cma.2007.05.032

16. Ma, C. (2007). Why is isotropy so prevalent in spatial statistics? *Proceedings of the American Mathematical Society*, vol. 135, 865-871.

17. Ma, C. (2007). Planar-temporal stationary correlation models that depend on the maximum norm. *IEEE Transactions on Signal Processing*, vol. 55, 889-896.

18. Ma, C. (2007). Stationary random fields in space and time with rational spectral densities. *IEEE Transactions on Information Theory*, vol. 53, 1019-1029.

19. Ma, C. (2007). Stochastic processes with a particular type of variograms. *Research Letters in Signal Processing*, vol. 2007, Article ID 61579, 5 pages.

20. Ma, C. (2008). Rational covariance functions for nonstationary random fields. *IEEE Transactions on Information Theory*, vol. 54, 895-987.

21. Ma, C. (2008). Recent developments on the construction of spatio-temporal covariance models. *Stochastic Environmental Research and Risk Assessment*, vol. 22 (Supplement 1) S39-S47.

22. Ma, C. (2008). Intrinsically stationary variograms in space and time. *Theory of Probability and Its Applications*, in press.

23. Ma, C. (2008). Construction of non-Gaussian random fields with any given correlation structure.*Journal of Statistical Planning and Inference*, in press.

24. Abhyuday Mandal, Pritam Ranjan and C.F. Jeff Wu "G-SELC: 'Optimization by Sequential Elimination of Level Combinations Using Genetic Algorithms and Gaussian Processes" to appear in *Annals of Applied Statistics*

25. Pahle, J. et al (2008) Information transfer in signaling pathways: a study using coupled simulated and experimental data. *BMC Bioinformatics* **9**:139.

26. Rempala, G. A., Kalbfleisch, T., Teneng, I. (2007) Validation of a mathematical model of gene transcription in aggregated cellular systems: application to L1 retrotransposition. *Journal of Computational Biology*, **14**(3): 339–349.

27. Reich, B., and Fuentes, M. (2007). A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields.*Annals of Applied Statistics*, 1, 249-264.

28. T. J. Santner, P. F. Craigmile, C. A. Calder, and R. Paul "Effect and Pathways Modifiers in a Bayesian Pathways Analysis of the National Human Exposure Assessment Survey for Arsenic in EPA Region 5," to appear in *Environmental Science & Technology*.

29. Singh, A. V., Rouchka, E., Rempala, G., Bastian, C., Knudsen, T. B. (2007) Integrative database management for mouse development: systems and concepts. *Birth defects research (Part C)*, **81**, 1–19.

30. Yamada, Y. R. (2007) *Quantitative models of transcriptional elongation*. PhD thesis, Cornell.

### 4.3 Submitted Papers and Technical Reports

1. M.J. Bayarri, J.O. Berger, E. Calder, K. Dalbey, S. Lunagomez, A.K. Patra, E.B. Pitman, E.T. Spiller, and R.L. Wolpert, Using Statistical and Computer Models to Quantify Volcanic Hazards. Submitted to the *Techometrics* special issue on computer models.

2. M. J. Bayarri, J. O. Berger, M. C. Kennedy, A. Kottas, R. Paulo, J. Sacks, J. A. Cafeo, C. H. Lin, and J. Tu (2005). Bayesian Validation of a Computer Model for Vehicle Crashworthiness. Tech. rep.163, National Institute of Statistical Sciences. (Submitted to Journal of the American Statistical Association)

3. K.J. DeVault, P.A. Gremaud, V. Novak, M.S. Olufsen, G. Vernières and P. Zhao (2008). *Blood flow in the Circle of Willis: modeling and calibration*, submitted for publication in SIAM MMS (Multiscale modeling and simulation: a SIAM interdisciplinary journal).

4. G. Han, T. J. Santner, W.I. Notz, and D. L. Bartel "Prediction for Computer Experiments Having Quantitative and Qualitative Input Variables," under revision for *Technometrics*.

5. Henderson, D. A., Boys, R. J., Krishnan, K. J., Lawless, C., Wilkinson, D. J. (2007) Bayesian emulation and calibration of a stochastic computer model of mitochondrial DNA deletions in substantia nigra neurons, under revision for *J*.A.S.A.

6. Henderson, D. A., Boys, R. J., Wilkinson, D. J. (2008) Bayesian calibration of a stochastic kinetic computer model using multiple data sources, submitted to the *Technometrics* special issue on computer models.

7. Kaufman, C. and Sain, S. Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions (2008), submitted.

8. K. L. Ong, T. J. Santner, and D. L. Bartel (2006) "Robust Design for Acetabular Cup Stability Accounting for Patient and Surgical Variability."

9. Reichert, P. and Mieleitner, J., Analyzing Input and Structural Uncertainty of a Hydrological Model with Stochastic, Time-Dependent Parameters, submitted to Water Resources Research.

10. J.C. Rougier (2008). Efficient Emulators for Multivariate Deterministic Functions, being revised for the Journal of Computational and Graphical Statistics.

11. J.C. Rougier, S. Guillas, A. Maute, A.D. Richmond (2008). Emulating the Thermosphere-Ionosphere Electrodynamics General Circulation Model (TIE-GCM), being revised for Technometrics.

12. Matthew Taddy, Herbert K. H. Lee, Genetha A. Gray, and Joshua D. Griffin, "Bayesian guided pattern search for robust local optimization," submitted for publication, available as tech report UCSC ams2008-02 at
http://www.ams.ucsc.edu/reports/trview.php?content=view&name=ams2008-02

13. White, G., Reichert, P., Bayarri, M.J., Santner, T.J. and Pittman, E.B., Bayesian Kalman Filter for Emulation of Complex Engineering Computer Models, submitted to Technometrics.

14. B. J. Williams, T. J. Santner, W. I. Notz, J. S. Lehman (2006) "Sequential Design of Computer Experiments for Constrained Optimization."

15. Yamada, Y. R. (2008) A look-ahead model for the elongation dynamics of transcription, submitted to the Biophysical journal.

### 4.4 Work in Progress

1. Pankaj Agarwal, Sukhendu Chakraborty, Jim Clark, Sean McMahon, and Hai Yu . Scalable Algorithms for Computing Understory Light and Dispersal in a Forest Growth Model.

2. M.J. Bayarri, J. Berger, F. Liu and R. Paulo. "Modeling issues in implementing Bayesian analyses of complex computer models: Modularization, confounding and marginalization"

3. M.J. Bayarri, J. Berger, F. Liu, R. Paulo and P. Reichert. "Using computer model derivatives to improve emulators and extrapolating predictions"

4. Bondell, H., Reich,B., Fuentes, M., Hacker, J., and Tardif, R. On incorporating a stochastic parameterization into a numerical weather prediction model, in preparation.

5. A. Cintrón-Arias, P. Reichert, A. L. Lloyd, and H. T. Banks, *An Estimation Methodology for Seasonal Transmissibility*, in preparation.

6. K.J. DeVault and P.A. Gremaud. *Analysis and numerical analysis of impedance boundary conditions for hemodynamics*,

7. K.J. DeVault, P.A. Gremaud, V. Novak, M.S. Olufsen. *Computational study of the effects of strokes on cerebral perfusion*,

8. Govindarajan, S., S. Chakraborty, P. K. Agarwal, M. C. Dietze, J. S. Clark, and Michael S. Wolosin. Light competition in forest canopies.

9. S. Guillas, A. Gelfand and S. Sahu, "Bayesian downscaling of an air quality model". In this work, uncertainties are naturally assessed by the Bayesian approach. Meteorological variables are used to locally improve the regional forecasts, using strong priors to give preference to the chemistry transport model.

10. S. Guillas, L. Lefton, Y. Choi, Y. Wang. "Calibration of a 3-D air quality model". We use the Bayesian calibration method to find the best values for 5 parameters of interest in a pollution model.

11. S. Guillas, Chunsheng Ma. "Space-time downscaling of regional ozone forecasts with nonseparable covariance models". In this work we introduce a new type of space-time covariances that enable us to better fit the differences between model outputs and observations. The resulting downscaling should improve on the result for time series.

12. Hacker, J., Tardif, R., Shows, J., Fuentes, M., and Bondell, H. Stochastic parameterization of WRF-1D, in preparation.

13. G. Han and T. J. Santner, "Simultaneous Determination of Calibration Inputs and Tuning Parameters"

14. Gang Han, T. J. Santner, W. I. Notz, and D. L. Bartel. “Prediction for Computer Experiments Having Quantitative and Qualitative Input Variables.”

15. Henderson, D. A., Boys, R. J., Proctor, C. J., Wilkinson, D. J. (2008) Linking systems biology models to data: a stochastic kinetic model of p53 oscillations. In preparation for the Handbook of Applied Bayesian Analysis.

16. Cari Kaufman and Derek Bingham. “Efficient Emulators of Computer Experiments Using Compactly Supported Correlation Functions.” In preparation.

17. Kaufman, C. and Sain, S. Functional ANOVA modeling of regional climate model experiments, in preparation.

18. Kaufman, C., Sain, S., and Fowler, H. Attributing Sources of Uncertainty in Regional Climate Projections for Europe, in preparation.

19. McMahon, S. M., J. S. Clark, P. Agarwal, and H. Yu. (in prep). Long-distance dispersal in forests cannot explain or sustain community dynamics.

20. McMahon, S. M., J. S. Clark, P. Agarwal, and H. Yu. (in prep). The role of individual differences in forest community inference: experiments using stochastic simulations.

21. H. Moon, Angela Dean, and T. Santner “Two-Step Group Screening for Computer Experiments”

22. Reichert, P., White, G., Bayarri, M.J., Pitman, E.B., and Santner, T.J., Mechanism-Based Emulation of Dynamic Simulators: Concept and Application in Hydrology, to be submitted in May 2008.

23. Shows, Fuentes, Bondell, Hacker and Tardif. “Stochastic parameterization of WRF-1D” (manuscript in preparation).

24. Yamada, Y. R. Error analysis for the look-ahead model.

25. Yamada, Y. R., Wilkinson, D. J. Parameter estimation for a stochastic model of transcription elongation dynamics.

# 5 Workshops

This SAMSI Program also organized many Workshop, several in collaboration. Similarly it had an instrumental role to play in the organization or development of several other meetings. They had different specific roles to play, as detailed below, although a common goal was to foster interaction among different group of researchers that might otherwise not be exposed to this enriching collaboration.

**Kickoff Workshop** was held September 10-14, 2006. Its principal goal was to engage a broadly representative segment of the statistical, applied mathematical and computer modeling communities in formulation and pursuit of specific research activities to be undertaken by the Program Working Groups, which include: *i)* Formulation of central research issues, *ii)* Identification of testbed computer models and data, *iii)* Formation of external partnerships between the Working Groups and others, especially Kickoff Workshop participants, interested in the Program. These goals were successfully attained, as can be seen in the rest of this report.

**Joint Engineering and Methodology Subprograms Workshop** was held on October 26-27, 2006 at SAMSI. While the kickoff workshop intended to have the working groups started, this one attempted to focus their orientation, profile the main issues and interest, get the working group exposed to insights, methods and ideas from experts in the field. The emphasis was on discussions, interaction, problem solving and the detailed specification of problems of engineering and more general methodology. As part of the workshop, Anthony O'Hagan (U. Sheffield) delivered a SAMSI distinguished lecture.

**Joint NCAR and SAMSI workshop. Geophysical Models at NCAR: A Scoping and Synthesis Workshop** (Opening workshop for the climate modeling SAMSI group.) The SAMSI climate/weather modeling group and NCAR jointly hosted a workshop at NCAR on November 13-14. The objective of the workshop was to introduce researchers to the atmospheric numerical models that are being developed and used at NCAR.

It was intended as a scoping and brainstorming meeting where four NCAR modeling groups interacted with a large group of SAMSI statisticians and applied mathematicians interested in the design and analysis of computer experiments. The NCAR geophysical models/groups were:

- Upper atmosphere model (TIEGCM) (HAO: Maura Hagan, Ray Roble, Art Richmond)
- Single column boundary layer model (RAL/MMM: Josh Hacker)
- Two dimensional turbulence in Navier Stokes flows (IMAGe: Annick Pouquet)
- Land component of the NCAR climate model (CGD: Gordon Bonan; U Kansas: Johannes Feddema)

The intention was to idenify concrete problems, to help structure the statistical climate modeling working group activity at SAMSI. For each modeling group a statistical researcher served as a liason to guide collaboration among the modeling group and the statistical working groups.

**Mathematical and statistical challenges in systems biology modeling** March 5–7, 2007. The workshop gathered many of the world's leading experts in systems biology modeling to present their view of the key challenges and potential strategies for their solution. It was a stimulating

3 days, with many highlights, including a SAMSI Distinguished Lecture by Daniel Gillespie on the importance of stochastic effects in the modelling of genetic and biochemical networks. There was an excellent mix of disciplines present, including computational biologists, physicists, applied mathematicians, statisticians and probabilists. The two key problems highlighted by the workshop were parameter estimation issues (especially in the context of stochastic models), and multi-scale modelling of stochastic kinetic biochemical systems.

**Joint SAMSI/MUCM Mid-Program Workshop** was held April 2–3, 2007. This joint workshop brought together methodologists from the US NSF-sponsored SAMSI program on the development, assessment and utilization of complex computer models with those in the Research Councils UK (RCUK)-funded MUCM (Managing Uncertainty in Complex Models) project, offering opportunities for cross-fertilization of these large-scale parallel international efforts. Main goals were to share information on the progress of ongoing projects annd to coordinate research efforts that have been stimulated by the close collaboration between the MUCM and SAMSI research teams.

The workshop drew a stellar collection of speakers and participants, offering a unique opportunity for collaboration among an international community of scholars studying related problems. The format featured a series of talks, each followed by extended discussion of the issues it highlighted, in the context of previous presentations. Selected outside experts offered invited discussion and criticism, helping to bring perspective and often new direction to these lines of research. A number of scholars were able to particimate remotely through our offering of WebEx and Teleconference access.

**Satellite Meeting. Terrestrial Models working group** had an intensive one-day satellite meeting on April 4, 2007. It provided an update of the progress within the working group and to elicited input from and generated collaborations with a small group of interdisciplinary scientists with expertise in one or more of the relevant fields (ecology, climatology, statistics, computer science). It consisted in short presentations and discussions on the work that was done to date, feedback/input from all, and plans for potential new research to extend th2 effort. Invitees to the workshop included:

Pankaj Agarwal, Duke; Jim Clark, Duke; Benoit Courbaud, CEMAGREF; Jim Crooks, SAMSI; Mike Dietze, Harvard; Ken Feeley, Wake Forest; Alan Gelfand, Duke; Cari Kaufman, NCAR, SAMSI; Sean McMahon, Duke; Paul Moorcroft, Harvard; Miles Silman, Wake Forest; Steve Sain, NCAR; Maria Uriarta, Columbia; Wei Wu, Duke. Duke Grads: Dave Bell and Carl Salk.

**Atlantic Coast Conference on Mathematics in the Life and Biological Sciences** . This Conference had a large group of SAMSI researchers supported by the *Dynamic of Infectious Desease* working group. In particular, the funded participants were Tom Banks (NCSU), Ariel Cintrón-Arias (SAMSI/NCSU), Marie Davidian (NCSU), Jimena Davis (NCSU), Sava Dediu (NCSU), Stacey

Ernstberger (NCSU), Sarah Grove (NCSU), Shuhua Hu (NCSU), Grace Kepler (NCSU), Alun Lloyd (NCSU), Johnny Samuels (NCSU), and Karyn Sutton (Arizona State).

The Conference was held in May 3-5, 2007, in the Virginia Bioinformatics Institute of Virginia Tech (Blacksburg, Virginia)

**Joint SAMSI-NCAR Workshop. Application of Random Matrices Theory and Methods** . Held at NCAR on 7-10 May 2007. This workshop served as the transition workshop for the Random Matrices group of the SAMSI program. The structure was traditional, and included a blend of tutorial and research talks. Ample time was reserved for discussion.

**Transition Workshop** May 14–16, 2007. The Transition Workshop was held to present initial results developed by this program and to formulate follow-up plans for continuing the research in this interdisciplinary area. Program members Susie Bayarri (U Valencia), Jim Clark (Duke U), Cari Kaufman (SAMSI/NCAR), Fei Liu (Duke U), Bruce Pitman (U Buffalo), Peter Reichert (EAWAG), Jonathan Rougier (U Bristol, UK), Elaine Spiller (SAMSI), and Robert Wolpert (Duke U) presented research developed at the Working Groups; guest Dave Higdon (LANL) presented related work seeking low-dimensional structure in high-dimensional datasets. The Transition Workshop offered an opportunity for all the Working Groups to learn of each others' progress and to foster research collaborations with interested scientists outside the Working Groups; the strength of its program made the Workshop a striking success.

**Transition/Satellite Meeting.** The Calibration of Computational Models of Cerebral Blood Flow working group followed the previous workshop with an intensive satellite meeting on blood flow modeling. The outside speakers were Jordi Alastruey Arimon, Imperial College (modeling, computation), Suncica Canic, Dept. of Mathematics, U. of Houston (modeling and analysis), Vera Novak, Beth Israel Medical Center, Harvard (experiments), Brooke Steele, Biomedical Engineering, NCSU/UNC (modeling), Charles Taylor, Dept. of Surgery and Dept. of Mechanical Engineering, Stanford U. (modeling, calculations, experiments). We expected this workshop to

1. consolidate some informal collaboration between the various groups,
2. identify remaining key issues in a "holistic" (modeling, computation, experiments) approach to the problem,
3. give some exposure to the junior members of our group (K. DeVault, graduate student and G. Vernieres, SAMSI postdoc),
4. foster new collaborations between MDs, engineers, mathematicians and statisticians.

The first three goals were successfully achieved; contributions to the fourth one were made.

**Joint SAMSI-NCAR Workshop. Application of Statistics to Numerical Models: New Methods and Case Studies.** This was held at NCAR on 21-24 May 2007. This workshop served as the transition workshop for the climate modeling group of the Computer Models SAMSI program. As such, one of the objectives of this workshop was to discuss and present results of the research conducted by the climate/weather working group. The structure was traditional, and included a blend of moderated discussions, tutorials, and research talks. Also, ample time was reserved for discussion and for presentations on progress on the specific modeling projects initiated in the first workshop.

**Joint Statistical Meetings JSM07** . Salt Lake City, July 29-August 2, 2007. Three Topic Contributed Sessions were organized under the SAMSI umbrella for the 2007 Joint Statistical Meetings under the Complex Computer Models Program:

1. *Methodological Issues in Engineering Applications of Computer Models: A SAMSI Program.* organized by Tom Santner. Sponsored by the Section on Bayesian Statisticis (SBSS) of ASA. Speakers were Herbert Lee, Fei Liu, Gentry White, Dianne Bautista and Laura Swiler

2. *The role of statistics in ecological and climate modeling.* Organized by M. Fuentes. Sponsored by the climate and weather modelling SAMSI working group and the Environmental Statistics (ENVR) Section of the ASA.

3. *Using Computer Models for Ecological, Environmental, and Biological Applications: A SAMSI Program.* Organized by M.J. Bayarri. Sponsored by the Section on Bayesian Statisticis (SBSS) of ASA. Speakers were Leonard Smith, Peter Reichert, Ariel Cintron-Arias, Bruce Pitman, Serge Guillas.

**IMA Hot Topics workshop on "Stochastic models for intracellular reaction networks"** organized by Lea Popovic, Greg Rempala and John Yin, to take place in Minneapolis in May 2008.This workshop features many of the Systems Biology working group participants as invited speakers (including Samuel Kou, Greg Rempala and Darren Wilkinson).

**Calibration and Validation of Complex Computer Models: Bayesian Approaches, Bayesian Solutions.** A Satellite Meeting to ISBA 2008. Sydney, Australia, July 27-28, 2008. This is a workhsop organized following previous meetings and preliminary planning during the SAMSI Program. It is organized by the Program Leader, one of the invited participants is SAMSI director James Berger, and most speakers were participants in the Program. We will hear about the follow up and results of the Program research projects that were very advanced when the Program ended and are producing promising results. This will further foster long lasting collaborations initiated during the SAMSI Program. Invited Talks will be delivered by M.J. Bayarri, D. Higdon, J. Rougier, D. van Dick; J. Berger is invited to master the Panel Discussion Session. There will be two Invited to Contribute Sessions with speakers: J. Oakley, S. Reese J. Gattiker and J. Loeppky and two New Researcher Sessions with (invited) participants L. House, R. Wilkinson, L. Soares Bastos and T. Cui.

# 6 Other Activities

## 6.1 Research sessions, working get-together, brainstorming sessions ...

**Calibration of computational models of cerebral blood flows.** Dr. Novak and Dr. Zhao (Beth Israel Deaconess Medical Center, Harvard University) visited the working group at SAMSI in Fall 06 and Spring 07 to discuss various issues related to data acquisition, measurement procedures and interfacing with the numerics. DeVault and Gremaud will visit them in Spring 08.

**Systems Biology. March Intensive Research Sessions** involving a select group of contributors from the US and abroad. The main theme was on parameter inference for deterministic biochemical network models. The main focus was on a metabolic network model in fish, of interest to toxicologists at the Environmental Protection Agency. A second application was a stochastic model of mtDNA accumulation, involving collaborators from Newcastle University.

**Systems Biology. April Intensive Research Sessions.** In April there was an intensive research session on parameter inference for stochastic kinetic models of molecular motor dynamics, based on models and data originating from Cornell.

**Granular Materials - Engineering Applications** Several memebers of this working group are planning an intensive get together working sessions in June 2008 to plan future collaboration and research activities.

**Terrestrial Models** .- The April 2007 Workshop of this working group will be integrated with the wireless sensor Program in 2007/08, providing important background on modeling.

## 6.2 Submitted Research Projects

Several Research Projects have been also submitted as a direct consequence of the work and collaboration during the Program:

1. M.J. Bayarri (co-PI), J.Berger, E.B. Pitman (co-PI), E. Spiller, R. Wolpert (PI), have obtained funding under the NSF Focused Research Group program to extend their early work on hazard mapping. *RG Collaborative Research: Prediction and Risk of Extreme Events Utilizing Mathematical Computer Models of Geophysical Processes. Total amount of this three-institution proposal was approximately $980,000.*

2. J. Clark (PI) and collaborators. Assessing forest dynamics under global change: A model linking physical and biological processes across multiple scales. Submitted to the MacDonnell Foundation, 2008.

3. M. Fuentes, Shane Reese, Astrid Maute, Mike Wiltberger, and Art Richmond."Models, Tools and Analysis for Studies of the Upper Atmosphere and the Magnetosphere ". CMG Collaborative Research.

4. M. Fuentes (PI), L. Xie and B. Reich (co-PI). NSF-DMS. 2007-2010. $260,000. Multivariate space-time models and methods to combine large disparate spatial data and numerical models.

5. S. Guillas. College of Sciences Faculty Research Developement Grant-Georgia Tech Foundation, Statistical calibration of RAQAST (2006). Funded, $10,434.00

6. S. Guillas. NASA-Atmospheric Composition Modeling and Analysis: High Resolution Assessment of Stratosphere-Troposphere Exchange and its Impact Upon the Background Tropospheric Ozone Concentrations. Program: Atmospheric Composition Modeling and Analysis, co-PI, pending, $55,849.00

7. S. Guillas. NSF-DMS: CMG–Statistical Improvements of Three-Dimension (3-D) Model Air Quality Forecasts. PI, pending, $704,562.00

8. Tom Loredo (Cornell U), Carlo Graziani (U Chicago), and Michael Stein (U Chicago). A DOE PSAAP (Predictive Science Academic Alliance Program) proposal to simulate Type Ia Supernova explosions emerged from the Methodology Working Group. Finalist but not funded; in preparation for resubmission to NSF CDI program.

9. A renewal of NSF grant NSF-DMS-0616597 (Olufsen & Gremaud) will be submitted shortly.

10. E. B. Pitman, and U. Buffalo geologist M.F. Sheridan, with support from M.J. Bayarri, J. Berger and L. Pericchi, submitted a proposal to use investigate debris flows at Tungurahu in Equador and Ruapehu Volcano in New Zealand.

11. E.B. Pitman, A. Patra and U. Buffalo geologist M. Bursik have obtained funding to develop advanced simulation methodology for mass flows.

12. T. Santner. The Cornell University/Hospital for Special Surgery Biomechanics Program awarded GRA support for the calendar year 2008 in order to further the development of statistical methodology to facilitate designing prosthetic devices.

13. T. Santner, Robert Spilker and Timothy Wright. Submitted a proposal (December of 2007) to the National Science Foundation (Cyber-Enabled Discovery and Innovation) to fund a program in tissue engineering that will develop replacement cartilage as the next generation of

joint replacement treatments. The interdisciplinary team consists of a computational modeler from RPI (Robert Spilker) and a experimental team at the Cornell University/Hospital for Special Surgery (headed by Timothy Wright).

14. Robin Tokmakian (PI, Naval Postgraduate School), James Gattiker, Peter Challenor, "Uncertainties within a Climate Model's Ocean Component", submitted to US Department of Energy Office of Biological and Environmental Research. Jan 2007.

15. Robert Wolpert (Duke U Statistical Science), Berndt Mueller (Duke U Physics), and several other investigators. DOE PSAAP proposal to develop simulators and emulators for high-energy heavy ion collisions to deepen our understanding of novel states of matter (quark-gluon plasma) grew in part out of the Methodology Working Group's studies. Finalist but not funded; in preparation for possible resubmission to another agency.

### 6.3 Other Collaborative Research and tentative Projects

Apart from the collaborations in the submitted research projects and publications, other collaborative work grew out of this SAMSI Program and tentative research projects were initiated and are going on as a direct consequence of the Program:

- Terrestial working group and the air quality modeling group at EPA. Jim Crooks has developed an emulator for stochastic simulators, the specific application being the large forest simulator. The emulation method, based on the Kernel Stick-Breaking Process model of Dunson and Park, is extremely flexible and is able to infer highly non-Gaussian output distributions without a large number of simulator runs. Crooks presented this method at the 9th Case Studies in Bayesian Statistics workshop in October 2007 and at the Environmental Protection Agency - RTP in March 2008. The air quality modeling group at EPA was extremely interested in the emulator in particular and in statistical methods for addressing various other issues in computer experiments. Given funding Crooks will be working with this air quality group full time starting in July.

- A. Cintrón-Arias, P. Reichert, A. L. Lloyd, and H. T. Banks, are ensuing collaboration in a research project pertaining to model validation (using a Bayesian methodology) in the context of seasonal infectious disease epidemiology. Three main contributions are conveyed: (i) to model seasonal transmission as a continuous-time stochastic process known as a mean-reverting Ornstein-Uhlenbeck process, (ii) to carry out the parameter estimation using a Markov chain Monte Carlo (MCMC) scheme, (iii) to construct sample trajectories of the effective reproduction number (a quantification of transmissibility) from the joint posterior density calculated by the MCMC scheme. Tentative title: *An Estimation Methodology for Seasonal Transmissibility*, in preparation.

- M.J. Bayarri (U Valencia), James O. Berger (Duke U Statistics), Eliza S. Calder (U Buffalo Geology), E. Bruce Pitman (U Buffalo Math), Elaine T. Spiller (SAMSI, Duke U Math), and Robert L. Wolpert (Duke U Statistics), began a collaboration on the simulation and emulation of geophysical flows.

- James L Crooks (Duke Statistics), Jim Clark (Duke Ecology), Sean McMahon (Duke Ecology), Pankaj Agarwal (Duke CS), began a collaboration studying the emulation of forest simulation;

- S. Guillas and Yuhang Wang.- A joint effort to calibrate pollution models at Georgia Tech.

- S. Guillas, Alan Gelfand and Sujit Sahu. Bayesian downscaling of regional ozone forecasts.

- S. Guillas and Chunsheng Ma, Space-time downscaling of regional ozone forecasts with non-separable covariance models.

- Cari Kaufman (SAMSI/NCAR), Derek Bingham (Simon Fraser U), began a collaboration studying compactly supported covariance functions.

- Cari Kauffman and Jonty Rougier (Terrestial Models working group) are building a model to predict the spatial and temporal evolution of soil moisture under varying climatic conditions. Soil moisture is an important factor in tree growth and fecundity, and these predictions are part of the Clark group's larger goal of studying the effect of climate change on forest dynamics. The approach used was to build a state-space model in which the temporally evolving soil moisture field is the unknown state vector. The initial predictions used a simple linear model, which could be fit using a Kalman filter, but they did not take advantage of valuable prior knowledge about the evolution of soil moisture over time. In our new model, this process is governed by a simple, catchment-level hydrology model we developed based on the literature, combined with a stochastic model for distributing total water according to local topography. We are now developing an MCMC algorithm for sampling the non-time-varying parameters, with an embedded smoothing algorithm to sample the temporally evolving soil moisture fields conditional on these parameters.

- H. Lee, G. Gray, and M. Taddy are working on a project integrating pattern search optimization with a treed Gaussian process oracle.

- Chunsheng Ma (Wichita State U, Math & Stat) and Serge Guillas (Ga Tech U, now U Kentucky), began a collaboration on the Space-time downscaling of regional ozone forecasts with nonseparable covariance models.

- Greg Rempala and John Yin collaborating on virus reproduction models.

- Greg Rempala and Enrico Capobianco collaborating on multiscale modeling of biochemical networks.
- J. Rougier, S. Conti, and A. O'Hagan on working on emulators for dynamical systems.
- Tom Santner and Angela Dean (OSU) begun a research project in the computer experiments area, ongoing since March 2007.
- Tom Santner (Ohio State U Statistics) is working with Cornell University PhD students Jason Long and Garry Hayeck on designing experiments for computational simulators that they are building to explore a hip resurfacing system and an elbow prosthesis, respectively. For example, in the first project, they seek to understand how design and environmental variables have an influential effect on strain magnitudes near the rim of a surface implant. In particular, they wish to identify variables cause strains large enough to lead to rapid damage accumulation in the bone and possible neck fracture?
- Tom Santner (Ohio State U Statistics) has begun to work with Tim Wright (Hospital for Special Surgery, NYC) and Robert Spilker (RPI, Troy NY) on tissue engineering.
- Darren Wilkinson continuing to collaborate with Richard Yamada on calibration of molecular motor models.
- Brian Williams, Max Morris, and TJ Santner are working on methodology for using multiple computer models and multiple data sources simultaneously to infer calibration parameters.
- Robert Wolpert (Duke Statistics), Berndt Mueller (Duke Physics), Steffen Bass (Duke physics) began a collaboration on the emulation of high-energy ion collision simulation.
- Robert Wolpert (Duke Statistics), Doug Nychka (NCAR), Dan Cooley (CSU), began a collaboration on the spatial extreme and Lévy random fields;

# 7 Education and Outreach

## 7.1 Credit Courses

The Program offered two courses, one in the Fall and one in the Spring semester, which were credited 3 credits/units at each of the participating Universities.

**Granular Flow Course** to help educate the next generation of physicists, engineers, mathematicians and statisticians, a course (3 credits/units) on the flow of granular materials was offered during the Fall 2006 semester. Seven students (4 women, 3 men) and 3 post-doctoral fellows (1 woman, 2 men) participated in the course. The instructor was Bruce Pitman (U. Buffalo).

**Course on Environmental Modelling** The Subprogram on Ecological, Environmental and Climate models offered a course during Spring 2007, which included not only detailed study of these models, but also statistical inference and uncertainty analysis for them. It was held at SAMSI, on Wednesdays 4:30 to 7pm. Instructors: M. Fuentes, J. Clark, Peter Reichert, G. Hegerl.

## 7.2 Workshops for Students and New Investigators

The following workshops were mainly oriented to graduate students, post-docs and new researchers, with a primarily educational goal, to make these new researchers familiar with the models, problems and issues in the development, design and utilization of Computer Models.

**Summer School on the Design and Analysis of Computer Experiments** August 11-16, 2006 at IRMACS, Simon Fraser University. This Summer school was conducted jointly between SAMSI and the Canadian National Program on Complex Data Structures. It was and excellent opportunity for students, new researchers, and others interested in becoming involved with the study of computer models to learn many of the latest methodological developments in the area. It consisted on a two-days Short Course, two-days Software Symposium, and two days Hands-on Problem Solving.

An Undergraduate Workshop was held at SAMSI on March 2-3, 2007. Approximately 25 undergraduate students from ?? undergraduate majors and ?? colleges and universities were shown how computer models are used in the engineering design of consumer products, the description of watershed activity, the prediction of climate and weather, and the temporal evolution of terresterial land cover. A computer laboratory exercise was used to illustrate aspects of the complexities of using such models.

**SAMSI Two-Day Undergraduate Workshop** March 2-3, 2007 at SAMSI. Approximately 30 undergraduate students from undergraduate majors colleges and universities were shown how computer models are used in the engineering design of consumer products, the description of watershed activity, the prediction of climate and weather, and the temporal evolution of terresterial land cover.The workshop presented the idea of mathematical models, their numerical computer implementation, and their statistical utilization, in a wide variety of scenarios and at a level adequate for the wide range of students present. Presentations were given by both new and senior researchers, and hand-on computer practices helped students grasp the basics of computer simulators, understand the role of the inputs and input sensitivity, as well as served to illustrate aspects of the complexities of using such models. Big emphasis was given to discussion, which was alive and insightful. The workshop was very well attended (the pre-registrations amply exceeding the quota for this workshop) with students from all over USA. The Workshop accomplished the goals of exposing and interesting a wide diversity of bright students to the area of Complex Computer Models, their development, assessment and utilization.

**Summer Graduate Workshop on Data Assimilation for the Carbon Cycle** . Held at NCAR on July, 8-13, 2007. This summer school exposed students in the geosciences, ecology, and mathematics to multidisciplinary science through a focus on estimating the sources and sinks of carbon for the Earth system. One goal was to train the next generation of researchers to work within a multidisciplinary science team that combines geoscientists, ecologists, applied mathematicians, and statisticians. Participants obtained an overview of this problem but also some specific skills in tackling inverse problems and working with geophysical and biogeochemical models.

## 7.3 Presentations at Professional Meetings

Apart from the meetings, workshops, presentations at working groups, etc., directly organized by the Program, the following presentations at Professional Meetings have been held or are planned for the immediate future:

1. M.J. Bayarri, J.O. Berger and G. Molina. "Incorporating Uncertainties into traffic Simulators", Invited Talk. Cost Action 285 Final Symposium on Recent Advances for Modeling and Simulation Tools for Communications Networks and Services. University of Surrey, United Kingdom, 28-29 March 2007.

2. M.J. Bayarri, J.O. Berger, F. Liu, and R. Paulo. "Invited talk: Some intriguing methodological issues when statistically analyzing computer models data." Transition Workshop. Research Triangle Park, NC, May 14–16, 2007. 86.

3. M.J. Bayarri, J.O. Berger, F. Liu, and R. Paulo. "Modeling issues when combining field and computer model data for prediction". Invited talk. International Workshop on Statistical Modelling 2007. Barcelona, July 2–6, 2007

4. M.J. Bayarri, J.O. Berger, F. Liu, R. Paulo and J. Sacks. Invited Plenary Session: Validation of complex computer models with functional outputs. IMS/ASA/SPES 14th Annual Spring Research Conference (SRC) on Statistics in Industry and Technology. Ames, IA, May 21-23, 2007. Iowa State, Ames, IA (USA), May 21-23 2007.

5. M.J. Bayarri, J.O. Berger, F. Liu, J. Sacks. opening talk: Computer Model Validation with Functional Outputs. Joint Engineering and Methodology Subprograms Workshop. Research Triangle Park, NC, October 26-27.

6. Jim Berger. "Analysis of Complex Computer Models of Processes". Invited talk. BISP5, Fifth Workshop on Bayesian Inference in Stochastic Processes. Valencia (Spain), June 14-16, 2007

7. Jim Berger. "The world of computer modelings: Dealing with imperfect models" Wald Lecture III. Joint Statistical Meetings, Salt Lake City (USA), July 29-August 2, 2007.

8. Richard Boys. " Bayesian inference for stochastic epidemic models with time-inhomogeneous removal rates". Invited talk. BISP5, Fifth Workshop on Bayesian Inference in Stochastic Processes. Valencia (Spain), June 14-16, 2007.

9. Enrico Capobianco. "A multiscale tour in protein interactomics" and "A multiscale look at protein interactomes". Contributed posters. IMA workshop on Organisation of Biological Networks, Minnesota, March 1–3, 2008.

10. Ariel Cintron-Arias. Analysis of Seasonal Epidemiological Parameters. Presentation at JSM 2007, Salt Lake City, July 29 - Aug. 2.

11. J. Clark Opening plenary, International Statistical Ecology Conf, St Andrews

12. J. Clark Keynote speaker, Annual Spring Ecology Symposium, Madison

13. J. Clark Invited speaker, International Society for Bayesian Analysis, Hamilton Island (July 2008)

14. J. Clark. Ecological Forecasting: Applications of model-data fusion techniques, ESA Symposium, Milwaukee

15. J. Clark. Prediction of biome boundary shifts in regional and global dynamic vegetation models, Yokohama

16. J.L. Crooks. "Bayesian Analysis of Stochastic Computer Model Output: A Forest Simulator Response to Future Climate Change." Invited talk, 9th Case Studies in Bayesian Statistics (CMU), Pittsburgh PA, 2007 Oct 20.

17. J.L. Crooks. "The Forest of the Future: Emulating a Stochastic Simulator." U.S. Environmental Protection Agency (RTP), 2008 March 11.

18. M. Fuentes, "M. Nonparametric spatial models", invited speaker for the ISI international conference 2007 (August 2007, Portugal).

19. M. Fuentes, "Likelihood approximation for spatial temporal processes", invited speaker for ENAR 2007 (March 2007, Atlanta).

20. M. Fuentes, "Spatial temporal modeling framework for hurricane wind fields", plenary speaker for the CLAPEM international conference (February 2007, Lima, Peru).

21. M. Fuentes, "Spatial nonparametric models for wind fields", plenary speaker for the International Conference on Advances in Interdisciplinary Statistics and Combinatorics,(October 2007, North Carolina).

22. J. Gattiker, "Calibrating Climate Models: a case study and comparison of DACE approaches.", Environmental Statistics Section of the Royal Statistical Society meeting in Southampton, 12 Jan 2007.

23. S. Guillas. *SAMSI/MUCM Mid-Program Workshop*, SAMSI, Research Triangle Park, NC, April 2007.

24. S. Guillasl. *SAMSI/NCAR Workshop on Application of Random Matrices Theory and Methods*, National Center for Atmospheric Research, Boulder, CO, May 2007.

25. S. Guillas. *Transition Workshop*, SAMSI, Research Triangle Park, NC, May 2007.

26. S. Guillas. *SAMSI/NCAR Workshop on Application of Statistics to Numerical Models: New Methods and Case Studies*, National Center for Atmospheric Research, Boulder, Colorado, May 2007.

27. S. Guillas. *6th Spatial Econometrics and Statistics workshop*, Dijon, France, June 2007.

28. S. Guillas. *The 27th Annual International Symposium on Forecasting*, New York, NY, June 2007.

29. S. Guillas. *Platinum Jubilee Conference* of the Indian Statistical Institute, Kolkata, India, January 2008.

30. S. Guillas. *Joint Meeting of the Statistical Society of Canada and the Société Française de Statistique*, Ottawa, Canada, May 2008.

31. S. Guillas. *International Workshop on Recent Advances in Time Series Analysis*, Cyprus, June 2008.

32. C. Kaufman, "Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions", American Geophysical Union Fall Meeting, San Francisco, CA (December 2007).

33. H. Lee. "Treed Gaussian processes and adaptive sampling." Invited talk, Joint Engineering and Methodology Subprograms Workshop on Development, Assessment and Utilization of Complex Computer Models, SAMSI, Durham, NC, October 27, 2006.

34. H. Lee. "Treed Gaussian processes for surrogate modeling under uncertainty." Invited talk, SIAM Conference on Computational Science and Engineering, Costa Mesa, CA, February 22, 2007.

35. H. Lee. "Global optimization using pattern search and treed Gaussian processes." Invited talk, Mid-program Workshop on Development, Assessment and Utilization of Complex Computer Models, SAMSI, Durham, NC, April 2, 2007.

36. M. Olufsen SIAM Conference on the life sciences Aug. 2008,

37. Juergen Pahle. "Information transfer in signaling pathways: a study using coupled simulated and experimental data. Contributed talk. ECCS 2007, Dresden (Germany). October 1–6, 2007.

38. R. Paulo, M.J. Bayarri, J.O. Berger, and P. Reichert. Invited Talk: An Alternative Approach to Modelling the Bias Term. 84. Joint SAMSI/MUCM Mid-Program Workshop (2006-2007 Program on Development, Assessment and Utilization of Complex Computer Models). Research Triangle Park, NC (USA), April 2-3, 2007.

39. P. Reichert Joint SAMSI/MUCM Mid-Program Workshop talk on Computer Emulation of Dynamic Models.

40. Peter Reichert. Analyzing Input and Structural Uncertainty of a Hydrological Model with Stochastic Time-Dependent Parameters. Presentation at JSM 2007, Salt Lake City, July 29 - Aug. 2.

41. Peter Reichert. Mechanism-Based Emulation of Dynamic Simulators: Concept and Application in Hydrology. Presentation to be held at JSM 2008, Denver, Aug. 2 - 7, 2008.

42. J. Rougier, "Multivariate Emulation", given at 2nd Intl PUCM Meeting (Durham, Sept 2007); MUCM one-day meeting on Multivariate Emulation (Aston, March 2008).

43. J. Rougier, "Avalanche prediction", given at statistics@bristol meeting (Nov 2007).

44. T. Santner SAMSI Workshops (March 2007 and May 2007)

45. T. Santner Joint Statistical Meetings (July 2007) talk on Sequential Design of Experiments to Find Pareto Optimal Sets

46. T. Santner Sandia CSRI Workshop on Mathematical Methods for Verification and Validation (August 2007)

47. T. Santner MBI Workshop on Tissue Engineering (October 2007) talk on interdisciplinary integration of computational, experimental, and statistical methods for engineering design.

48. E. Spiller, "Models for Volcano Avalanches. Constructing a Risk Map for Pyroclastic Flows: Using simulations and data to predict rare events," invited talk at NCAR Applications of Statistics to Numerical Models: New Methods & Case Studies. Boulder, CO May 2007

49. E. Spiller, "Models for Volcano Avalanches. Constructing a Risk Map for Pyroclastic Flows: Using simulations and data to predict rare events" invited talk at SAMSI Workshop on EXTREMES: Events, Models and Mathematical Theory. RTP, NC 2008 January 22–24

50. M. Taddy "Pattern search optimization with a treed Gaussian process oracle," Kickoff Workshop on Development, Assessment and Utilization of Complex Computer Models, SAMSI, Durham, NC, September, 11, 2006 (poster)

51. M. Taddy "Emulator guided pattern search," SIAM-CSE, Costa Mesa, CA, February, 2007 (invited talk)

52. M. Taddy "Pattern search optimization with a treed Gaussian process oracle," Joint Statistical Meetings, Salt Lake City, UT, July 30, 2007 (contributed talk)

53. M. Taddy "Optimization and sensitivity analysis with help from statistics," ICCOPT-II, Hamilton, ON, August, 2007 (invited talk)

54. M. Taddy "Bayesian guided pattern search for robust local optimization," MCMSki, Bormio, Italy, January 10, 2008 (poster)

55. G. White Joint SAMSI/MUCM Mid-Program Workshop talk on Computer Emulation of Dynamic Models

56. Gentry White. Bayesian Kalman Filter for Emulation of Complex Computer Models. Presentation at JSM 2007, Salt Lake City, July 29 - Aug. 2, 2007.

57. Darren Wilkinson. "The Chemical Langevin Equation: bridging the gap between high and low level Systems Biology models". Invited talk. BIRS workshop on "Bioinformatics, Genetics and Stochastic Computation: Bridging the Gap" at the Banff International Research Station, Canada. July 2–6, 2007.

58. Darren Wilkinson. "High-thoughput data for Systems Biology modelling". Invited talk. CRISM/INI workshop on "Bayesian analysis of high-dimensional data". Warwick (UK), April 14–16, 2008.

59. Darren Wilkinson. "Parameter estimation for a stochastic kinetic model of p53 oscillations". Opening keynote presentation. Workshop on Computational Systems Biology. Leipzig (Germany), June 11–13, 2008.

60. R.L. Wolpert. "Pyroclastic Flow Models." Invited talk, SAMSI Computer Modeling Workshop, 2007 May 5–14; RTP, NC.

61. R.L. Wolpert. "Modeling with infinitely-divisible moving averages." Invited talk, Joint Statistical Meetings, Salt Lake City, UT, 2008 July 29–Aug 02.

62. R.L. Wolpert. "Bayesian Semiparametric Space-time Models." Invited talk, International Biometric Society Meeting, Coffs Harbour, NSW, AU. 2007 December 02–06.

**BIRS** Darren Wilkinson and other participants of the system Biology working group were involved in the BIRS workshop on "Bioinformatics, Genetics and Stochastic Computation: Bridging the Gap" at the Banff International Research Station, Canada. July 2–6, 2007.

http://www.pims.math.ca/birs/birspages.php?task=displayevent\& event_id=07w5079 and this therefore provided a natural venue for continuing face-to-face discussion among the community that is developed, as well as to disseminating work and results from the SAMSI Program.

### 7.4 Seminars

A number of presentations at Department seminars have been given or are planned for the near future:

1. M.J. Bayarri. Statistical Validation and Calibration of Computer Models. Department of Statistics, University of Georgia, April 12, 2007.
2. M.J. Bayarri. Assessing the probability of a catastrophic (rare) event using computer models and field data. Working group Bayesian Analysis of Extremes of the 2007-2008 SAMSI Program on Risk Analysis, Extreme Events and Decision Theory. March 24, 2008.
3. M.J. Bayarri. Constructing a Risk Map for Pyroclastic Flows: Using simulations and data to predict rare events, Part II. Department of Statistical Science, Duke University. March 24, 2008.
4. M.J. Bayarri. Quantifying risk of catastrophic events using computer and statistical models. Dep. de Estadística e I.O., Universidad Pública de Navarra, May 9, 2008
5. J. Clark. Special Seminar, Data Assimilation for the Carbon Cycle, NCAR
6. J. Clark. Biomathematics, NCSU
7. J. Clark. CEMAGREF, Grenoble
8. DeVault: *Data calibration for numerical simulation of blood flow in the Circle of Willis*, Graduate Student Seminar, Department of Mathematics, NCSU, Fall 07.
9. DeVault: *Blood flow in the Circle of Willis*, Biomathematics Seminar, NCSU, Fall 07.
10. DeVault: *Blood flow in the Circle of Willis*, Department of Mathematics, Michigan Tech. Univ., Spring 08.
11. DeVault: *Blood flow in the Circle of Willis*, Department of Mathematics, University of Pittsburgh, Spring 08.
12. DeVault: *Simulation of stroke of strokes in the Circle of Willis*, Graduate Student Seminar, Department of Mathematics, NCSU, Spring 08.

13. Gremaud: *Numerical simulation of cerebral blood flows*, Seminar of Numerical Analysis, Department of Mathematics, Texas A&M, Spring 08.

14. Gremaud: *Numerical simulation of cerebral blood flows*, presentation the Biomechanical Engineering Research Group at the Department of Mechanical and Aerospace Engineering and Department of Biomedical Engineering at NCSU (to be given later this Spring 08).

15. Gremaud: *Issues and challenges in the numerical simulation of blood flow in the Circle of Willis*, working seminar at the Harvard Medical school, (to be given later this Spring 08).

16. Gremaud: *Numerical simulation of cerebral blood flows*, Seminar of Numerical Analysis, Department of Mathematics, Université Blaise Pascal, Clermont-Ferrand, France, (to be given later this Spring 08).

17. Serge Guillas. Calibration of a 3-D air quality model. U. of Chicago, CISES seminar, February 2007.

18. Serge Guillas. Applied mathematics, University of Georgia, April 2007.

19. Serge Guillas. Statistics, University of Georgia, August 2007.

20. Serge Guillas. University of Bristol, UK, May 2008.

21. C. Kaufman. "Efficient Emulators of Computer Experiments Using Covariance Tapering." Statistics and Actuarial Science Seminar, Simon Fraser Univ, Burnaby, BC; 2007 March 30.

22. C. Kaufman, "Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions", Colorado State University, Fort Collins, CO (August 2007), Colorado School of Mines, Golden, CO, (November 2007), UC Santa Barbara, Santa Barbara, CA (January 2008), Florida State University, Tallahassee, FL (January 2008), Virginia Commonwealth University, Richmond, VA (January 2008), UC Berkeley, Berkeley, CA (January 2008) Virginia Tech, Blacksburg, VA (January 2008), North Carolina State University, Raleigh, NC (January 2008), Duke University, Durham, NC (February 2008), UC Davis, Davis, CA (February 2008).

23. Chunsheng Ma, "Space-time downscaling of regional ozone forecasts with nonseparable covariance models" (May 2007).

24. Ma Chunsheng. Space-time downscaling of regional ozone forecasts with nonseparable covariance models. May 2007.

25. Doug Nychka (NCAR)."The ensemble Kalman filter: The movie". NCSU Seminar and special presentation for the climate modeling group. March 15, 2007.

26. Olufsen: *The cardiovascular system*, Seminar of Applied Mathematics, Department of Mathematics, UNC, Spring 08.

27. E.B. Pitman "A New Approach to Volcanic Hazard Mapping". University at Buffalo, April 2008

28. Greg Rempala. Seminar at the Bioinformatics research and development lab, Pula, Sardina (Italy), July 2007.

29. J. Rougier, "Multivariate Emulation", Statistics Dept, Oxford (October 2007).

30. J. Rougier, "Avalanche prediction", Physics Dept, U. Catholique Louvain (Belgium, Feb 2008), Geography Dept, Bristol (March 2008).

31. T. Santner. "Prediction for Computer Experiments that include Quantitative and Qualitatie Inputs." Statistical Science, Duke U, Durham NC, 2007 January 12.

32. T. Santner. "Designing Computer Experiments to Determine Robust Control Variables." NC State Univ Biostatistics/BSWG Seminar, 2007 January 25.

33. T. Santner Cornell University, July 2007 (series of 3 talks)

34. T. Santner Iowa State University, November 2007

35. E. Spiller. "Models for Volcano Avalanches. Constructing a Risk Map for Pyroclastic Flows: Using simulations and data to predict rare events." Statistics, CMU, Pittsburgh, PA; 2007 October 22.

36. Darren Wilkinson. Biostatistics seminar at the NIEHS in RTP, April, 2007.

37. Darren Wilkinson. Warwick Systems Biology seminar, Warwick (UK), October 2007.

38. Darren Wilkinson. Launch of the Oxford Centre for Integrative Systems Biology, Oxford (UK), January 2008.

39. Darren Wilkinson. Quantitative Biology and Applied Statistics seminar, Reading (UK), February 2008.

40. Darren Wilkinson. Institute for Mathematical Statistics seminar, Kent (UK), February 2008.

41. R.L. Wolpert. "Bayesian Semiparametric Space-Time Models." OR & IE, Cornell Univ, Ithaca, NY; 2006 October 13.

42. Richard Yamada. Mathematical biology seminar, Michigan, September 2007.

43. Richard Yamada. Bioinformatics seminar, Michigan, December 2007.

44. Richard Yamada. Bio-mathematics seminar, NYU, December 2007.

### 7.5 Graduate students.

The Program is contributing in the achievements, education, and Ph.D. projects of many undergraduate students, both locals and non locals:

**Ms. Dianne Bautista** (OSU) visited SAMSI from September 1, 2006 until December 15, 2006 to participate in the Program on Development, Assessment and Utilization of Complex Computer Models. She is a graduate student in the Department of Statistics at The Ohio State University working on her PhD thesis under the direction of Tom Santner on the design and analysis of computer experiments, spatial statistics, and statistical machine learning. Her research is on metamodel-based multiobjective optimization under uncertainty. Its applications are mainly in the areas of robust and reliable engineering design.

She participated in several working groups, mainly in the *Engineering Methodology* and *Methodolgy* working groups. During her visit, Dianne Bautista worked on two projects related to her thesis. The first of these studied a non-parametric method of estimating (a valid) correlation function as part of the process of predicting the output of a computer code. The second project was to develop a method for sequentially designing a computer experiment with multivariate output to find the Pareto Frontier of the set of codes. Both projects are continuing upon her return to Ohio State.

**Miyuki Breen** (NCSU) Estimating biological models. Presented a poster at the opening workshop. Involved in the *System Biology* working group. Collaborates closely with Michael Breen on estimation of computational toxicology models. Still at NCSU.

**Keith Dalbei** (Buffalo U) is completing his dissertation at the University of Buffalo under Abani Patra, both heavily involved in the *Granular Materials: Engineering Applications* working group. Dalbey is experienced in high performance computing, and was heavily involved in performing computations needed in several projects in this working group as well as running TITAN2D emulator for volcanic flows. In addition, Dalbey's thesis work explore a Bayes Linear approach to constructing emulators. He gave a presentation at the working group meetings.

**Jimena Davis** (North Carolina State University, Mathematics), graduate student, participated in the *Dynamics of Infectious Disease* working group. Supported to attend the *Atlantic Coast Conference on Mathematics in the Life and Biological Sciences* (Blacksburg, Virginia, May 3-5, 2007). She gave the oral presentation *A Computational and Statistical Comparison of Approximation Methods for the Estimation of Probability Distribution on Parameters.* Ms. Davis evaluated the working group as follows: "By participating in this working group, I was exposed to various methods and techniques that are applicable to modeling infectious diseases. I think this working group was very valuable in that it provided me with the opportunity to expand my basic knowledge of the different approaches one could take in modeling infectious diseases on different levels.".

**Kristen DeVault.** (NCSU graduate student) Kristen was a key participant in the *Calibration of Computational Models of Cerebral blood flows* working group. She was in charge of processing the data from their medical coworkers (Dr. V. Novak and her team, Harvard Medical School). She also implemented the numerical approach; discretization methods were discussed with her adviser (Gremaud) while data calibration methods were discussed with Gremaud, Olufsen and Vernières. Kristen has passed all the requirements for a PhD and will graduate in May 2008.

**Stacey Ernstberger** (North Carolina State University, Mathematics), graduate student, participated in the *Dynamics of Infectious Disease* working group. Supported to attend the *Atlantic Coast Conference on Mathematics in the Life and Biological Sciences* (Blacksburg, Virginia, May 3-5, 2007). She gave the oral presentation *Mathematical Modeling in Biomass and Vaccine Production Systems for Shrimp.*

**Mr. Gang Han** (OSU) visited SAMSI from September 1, 2006 until December 15, 2006 to participate in the Program on Development, Assessment and Utilization of Complex Computer Models. He is a graduate student in the Department of Statistics at The Ohio State University working on his PhD thesis under the co-direction of Tom Santner and William Notz in the statistical design and analysis of complex computer models. He participated in several working groups, mainly in the *Engineering Methodology* and *Methodology* working groups.

Gang Han completed work on developing methodology for predicting computer experiment output that has inputs which are either, by their nature, nominal-valued *or* should be treated as such. As an example, qualitative inputs in biomechanics applications are the mesh density of a finite-element code, and the type of loading during a biomechanical task (say, normal gait during walking on a level surface versus stair climbing). He has developed software that allows prediction of computer output at an arbitrary vector of qualitative-quantitative inputs based on training data.

Gang also developed methodology (and MATLAB software) for simultaneously setting the values of calibration and tuning-parameter inputs to computer codes. This problem occurs in settings which both computer- and physical experiments have been conducted and, in addition to common engineering design inputs, the computer code has calibration inputs and tuning parameters (inputs). The calibration inputs are inputs present in both the computer code and the physical experiment but whose values are unknown in the computer code. Tuning-parameters are (numerical) inputs present only in the computer code, that are used to move the computer code output nearer to that of the physical experiment.

This work is the core of Gang's Ph.D. Thesis.

**Fei Liu** (Duke). Fei Liu was a graduate student at Duke University. She defended her Ph.D. thesis in May 2007 and is now Assistant Professor at University of Missouri, Columbia. She actively participated in a number of working groups: *Air Quality*, where she gave a presentation: "Calibration for spatial and spatio-temporal model ouputs"; *Engineering Methodology*, where she gave two presentations:"Discussion of a Thermal Model", and "Dynamic Linear Models as emulators"; *Terrestrial models*, and also in the *Methodology* working group, where she was the web-master. She presented a talk at the Transition Workshop in May, and also at the JSM07 SAMSI Topic Contributed Sessions in July 2007.

Fei also was working on a thesis in the computer modeling area, as it applies to functional data. The approach developed for such data consists of representing the function in the wavelet domain; reducing the number of nonzero coefficients by thresholding; modeling the nonzero coefficients as functions of the associated inputs using nonparametric Bayesian methods; and reconstructing the functions (with confidence bands) in the original (time) domain. For computational reasons, an extension of this approach is considered to an eigen-space whose basis elements are linear combinations of the wavelet basis elements. The number of nonzero coefficients is greatly reduced in this eigen-space, as consequently is the computational expense for the statistical inverse problem. Finally, the thesis considers an approach to representing functions as multivariate Dynamic Linear Models. This approach is useful when the functions are highly variable and, as opposed to attempting to represent the functions exactly, one seeks primarily to capture relevant stochastic structure of the functions. The method has been tested with a simulated data set, and will be applied to validate the Community Multi-scale Air Quality model, considered in the Air Quality working group.

**Simon Lunagomez** of Duke University DSS is working on his Ph.D thesis under the guidance of Robert Wolpert (also of Duke DSS), and both involved in the *Methodology* working group. He work in the development of hierarchical Bayesian models for pyroclastic flows, intended to help predict the frequencies of large volcanic eruptions over decade- or century-long periods. The models have underlying Pareto components for large individual flow eruptions and $\alpha$-stable components for the aggregation of many smaller flows, all tailored to the kinds of data emerging from the Montserrat Volcano Observatory (MVO), established in 1995 by the British Geological Survey and the University of the West Indies' Seismic Research Unit to study the ongoing volcanic activity of the Soufrière Hills volcano. The models have been used in the risk assessment of catastrophic pyroclastic flows in collaboration with the *Granular Materials: Engineering Applications* working group. A range of modeling, numerical, and methodological issues arose in this work, giving Lunagomez a remarkable educational opportunity. This work will be part of his Ph.D. Thesis.

He is also developing new methods for detecting dependence of heavy-tailed random variables, intended for the study of Gamma Ray Burst (GRB) data, working in collaboration with Cornell University astronomer and active participant of the *Methodology* working group Tom Loredo.. This work will also be part of his Ph.D. Thesis.

Simon Lunagomez gave a presentation to the Methodology working group on March 12, 2007.

**Heyjung Moon** of the OSU Dept of Statistical Science was involved in the *Engineering Methodology* working group. He is working on a Ph.D. thesis under the co-direction of Tom Santner and Angela M. Dean, studying the use of group screening in computer experiments.

**Jarad Niemi** (Duke) worked on parameter estimation for stochastic biochemical network models. She was involved in the *System Biology* working group. Presented a poster at the opening workshop. Developing discrete-time stochastic models for biochemical network dynamics. Still at Duke.

**Johnny Sammuels** (North Carolina State University, Mathematics), graduate student, participated in the *Dynamics of Infectious Diseases* working group. Supported to attend the *Atlantic Coast Conference on Mathematics in the Life and Biological Sciences* (Blacksburg, Virginia, May 3-5, 2007). She gave the oral presentation *Inverse Problem Methodology in Shear Wave Propagation in Biotissue.* Mr. Sammuels said the following about the working group: "The working group was valuable in attaining an introduction to various modeling approaches. In particular, most of the work done by the participating students and faculty at VT was unknown to me, and hence the exposure was very valuable and instructive."

**Justin Shows.** Graduate student at NCSU, working under Dr. Fuentes' supervision. Mr. Shows has passed the written and oral Ph.D. exams, and was involved with the *Climate and Weather* working group.

Current mesoscale numerical weather prediction (NWP) models use complex, multi-layer, soil and canopy models to specify time-dependent lower boundary conditions for atmospheric solutions. Parameters, with values typically constrained by empiricism or heuristic physical arguments, are ubiquitous in these models. The literature shows that atmospheric solutions can be sensitive to these choices. In practice, many of these parameters can be viewed as tuning knobs, and subjective tuning is acknowledged practice in NWP implementations. Many parameters are not necessarily constant, and may very slowly in time. Complex soil models attempt to account for a wide range of physical processes. Because these models provide lower boundary conditions for atmospheric models, and the metrics for success and utility are usually based in the atmospheric component, an argument can be made that simpler models should be constructed and objectively tuned. Simpler models usually have fewer parameters that control the atmospheric response, and their functional relationship to the atmosphere is usually more accessible. The goal of Shows's project, in collaboration with NCAR and under the supervision of Fuentes (NCSU), was to design and explore optimal methodologies for finding distributions of parameters. Experience shows that ensemble data assimilation is a useful paradigm to approach this problem, where covariance between prior distributions in observation space and parameter distributions are readily available. The 1D model described above also enables efficient research on this topic. Some references for ensemble data assimilation within this modeling framework are Hacker et al. (2007) and Hacker and Rostkier-Edelstein (2007).

Shows presented his research at the May 21-24 SAMSI/NCAR workshop.

**Karyn Sutton** (Arizona State University, Mathematics), graduate student, participated in the *Dynamics of Infectious Disease working group.* Supported to attend the *Atlantic Coast Conference on Mathematics in the Life and Biological Sciences* (Blacksburg, Virginia, May 3-5, 2007). She gave the oral presentation *The Impact of Vaccination on Pneumococcal Disease Dynamics.* Ms. Sutton noted the following concerning the working group: "Well it was good to see the approaches from the algebraists. It's always useful to expand your repertoire of knowledge and to increase your awareness of the current ideas in your field, even if they are not directly related to or can be applied to your own research. I think it is important that researchers of various backgrounds learn about what one another does, even if at a rather superficial level, to facilitate communication and sometimes perhaps collaboration. I don't think that what I've learned from the other groups during the working group will be applied to my research, nor will I probably look any further into the ideas presented. However, I'm glad to have had the exposure to their train of thought, and I think there's always something to be gained from looking at your own research and techniques from a different perspective".

**Daniela Valdez-Jasso** works under the direction of Olfusen (*Calibration of Computational Models of Cerebral blood flows* working group) and was involved in the study of proper visco-elastic models for the arterial walls.

**Richard Yamada** (Cornell) worked on stochastic kinetic modeling of molecular motor dynamics. He was involved in the *System Biology* working group. Interested in developing stochastic kinetic models for molecular motor dynamics, and estimating the parameters from experimental data. Now successfully completed his thesis, "Quantitative models of transcription elongation", and moved to an Assistant Professor position at Michigan. Continuing to collaborate with Darren Wilkinson in this area.