SAMSI Final Report:
Large-scale Computer Models for Environmental Systems

April 12, 2004

1 Introduction

The SAMSI program on "Large-scale Computer Models for Environmental Systems" took place during January–June 2003. In Section 2, we summarize the main activities of the program. The remainder of the report is divided into two sections corresponding to the main breakdown of activity during the program. Section 3 is concerned with applications in the atmospheric and ocean sciences, while Section 4 is concerned with applications in porous media science.

2 Outline of the Program

The main activities of the program were as follows:

1. Three workshops, on "Multiscale Modeling" (at SAMSI, February 2–7), "Optimization" (at SAMSI, April 28–30) and "Spatial-Temporal Statistics" (at the National Center for Atmospheric Research, jointly organized by SAMSI and the Geophysical Statistics Project at NCAR, June 1–6);

2. Two one-day workshops, on "Physical-Statistical Modeling and Data Assimilation" on March 25, and on "Closure Relations for Evolving Models of Multiphase Flow" on May 16;

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

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

5. A series of working groups that met at regular intervals throughout the program.

The working groups were divided into two subgroups, reflecting the interests of the participants in two main areas of application. The remainder of this report covers the main achievements of these working groups.

3 Applications in the Atmospheric and Ocean Sciences

3.1 Spatial-Temporal Statistics

A major theme of the program was the development of statistical methods for spatial-temporal data arising in atmospheric and ocean sciences. This was one of the subthemes of the February workshop on Multiscale Modeling and of the one-day workshop in March on “Physical-Statistical Modeling and Data Assimilation”. It was also the principal theme of the June workshop at NCAR. Visitors who contributed strongly to this theme included James V. Zidek of the University of British Columbia (who visited SAMSI from January 1 to May 30), Mark Berliner of Ohio State and Chris Wikle of the University of Missouri (who gave talks at all three of the workshops just mentioned), Doug Nychka of NCAR (invited speaker at two workshops, co-organizer of the Boulder workshop and also a SAMSI seminar speaker) and Michael Stein of the University of Chicago (SAMSI seminar speaker and a speaker at the Boulder workshop). After the February workshop, a regular weekly meeting was organized chaired by Dr. Zidek; this spawned various other activities including one on network design (reported separately here in Section 3.2).

One outcome of these working group meetings was a collaboration involving Dr. Zidek and his co-workers at British Columbia, and Dr. Prasad Kasibhatla (Duke School of the Environment) on the relationship between, and means of combining, physical and statistical models for large-scale environmental processes. In particular, Dr. Kasibahtla supplied large simulated and real data sets for hourly ozone concentrations over much of the United States. The overall goal of the work was a better understanding of the relationship between the two sorts of data and the advantages if any that might accrue from their integration. The outcome would be intended to complement the substantial body of work on this general topic over the past few years.

To date the following have been achieved: (1) the large data set has been copied to a UBC site. (2) a research group involving Dr. Nhu Le, Dr. Zidek plus three students has been set up with NSERC funding and has been holding regular meetings for some time now. (3) two projects are well underway.

The first consists of a study of the physical models that generated the simulated data with a view to redesigning them as or incorporating them directly into a statistical model. A lot of analysis of both simulated and real data has been done. The second, nearing completion, uses a “forward-filtering backward sampling” Bayesian hierarchical approach to model the real data. It is hoped that this model can be combined with that developed over the past decade by Drs. Le, Zidek and their co-investigators.

Another activity arising from the presentations by Mark Berliner and Chris Wikle at the February workshop was a weekly working group on the subject of “physical-statistical modeling” (spatial-temporal modeling in which the form of the temporal dynamics is influenced by physics) and the connections between this work and the applied mathematics activity in data assimilation. To pursue these topics, a one-day workshop was organized at SAMSI on March 25, with Berliner, Wikle and Chris Jones (UNC, Applied Mathematics) as the three presenters. Approximately 30 people attended this. One of the outcomes of this meeting has been work by another long-term visitor, Dr. Bent Natvig of the University of Oslo and his graduate student Ingunn Fride Tvete, on the analysis of earthquake data. Ms. Tvete has given a talk about this at the 13th European Young Statisticians Meeting, Ovrannaz, Switzerland, September 2003, and a poster at the Spruce Advanced Workshop on Spatial/Temporal Models and Methods, Lisbon, Portugal, March 2004, as part of an “Evaluation of Bayesian Hierarchical Models” project at the Department of Mathematics, University of Oslo. An abstract of a forthcoming paper is presented in Section 3.9.

A third area of research was in the application of spatial-temporal models to data on fine particulate matter ($PM_{2.5}$). This involved long-term visitor Dr. Sujit K. Sahu (School of Mathematics, University of Southampton, UK) as a visiting research scholar at SAMSI and at Duke University during March–July, 2003. During this time Dr. Sahu developed models and methods for analyzing spatio-temporal data obtained from environmental monitoring networks. In particular his research work focused on modelling $PM_{2.5}$ data.

Dr. Sahu's collaborative work [17] with Professor Kanti Mardia (University of Leeds) presents a short-term forecasting analysis of $PM_{2.5}$ data in New York City during 2002. Within a Bayesian hierarchical structure, they model the spatial structure with principal kriging functions and the time component is modeled by a vector random-walk process. This work was presented by Dr. Sahu in the Joint Statistical Meetings, San Francisco, August, 2003.

In close collaboration with Professor Alan Gelfand (Duke University) and Dr. David Holland (EPA), Dr. Sahu developed a new spatio-temporal model for analyzing spatio-temporal $PM_{2.5}$ data observed over large spatial domains containing urban and rural areas. They introduce two random effects components, one for rural or background levels and the other as a supplement for urban areas. These are specified in the form of spatio-temporal processes. Weighting these processes through population density results in nonstationarity in space. They analyze a dataset on observed $PM_{2.5}$ in three states in the U.S. — Illinois, Indiana and Ohio. A fully Bayesian model is implemented using MCMC techniques which enable them to obtain full inference in terms of predictions in time and space. This paper [16] has been submitted for publication and can be downloaded from Dr. Sahu's website.

Dr. Sahu presented the above paper as an invited paper in the Advanced Workshop on Spatial/Temporal Models and Methods organized by SPRUCE in Lisbon, Portugal during March 2004.

After successful completion of the above project, Dr. Sahu is currently looking to model other suitable spatio-temporal data sets. As mentioned above, this SAMSI program has opened up an excellent opportunity for collaboration between researchers in SAMSI and Dr. Sahu, and he intends to return to SAMSI for short visits to foster the existing links further.

3.2 Network Design

This activity began as a subgroup of the working group on spatial-temporal statistics (see Section 3.1) but soon organized its own meetings under the chairmanship of Dr. David Holland of the Environmental Protection Agency. Those participating included Dr. James Zidek as well as several local faculty and EPA researchers.

A number of meetings of this group took place at both EPA and SAMSI. There was also an invited session at the ENAR meeting in Tampa, FL, during March 2003, which included Zidek and Holland as invited speakers and Smith as discussant. Dr. Zidek also gave an invited talk at the SAMSI Optimization workshop in April.

Subsequent activities included a review paper by Zidek and co-authors [12], which has been submitted for publication. A second project is nearly finished concerning the measurement of random fields of spatial extreme values. Work on that project wrapped up in February 2004 and a manuscript will soon be submitted for publication as a SAMSI Tech Report and elsewhere. This manuscript will describe the many challenges facing designers attempting to develop monitoring networks, that do not present themselves when interest focuses on conventional environmental fields. In particular, spatial correlation declines between spatial sites as the degree of the extremes increases. Finally, a framework for approximating these fields is developed, one which does not rely on intractable multivariate extreme value models. When implemented, that framework will allow an assessment of current urban monitoring strategies and may well reveal their deficiencies, at least where the measurement of extremes is concerned.

Another focus of discussion was the approach developed by Dr. Zhengyuan Zhu (University of North Carolina) in his thesis at the University of Chicago, and continued in subsequent work with Dr. Richard Smith and other participants during and following the SAMSI program. A common problem in spatial statistics is to observe a random process Z at a set of sample locations $S = \{s_1, \ldots, s_n\} \subset D$, and then make inference about the unobserved $Z(x)$ for $x \in D$, where D is the region of interest. Specific prediction objectives vary according to the particular applications. For example, in pollutant monitoring applications, one is most likely interested in predicting the region where $Z(x)$ is greater than some critical value, while in mining applications it is the average of $Z(x)$ over D that is of economic interest. For fixed sample size n, it is natural to consider choosing the sample location network $S \subset D$ so that we can have the most accurate prediction (point prediction and/or prediction interval) in D. One critical issue is to decide how to quantify this notion.

The research focusses on spatial sampling design for a stationary isotropic Gaussian model. In many applications, we have to use the same data for both estimation and prediction. The problem of finding a design that performs both these functions simultaneously has not been adequately addressed in previous research on optimal design. Both Bayesian and frequentist approaches are used. To reduce some of the computational challenges, a two-step optimization method was developed, in which a part of the design is set aside for estimation and the remainder used to optimize predictive criteria conditional on the estimative part of the design.

In work during and following the SAMSI program, Zhu and Smith have outlined an alternative approach to combining estimative and predictive criteria, using a Bayesian prediction interval to take the uncertainty of estimating parameters into account, but also using analytic approximations (deriving from the Laplace approach to approximating integrals) to avoid the computational complexity of a MCMC approach to Bayesian inference. They have also explored ways of extending this approach to designs for generalized linear models and for designs to estimate extreme values.

Zhu and Smith submitted a joint proposal for this research to NSF in the Fall of 2003. They are continuing to collaborate at UNC where Smith is also serving as advisor to a PhD student on a closely related topic. Zhu and Smith will each give invited talks at the International Environmetric Society's annual conference in Portland, Maine (June 2004), and Smith will give an invited talk at a session on optimal design in the Joint Statistical Meetings in Toronto (August, 2004).

3.3 Source Apportionment

One of the growing themes of environmental research in recent years has been the identification of sources and (where appropriate) sinks of atmospheric environmental contaminants. Particular attention has been paid to the global carbon budget where the transport of both carbon monoxide (CO) and carbon dioxide (CO_2) are of interest.

In a typical problem, a vector of sources or sinks (x) is related to a vector of observations (y) by a relationship of the form

$$y = Kx + \epsilon, \tag{1}$$

where K is an approximate matrix derived from an atmospheric chemistry model and ϵ is a vector of errors, typically assumed to have a normal distribution $N[0, S_\epsilon]$ subject to some suitable specification of the covariance matrix S_ϵ. At its simplest level, the problem is about inversion of (1) to make inferences about x from the vector of observations y. However there are many uncertainties in the problem, one of the main ones being that (1) is typically ill-conditioned, large changes in x potentially resulting in only small changes in y. To get around this question, it is common to combine (1) with a “prior distribution” of form

$$x \sim N[x_a, S_a] \tag{2}$$

where the mean vector x_a and the covariance matrix S_a represent prior information about x (such as information derived from inventory reports).

One analysis of equations (1) and (2) is due to [9], whose lead author, Dr. Prasad Kasibhatla, is in the Nicholas School of the Environment at Duke. They showed that if x_a, S_a, S_e and K are all treated as known, there is a simple exact solution to the posterior distribution of x given y. The ramifactions of this were explored for the modelling of global CO patterns. Other papers have explored similar ideas for source-sink modeling of CO_2.

This topic formed the focus of a number of discussions at SAMSI. Dr. Tapio Schneider of Cal Tech gave a talk at the February workshop, focussing on the CO_2 modeling problem and associated controversies. Another group working on similar problems (with particular focus on the attribution of ammonium sources) is at the EPA, headed by Dr. Robin Dennis and Dr. Alice Gilliland; Dennis gave an invited talk on this work at the Boulder workshop. In work continuing at SAMSI, Kasibhatla, Zidek, Smith and a number of others met to discuss statistical ramifactions of this method. Particular ideas that were explored include:

- Introducing spatial covariances into (1) and (2); in particular, allowing for spatial correlation in the error vector ϵ. B. Lopes (student of R. Smith) gave a poster about this at the ISI International Conference on Environmental Statistics and Health (Santiago de Compostela, Spain, July 2003) and is continuing to work on this for his PhD thesis.
- J. Zidek wrote a preliminary paper on a hierarchical Bayesian approach that more fully reflects model uncertainties than do conventional physical models developed for that purpose. The eventual goal is a predictive model for the unmeasured inputs (emissions) with predictive intervals truly reflecting that uncertainty, given the measured outputs.
- Current discussions are related to extending this methodology to a more comprehensive hierarchical Bayes model, and extending it to broader data bases (in particular, those based on satellite data, which are potentially much more extensive and informative than ground-based monitor data).

Future plans are for Kasibhatla and Smith to submit a joint grant proposal later this year. In addition, Dr. M. Fuentes (NCSU Statistics) received funding to be a visitor at Duke’s Center for Global Change where she has worked with Kasibhatla and other researchers.

3.4 Intermittency and sub-grid-scale processes

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

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

Anne Bourlioux (Samsi visitor), Roberto Camassa, and Richard McLaughlin focussed upon a passive scalar diffusing in the presence of a mean gradient, and a (random) shear layer with transverse unsteady wind field in a specific regime known as the quasi-steady regime for which the scalar is shown to develop a long tail. This work is in preparation for publication.

A working sub-group (Alberto Scotti, Roberto Camassa, Chris Jones, and Richard McLaughlin) was awarded a 3 year NSF CMG (Collaborations between Mathematics and Geocsciences) grant in Sept of 2003 (McLaughlin, PI) to explore the physical origins of Non-Gaussian statistics in geophysical sciences ($480,000), and to explore non-Gaussian data assimilation (a joint effort with a UCLA group of atmospheric scientists). Interactions between local statistician Steve Marron will also be continued with goal to develop more stringent statistical tests to assess non-Gaussian time series.

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

Along the lines of sub-grid scale parametrizations, Camassa and McLaughlin have continued work studying the effective diffusivity for a passive scalar diffusing in the presence of a time varying shear profile using non-standard stationary phase arguments. This work (with James Bonn, Roberto Camassa, Ken McLaughlin, and Rich McLaughlin) is in final preparation. Additionally, McLaughlin performed experiments at the UNC Fluid dynamics laboratory within the mathematics department and identified a new phenomenon involving the motion of a falling sphere in a strongly stratified fluid. The physics of this new phenomenon involves a sub-grid scale process which was modelled using an idealized, effective fluid particle. This work has just appeared [1].

3.5 Combining networks

A working group on this subject was co-chaired by Dr. Peter Finkelstein of EPA and Dr. Montserrat Fuentes of NCSU. The goal was to investigate and evaluate methods that could be used to combine information from individual air pollution monitoring networks to gain a better understanding of the spatial distribution of pollutants over large geographic areas.

Individual air monitoring networks have different design philosophies and operational protocols, based on their design criteria. Combining information from more than one network requires understanding and overcoming the differences which bring about biases between the networks. Our goal is to develop an optimal model that accounts for the spatial dependence of combined data and measurement errors of each network, thus enabling the analysis of multivariate measurement data.

The group has chosen for a study example, the problem of combining aerosol nitrate (ANO_3) data from two EPA networks, Castnet and the Speciated Trends Network (STN). Castnet was designed to measure the total input of pollutants to the non-urban areas of the country, with an emphasis on the effects of pollutants on natural and managed ecosystems. Thus the sites are in rural areas, and the monitoring protocol is designed for the complete capture of all Nitrogen species. STN was designed to measure the amount of fine aerosols to which a majority of the population may be exposed. Thus measurements are in cities, and the samplers are designed to measure only fine ANO_3 and not other N compounds. Other network operation protocols also differ considerably. We expect the bias between these networks to be a complex spatio-temporal function which involves the concentrations of other atmospheric gas and particle spe

The discussions and frequent meetings of the SAMSI working group has motivated the dissertation topic of Darryl Cooney, a Ph.D. student at the Statistics Department at NCSU. Mr. Cooney's research focuses on the development of new methodology to combine spatial disparate data, and he is currently working on his dissertation under the supervision of Dr. Fuentes of NCSU; and Drs. Swall and Finkelstein of EPA. Fuentes and Finkelstein initiated this productive collaboration under the SAMSI umbrella.

D. Cooney and K. Foley, another member of this SAMSI working group gave a joint talk at NCSU on September 18, 2003, about the research done by this SAMSI group.

3.6 Human exposure and health effects

A vast amount of research on the adverse impact of air pollution on human health (with growing societal concern) has generated a critical need for regulatory and control criteria. In particular, EPA air pollution regulations are seen as vital in protecting human health. However, the complexity of the processes that determine human exposure and its associated health impacts, makes the formulation of effective regulations difficult, leading to a need for computer models that simulate those processes. Such models have been developed in response to that need and with them an associated literature on personal exposure. However, that work has not been early as extensive as the physical/chemical models that have been covered in other parts of this program.

Projects started or continued at SAMSI, and subsequent work, are summarized in the following.

- *Health Effect Models.* By their daily, random time activity patterns, humans determine their exposures to random environmental hazard's space-time field. This poses a dilemma for regulators who try to control or even reduce ambient levels of that hazard. After all, it may well be that people would spend more time out-of-doors if regulators were to reduce the level of that hazard and thereby sustain even higher levels of exposure, not lower as might naïvely be expected. To predict those exposures better by accounting for random time-activity patterns of population and subpopulation members, a large internet based exposure simulation model, pCNEM, has been put online at the University of British Columbia. While visiting SAMSI, J. Zidek completed a study of the performance of that simulator [20]. A manuscript based on part of that study has been provisionally accepted by Atmospheric Environment. A second has been submitted as a special invited paper for Environmetrics.

- The performance of pCNEM and other models like it have been little assessed. Zidek's SAMSI visit generated a collaborative research project with Sandra McBride and co-investigators at the EPA's National Health and Environmental Effects Research Laboratory and the National Exposure Research Laboratory. Memoranda of understanding were developed to allow analysis of USEPA exposure-epidemiology particulate matter panel studies.

  Two papers are in progress as a result of the SAMSI subgroup's efforts, under the titles "A Bayesian Hierarchical Model for Health Effects due to Personal Exposure to Particulate Matter" and "Assessing a Computer Model for Predicting Human Exposure to $PM_{2.5}$". Abstracts follow below in Section 3.9. Both papers are expected to be completed by July, 2004. Preliminary results of the model evaluation study of pCNEM were presented at the International Society for Exposure Analysis 13th Annual Conference, September 21-25, 2003, Stresa, Italy. Preliminary results of the Bayesian hierarchical modeling will be presented at an invited talk at the 11th Annual Spring Research Conference on Statistics in Industry and Technology, May 19-21, 2004, hosted by the National Institute of Standards and Technology (NIST).

- A central issue in modeling the relationships among health effects, human exposure and ambient air pollution is the integration of data at different levels of spatial and temporal variation. Following the approach of Best, Ickstadt and Wolpert (JASA, 2001), the pollutant space-time field may be modeled by kernel smoothing a Lévy process, so making it possible to construct a rich class of non-Gaussian fields. This approach is being developed on EPA data by Robert Wolpert (Duke); preliminary results were presented at the June Boulder workshop.

3.7 Extreme Values

In recent years, much of the focus of research on global climate change has moved beyond the rise in average temperature to focus on more complex measures of climate change. There has been much attention to the question of whether climate extremes are becoming more frequent, as many climatologists claim. In particular, there has been much attention to the question of rainfall extremes.

A critical question is to what extent global and regional climate models correctly capture the distribution of rainfall extremes. Such models typically predict an increase in rainfall extremes over the next century, but the extent to which they successfully reproduce current data is uncertain. If one could come up with quantitative measures of this uncertainty, one might be able to create meaningful probabilistic predictions of future climate extremes.

One of the major issues in this kind of analysis is the mismatch of spatial scales between climate models and data: climate models recreate meteorological variables averaged over grid cells, typically of 3–5 degrees in extent, whereas most observational data is obtained from point data. An intermediate source of data is from re-analyses such as those produced by the National Center for Environmental Prediction (NCEP), which take observational data and run them through a numerical weather forecasting model to recreate grid-cell averages from the real data. However, there is still a potential mismatch between what a reanalysis model is doing and a corresponding climate model. Also, extreme rainfall events may not be realistic from all reanalysis products. For example, there is a substantial difference between the wettest 5-day event in the NCEP reanalysis, the European Center reanalysis and a blended satellite-station pentad rainfall dataset [19] as studied in [8].

A working group at SAMSI consisting of R. Smith of UNC, Dr. Amy Grady of NISS and climatologist Dr. Gebriele Hegerl of the Duke School of the Environment, met to discuss ideas on the spatial modeling of extremes and the related problems of validating climate models with observational data. Two papers are planned on this work and will likely be completed during the summer of 2004. This group also established a link with extreme value researchers working at NCAR, in particular Dr. R. Katz (NCAR) and Dr. P. Naveau (University of Colorado), who are developing an extremes toolkit for the climate community. R. Smith is also working with the Climate Assessment Initiative at NCAR, where there is a major thrust on the ability of climate models to model extremes. Initial collaborations with that group led to some novel Bayesian methods for combination of data from different climate models [18] and work is continuing on extending that approach to the modeling of extremes.

Separately from this line of work, J. Zidek and his co-investigators worked on extreme values in the simulated precipitation fields generated by the Canadian Global Climate Model. A paper on that work is just being prepared for submission for publication. However, considerable impact from the work could derive from the discovery that the joint distribution of log-transformed annual daily precipitation maxima over the 319 grid cells covering Canada generated by the model is well modeled by a multivariate t distribution. For example, the distribution of the number of these cells that achieve their 100 year levels or more in a single year can be simulated quite simply.

3.8 Data Assimilation

Dr. Jie Yu (SAMSI postdoc) worked extensively with Dr. Chris Jones (UNC, Applied Mathematics) on data assimilation. This is to be the subject of a separate SAMSI program in 2005, with Jones and Dr. Doug Nychka as program leaders. Dr. Jones also gave a talk at the one-day March workshop on physical-statistical methods.

Drs. Jones and Kayo Ide (SAMSI visitor) were awarded a joint ONR grant on Lagrangian data assimilation. Following from Dr. Yu's work as a SAMSI postdoc, there is the possibilty of UNC obtaining a DARPA/ONR award on predicting wave fields.

3.9 Abstracts of Papers in Preparation

1. *Comparison of Bayesian Hierarchical Models of Spatial and Temporal Dependencies between Earthquakes*, by Ingunn Fride Tvete and Bent Natvig, Institute of Mathematics, The University of Oslo, 0316 Oslo, Norway.

   We consider parts of an earthquake catalog provided by The Northern California Earthquake Data Center limited to the area 32 - 37 degrees north, 115 - 120 degrees west and to the time period January 1981 until August 1992.

   Different alternative Bayesian hierarchical space-time models in the spirit of Wikle, Berliner and Cressie (1998, Environmental and Ecological Statistics) are compared. In these models the earthquakes are represented as potentials on a grid. The potentials at each time and grid point are decomposed into a time independent Markov Random Field term and a time dependent term also taking earthquakes at the previous time point into account. The models are implemented within a MCMC framework in Matlab.

   Keywords: Markov random fields; Markov chain Monte Carlo methods.

2. A Bayesian Hierarchical Model for Health Effects due to Personal Exposure to Particulate Matter, by Sandra McBride, ISDS, Duke University and Ron Williams, NERL, US Environmental Protection Agency.

   In the USEPA's 1998 Baltimore Epidemiology-Exposure Panel Study, a group of 21 residents of a single building retirement community wore personal monitors recording personal fine particulate air pollution concentrations ($PM_{2.5}$) for 27 days, while other monitors recorded concurrent apartment, central indoor, outdoor and ambient site $PM_{2.5}$ concentrations. Daily health measurements including heart rate variability and blood pressure were also monitored for study participants. Using the panel study data, we first develop a Bayesian hierarchical model to characterize the relationship between personal exposure and concentrations of $PM_{2.5}$ indoors and outdoors. Personal exposure is expressed as a linear combination of time spent in microenvironments and associated microenvironmental concentrations. The model accounts for missing data and sources of uncertainty such as measurement error and individual differences in exposure. Second, we investigate the relationship between personal exposures and associated health effects. We discuss the implications of using personal versus ambient $PM_{2.5}$ measurements in characterization of the posterior distribution of health effects due to $PM_{2.5}$.

3. *Assessing a Computer Model for Predicting Human Exposure to $PM_{2.5}$*, by Sandra McBride, Duke University, James V Zidek, University of British Columbia and Ron W. Williams, U.S. Environmental Protection Agency.

   We present the results of an evaluation of a probabilistic exposure model for predicting personal exposure to $PM_{2.5}$. The exposure model is built on a WWW platform called pCNEM, "A PC Version of pNEM." The pCNEM platform enables users to develop exposure models online for any environmental hazard for which requisite pollutant, meteorological and source information is available. Our study population is defined as participants in the USEPA 1998 Baltimore Particulate Matter Epidemiology-Exposure Study, which examined personal exposures to $PM_{2.5}$ of a group of senior citizens living in a retirement facility during the summer of 1998. First, we examine the sensitivity of predicted personal exposures to parameter values used to calculate indoor residential concentrations under the mass balance equation. We use a fractional factorial design to examine the impact of parameters on predicted personal exposure means and variability. We next evaluate model performance by comparing modeled exposures to measured personal exposures and develop statistical tests to quantify the discrepancies. These tests focus on the ability of the model to capture variability between individual exposures and across the study participants. This analysis also demonstrates the utility of the freely available pCNEM platform for application in other contexts by the exposure research community.

4 Applications in Porous Media Science

4.1 Model formulation and closure

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

- to educate researchers and visitors in advanced methods for model formulation to describe flow and transport in porous medium systems;
- to catalyze an interdisciplinary group working to develop closure relations for evolving models; and
- to evolve improved pore-scale modeling and analysis methods to aid the development of closure relations.

Accomplishment on these research objectives includes the following:

1. An advanced-level course in multiphase porous medium model formulation was taught at SAMSI, by Gray and Miller, during the spring semester of 2003.
2. Two interdisciplinary NSF proposals were developed and funded as an outgrowth of this aspect of the work:
   - CMG: Multiphase Porous Medium Dynamics: Pore to Field Scale. Sponsored by: National Science Foundation (DMS-0327896, 8/03–7/07). Amount: $662,000.

- Collaborative Research: Interfacial Dynamics in Multi-phase Flow and Transport Processes. Sponsored by: National Science Foundation (EAR 0337535, 6/4–5/7), Amount: $175,000 (among UNC, Oregon State University, and U.S. Salinity Laboratory).

3. Work was completed and published on lattice-Boltzmann modeling of closure relations for porous medium models [14, 15].

4. A manuscript was written to show conditions under which Darcy's law is flawed and to present a consistent form [6].

5. A series of manuscripts was initiated to present a theoretical framework for thermodynamically constrained averaging theory based model development and closure. The first manuscript in this series has been submitted [7], two others are in development, and some additional work that is expected to contribute additional manuscripts in this series has been completed. The current plan is to finish this set of manuscripts detailing the theoretical framework and then to develop a monograph based upon this work.

6. A manuscript was written [2], and will be submitted shortly, on measuring interfacial areas in porous medium systems, which is an important component of the theory developed.

7. Additional work on interfacial area computation, improved lattice-Boltzmann methods, and viscous coupling was catalyzed by work started at SAMSI and additional manuscripts in each of these areas will be forthcoming.

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

4.2 Model solution approaches

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

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

Accomplishment on these research objectives includes the following:

1. A proposal was funded to help support the optimization and simulation workshop and a second proposal on nonlinear solvers is pending review:

- Workshop on Simulation and Optimization of Porous Media. Sponsored by: U.S. Army Research Office (DAAD19-03-1-0115, 6/1/03 – 5/31/04). Amount: $9,986.
- Iterative Methods for Nonlinear Equations. Submitted to: National Science Foundation (DMS, 6/4–5/9). Amount: $420,992.

2. A paper was published comparing methods to solve the steady-state form of Richards' equation [3].

3. Work has been completed and drafted on a spatially and temporally adaptive method to solve Richards' equation and on a comparison of algorithm and computational advances over the last decade for solving Richards' equation. Two manuscripts from these aspects of our work are in draft form and will be submitted in the near future.

4. We have done significant work exploring the joint use of higher order discontinuous Galerkin methods in space and a variety of time integration methods for a range of multiphase model formulations. We hope that this work will mature over the coming months to yield publishable results, but this result is not yet certain.

5. We have completed a skeleton of a problem solving environment to facilitate the rapid building of environmental models. A manuscript on this work is nearing completion and will be submitted in the near future.

6. SAMSI support was also instrumental in catalyzing work on simulation of porous medium models that contain non-differentiable components [4, 11].

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

4.3 Optimal design

This working group investigted the evolution of improved optimization methods for the design of water supply and contaminant restoration systems. Previous work by [13] has produced a suite of "benchmark problems" for optimal design in hydrology, with the intention of enabling both the optimization and simulation communities to improve design and evaluations methods for the very difficult problems in this field. These community problems have been used as a focal point for objectives of this work:

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

Accomplishment on these research objectives includes the following:

1. In [10] we report on the implementation, solution, and distribution of two of the community problems.

2. Additional work on the hydraulic capture problems from taken from the Community Problems has been completed and a manuscript submitted [5].

3. We have also catalyzed a group of mathematicians working in optimization to solve a set of water supply problems taken from the Community Problems to compare a variety of methods. This work is in draft form, and we expect to submit a manuscript on this for review in the near future.

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

References

[1] N. Abaid, D. Adalsteinsson, A. Agyapong, and R. M. McLaughlin. An internal splash: Levitation of falling spheres in stratified fluids. *Physics of Fluids*, 16(5):1567–1580, 2004.

[2] K. A. Culligan, D. Wildenschild, B. S. B. Christensen, W. G. Gray, and A. F. B. Tompson. Interfacial area measurements for unsaturated flow through a porous medium. *to be submitted*, 2004.

[3] M. W. Farthing, C. E. Kees, T. S. Coffey, C. T. Kelley, and C. T. Miller. Efficient steady-state solution techniques for variably saturated groundwater flow. *Advances in Water Resources*, 28 (8):833–849, 2003.

[4] K. R. Fowler and C. T. Kelley. Pseudo-transient continuation for nonsmooth nonlinear equations. Technical Report CRSC-TR03-29, Center for Research in Scientific Computation, July, 2003.

[5] K. R. Fowler, C. T. Kelley, C. E. Kees, and C. T. Miller. A hydraulic capture application for optimal remeidation design. Technical Report CRSC-TR04-04, Center for Research in Scientific Computation, February, 2004. To appear in the proceedings of Computational Methods in Water Resources XIV.

[6] W. G. Gray and C. T. Miller. An examination of darcy's law for flow in variable porosity porous media. *in review: Environmental Science & Technology*, 2004.

[7] W. G. Gray and C. T. Miller. Thermodynamically constrained averaging theory approach to modeling of flow in porous media: 1. motivation and overview. *in review: Advances in Water Resources*, 2004.

[8] G.C. Hegerl, F. Zwiers, S. Kharin, and P. Stott. Detectability of anthropogenic changes in annual temperature and precipitation extremes. *Journal of Climate, provisionally accepted*, 2004.

[9] P. Kasibhatla, A. Arellano, J. A. Logan, P. I. Palmer, and P. Novelli. Top-down estimate of a large source of atmospheric carbon monoxide associated with fuel combustion in asia. *Geophys. Res. Lett.*, 29(19):1900, doi:10.1029/2002GL015581, 2002.

[10] K. R. Kavanagh, C. T. Kelley, C. T. Miller, C. E. Kees, R. W. Darwin, J. P. Reese, M. W. Farthing, and M. S. C. Reed. Solution of a well-field design problem with implicit filtering. *in press: Optimization and Engineering*, 2004.

[11] C. T. Kelley, K. R. Fowler, and C. E. Kees. Simulation of nondifferentiable models for groundwater flow and transport. Technical Report CRSC-TR03-47, Center for Research in Scientific Computation, December, 2003. To appear in the proceedings of Computational Methods in Water Resources XIV.

[12] N. D. Le, L. Sun, and J. V. Zidek. Designing networks for monitoring multivariate environmental fields using data with monotone pattern. Technical Report 2003-5, SAMSI, May 23, 2003.

[13] A. S. Mayer, C. T. Kelley, and C. T. Miller. Optimal design for problems involving flow and transport in saturated subsurface systems. *Advances in Water Resources*, 25(8–12):1233–1256, 2002.

[14] C. Pan, M. Hilpert, and C. T. Miller. Lattice-boltzmann simulation of two-phase flow in porous media. *Water Resources Research*, 40(1), 2004.

[15] C. Pan, J. F. Prins, and C. T. Miller. A high-performance lattice boltzmann implementation to model flow in porous media. *Computer Physics Communication*, 158(2):89–105, 2004.

[16] S.K. Sahu, A.E. Gelfand, and D.M. Holland. Spatio-temporal modeling of fine particulate matter. Technical report, University of Southampton, 2004. Available from http://www.maths.soton.ac.uk/staff/Sahu/.

[17] S.K. Sahu and K.V. Mardia. A bayesian kriged-kalman model for short-term forecasting of air pollution levels.

[18] C. Tebaldi, R.L. Smith, D. Nychka, and L.O. Mearns. Quantifying uncertainty in projections of regional climate change: A bayesian approach to the analysis of multimodel ensembles. Technical report, National Center for Atmospheric Research, January 2004.

[19] P. Xie, J.E. Janowiak, P.A. Arkin, R. Adler, A. Gruber, R. Ferraro, G.J. Huffman, and S. Curtis. Gpcp pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *Journal of Climate*, 16:2197–2214, 2003.

[20] J.V. Zidek, J. Meloche, G. Shaddick, C. Chatfield, and R. White. A computational model for estimating personal exposure to air pollutants with application to london's pm10 in 1997. Technical Report 2003-3, SAMSI, March 27, 2003.