Final report: SAMSI Program on Uncertainty Quantification (2011-12)

Background

Uncertainty Quantification (UQ) and its recent developments are closely intertwined with several applications, most notably climate and engineering. To properly explore not only methodological issues but also applications to specific fields, SAMSI devoted its entire resources in 2011-2012 to the UQ program. The structure of the 2011-12 UQ program is indeed unique among SAMSI programs in that, instead of running two year-long programs in parallel, we ran one UQ program with not only (i) a main methodological theme but also, stemming from it, three distinct application fields, namely, (ii) climate modeling, (iii) engineering and renewable energy and (iv) geosciences. This structure was aimed at fostering the seamless integration into the program of these fields and their corresponding research communities. For the sake of simplicity, we hereafter refer to each of these four themes as "programs", each with its own thematic areas.

Overall Program Leaders: Amy Braverman (JPL, Cal. Tech and UCLA), Don Estep (Colorado State), Roger Ghanem (USC), David Higdon (LANL), Christine Shoemaker (Cornell), Jeff Wu (Georgia Tech)

Methodology Program Leaders: Dan Cooley (Colorado State), Don Estep (Colorado State), Max Gunzburger (Florida State), Jan Hannig (UNC), David Higdon (LANL), Jeff Wu (Georgia Tech)

Climate Program Leaders: Amy Braverman (JPL, California Institute of Technology, and UCLA), Xabier Garaizar (LLNL), Dave Higdon (LANL), Gardar Johanneson (LLNL), Donald Lucas (LLNL)

Engineering Program Leaders: Don Estep (Colorado State), Roger Ghanem (University of Southern California), Miriam Heller (MHITech Systems)

Geosciences Program Leaders: Omar Ghattas (Univ. of Texas-Austin), Christine Shoemaker (Cornell University), Daniel Tartakovsky (University of California - San Diego)

Local Scientific Coordinators: James Nolen (Duke), Jan Hannig (UNC)

Directorate Liaison: Pierre Gremaud (NCSU)

National Advisory Committee Liaison: Max Gunzburger (Florida State)

1 Workshops

1.1 Opening workshops

Each of the above program has its own opening workshop. Tutorials were offered on the first day of the Methodology Program Opening Workshop, see below.

1.1.1 Methodology Program Opening workshop

The Opening Workshop for the Methodology part of the SAMSI UQ program was held on Wednesday-Saturday, September 7-10, 2011, at the Radisson RTP in Research Triangle Park, NC. There were 156 participants. Tutorial sessions were held on Wednesday, September 7 on

- Habib Najm (Sandia), *Foundations*,

- Susie Bayarri (University of Valencia), *Representation and propagation of Uncertainty*,
- Adrian Sandu (Virginia Tech), *Data assimilation*,
- Peter Kitanidis (Stanford), *Inverse problems and calibration*,
- Dan Cooley (Colorado State), *Statistics of Extremes.*

Videos of these presentations are available at the SAMSI website. Invited talks were presented Thursday through Saturday. There was a poster session and reception on Thursday, September 8. Working groups were finalized on Saturday afternoon. The workshop focused on five complementary themes at the forefront of current research in UQ. The five focus areas were

Foundations: theory and methods that are related to uncertainty quantification, including treatment of model error, inference with multi-models, treatment of multiscale models, numerical error estimation, reduced order modeling, and theoretical foundations of various UQ problems. This session featured the following presentations

- Serge Prudhomme (Univ. of Texas, Austin) *Validation and Uncertainty Quantification*,
- Jeff Wu (Georgia Tech) *Modeling Mesh Density in Computer Experiments*,
- Clayton Webster (ORNL) *An Adaptive Sparse Grid Generalized Stochastic Collocation Method for High- Dimensional UQ*
- Panel discussion: Panel Chair: Don Estep Panel: Jim Berger Duke, Habib Najm (Sandia)

Representation and propagation of uncertainty: efficient representations of uncertainty, integration of different representations of uncertainty and error, propagation of uncertainty through complex physical models, exploiting mathematical structure of models for propagation, and identification of sensitivity of models. The following talks were given

- Daniel Tartakovksy (UCSD) *PDF Methods for Uncertainty Quantification in Hyperbolic Conservation Laws*
- Antonio Possolo (NIST) *Assessment and Expression of Measurement Uncertainty*
- Dongbin Xiu (Purdue) *Numerical Methods for UQ: Algorithms Beyond Polynomial Chaos*
- Panel Chair: Susie Bayarri (Valencia) Panel: Youssef Marzouk (MIT), Ilya Timofeyev (Houston)

Data assimilation: incorporating information from disparate and sparse data sets, filtering and forecasting, response surface techniques, sampling based approaches, derivative/variational-based approaches. the following talks were offered

- Jeff Anderson (NOAA) *Ensemble Data Assimilation and Uncertainty Quantification*
- Pierre Lermusiaux (MIT) *Non-Gaussian Data Assimilation with Stochastic PDEs: Bye-Bye Monte-Carlo?*
- Juan Restrepo (Univ. of Arizona) *An Introduction to Global and Sequential Estimation*
- Panel Chair: Adrian Sandu (Virginia Tech) Panel: Tony O'Hagan (Univ. of Sheffield), Elaine Spiller (Marquette)

Inverse problem and calibration: treating sparse and noisy data, use of reduced order models, regularization and optimization, integration of parameter and prediction uncertainty, use of derivative information from physics-based models. This session included the following presentations

- Tony O'Hagan (Univ of Sheffield) *Calibration and Model Discrepancy*
- Lulu Kang (Illinois Inst. of Technology) *Analysis of Large-Scale Computer Experiments*
- Karen Wilcox (MIT) *Model Reduction and Surrogate Modeling for Uncertainty Quantification*
- Panel Chair: Jeff Wu (Georgia Tech) Panel: Roger Ghanem (USC), Peter Kitanidis (Stanford)

Rare events: use of reduced order models, use of computational and physics models to detect rare events, investigation of relevant statistical methodologies, optimization techniques. This session featured the following presentations

- Debbie Dupuis (HEC Montreal) *Extremes of Dependent Sequences*
- Jose Blanchet (Columbia) *Monte Carlo Methods for Rare Events in Stochastic Processes*
- Laura Swiler (Sandia) *Reliability Analysis Methods: Determining Failure Probabilities*
- Panel Chair: Richard Smith (SAMSI) Panel: Dan Cooley (Colorado State), Mircea Grigoriu (Cornell)

1.1.2 Climate Program Opening workshop

The Opening Workshop for the Climate theme of the SAMSI program on Uncertainty Quantification was held on Monday-Wednesday, August 29-31, 2011, at the Marriott Pleasanton, in Pleasanton, CA. The location is in close proximity with Lawrence Livermore National Laboratory (LLNL) which co-sponsored the event. From Monday to mid-day Wednesday, invited speakers gave presentations. For each of the five themes described below, one presentation was introductory. There was a poster session and reception on Monday, August 29. There were 99 participants. The workshop focused on four complementary themes at the forefront of current research in climate modeling

Observations are key to uncertainty quantification in climate research because they provide a corroborating source of information about physical processes being modeled. However, observations have uncertainties and this poses a set of methodological and practical issues for comparing them to model simulations: (i) quantifying observational uncertainty when the observations are themselves inferences based on other quantities, (ii) change of support between model resolution and the resolution of remote sensing or in-situ data, (iii) rectifying or accounting for spatial and temporal inconsistencies, (iv) coping with dependence between observations used in model construction and observations used for UQ, and (v) leveraging massive volumes of distributed data. This session was chaired by Amy Braverman (JPL, California Institute of Technology, and UCLA) and featured the following presentations

- Gabi Hegerl (University of Edinburgh) *Deriving Observational Constraints on Climate Model Predictions*,
- Noel Cressie (Ohio State University) *The Statistical Nature of Satellite Retrievals*,
- Robert Pincus (University of Colorado) *Spatial Scale and Uncertainty in Observing the Distribution of Cloud Properties.*

Climate models remain our best tool for understanding past, present and future climate change. However, state-of-the-art climate models still contain many sources of uncertainty, including uncertainties from physical processes that are poorly known or are not resolved at the temporal and spatial scales represented in climate models. These uncertainties may cloud the analysis and interpretation of climate simulations. This theme focuses on the applications of UQ to characterize uncertainties in climate model simulations. Don Lucas (LLNL) chaired this session which included the following presentations

- Karl Taylor (LLNL/PCMDI) *A Multi-Model Perspective of Climate Uncertainties*,

- Don Wuebbles (University of Illinois) *Climate Models and Their Uncertainties*,
- Dorian Abbot (University of Chicago) *Modeling Paleoclimate to Reduce Climate Uncertainty*,
- Linda Mearns (NCAR) *Credibility of Climate Model Projections of Future Climate: Issues and Challenges*.

Assimilation/calibration/forward UQ. With computational models that simulate climate 10s to 100s of years in the future, there comes the need to quantify the uncertainties of the predictions they produce. Uncertainties in these large-scale computational models can stem from a variety of sources including numerical approximations, unknown initial conditions, unknown model parameter settings, missing physics and other inadequacies in the model. Some of these uncertainties can be reduced by constraining the model be consistent with physical observations. This theme focuses on approaches for uncertainty propagation, data assimilation and model calibration that help estimate and constrain uncertainties in model-based predictions. The use of large-scale models make such approaches challenging due to their computational burden, their complexity, and their inadequacies. The incorporation of physical data leads to challenges as well — data are recorded at different scales, the volume of data can be quite large, while their spatial and temporal coverage can be quite small. Dave Higdon (LANL) chaired the session which included the following contributions

- Charles Jackson (University of Texas) *Assessing Which Climate Model Biases Affect Predictions*,
- Nathan Urban (Princeton University) *Climate Uncertainty and Learning*,
- Bruno Sanso (Univ. of California-Santa Cruz) *Blending Ensembles of Regional Climate Model Predictions*,
- Ben Sanderson (NCAR) *Interpretation of Constrained Climate Model Ensembles*.

Multiscale inference. Uncertainty in climate model projections is often represented by an ensemble of plausible simulations, which can either be a collection of simulations from multiple models or from a single climate model. Drawing inference from such ensembles can be challenging; for example, conclude about changes in trends and extreme events. The theme of this topic is statistical analysis of climate model simulations, in particularly methods to draw inference from an ensemble of simulations and across different spatial and temporal scales. This session was chaired by Gardar Johanneson (LLNL) and featured the following talks

- Cari Kaufman (Univ. of California-Berkeley) *Functional ANOVA Models for Comparing Sources of Variability in Climate Model Output*,
- Dan Cooley (Colorado State University) *A Comparison Study of Extreme Precipitation from Six Regional Climate Models via Spatial Hierarchical Modeling*.

LLNL session. Wednesday was devoted to description of various ongoing efforts at LLNL pertaining to Climate and UQ. Two related presentations were given in the morning

- Ben Santer (LLNL) *Accounting for Signal and Noise Uncertainties in Multi-Model Detection and Attribution Studies*,
- Don Lucas and Gardar Johanneson (LLNL) *The Climate UQ Project at LLNL*.

An additional visit to LLNL proper was organized on Wednesday afternoon. During that time, SAMSI program participants and LLNL were able to initiate further interactions.

1.1.3 Engineering Program Opening workshop

The Opening Workshop for the Engineering theme of the SAMSI program on Uncertainty Quantification was held on Monday-Wednesday, September 19-21, 2011, at the Radisson Hotel in Research Triangle Park, NC. There were 97 participants. From Monday to mid-day on Wednesday, five specific themes were explored. The workshop itself followed a nonstandard format in that each theme was introduced by two speakers with one speaker giving an application oriented presentation (Monday morning) and the other a methodology oriented talk (Tuesday morning). On Monday and Tuesday afternoons, participants were divided in breakout sessions and finally, on Wednesday morning, working groups were formed. The intent was for the themes themselves to mirror the working groups (see below). A poster session took place on Monday evening. The worksop featured the following presentations

Renewable energy

- Marija Ilic (Carnegie Mellon) *Toward Qualitative and Quantitative Sustainability Models of Electric Energy Systems*,
- Joseph DeCarolis (NCSU) *Approach to Uncertainty Analysis with an Energy Economy Optimization Model*.

Materials

- Greg Olson (Northwestern) *Probabilistic Materials Science: From Genome to Flight*,
- Wing Kam Liu (Northwestern) *Archetype-Blending Stochastic Multiresolution Theory for Materials Design*.

Nuclear energy

- Rose Montgomery (Tennessee Valley Authority) *Nuclear Fuel Performance, Simulation and UQ*,
- Hany Abdel-Khalik (NCSU) *Uncertainty Quantification and Sensitivity Analysis Methodologies in Nuclear Reactor Calculations*.

Engineered systems

- Sankaran Mahadevan (Vanderbilt),
- Robert Moser (UT Austin) *Quantifying Uncertainty in Simulations of Complex Engineered Systems*.

Sustainability

- Thomas Russell (NSF) *NSF Sustainability Activities from the Perspective of the Mathematical Sciences*,
- Anna Nagurney (U. Mass. Amherst) *Methodology with Applications*.

1.1.4 Geosciences Applications Opening Workshop

The Opening Workshop for the Geoscience theme of the SAMSI program on Uncertainty Quantification was held on Wednesday-Friday, September 21-23, 2011, at the Radisson RTP in Research Triangle Park, NC. From mid-day Wednesday to Friday evening, invited speakers gave presentations and interdisciplinary issues were considered. A poster session and reception took place on Wednesday, September 21. There were 58 participants. The following themes and corresponding talks were considered

Data Assimilation and Inverse Modeling in Geosciences (contamination transport, CO2 storage)

- Christine Shoemaker (Cornell University) *Uncertainty Quantification with SOARS for Computationally Expensive Simulations Including Applications to Carbon Sequestration*,
- Sergey Oladyshkin (University of Stuttgart) *Data-Driven Polynomial Response Surfaces As Efficient Tool For Applied Tasks Under Uncertainty*,
- Omar Ghattas (Univ of Texas-Austin) *A Stochastic Newton Method for Large-Scale Statistical Inverse Problems with Application to Geophysical Inverse Problems*,
- Peter Kitanidis (Stanford) *Recent Advances in the Geostatistical Approach to Inverse Problems*.

Mathematical Methods for UQ in Models of Contaminant Transport

- Margaret Short (University of Alaska) *Predicting Vertical Connectivity within an Aquifer System: A Process Convolution Approach to Reconstructing a Binary Field*,
- Daniele Venturi (Brown University) *A new approach for developing evolution equations for probability density functions*,
- Mina Ossiander (Oregon State University) *Stochastic Parameterizations of Heterogeneous Reservoirs Using Prior Information*,
- Gowri Srinivasan (LANL) *Uncertainty Quantification for Reactive Transport of Contaminants*,
- Vladimir Cvetkovic (Royal Institute of Technology) *A New Approach for Addressing Model Uncertainty in Hydrological Transport*,
- Mary Hill (USGS) *The How of Environmental Modeling: Toward Enhanced Transparency and Refutability*.

UQ in subsurface modeling

- Souheil Ezzedine (LLNL) *Uncertainty Quantification of Flow, Heat and Mass Transport in Discrete Fracture Network*,
- Lucy Marshall (University of Montana) *Advancing Multi-model Approaches for Dynamic Hydrologic Systems*,
- Sean McKenna (Sandia) *Estimation of Fine-Scale Conductivity Fields from Multi-Scale Observations*,
- Daniel Tartakovsky (UCSD).

1.2 Working groups

The above opening workshops worked as catalysts for the activities of the entire program. The opening workshop resulted in the formation of 15 active working groups which we now briefly describe. Report from each individual group are included in Section 3.

Methodology: Surrogate Models. This working group focused on the exploration of properties, utility, and performance of two classes of model surrogates, namely Polynomial Chaos and Gaussian Process surrogates. The study was done in the context of specific model problems with a range of difficulty involving nonlinearity and dimensionality. Test problems included both algebraic functions as well as simple ODE/PDE problems. The group leaders were Habib Najm (Sandia) and Jerome Sacks (retired).

Methodology: Inverse Function-Based Inference This working group studied the use of set-valued inversion of models for inference. Research issues included

- approximation of set-valued inverses in complex spaces,
- computation of inverse measures in parameter space,
- convergence and accuracy of computed inverse measures,
- theoretical issues regarding inversion of multiple observations,
- relation to fiducial inference and Dempster and Shafer calculus,
- intrusive and non-intrusive algorithms, dimension-benign computational algorithms.

The group leaders were Jan Hannig (UNC) and Don Estep (Colorado State).

Methodology: Approximating Computationally Intensive Functions and Sampling Design in High Dimensions. This is a core problem in constructing surrogates, with application to uncertainty propagation, inference, prediction, and design. The relevant research questions include (i) cross-examination of different methods (projection, regression, interpolation, L^1 minimization, Gaussian process/kriging), (ii) appropriate measures of performance/accuracy and their dependence on the intended use of the surrogate model, (iii) error analysis and convergence properties, (iv) sparse representations, (v) adaptive approaches (a posteriori error estimates, derivative properties, dimension reduction, additional optimization of nodes, ANOVA, relation to sequential design of experiments) and (vi) deriving optimality criteria and search algorithms that are good for high dimensions. The group leaders were Youssef Marzuk (MIT) and Dongbin Xiu (Purdue).

Methodology: Multiphysics. Multiphysics models comprising compositions of models of several physical processes, often at different scales, dominate many areas of science and engineering. This working group studied UQ topics for multiphysics models (including both forward and inverse topics). Research issues included (i) complex feedback between physical processes and highly nonlinear responses, (ii) complex and unresolved coupling mechanisms, (iii) representation of uncertainty for different components, and complex interactions between sources of uncertainty and error, (iv) complex, high dimension parameter space, (v) bifurcations and discontinuous model changes and (vi) high performance computational issues. The group leader was Don Estep (Colorado State).

Methodology: Model validation. Model validation refers to the process of assessing the accuracy with which mathematical models can predict physical events, or, more specifically, quantities of interest observed in physical phenomena. Validation should be a prerequisite for predictive modeling, which often forms the basis for decision-making. This working group studied the principles, merits, and limitations of various probabilistic approaches to model validation. Special emphasis was laid on methods for splitting datasets for calibration and validation purposes, on the analysis of model discrepancies, on the development of rejection metrics, and on any other issues of interest raised during the working group meetings. The group leaders were Sujit Ghosh (NCSU), Jan Hannig (UNC) and Serge Prudhomme (UT Austin).

Methodology: Stochastic to Deterministic Models and Back Again. Models of complex multiscale and/or multiphysics phenomena often require combining stochastic and deterministic models. Examples include direct coupling of stochastic and deterministic models, e.g. molecular dynamics with a continuum model and stochastic parameterization, with parameters determined by a stochastic model simulation or other statistical models. Such models are often used to predict "engineering scale" questions from limited microscale information. Of course, this is a classic analysis/modeling problem. This working group focused on computational issues, including (i) rigorous formulation and analysis of coupling mechanisms and their discretizations, (ii) numerical treatment of averaging and computed expectations and the effect of approximations, (iii) a posteriori error analysis and (iv) feedback between stochastic and deterministic models, e.g. nonlinear iterative methods and convergence. The group leaders were Mansoor Haider (NCSU) and Daniel Tartakovsky (UCSD).

Methodology: data assimilation. Data assimilation is the process of fusing information from imperfect models, noisy measurements, and priors, to produce an optimal representation of the state of a physical system. Data assimilation can be interpreted and carried out in a Bayesian framework. Practical methods for large-scale systems include suboptimal and the ensemble Kalman filter approaches, optimal interpolation, and three and four dimensional variational methods. This working group focused on emerging problems that included, but were not limited to: new computational algorithms, modeling of model errors, impact of observations, and quantification of posterior uncertainties. The group leader was Adrian Sandu (Virginia Tech).

Methodology: simulation of rare events. This group focused on methods for simulating rare events in high-dimensional physical systems, especially PDE models. The group explored the use of importance sampling and large deviation theory in order to identify important mechanisms or configurations of the parameters that lead to rare events. They also considered the role of asymptotic analysis in constructing effective sampling weights for such computations. The group leader was Jim Nolen (Duke).

Climate: Data Assimilation in IPCC Level Models. In the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), the results were based on runs from ≈24 models. These were built and run at climate research centers around the world and are each integrated Earth system models that comprise many components, including atmosphere, ocean, ice and land. The so-called dynamical core of such models is a computational model covering both the atmosphere and ocean and based on the primitive equations of GFD. While this is essentially computational, data come into the process of forming the final model. This incorporation occurs at a number of stages of the model development, including parametrization of sub-grid scale effects and model tuning. The process is not, however, done systematically, and current practice is not thought of as "data assimilation." There seems to be a growing realization that DA will have a significant role to play in future climate model development. This is, in part, driven by the need to quantify uncertainty in the model predictions. Nevertheless, there is not a consensus as to how DA should be used in these large-scale climate models. This working group considered the issues involved in formulating a plan for DA in such models. The first step is to understand how such a model is put together and to uncover all the steps where data is currently used in the model formation. For this purpose, the group looked at the latest Community Earth Syetem Model (CESM) from the National Center for Atmospheric Research (NCAR). The group leader was Chris Jones (UNC).

Climate: Parallel Computing Issues. This group looked at adaptive design of experiments, and resolution issues. The group also considered the embedding of emulated sub-models to resolve sub-processes that are now computationally prohibitive. The group also worked towards a python open source software platform. The group leader was Serge Guillas (University College London).

Climate: Statistics of Extremes. This group examined the characterization of extreme events from a statistical point-of-view. The overall group wascomposed of several project groups, each with a specific focus. Some project groups were application-driven, others workedto develop new statistical methodologies, and still others worked to further the theory on which extreme value analyses rely. The overall group aimed to provide a structure for the project groups as well as to provide an environment for investigating aspects which had general interest. The group leader was Dan Cooley (Colorado State).

Engineering: Nuclear energy. This highly interdisciplinary group considered several issues related to uncertainty and propagation of errors in complex, multiscale and multi physics models pertaining to nuclear fuel. The group leader was Don Estep (Colorado State).

Engineering: Renewable energy. The group considered uncertainty quantification issues arising in specific applications linked to renewable energy. In particular, the group worked towards the construction of surrogate problems for large blackbox biofuels problems. The group leader was Ralph Smith (NCSU).

Engineering: sustainability. This group aimed to survey, characterize and discover generalizations of the various methods for treating uncertainty associated with different sustainability dimensions. The group focused on the resilience of built and manufactured systems, exploring the propagation and convolution of uncertainty through system components and processes. The group approached this goal through two subtasks: (i) state-of-the-art study of sustainability mathematics, models and metrics, and (ii) state-of-the-practice case study of sustainability models, mathematics and metrics. The group leader was Miriam Heller (MHITech Systems).

Geosciences working group. This group considered several aspects of UQ pertinent to models in geosciences. In particular, it focused on uncertainty quantification of fault activation due to CO_2 sequestration or geothermal heat extraction. The group leader was Souheil Ezzedine (LLNL).

1.3 Workshops and other meetings

1.3.1 Planning workshop, April 6, 2010, SAMSI

A planning workshop was held at SAMSI on April 6, 2010. The goal of that meeting was twofold. First, SAMSI gathered about 30 experts who helped design the program. Second, the meeting was the initiation point of both a new SIAM activity group on uncertainty quantification, see `http://www.siam.org/activity/uq/`, and a sister interest group within the ASA. The SIAM activity group is now fully operational. Its three main officers — Don Estep (Colorado State), Max Gunzburger (Florida State) and Habib Najm (Sandia) — all had major organizational responsibilities with the SAMSI UQ program. The efforts within the ASA are led by Dave Higdon (LANL) who was also heavily involved in the preparation of the program. A new SIAM/ASA journal (SIAM/ASA Journal on Uncertainty Quantification (JUQ)), `http://www.siam.org/journals/juq.php`, is now also thriving.

1.3.2 UQ Summer School, June 20-24, 2011, Albuquerque, NM

In close collaboration with Sandia (Jim Stewart), SAMSI organized and co-sponsored a UQ Summer School, see `http://www.samsi.info/workshop/samsisandia-summer-school-uncertainty-quantification`. The summer school took place in Albuquerque, NM, next to Sandia, June 20-24, 2011. The four main presenters are Dan Cooley (Colorado State), Doug Nychka (NCAR), Adrian Sandu (Virginia Tech) and Dongbin Xiu (Purdue). They respectively addressed the issues of statistical analysis of rare events, data assimilation and applications in climate modeling, variational data assimilation and sensitivity analysis and polynomial chaos for differential equations. These main presentations were complemented by talks from UQ practitioners, namely Mihail Anitescu (ANL), Mike Eldred (Sandia), Jon Helton (Sandia), Dave Higdon (LANL), Gardar Johannesson (LLNL), Laura Swiler (Sandia), Charles Tong (LLNL).

The Summer School was a resounding success and brought in around 80 participants; not only graduate students but also non expert faculty and researchers took part in the event.

1.3.3 Engineering: Scientific Problems for the Smart Grid, October 3-5, 2011, SAMSI

This two and half day workshop featured talks and working group discussions. The invited speakers both introduced the main technical problems related to the Smart Grid and discussed cutting edge research developments. The discussions formulated some of the pressing problems relevant to smart grid that require the attention of the statistics and applied mathematics communities.

Organizer

George Michailidis (U. of Michigan)

Talks

- Bill Booth (US Energy Information) *What is the Smart Grid?*
- Marija Ilic (Carnegie Mellon University) *Dynamic Monitoring and Decision Systems: The Missing Piece of the Smart Grid Puzzle*
- Russell Bent (Los Alamos National Laboratory) *Expansion Planning for the Smart Grid*
- Zhifang Wang (UC-Davis) *A Stochastic Approach to Studying Cascading Failures in Electric Power Grids*
- Jim Thorp (Virginia Tech) *Synchrophasor Detectives*
- Aranya Chakrabortty (North Carolina State University) *Wide-Area Modeling, Monitoring and Control of Large Power Systems Using Phasor Measurement Technology*
- Sean Meyn (University of Illinois) *The Dynamics of Power*
- George Kesidis (Pennsylvania State University) *Avoiding Overages by Deferred Demand for PEV Charging on the Smart Grid*
- Xiaoming Feng (ABB) *Challenges for Optimization Technology in Smart Grid Application - A Look at Volt and Var Optimization*
- Mesut Baran (North Carolina State University) *Smart Distribution System Research at the FREEDM Systems Center*
- Taufiquar Khan (Clemson University) *PDEs on Graphs: Applications to Complex Power System Networks*

In addition, a poster session was held on October 4, 2011.

1.3.4 Methodology: High Dimensional Approximation for Uncertainty Quantification, November 9, 10, 2011, SAMSI

The field of uncertainty quantification (UQ) has undergone tremendous growth in recent years. Many new methods and algorithms have been developed and used successfully to understand uncertainties in the computational simulation of large and complex problems. Though significant progress has been made,the curse of dimensionality remains a long-standing and cross-cutting challenge. Solving UQ problems(including uncertainty propagation, parameter estimation, probabilistic prediction and design) in high-dimensional parameter spaces requires efficient approaches to high-dimensional approximation. A host of ideas and techniques are relevant to this problem, including sparse grids, low-rank tensor approximations, L1 minimization, nonparametric regression, error estimation and adaptivity, and experimental design. The purpose of this workshop was to bring together a group of researchers, who have made notable contributions to this challenge, to present and compare different strategies and to discuss new ideas. A particular emphasis of the workshop was to bridge gaps of viewpoint and methodology between applied mathematicians and statisticians, and to foster intellectual interactions and cross-disciplinary collaborations.

Organizers:

Don Estep (Colorado State), Youssef Marzouk (MIT) and Dongbin Xiu (Purdue)

Talks

- Alireza Doostan (University of Colorado) *A Compressive Sampling Approach to Sparse Polynomial Chaos Expansions*
- Jeff Wu (Georgia Tech) *Sequential Design of Experiments: A Statistical Perspective*
- Peter Qian (University of Wisconsin-Madison) *Multi-Step Fitting of Large-Scale Emulators: Exploring the Interface between Design and Modeling*
- Myung-Hee Lee (Colorado State University) *HDLSS Discrimination with Adaptive Data Piling*
- Claude Gittelson (Purdue University) *Sparse Tensor Discretization of High-Dimensional Parametric and Stochastic PDE*
- Houman Owhadi (California Institute of Technology) *Optimal Uncertainty Quantification for Legacy Data Observations of Lipschitz functions*

A poster reception took place on Nov. 9, 2011.

1.3.5 Climate: observations workshop, January 17-19, 2012, CISC-NC, Asheville, NC

The workshop was co-sponsored by the Cooperative Institute for Climate and Satellites (CICS-NC) and by SAMSI. It was organized in cooperation with the Program in Spatial Statistics and Environmental Statistics (SSES) at the Ohio State University. The event was hosted by CICS at its Asheville, NC, facilities. Observations are key to uncertainty quantification (UQ) in climate research because they provide a corroborating source of information about physical processes being modeled. However, observations have uncertainties and this poses a set of methodological and practical issues for comparing them to model simulations. The scientific themes of the workshop include: (i) experimental design aspects of collecting ground-based and remotely-sensed observations, (ii) data fusion and homogenization: how to combine heterogeneous observational data sources to get a clearer picture of the true physical process and (iii) link to Uncertainty Quantification for projection of future climate based on past and present observations.

Organizers:

Jessica Matthews (CICS-NC) and Elizabeth Mannshardt (Ohio State U.)

Talks

- John Bates (NCDC) *Uncertainty Challenges for Scientists and Policymakers*
- George Huffman (NASA) *Approaches and Data Quality for Global Precipitation Estimation*
- Dick Reynolds (CICS) *Combining in Situ and Satellite Observations in an SST Analysis*
- Peter Guttorp (University of Washington) *Global Temperature: Data, Methods, Sensitivity*
- Ying Sun (SAMSI) *Functional Median Polish*
- Doug Nychka (NCAR) *Multi-Resolution Spatial Methods for Large Data Sets*
- Amy Braverman (JPL) *Likelihood-based Climate Model Evaluation*
- Brian Reich (N.C. State University) *Extreme Value Analysis for Evaluating Ozone Control Strategies*
- Jenny Brynjarsdottir (SAMSI) *Downscaling Temperatures Over the Antarctic Using a Dimension Reduced Space-Time Modeling Approach*
- Hai Nguyen (JPL) *Spatio-Temporal Data Fusion for Large Remote Sensing Datasets*
- Juan Restrepo (University of Arizona) *Sensitivity, Bred Vectors and Dynamics*
- Peter Thorne (CICS) *GCOS Reference Upper Air Network: Assuring the 21st Century Climate Record?*
- Matt Menne (NCDC) *Uncertainties in the Surface Temperature Record*
- Claude Williams (NCDC) *Benchmarking the Performance of the Pairwise Homogenization of U.S. Surface Temperatures*

In addition, a poster session was held on January 17, 2012. This workshop has also led to a publication: J.L. Matthews, E. Mannshardt and P.A. Gremaud, *Uncertainty Quantification for Climate Observations*, Bull. Amer. Meteo. Soc., `http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-12-00042.1`.

1.3.6 Engineering: Multiscale Modeling and Uncertainty Quantification for Nuclear Fuel Performance, January 25, 26, SAMSI

The Office of Nuclear Energy in the Department of Energy is undertaking a concerted effort to accelerate the time-consuming and expensive process of developing new types of nuclear fuel that increase sustainability. The acceleration depends heavily on the development of improved mathematical and statistical models of potential fuel types that provide robust reliability in behavior predictions. Existing fuel performance models are most semi-empirical descriptions calibrated by the limited observational data, which leads to serious limitations in terms of predicting the behavior of new types of fuel. Moreover, there are significant microstructural and compositional changes in fuels in the extreme environment of a reactor, which makes the modeling all the more difficult. Recently, a team of three universities (Colorado State, Wyoming, and Purdue) and one laboratory (Idaho National Laborator) were awarded a Nuclear Energy University Programs (NEUP) grant in this area. The team addressed the challenges of multiscale modeling of nuclear fuel performance on both the level of creating new modeling paradigms and carrying out systematic uncertainty quantification for multiscale models. During the workshop, the participants surveyed current practices, challenges, and ongoing research pertaining to nuclear fuel performance. Further, they also discussed new multiscale approaches to nuclear fuel modeling.

Organizer

Don Estep (Colorado State)

Talks

- Anter El-Azab (Purdue University) *Modeling thermal transport in nuclear fuels*
- Peter Polyakov (University of Wyoming) *Cahn-Hilliard equations, theory and computation*
- Yushu Yang (SAMSI) *Decoupling coupled stochastic systems*
- William Newton (Colorado State University) *Progress on a mathematical framework for coupling stochastic models*
- Don Estep (Colorado State University) *Using set-valued inverse solutions in stochastic inverse problems for parameter determination*
- Michael Pernice (Idaho National Lab) *Multiscale modeling and UQ for nuclear fuel performance: project context*
- Anter El-Azab (Purdue University) *On the theory of radiation effects in nuclear fuel*

1.3.7 Methodology: Simulation of rare events, February 13-14, 2012, SAMSI

The main aim of the workshop was to contribute to the development of efficient algorithms to simulate complex stochastic systems conditioned on the occurrence of a rare event. Although they may occur with small probability, such events may be of great practical interest because they represent very undesirable system behavior. Simulation algorithms give an estimate of these small probabilities and an understanding of the likely mechanism by which the event occurs. Applications of rare event simulation methods include many engineered systems such as noisy optical laser systems, queuing networks, and mechanical reliability analysis, as well as financial risk, climate dynamics, and random graph and network models. Speakers addressed both theoretical and computational issues related to rare event simulation. Specific topics include large deviation theory, heavy-tailed processes, importance sampling, particle methods, and other novel methods. The workshop took place over two days and provided an ideal place for beginning researchers in the field to interact with experts. The outcomes included opportunities to understand and contribute to the state of the art of the field as well as some of the core questions in the area.

Organizers:

Shankar Bhamidi (UNC), Jose Blanchet (Columbia U.) and Jim Nolen (Duke)

Talks

- Peter Glynn (Stanford University) *Small-sample Behavior for Importance Sampling Rare-event Estimators*
- Maria Cameron (University of Maryland) *Finding Quasipotential for Nongradient SDE's*
- Henry Lam (Boston University) *Rare-Event Simulation for Many-Server Queues*
- Jingchen Liu (Columbia University) *Rare-event Analysis and Simulations for Gaussian and Its Related Processes*

- Kavita Ramanan (Brown University)
- Jonathan Weare (University of Chicago) *A Modified Diffusion Monte Carlo for Rare Event Simulation and More*
- Kay Kirkpatrick (University of Illinois, Urbana-Champaign) *Bose-Einstein Condensation and Quantum Many-Body Systems*
- Paul Dupuis (Brown University) *Importance Sampling for a Multiscale Diffusion*
- Kevin Leder (University of Minnesota) *Analysis of a Splitting Estimator for Rare Event Probabilities in Jackson Networks*

In addition, a poster session took place on Feb. 13, 2012.

1.3.8 Methodology: Models with Complex and Uncertain Domains, March 22, 23, 2012, SAMSI

Classic error quantification for differential equations focused on the effects of discretization on the accuracy of computed solutions. In modern uncertainty quantification, the focus widened considerably to account for the effects of (stochastic) uncertainty and variation in data, boundary conditions, and coefficients of differential equations. This reflects the practical issues that arise in the practical application of differential equation models, which involves the use of experimental and observational data to populate initial data, boundary conditions, and coefficients. However, the vast majority of the theoretical work has considered models posed on simple domains with simple boundaries; domains that are rarely seen in practical applications. The goals of this SAMSI Workshop were to (i) expose problems in which uncertainty and variation in the domain and/or domain boundary geometry is an important issue, (ii) discuss ways that these problems have been tackled mathematically and statistically, (iii) identify areas of future need and opportunity.

Organizers:

Don Estep (Colorado State) and Dongbin Xiu (Purdue)

Talks

- Olivier Pinaud (Stanford University) *Waves in Random Environments and Imaging*
- Nate Burch (SAMSI) *Nonparametric Density Estimation for Elliptic Problems on Uncertain Domains*
- Ilenia Battiato (Clemson University) *Multiscale Models for Transport Processes in Complex Geometries: Applications to Porous and Granular Media*
- Alan Demlow (University of Kentucky) *Adaptive Surface Finite Element Methods*
- Will Newton (Colorado State University) *Adjoint-Based a Posteriori Error Estimate For The Poisson Problem on Closed, Two-Dimensional Surfaces*
- Will Cousins (North Carolina State University) *Outflow Boundary Conditions For Hemodynamic Modeling*
- Eugene Morgan (SAMSI) *Mapping Landslide Hazard over a Nonstationary Space*
- Daniel Tartakovsky (University of California-San Diego) *Differential Equations on Random Domains*

In addition, a poster session was held on March 22, 2012.

1.3.9 Methodology: Uncertainty quantification for high-performance computing, May 2-4, Oak Ridge National Lab, Oak Ridge, Tennessee

The workshop was co-organized and co-sponsored between the Oak Ridge National Laboratory and SAMSI. It was on location at Oak Ridge National Laboratory and included practical workshop guidance in the use of resources from the Oak Ridge National Computing Facility. The future designs of large peta and exascale supercomputing systems will require a paradigm shift in the development of mathematical algorithms and theory. The field of computational uncertainty quantification will have a unique role to play in maximizing the knowledge that can be gained through the full utilization of these supercomputing systems. Key drivers to developing effective mathematical methods for these systems will be achieved through algorithms and theory that expose hierarchies of parallel work while minimizing the power cost of data movement and communication. Speakers addressed both theoretical and computation issues involved with UQ in high performance computing (HPC). More specifically, this three day workshop covered (i) scalable algorithms for UQ, (ii) application, software and data-driven reduced order models, (iii) resources and opportunities at ORNL and the DOE and (iv) calibration, estimation and identification.

Organizers:

Rich Archibald (ORNL) and Clayton Webster (ORNL).

Talks

- Max Gunzburger (Florida State University)
- Dongbin Xiu (Purdue University) *Practical Stochastic Computation Algorithms For Large-Scale Systems*
- Eric Phipps (Sandia National Laboratories) *Exploring Embedded UQ Approaches for Improved Scalability and Efficiency*
- Mihai Anitescu (ANL) *Scalable Gaussian Process Analysis*
- Daniel Tartakovsky (University of California-San Diego) *Uncertainty Quantification for Nonlinear Parabolic & Hyperbolic Conservation Laws*
- Habib Najm (Sandia National Laboratories) *Bayesian Parameter Estimation with Partial Information*
- John Burkardt (Florida State University) *Extending the Power of Sparse Collocation*
- Kate Evans (Oak Ridge National Laboratory) *Scalable Algorithms for Climate Modeling*
- Serge Guillas (University College, London) *Bayesian Calibration and Emulation of Geophysical Computer Models Using High Performance Computing*
- Ernesto Prudencio and Karl W. Schulz (University of Texas) *HPC Challenges Related to UQ Algorithms*
- Marta D'Elia (Florida State University) *Data Assimilation in Hemodynamics,Bayesian Inversion and Model Reduction*
- Karen Pao (ASCR - DOE) *Uncertainty Quantification and the March to Exascale*
- Jeff Nichols (Oak Ridge National Laboratory)

- Don Estep (Colorado State University) *Stochastic Inverse Problems for Parameter Determination*
- Youssef Marsouk (MIT) *Large-scale Bayesian inference without MCMC*
- Omar Ghattas (University of Texas) *Extreme-scale UQ for inverse problems with applications to global seismic inversion*
- Guannan Zhang (Florida State University) *Stochastic Model Calibration with Sparse-grid Bayesian Method for Computationally Expensive Simulations*

In addition, there was a poster session on May 3, 2012.

1.3.10 Engineering: Mini-workshop/special seminar series on multiscale, multiphysics models in materials, energy and sustainability issues, SAMSI

The purpose of this mini-workshop/seminar series was to explore issues of uncertainty and error that arise in complex multsicale and multiphysics models used in materials and energy research. We hosted leading application domain scientists and engineers to describe their multiphysics models in detail, and then describe the UQ issues that are relevant for the context of their models. The talks were followed by several hours of discussion aimed at discovering the important details for UQ research. One goal of the series was to foster new research collaborations. The series focused on materials and energy.

Organizer

Don Estep (Colorado State)

Talks

- November 3, 2011: Khalil Elkhodary (Northwestern) *Construction of multiscale materials models*
- November 17, 2011: Hany Abdel-Khalik (N.C. State and Idaho National Laboratory) *Multiscale modeling of neutronics*
- December 15, 2011: Ayetkin Gel (National Energy Technology Laboratory), UQ in upscaling goal-gas plant prototypes
- February 23, 2012: Wei Chen (Northwestern University) *Stochastic Multiscale Analysis and Design*

1.3.11 UQ Program Transition Workshop, May 21-23, 2012, Radisson RTP, Research Triangle Park, NC

As is customary, the transition workshop signified the formal end of the program under SAMSI but not the end of the research collaborations that have been established. We look forward to many more "products" to emerge from this program. Indeed, the UQ program has explored new formats and ways to engage various groups and communities. The purpose of this workshop was two-fold. First, we reviewed and discussed progress made so far in the various projects of the program, discussed their significance and synergies. Second, future directions of research in the program areas were assessed. The workshop featured sessions from the active working groups and brought together participants across the various areas of the program.

Organizer

Don Estep (Colorado State)

Talks

Model validation

- Maarten Arnst (University of Liege, Belgium) *Description of the Model Problem Considered by the Model Validation Working Group*
- Ed Boone (Virginia Commonwealth University) *Rethinking Model Validation*
- Yifang Li (North Carolina State University) *A Simulation Study of the model considered by the Model Validation Working Group*
- Serge Prudhomme (University of Texas at Austin) *Summary, conclusions, and future directions -Model Validation Working Group*

Inverse functions

- Philip Stark (Univ. of California, Berkeley) *UQQ*
- Troy Butler (Colorado State University) *A Non-intrusive Alternative to a Computational Measure Theoretic Inverse*
- Jan Hannig (Univ. of North Carolina, Chapel Hill) *Statistical Inference Based on Inverse Functions*
- Jessi Cisewski (Univ. of North Carolina, Chapel Hill) *Inverse Function-Based Methodology for Inverse Sensitivity Analysis*

Nuclear engineering

- Rod Schmidt (Sandia) *Nuclear Reactor Safety Analysis: A perspective on and survey of practices, challenges, and needs for UQ*
- Michael Pernice (INL) *Validation and Uncertainty Quantification in the Consortium for Advanced Simulation of LWRs (CASL)*
- Hany Abdel-Khalik (North Carolina State University)
- Don Estep (Colorado State University) *Summary, conclusions and future directions*

Computing

- Lulu Kang (Illinois Institute of Technology) *Functional Data Analysis for the WACCM Computer Simulator*
- Rick Archibald (ONRL) *Uncertainty Quantification Methods for High Performance Computing: Scalable Approximation and Error Estimation for Stochastic Collocation*

Stochastic to deterministic and back

- Mansoor Haider (North Carolina State University) *Bridging Cell and Tissue Scale Models for Cell-Matrix Interactions in Articular Cartilage*

Rare events

- Jim Nolen (Duke University) *Simulation of Rare Events*
- Richard Moore (New Jersey Insitute of Technology) *Importance Sampling Using Low-dimensional Reductions*

- Chia Ying Lee (SAMSI) *Designing Importance Sampling Schemes for Simulating Rare Random Graphs*

Data assimilation in IPCC level models

- Elaine Spiller (Marquette University) *En-Route Observations and 3D Lagrangian Data Assimilation*
- Zav Kothavala (New York University) *Does Data Assimilation Improve the Simulation of Polar Climates in a GCM?*
- Collaborative panel: Chris Jones (UNC), Jesse Berwald (College of William and Mary), Lewis Mitchell (University of Vermont), Chris Orum (University of Utah)

Surrogate models

- John Jakeman (Sandia) *Gaussian Process Models and Polynomial Chaos Approximations for Uncertainty Quantification*

Sustainability

- Alireza Yazdani (Rice University) *Quantification of Uncertainty to Support Sustainability in Water Supply Infrastructure*
- Miriam Heller (MHITech Systems/Johns Hopkins University) *Summary, Conclusions and Future Directions*

Geosciences

- Eugene Morgan (SAMSI) *Geospatial Modeling of Landslide Hazard Offshore the U.S. East Coast*
- Amilcare Porporato (Duke University) *Stochastic Soil Moisture Dynamics: from Soil-Plant Biogeochemistry and Land-Atmosphere Interactions to Sustainable Use of Soil and Water Resources*
- Shouheil Ezzedine (LLNL) *Uncertainty Quantification of Fault Activation Due to CO2 Sequestration or Geothermal Heat Extraction*

Statistics of extremes

- Grant Weller (Colorado State University) *An Alternative Representation of Hidden Regular Variation in Multivariate Extremes*
- Ying Sun (SAMSI) *Estimating Change-points in Extreme Values*
- Brian Reich (North Carolina State University) *A Hierarchical Bayesian Analysis of Max-Stable Spatial Processes*
- Dan Cooley (Colorado State University) *Summary, Conclusions and Future Directions-Statistics of Extremes Working Group*

Data assimilation

- Ralph Smith (North Carolina State University) *Parametric and Model Uncertainty Quantification for Nonlinear Smart Materials*
- Adrian Sandu (Virginia Tech) *Activities of the Data Assimilation Work Group*
- Milija Zupanski (Colorado State University) *Quantifying Observation Impact in Data Assimilation*
- Dan Cacuci (North Carolina State University) *High-Order Uncertainty Quantification and Data Assimilation for Multi-Physics Systems*

In addition, a poster session was held on May 21, 2012.

1.3.12 SAMSI UQ colloquium talks

Throughout the UQ program, SAMSI participants were invited to present their work at the SAMSI colloquium. The following presentations were given (only speakers with direct link to the UQ program are listed).

- September 15, 2011, Tony O'Hagan (U. of Sheffield) *Total UQ*
- September 29, 2011, Maartan Arnst (U. of Southern California) *Dimension reduction and measure transformation in stochastic analysis of coupled systems*
- October 13, 2011, Ruchi Choudary (U. of Cambridge) *Uncertainty quantification of future energy consumption of buildings*
- October 27, 2011, Susie Bayarri (U. of Valencia) *UQ for geophysical risk assessment*
- November 3, 2011, Devon Lin (Queen's University) *Designs for data pooling*
- November 17, 2011, Ricardo Todling (NASA) *A fixed-lag approach to estimate system error*
- December 2, 2011, Paul Dupuis (Brown) *Subsolutions for the design and analysis of rare event Monte Carlo*
- December 15, 2011, Richard Moore (NJIT) *Rare event simulations in nonlinear optics*
- February 2, 2012, Juan Restrepo (U. of Arizona) *Climate variability, when data fail us*
- February 9, 2012, Xiaoying (Maggie) Han (Auburn U.) *Asymptotic behavior of stochastic lattice differential equations*
- February 16, 2012, Fabrizio Ruggeri (CNR IMATO, Milano, Italy) *On the choice of the prior distribution in Bayesian statistics*
- March 1, 2012, Alessandra Mattei (Università degli Studi di Firenze, Italy) *Quantifying uncertainty exploiting multiple outcomes in Bayesian inference for causal effects with noncompliance*
- March 15, 2012, Liang Peng (Georgia Tech) *Empirical Likelihood Tests for High Dimensional Data*
- March 29, 2012, Alexey Chernov (U. of Bonn, Germany) *Sparse space-time Galerkin BEM for the nonstationary heat equation*
- April 12, 2012, Peng Zeng (Auburn U.) *Linearly Constrained Lasso with Application in Glioblastoma Data*
- April 19, 2012, Don Estep (Colorado State) *Efficient approximation, error estimation, and adaptive computation for randomly perturbed elliptic problems*

1.3.13 UQ conference in cooperation with SIAM

In addition to the above, it is worth noting that the first UQ meeting organized by SIAM (and its new UQ activity group) took place in Raleigh, April 2–5, 2012. The location was chosen to coordinate activities with the ongoing UQ SAMSI program. The SIAM conference was in fact held in cooperation with the American Statistical Association (ASA), the Statistical and Applied Mathematical Sciences Institute (SAMSI), and the United States Association for Computational Mechanics (USACM). The three co-chairs (Don Estep (Colorado State University), Dave Higdon (Los Alamos National Laboratory) and Habib Najm (Sandia)) were all key participants in the SAMSI program and so were most of the members of the Organizing Committee (Susie Bayarri, University of Valencia, Spain; Jim Berger, Duke University; Roger Ghanem, UCLA (USACM Representative); Albert Gilg, Siemens, Germany; Pierre Gremaud, SAMSI (SAMSI Representative); Max Gunzburger, Florida State University; Gardar Johannesson, Lawrence Livermore National Laboratory; Laura Lurati, Boeing Inc.; Youssef Marzouk, MIT; Anthony O'Hagan, University of Sheffield, United Kingdom; Stephan Sain, National Center for Atmospheric Research; Philip Stark, University of California, Berkeley; Christian Soize, Université Paris-Est Marne-la-Vallée, France; Dongbin Xiu, Purdue University).

A large number of SAMSI program participants, both mathematicians and statisticians, presented their work and/or organized mini-symposia at the SIAM conference.

2 Courses and workshops for students

2.1 Graduate course: Uncertainty Quantification

A two semester course was offered as part of the program. During the fall, 16 graduate students from Duke University, North Carolina State University and the University of North Carolina at Chapel Hill took the course for credit while around 20 others attended the class on a regular basis. The course continued during the Spring Semester with 13 students completing the course for credit. Lectures were given at weekly SAMSI on Wednesdays, 4:30-7:00 p.m. The main topics covered in the first semester included

- basic probability theory
- basic approximation theory
- basic numerical methods and error estimation
- formulation of stochastic system
- sensitivity analysis
- generalized polynomial chaos
- stochastic Galerkin and collocation method.

The main topics covered in the second semester included

- advanced numerical techniques for SPDE: adaptive methods and compressive sensing
- propagation of probability distributions
- Bayesian inference
- data assimilation
- model calibration

The main instructors were Don Estep (Colorado State) and Dongbin Xiu (Purdue) plus guest lecturers.

2.2 Undergraduate student workshops

Two two-day Undergraduate Workshop were held at SAMSI on October 29-30, 2010, on the general theme of UQ. In addition, a week long modeling workshop focused on the same theme. These workshops were part of SAMSIs Education and Outreach Program for 2011-2012. All three workshops had over 30 student participants each. The presenters were SAMSI postdocs and faculty and grad fellows as well as senior SAMSI program participants.

2.2.1 Fall UQ outreach workshop, October 28, 29, 2011, SAMSI

Organizer

Pierre Gremaud (SAMSI)

Talks and activities

- Dan Cooley (Colorado State University) *Spatial Analysis of Return Levels for Extreme Precipitation*
- Andreas Aristotelous (SAMSI) *MATLAB demo*
- Maggie Han (Auburn University) *An Introduction to Stochastic Differential Equations*
- Chia Ying Lee (SAMSI) *Random Walking to the Brownian Motion*
- Ying Sun (SAMSI) *R demo*
- Jenny Brynjarsdottir and Nate Burch (SAMSI) *Data Assimilation and the Kalman Filter*
- Pierre Gremaud (SAMSI) *career options*
- Ralph Smith (N.C. State University) *Mathematics of Uncertainty Quantification*
- Lulu Kang (Illinois Institute of Technology) and Devon Lin (Queen's University) *Introduction on Gaussian Process*
- Lulu Kang (Illinois Institute of Technology) and Devon Lin (Queen's University) *R lab*

2.2.2 Spring UQ outreach workshop, February 24, 25, 2012, SAMSI

Organizer

Pierre Gremaud (SAMSI)

Talks and activities

- Dongbin Xiu (Purdue University) *Numerical Integration*
- Chia Ying Lee (SAMSI) *MATLAB demo*
- Andreas Aristotelous (SAMSI) *From Richardson's Model to Cancer Growth Simulation*
- Richard Moore (New Jersey Institute of Technology) *Importance Sampling in Optical Communication*
- Jenny Brynjarsdottir (SAMSI) *R demo*
- Ying Sun (SAMSI) *descriptive statistics*
- Pierre Gremaud (SAMSI) *career options*
- Nate Burch (SAMSI) *Sensitivity Analysis Based Uncertainty Quantification*
- Alex Chen (SAMSI) *This Means War! Modeling Combat with Applications to Real Time Strategy Games*
- Alex Chen (SAMSI) *MATLAB lab*

2.2.3 Interdisciplinary Workshop for Undergraduate Students - May 14-18, 2012, SAMSI and NCSU

This week long workshop provided an introduction to the field of uncertainty quantification (UQ) to students from the mathematical and statistical sciences. The overall goal of the workshop was to illustrate to need for and power of quantitative methods to confront the very hardest and most important data- and model-driven scientific challenges. The students worked on controlling specific stochastic dynamical systems. At the end of the workshop, they had acquired a working knowledge of variational techniques within the above framework, simple Kalman filters and Monte Carlo approaches. Emphasis was put on practical issues and implementation. The students worked in groups of three under the supervision of SAMSI researchers and postdocs.

Organizers

Cammey Cole Manning (Meredith College), Pierre Gremaud (SAMSI) and Juan Restrepo (U. of Arizona)

Talks and introductory material

- Juan Restrepo (U. of Arizona) *Project overview*
- James Nance (NCSU) *MATLAB demo*
- Ying Sun (SAMSI) and Mi Zhou (NCSU) *basic statistics*
- Nate Burch (SAMSI) Bevin Maultsby (Duke University) Mami Tonoe (N.C. State University) *Linear algebra I, II and Kalman filters*
- Andreas Aristotelous (SAMSI) and Esteban Sanchez (Duke) *ODEs*
- Jenny Brynjarsdottir (SAMSI) and Yifang Li (NCSU) *introduction to Bayesian statistics*
- Alex Chen (SAMSI), Jessi Cisewski (UNC) and Chia Ying Lee (SAMSI) *Monte Carlo*

At the end of the workshop, the participating students presented the results of their group project.

2.2.4 Graduate Fellow Presentation Poster Session and Reception - April 18, 2012, SAMSI

In addition to being attending the program courses and being involved in working groups (see below), the SAMSI grad fellows were also given the opportunity to present their work to the entire SAMSI community during a poster reception devoted to them.

3 Working group reports

3.1 Methodology: surrogate models

Leaders: Habib Najm (Sandia) and Jerome Sacks (retired)

3.1.1 Participants

Susie Bayarri (U. of Valencia), Jim Berger (Duke), Jenny Brynjarsdottir (SAMSI), Ruchi Choudary (U. of Cambridge), Maggie Han (Auburn), Alan Lenarcic (UNC), Devon Lin (Queen's U.), Danilo Lopes (Brasil), Silvia Montagna (Duke), Habib Najm (Sandia), Tony O'Hagan (U. of Sheffield), Facundo Munoz (U. of Valencia), John Jakeman (Sandia), Jerome Sacks, Surya Tokdar (Duke), Robert Wolpert (Duke)

3.1.2 Research activities and results

The Surrogates working group was initiated to explore comparisons between Polynomial Chaos (PC) related methods and Gaussian process (GP) based methods for creating surrogates for computer models. Collaboration with John Jakeman led to a series of working group reports and, with Jason Loeppky, a manuscript is under preparation that will expand and report on the issue.

A subgroup was formed with James Berger, M. J. Bayarri, Robert Wolpert and Danilo Lopes to study the use of Bayesian model selection techniques for dimension reduction in analyzing computer models with large dimensional input.

The work of Surya Tokdar and Silvia Montagna on developing surrogates for computer models whose output has "non-stationary" characteristics, creating barriers for the use of standard GP surrogates, was abetted through discussions and presentations at working group meetings.

The project on comparison of PC and GP methods led to a collaboration with John Jakeman, an applied mathematicianat Sandia in New Mexico. Initiating the study also involved collaboration with Habib Najm and Omar Knio. The work on dimension reduction advanced ongoing collaborations with the described team. Collaboration with Habib Najm has led to a new effort (in collaboration with William Welch of British Columbia) for dimension reduction in a component of a climate model.

3.1.3 Working papers and publications

1. Brynjarsdottir, J. and O'Hagan, A. (2013). Learning about physical parameters: The importance of model discrepancy. Submitted.
2. J.D. Jakeman and M.S. Eldred, Enhancing compressive sensing for Polynomial Chaos Expansions through basis selection. In preparation.
3. J.D. Jakeman, J. Leopky and J. Sacks, Gaussian Process Models and Polynomial Chaos Expansions for Uncertainty Quantification: A Discussion. In preparation.

3.2 Methodology: inverse function-based inference

Leaders: Jan Hannig (UNC) and Don Estep (Colorado State)

3.2.1 Participants

Troy Butler (UT-Austin), Jessi Cisewski (UNC), Don Estep (Colorado State), Jan Hannig (UNC), Bernard Omolo (USC Upstate), Adrian Sandu (Virginia Tech)

3.2.2 Research activities and results

This working group was interested in developing methodology for set-valued inversion of models for the purpose of inference. The group focused on several issues, including (i) approximation of set-valued inverses in complex spaces; (ii) computation of inverse measures in parameter space; (iii) convergence and accuracy of computed inverse measures; (iii) theoretical issues regarding inversion of multiple observations; (iv) intrusive and non-intrusive algorithms; and (v) dimension-benign computational algorithms. In addition, the group investigated the relations of these issues to generalized fiducial inference. The group members were introduced to the two main themes of the working group: 1. Generalized fiducial inference (by Jan Hannig) 2. Stochastic inverse sensitivity analysis (by Don Estep and Troy Butler) The next step was to see how generalized fiducial inference could be applied to the inverse problem as presented by D. Estep and T. Butler for smooth deterministic maps. Previous work by D. Estep, T. Butler and co-authors developed and analyzed a measure-theoretic approach to solve this problem. There are also Bayesian solutions to this inverse sensitivity problem when the map is statistical rather than deterministic. The problem considered by this group, however, is the following: given the deterministic map, what parameter distribution (i.e. distribution on the input space) produced it? In the Bayesian setting, this parameter distribution would be obtained by assuming a prior distribution and computing a posterior. The objective was to use the inverse of the physical map and obtain such a distribution without an explicit prior distribution.

3.2.3 Working papers and publications

1. J. Cisewski and J. Hannig, Generalized fiducial inference for logistic regression with mixed effects (In preparation).
2. A posteriori error estimates for explicit time integration methods, J. Collins, D. Estep and S. Tavener, BIT Numerical Mathematics, 2012, submitted
3. Density estimation for a class of elliptic problems on stochastically perturbed domains, N. Burch and D. Estep, in preparation.
4. A posteriori error estimation for the Lax-Wendroff method, J. Collins, D. Estep, and S. Tavener, in preparation.
5. A measure theoretic approach to inverse sensitivity analysis III: disintegration and multiple quantities of interest, T. Butler, D. Estep, S. Tavener, in preparation.

3.3 Methodology: high dimensions and multi physics

Leaders: Don Estep (Colorado State), Mansoor Haider (NCSU) and Dongbin Xiu (Purdue)

3.3.1 Participants

Lulu Kang (IIT), Nate Burch (SAMSI), Devon Lin (Queens U.), Susie Bayarri (Valencia), Andreas Aristotelous (SAMSI), James Nance (NCSU), Daniel Tartakovsky (UC San Diego), Michael Pernice (Idaho National Laboratory), Troy Butler (UT Austin), Will Newton (CSU), Simon Tavener (CSU)

3.3.2 Research activities and results

The major goal of the Multiphysics working group was to identify some of the key mathematical and statistical problems underlying the many challenges that arise in multiphysics simulations and to initiate research on these problem. Multiphysics systems incorporate different physical processes spanning a wide range of scales. The current state of the study and application of multiphysics systems is itself a significant challenge (Keyes *et al.*, 2012). Currently, the mathematical models of behaviors at various scales are often relatively crude and unverified, the mechanisms by which different processes interact are often unclear, and experimental data is sparse and available only on a limited number of scales. Yet, a perusal of current engineering and scientific literature reveals a rapidly burgeoning activity in the application of mathematical, computational, and statistical techniques to the study and application of multiphysics systems. Consequently, there is now a large body of ad hoc techniques and results that substitute (unfortunately) for rigorous theoretical development. This situation is fully described in the recent DOE report (Keyes *et al.*). The working group ended up concentrating on a few key issues:

1. A posteriori error estimation for multiphysics simulations. Unlike the case of simpler physics for which accurate simulations are achievable, numerical simulations of multiphysics, multiscale systems always have significant numerical error in practical situations. Numerical error is affected not only by the obvious discretization issues for each component physics, but also by the effects of finite iteration in the inevitable iterative schemes that must be employed, the error feedback between components, and the effects of transforming information from one component to be passed into other components. For the purpose of UQ and for making decisions on where to put computational resources, it is necessary to have a posteriori error estimates that detail the various contributions to the overall error. We pursued such estimates for elliptic systems coupled through the equations and elliptic and parabolic problems posed on different domains that share a common boundary. We pursued estimates both for deterministic and stochastic quantities.

2. The effects of error and stochastic variation in the domain geometry. Much of standard numerical analysis for numerical PDES is carried out on domains of simplistic geometry, e.g. the unit square. In multiphysics applications, domains are typically very complex, e.g. manifolds or with complicated boundaries, where the geometry is subject to uncertainty because of measurement area or natural variation. There has been relatively little investigation in a posteriori error estimation of these effects on the overall accuracy. The working group worked on algorithms for efficient simulation and a posteriori error estimation and UQ for (1) problems posed on manifolds in $\mathbb{R}^3$ where the manifold is represented as a piecewise polynomial surface determined by a set of measurement points and (2) domains in $\mathbb{R}^2$ where the boundary is a piecewise polynomial with knot values affected by stochastic variation.

3. Stochastic quantities in multiphysics coupling. In practical application of multiphysics models, the information that is passed between the physical components — the "coupling quantities" — are affected by stochastic variation. Current engineering UQ practice for multiphysics systems treats the models as black boxes and makes use of "uninformed" statistical models to model the output of a stochastic multiphysics model. This is very problematic however as (1) the output of a multiphysics model over a realistic parameter domain is extremely complex, hence not readily modeled by a standard statistical model, while there is no theoretical or asymptotic basis for employing typical statistical models; and (2) the way in which multiphysics models are simulated in practice means there is significant feedback between the stochastic variation produced in each component. Moreover, iterative solution methods are typically used for multiphysics models. There has been little investigation of the convergence of such approaches in the context of stochastic outputs. The group investigated the analog of fixed point iteration for coupled stochastic models, with convergence expressed in terms of convergence in the probability distributions of the coupling quantities, and developed a general theory.

4. Inverse problems for parameter determination. In practice, investigations of a typical multiphysics system must be based on a complex amalgamation of experimental observations of certain behaviors at certain scales coupled with mathematical models of of other behaviors at other scales, e.g. what is sometimes called "data-model fusion". The solution of inverse problems for physical parameters (as opposed to parameters in a statistical model) is quite often involved. Some approaches to this problem of (physical) parameter determination are based on statistical approaches, e.g. Bayesian methods. While these approaches are widely employed, there remain fundamental theoretical and practical issues that have been unaddressed. Some of these have to do with how the statistical models for physics models are constructed, and then validated. As an alternative, this group developed a new approach to solve the physical inverse problem. This approach provides a way to approximate the typical set-valued inverse and then use this inverse to produce a measure on the parameter space from a measure on the output. During the UQ year, the group focused on extending their previous results for a single observation to multiple observations.

3.3.3 Working papers and publications

1. Nonparametric density estimation for randomly perturbed elliptic problems III: Convergence, complexity, and generalizations, D. Estep, M. Holst, A. Malqvist, Journal of Applied Mathematics and Computing 38 (2012), 367-387

2. A computational measure theoretic approach to inverse sensitivity problems II: A posteriori error analysis, T. Butler, D. Estep and J. Sandelin, SIAM Journal on Numerical Analysis, 50 (2012)

3. A posteriori analysis of multirate numerical method for ordinary differential equations, D. Estep, V. Ginting, S. Tavener, 2012, Computer Methods in Applied Mechanics and Engineering, 223-224 (2012), 10-27

4. A numerical method for solving a stochastic inverse problem for parameters, T. Butler and D. Estep, Annals of Nuclear Energy, 2012, 86-94, 10.1016/j.anucene.2012.05.016

5. Multiphysics Simulations: Challenges and Opportunities, D. E. Keyes, L. C. McInnes, C. Woodward, W. Gropp, E. Myra, M. Pernice, J. Bell, J. Brown, A. Clo, J. Connors, E. Constantinescu, D. Estep, K. Evans, C. Farhat, A. Hakim, G. Hammond, G. Hansen, J. Hill, T. Isaac, X. Jiao, K. Jordan, D. Kaushik, E. Kaxiras, A. Koniges, K. Lee, A. Lott, Q. Lu, J. Magerlein, R. Maxwell, M. McCourt, M. Mehl, R. Pawlowski, A. Peters Randles, D. Reynolds, B. Riviere, U. Ruede, T. Scheibe, J. Shadid, B. Sheehan, M. Shephard, A. Siegel, B. Smith, X. Tang, C. Wilson, and B. Wohlmuth, Tech. Rep. ANL/MCS-TM-321, Revision 1.1, Oct. 2012, Argonne National Laboratory. To appear as a special issue of the International Journal of High Performance Computing Applications.

6. A posteriori analysis and adaptive error control for multiscale operator decomposition solution of elliptic systems II: Fully coupled systems, V. Carey, D. Estep, S. Tavener, International Journal of Numerical Methods in Engineering, 2011, in revision

7. A-posteriori error estimates for mixed finite element and finite volume methods for problems coupled through a boundary with non-matching grids, T. Arbogast, D. Estep, B. Sheehan, and S. Tavener, IMA J. Numerical Analysis, 2012, in revision

8. Density estimation for a class of elliptic problems on stochastically perturbed domains, N. Burch and D. Estep A posteriori error estimates for explicit time integration methods, J. Collins, D. Estep and S. Tavener, BIT Numerical Mathematics, 2012, submitted

9. A posteriori error estimates for the Poisson problem on closed two-dimensional surfaces, W. Newton, D. Estep, M. Holst

10. A posteriori analysis of an iterative multi-discretization method for reaction-diffusion systems, J. Chaudhry, D. Estep, V. Ginting, S. Tavener

11. A measure theoretic approach to inverse sensitivity analysis III: disintegration and multiple quantities of interest, T. Butler, D. Estep, S. Tavener

3.4 Methodology: model validation

Leaders: Sujit Ghosh (NCSU), Jan Hannig (UNC) and Serge Prudhomme (UT Austin)

3.4.1 Participants

Maarten Arnst (University of Lige, Belgium), Edward Boone (Virginia Commonwealth University), Mark Campanelli (NIST), Alex Capaldi (Valparaiso University), Qingyun Duan (Beijing Normal University), Sujit Ghosh (NCSU), Jan Hannig (UNC), Charles Jackson (UT Austin), Alan Lenarcic (SAMSI), Yifang Li (NCSU), Youssef Marzouk (MIT), Bernard Omolo (University of South Carolina Upstate), Zeng Peng (Auburn University), Serge Prudhomme (UT Austin), Zhiguang Qian (University of Wisconsin-Madison), Fabrizio Ruggeri (CNR IMATI, Italy), Ankur Srivastava (MIT), Ilya Timofeyev (University of Houston), Sylvie Tchumtchoua (Duke), Richard Wilkinson (University of Nottingham, UK)

3.4.2 Research activities and results

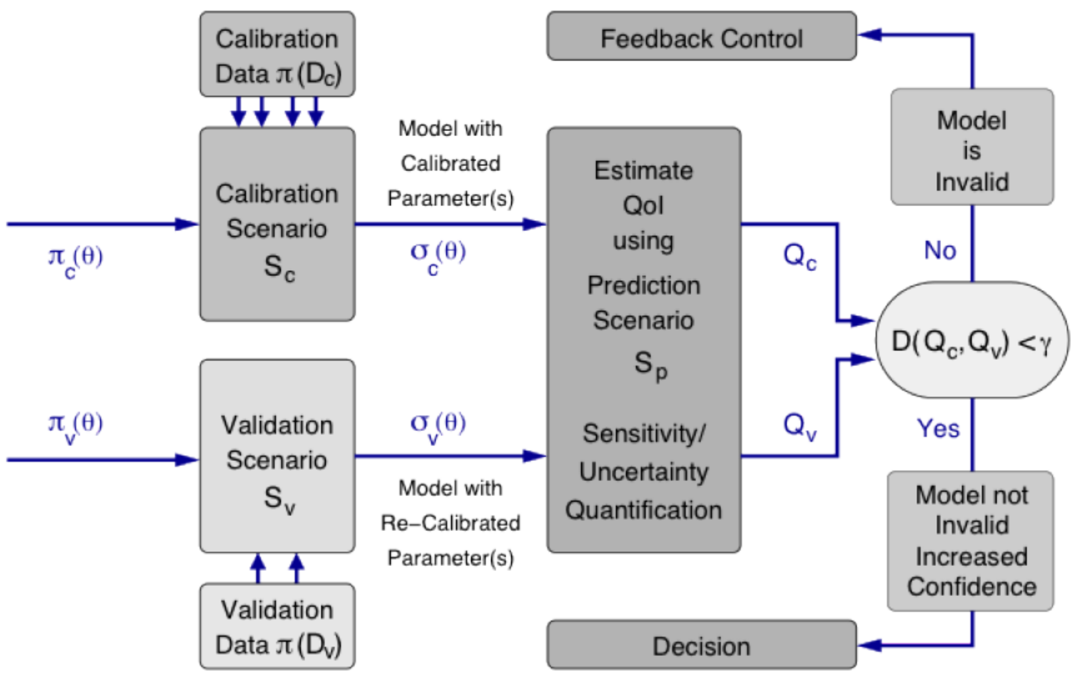

Figure 1. Flowchart of the validation process to test the validity of models. Prior pdfs on parameters θ are introduced for calibration scenarios S_c and validation scenarios S_v to yield posterior pdfs via Bayesian updates. The resulting QoIs, Q_c and Q_v, are compared using a metric D(.,.) against a tolerance γ. If the tolerance is met, the model is declared not invalid; otherwise, the data and model must be updated to deliver a reliable prediction.

Model validation refers to the process of assessing the accuracy with which mathematical models can predict physical events, or, more specifically, quantities of interest (QoIs) observed in physical phenomena. It addresses the question "Are we solving the right equations?" Validation should be a requirement for predictive modeling as computer simulations are increasingly relied upon to support critical decisions. The goals of this working group were to study the principles, merits, and limitations of various probabilistic approaches to model validation. One of the main issues that the group addressed was: How well can we expect a given model to perform on different or future scenarios? There is unfortunately no straightforward answer to this question as any response to this fundamental question of model validation rests upon several other factors including, but not limited to, (a) the choice of metric to be used to measure model discrepancy, (b) the choice of predictor variables to be included in the simpler model and (c) the complexity of the model. Hence, special emphasis was laid on methods for splitting datasets for calibration and validation purposes, on the analysis of model discrepancies, on the development of appropriate metrics to measure model discrepancies, and on other issues of interest raised during the working group meetings. One validation approach that was studied in the working group is based on Bayesian inference and is described in Figure 1. One advantage of the process is that it addresses right on the assessment of a model with respect to the QoI on a given prediction scenario. This is extremely important as it recognizes that a model may be valid for the prediction of a given QoI but invalid for a different QoI. Given a mathematical model, the first step of the framework is concerned with the statistical calibration of the model parameters using data from "simple" calibration experiments. The solution to the calibration (inverse) problem is the posterior pdf $\sigma_c(\theta)$ of the model parameters θ updated from a prior pdf $\pi_c(\theta)$. In the second step, a new experimental scenario, selected to corroborate the model assumptions, gives new data to update the prior $\pi_v(\theta)$ into a new posterior pdf $\sigma_v(\theta)$. The third step consists in solving the stochastic forward model on the scenario of interest using the two pdf's $\sigma_c(\theta)$ and $\sigma_v(\theta)$. This provides the two pdf's $Q_c(\theta)$ and $Q_v(\theta)$ for a selected QoI, which are then compared through a predefined metric; if the model passes the criterion, it is said to be "not invalidated", meaning that more confidence in the predictive capabilities of the model has been gained. Otherwise, the model is said to be invalid.

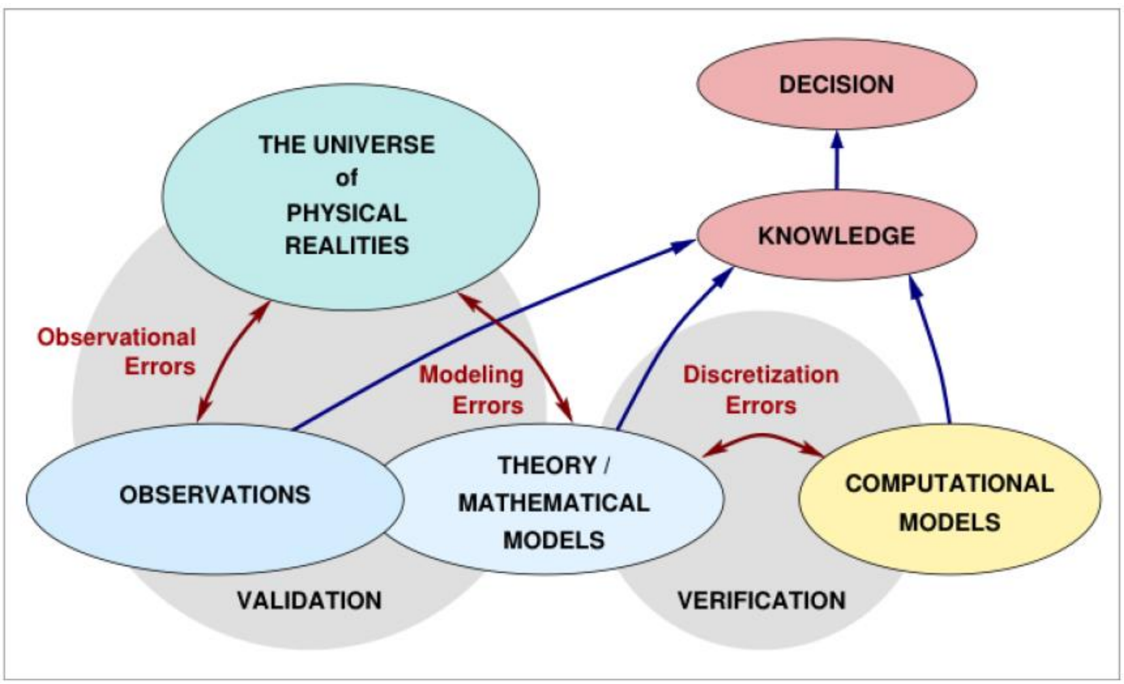

Figure 2. Reproduced from J. T. Oden, R. Moser, G. Ghattas, Quantification of uncertainty in computer predictions, Parts I, *SIAM News* **43**, no. 9-10, 2010.

Model validation may look deceptively simple. Yet it is a complex and time-consuming process, in critical need of new technological advances to better control the outcome of the process. The reasons are multiple:

validation requires cumbersome planning (description of prediction goals, selection of models, data collection, sensitivity analysis, selection of calibration and validation datasets, see figure to the right) based on many subjective choices; a model is never validated, it is at best not invalidated; in the case where a model is rejected, the process calls for feedback loop criteria that instruct whether one should modify the theory or ask for more data, etc. On the other hand, validation is becoming crucial for decision-making (see Figure 2).

In order to begin with a simple yet representative example based on physical systems, the group considered the simulation of elasto-plastic phenomena using a simple elasto-plastic spring-mass oscillator (one degree-of-freedom system) (see for example, Tetsuhiko Miyoshi, Foundations of the numerical analysis of plasticity, *Lecture Notes in Numerical and Applied Analysis*, Vol. 7, North Holland, 1984). The idea is to generate simulated (manufactured) data using a classical constitutive model for the plastic deformation and to use these data to calibrate the parameters of simpler competing models defined in terms of polynomial expansions (approximations). Several issues were explored. For example, there are clear differences between the calibration, validation, and operating regimes. If the model is calibrated in a regime that only involves elastic deformation, the elasto-plastic behavior will not be captured and represented correctly whenever the operating regime will involve elasto-plastic behavior. There is also an issue in how the model can be made more faithful to data by increasing the polynomial degree (model selection). However, it is well known within statistics literature that if a model is made too faithful to a given set of data (and hence reducing the potential bias) by using more complex models, such complex models may lead to inflated predictive variance. The objective is to seek models that would not only reduce potential predictive bias but also account for inflated predictive variance due to model complexity. Moreover, the choice of metric to discriminate the models is also important. A model that may perform well in terms of the squared residual error metric may not perform well if the metric is replaced by other choices (e.g., absolute errors, Kullback-Leibler discrepancy or Hellinger distance). By using manufactured/simulated data scenarios, we have the opportunity to consider situations in which one has small or large amounts of data available and hence detect departures from the assumptions of true model.

The Model Validation Working Group (MVWG) discussed several issues and model scenarios that pertain to various aspects of model validation. The MVWG consisted of about 20 members with backgrounds in computer science, building physical models, mathematics, and statistics. A significant number of the members participated each week and presented and shared their ideas. After several weeks, two core groups within the MVWG were formed: (i) the first core group consisted of mathematicians, computer scientists, biologists, etc. and led on creating representative physical/mathematical models to describe specific physical phenomena; (ii) the second core group consisted of statisticians, epidemiologists and biostatisticians, and led on developing relevant metrics for model validation using the test examples (created by the former core group) to measure the success (and failure) of different model validation techniques.

The group with mathematical modeling background provided a series of applications that were used as examples to test and illustrate new validation methodologies.

The group with statistical background, Frequentist and Bayesian, discussed theoretical developments in defense of cross-validation and out-of-sample based model validation and model comparison methods. The group theorized a defensible toy problem where we can state the error distribution of k-fold cross validation of many factors of k for several orders of model complexity. There are analogies with regression Bayes factors and Bayes model averaging. Furthermore, for procedures offering confidence interval coverage of parameters, such as fiducial and Bayes Gibbs selection, may be used to devise model validation metrics for fair comparison of such estimation procedures. For the benefit of the engineering project, which is comparing simpler differential equations models with data from a nonlinear equation, the statistical theorists in our group aimed to supply them validation metrics that aid in the search for out-of-sample valid models. The group sought to stringently define what strong hypotheses on out of sample/out of observed range data must hold to trust models, and to show that cross-validation tests these hypotheses for several model classes.

A new collaboration was initiated with J. Hannig, F. Ruggeri, S. Ghosh, E. Boone, M. Anherst and S. Prudhomme. The topic of the work was exploring the validity regions of misspecified pde models using statistical tools. The group worked on an expository paper on the issues involved, and planned to continue their collaboration.

3.5 Methodology: stochastic to deterministic models and back again

Leaders: Mansoor Haider (NCSU) and Daniel Tartakovsky (UCSD)

3.5.1 Participants

Andreas Aristotelous (SAMSI), Susie Bayarri (Valencia), Jim Berger (Duke), Don Estep (Colorado State), Mansoor Haider (NCSU), Xiaoying Han (Auburn), Chia Lee (SAMSI), James Nance (NCSU/SAMSI), Daniel Tartakovsky (UC-San Diego), Ilya Timofeyev (Houston)

3.5.2 Research activities and results

The focus of this group was the examination of computational techniques for applications in which there is a strong coupling between stochastic models (typically at the micro-scale) and deterministic models (typically at the meso- or macro-scale). In many cases, the development of robust, efficient and accurate system-level models requires the successful integration of diverse analytical and computational techniques that bridge the scales and reconcile stochastic and deterministic components. The regular meetings of the full working group were equally split between a discussion of techniques for passing from a discrete formulation of a stochastic model to an associated deterministic model (via a continuous limit), and the presentation of applications relevant to the working group theme. The group's activities could be categorized into three sub-topics, with the first under discussion by the entire group (often in joint meetings with the Multiphysics WG), and the latter two topics being considered by sub-groups.

(a) Methods for Bridging Stochastic and Deterministic Models.

The regular meetings of the full group primarily focused on the presentation and discussion of techniques for passing from a discrete model of a stochastic process to a deterministic model using a continuous limit. Earlier meetings focused on the connection between uncorrelated random walks, Brownian motion and the diffusion equation, and proceeded to the investigation of correlated random walks and the Telegrapher's equation. Additional meetings involved interaction with visiting applied scientists and engineers who develop and apply models with substantial multiscale and stochastic components. These interactions included topics such as multiscale modeling of neutronics in nuclear reactor calculations, and development of multiscale models for optimal material design.

(b) Agent Based Models.

Agent based modeling provides a natural framework for investigation of the topical theme of this working group. At finer scales in such models, individual "agents" have rules governing their responses that depend on "perception" of local variables in their surroundings as well as the other agents. At coarser scales, the resulting phenomena must be integrated into the framework of more traditional deterministic models. A primary focus of this sub-group was the development of agent-based computational methods within the context of tissue engineering applications for articular cartilage. In such systems, individual cells (the agents) are seeded in biomaterial scaffold materials saturated with nutrients and growth factors. Over time, the cellular biosynthesis, cell proliferation, and accumulation of extracellular matrix constituents results in tissue regeneration, with the role of modeling being to aid in realization of optimal functional outcomes in the tissue engineered construct. Initial models consider a reaction-diffusion system with individual cellular agents modeled via delta function sources that adjust their nutrient absorption and cellular proliferation rates based on nutrient concentrations in their local environment.

(c) Domain Decomposition in MC Particle Simulations.

Many micro-scale Monte Carlo-based computational models involve the direct numerical simulation of interactions between individual particles. Since the computational complexity of such simulation algorithms can be $O(N^2)$, ad hoc rules are often introduced to improve computational efficiency. However, such rules do not typically take account of the underlying probabilistic distributions that may be associated with physical aspects of the underlying phenomena. This sub-group considered computer simulations of interacting "ping-pong" balls with elastic or inelastic collisions with a focus on domain decomposition. The goal was to understand how the observed statistics of the simulations in each sub-domain could be used to identify probability distributions and patterns in these distributions within the sub-regions of the modeling domain. The aim was to then use these distributions to run simulations, having lower order computational complexity, that mimic or reproduce statistical properties of the full $O(N^2)$ computational model and also provide a quantification of the associated uncertainty (e.g. via confidence levels).

3.5.3 Working papers and publications

1. A. Chertock, A. Kurganov, A. Polizzi, and I. Timofeyev, *Pedestrian Flow Models with Slowdown Interactions*, submitted to Mathematical Models and Methods in Applied Sciences.
2. C. Hauck, Y. Sun, and I. Timofeyev, *On Cellular Automata Models of Traffic Flow with Look Ahead Potential*, submitted to Phys. Rev. E.

3.6 Methodology: data assimilation

Leader: Adrian Sandu (Virginia Tech)

3.6.1 Participants:

Ralph Smith (NCSU), Jenny Brynjarsdottir (SAMSI), X. Han (Auburn), A. O'Hagan (Sheffield), N. Burch (SAMSI)

3.6.2 Research activities and results

Data assimilation is the process of fusing information from imperfect models, noisy measurements, and priors, to produce an optimal representation of the state of a physical system. Data assimilation can be interpreted and carried out in a Bayesian framework. Practical methods for large-scale systems include suboptimal and the ensemble Kalman filter approaches, optimal interpolation, and three and four dimensional variational methods. The goals of the working group were to study emerging problems that included, but were not limited to: new computational algorithms, modeling of model errors, impact of observations, and quantification of posterior uncertainties. The group meetings hosted several tutorial presentations: introduction to data assimilation (A. Sandu), aspects of uncertainty in parameter estimation (R. Smith), introduction to particle filters (J. Brynjarsdottir and X. Han), model calibration and model discrepancy (A. O'Hagan). Beyond their informative role, the tutorials helped to shape a common language among group members coming from various backgrounds. They also provided a nice avenue to start meaningful discussions among group members and to identify problems of interest.

The group gathered and shared code for several test problems. They included the Lorenz 96 one- and two- scale models, and a quasi-geostrophic code from NERSC, Norway. The shared software also included an implementation of the ensemble Kalman filter.

The group identified, and explored, research directions of common interest for several members of the group. One was the problem of understanding and estimating model errors, roughly defined as systematic differences between model predictions and reality due to numerical errors and missing physics. (The model error was referred to as model discrepancy in A. O'Hagan's tutorial talk). A numerical investigation of the effects of model errors on data assimilation was carried out and presented to the group by N. Burch: the one scale Lorenz 96 model was constrained to calibrate a "reality" simulated by the two scale Lorenz model. This setting allows exploring the effects of missing physics in a model. Discussions with Dr. Todling of NASA GMAO were relevant in this regard.

The second topic of interest was particle filters. The group started with a test implementation of several particle filters on the Lorenz 96 model. The group additionally pursued the development of particle filters and data assimilation techniques for dynamic material and structural models using various time-dependent data sets.

The third topic of interest was goal-oriented data assimilation (in conjunction with D. Estep's work): instead of attempting to make the model state reproduce reality in a pointwise manner, the focus was on matching functionals of the state.

3.7 Methodology: simulation of rare events

Leader: Jim Nolen (Duke)

3.7.1 Participants

James Nolen (Duke), Chia Lee (SAMSI), Shankar Bhamidi (UNC), Richard Moore (NJIT), Esteban Chavez (Duke), Laura Swiler (Sandia), Jose Blanchet (Columbia Univ), Jan Hannig (UNC), John Jakeman (Sandia), Sergio Almeda (UNC), Jing Li (Purdue), Ilya Timofeyev (Univ. of Houston), Dan Cooley (Colorado State)

3.7.2 Research activities and results

The group studied algorithms and theory related to the simulation of rare events in probabilistic models. The group was interested in problems where there are many stochastic degrees of freedom in the input variables, yet the relation between these inputs and the quantity of interest is highly nonlinear. Even if all statistics of the input are known, estimating the probability of a rare event is very challenging. In addition to estimating probabilities, a successful theory and algorithm should give insight into the mechanism by which the rare event is most likely to occur.

Group members proposed model problems and topics for discussion. In particular, the group discussed problems in theory of large deviations, random graphs and networks, noisy lasers and optical communication systems, stochastic dynamical systems arising from climate models, and some engineering test problems from Sandia National Laboratory. Specific algorithms discussed include importance sampling, particle splitting methods, and cross entropy methods. Having surveyed several methods and problems, the group got to the stage of choosing specific problems on which to focus. Their goal was to better understand simulation techniques based on large deviation theory, importance sampling, and particle methods, and to apply these ideas to novel problems. In the end, the group made most progress on two problems. The first pertained to random graph models and the problem of estimating the probability of observing unusually large triangle counts in the graph (Jim Nolen, Chia Lee, Shankar Bhamidi, and Jan Hannig). The second problem pertained to jitter in noisy mode-locked lasers (Richard Moore).

3.7.3 Working papers and publications

1. S. Bhamidi, J. Hannig, C. Lee, J. Nolen, The importance sampling technique for understanding rare events in Erdös-Rényi random graphs, submitted.

3.8 Climate: data assimilation in IPCC level models

Leader: Chris Jones (UNC)

3.8.1 Participants

Chris Jones (UNC), Richard Moore (NJIT), Doug Nychka (NCAR), and others.

3.8.2 Research activities and results

Data assimilation is the mathematical area that aims to render an optimal state estimate, given both prognostic model output and observational data. As such, it involves the propagation of error under models where the error is accumulated from both model and data. In one natural formulation, this is exactly what uncertainty quantification is about. This working group was run jointly with MCRN (Mathematics and Climate Research Network). The working group named itself "DA in IPCC Level Models." The objective of the group was to assess the extent to which data assimilation is currently being used in models that are used as a basis for the IPCC Assessment Reports. Because of familiarity, research contacts and accessibility to information, the group focused on the NCAR Community Earth System Model. The group met using Webex every Thursday afternoon at 2pm for the entire program. The NCAR model is organized around working groups that, roughly, correspond to different components of the model (land ice, ocean atmosphere etc.) Each week, one group member represented a specific "working group" and gave an assessment of the use of DA in that part of the model. In some cases, the group had to dig deeply to expose any use of DA. The group's main conclusion was that DA is not being used systematically. Efforts at NCAR itself, mainly involving the Data Assimilation Research Testbed, led by Jeff Anderson, are growing. In terms of which contributions to the next (fifth) assessment report that will involve DA at some level, however, the group could only definitively point to the ocean model (POP). The working group also considered ways in which this might change, both in a prescriptive manner (how should DA be involved?) and as a prediction (how do we think DA will be used in the future by climate modellers?)

Chris Jones reinforced collaborations with two former SAMSI postdocs: Elaine Spiller (Marquette), who visited SAMSI a number of times during the program, and Amit Apte (TIFR-CAM, Bangalore.) Jones also mentored Kody Law, a visiting postdoc of Andrew Stuart (Warwick). Dr. Law was partly supported by SAMSI during his visit. Three of Jones's students were involved in SAMSI activities: Andrew Roberts and Bevin Maultsby were both Graduate Fellows from UNC, and each contributed to various SAMSI activities and were both involved in the working group. In addition, Jones's student Damon McDougall from Warwick also visited for most of the program and contributed in various ways. McDougall's research and that of another postdoc of Jones, Naratip Santitissekadorn, involves the assimilation of data that come from Lagrangian instruments in the ocean. Both projects benefitted from the SAMSI program in helping to frame the issues in terms of the wider context of uncertainty quantification. Various contacts were made by each of them with participants in the SAMSI program.

3.9 Climate: parallel computing issues

Leader: Serge Guillas (University College London)

3.9.1 Participants

Serge Guillas (University College London), Charles Jackson (University of Texas at Austin), Lulu Kang (Illinois Institute of Technology), Christopher Kuster (Caroll University), Vivien Mallet (INRIA-Paris), Robin Tokmakian (Naval Postgraduate School), Nina Glover (University College London)

3.9.2 Research activities and results

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. 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 simulations — not only by drastically reducing simulation time, thus permitting solution of larger and larger problems, but also by allowing simulation of previously intractable problems. Utilization of such computer models (also called simulators) of processes requires addressing Uncertainty Quantication, a greatly expanding field focusing on at least the following issues:

- Models are invariably not a completely accurate reflection of reality; this bias or discrepancy needs to be acknowledged in analysis and prediction.
- The models are often so expensive to run that model outputs may be available for only a few inputs; for practical purposes the model predictions are thus unknown at other inputs.
- There are often uncertainties in the initial conditions for the models.
- The models typically have numerous unknown parameters.
- The models may have stochastic features.
- Predictions combining model output and noisy data (data assimilation) is often required.

The intellectual content of computational modeling comes from a variety of disciplines, including mathematics, statistics, probability, operations research, and computer science. Despite a wide diversity in methodology and applications, there are a variety of common challenges in developing, evaluating and using complex computer models of processes.

The area of climate modeling is a quintessential field of application for UQ. In fact, a sizable part of the research done in the quantification of uncertainty in computer models has been driven by the pressing needs of climate modelers. Climate models are computer codes based on physical principles that simulates the complex interactions between the many parts of the Earth system such as atmosphere and oceans. A significant issue of interest is, for instance, the problem of regional climate as the global models cannot on their own give enough information at the regional/local scales. Various strategies (downscaling and upscaling) can be considered whereby various combinations of global and regional models are combined to gain information on model uncertainty. Other issues include: development of reliable atmospheric and ocean models and interface between the two, regional and local risk assessment from models, paleoclimate models and how to best deal with models that have huge uncertainties and biases.

This group targeted the NCAR Community Earth System Model (CESM). CESM is a coupled climate model for simulating the earth's climate. Composed of four separate models simultaneously simulating the earth's atmosphere, ocean, land surface and sea-ice, and one central coupler component, the CESM allows researchers to conduct fundamental research into the earth's past, present and future climate states. It is running at UCL on the supercomputer Legion, which was available for the effort. Obviously, these techniques can be employed in similar computer models and some researchers in the team (Mallet for chemistry-transport models, Sarri for tsunami models) contributed their own experience on UQ software and techniques for such models. Some members of the team (Jackson, Tokmakian, and researchers at Lawrence Livermore National Labs) have independently investigated the emulation and uncertainty analysis of the atmosphere and the ocean. They presented and discussed this research in the SAMSI working group "Parallel computing issues for UQ in climate models." Jackson and collaborators started an effort in this direction. They developed the Model Ensemble Control System (MECS) for the climate model CESM. It is a Python set of scripts that manages a large number of potentially logically linked jobs on High Performance Computing resources. MECS is a Python-based suite of functions that tackles design of experiments for climate models, and the UQ analysis, called here the "UQ pipeline." It was a starting block for the group's effort. While MECS is currently agnostic about sampling strategies, it can be coupled with a parallel version of a Markov Chain Monte Carlo called MVFSA (e.g. Villagran et al. (2008)). There is also a need to cope in the future with incomplete runs and on-the-fly assignment of parameter sets contingent on prior results. However, the software engineering, machine learning and statistical underpinnings require the joint effort of several experts. This is where the Harvest programme can yield enormous benefits. To improve each simulator, calibration aims to find the optimal parameters that will give the best outputs from simulator with respect to a certain measure.

A representation of model bias and uncertainty based on Kennedy and O'Hagan (2001) can help derive the best values of these parameters along with corresponding uncertainties using Gaussian Processes (GPs), see e.g. Rasmussen and Williams (2006). GPs are random interpolators that represent the simulator as finite-dimensional Gaussian vectors indexed by the inputs. GPs naturally allow for successful queries of the input parameters together with the incorporation of uncertainty quantification. Bayesian approaches are often preferred to maximum likelihood in this context because these methods allow scientific experts to include judgments about the parameter values as prior knowledge. Additionally, Bayesian methods enable better and faster calibration because the exploration of complex multivariate distributions is made possible. The type of representation of the outputs (e.g. in terms of Empirical Orthogonal Functions or summary statistics) is likely to yield potential ground-breaking knowledge about the uncertainties related to key inputs of climate models.

One obstacle to fully Bayesian implementations of the calibration of computer models is that MCMC algorithms can become slow. Manifold Metropolis Adjusted Langevin and Riemann Manifold Hamiltonian Monte Carlo algorithms (MMALA and RMHMC) (Girolami and Calderhead, 2011) potentially enable very fast computations. These methods allow the geometry of the distributions to be used in order to sample from the relevant regions. In the calibration of simulators, the number of parameters can be high, and more importantly the response is high dimensional since environmental models outputs are multivariate spatial (3-D) time series. These approaches will allow further parallelization on large clusters, saving more time and enabling more advanced studies.

An emulator is a simple statistical model that approximates a simulator. Given some inputs x, the simulator output is given by $y = f(x)$ and the emulator is denoted by $\hat{f}(x)$, which indicates that it is an approximation of the simulator. In most cases, running simulators is very time and resource consuming, so one can only afford a limited number of runs. The use of emulators comes as a solution to this problem, since emulators run almost instantaneously. However, due to the fact that they are approximations of the computer model, some uncertainty is introduced by using them. The amount of this uncertainty can be estimated since they can make probabilistic predictions of the output that the simulator would produce if it was exercised over certain regions of the input space. Therefore, the main use of statistical emulators is for fast predictions of the simulator output.

The most standard emulators use Gaussian Processes (GP). However, GPs suffer from the fact that the covariance matrix needs to be inverted to use the kriging algorithm. This brings: (1) numerical instability (the inversion of this matrix can be a challenge when inputs are close); (2) complexity (as inversions require

$O(n^3)$ operations where n is here the size of the sample of model runs); (3) difficulty in accommodating nonstationarity. Rougier (2008) and Rougier *et al.* (2009) showed that Outer Product Emulators (OPEs) can cut down the computations, but require strong assumptions on the separability of the covariance between inputs and outputs, as well as knowledge of the shape of the output functions. Joseph and Kang (2011) developed faster and flexible emulators: the inverse distance weighting (IDW) combined with sparse regression (e.g. LARS or LASSO) and now are investigating nonparametric kernel regression (iterated) that provide either much faster or more precise emulators than GPs. For moderately large dimensions of inputs and number/dimension of observations, it will be interesting to implement these models and fasten them through the use of parallelization. L. Kang was part of the team. She presented this approach, released the code and guided the group through it. Its use for complex models such as climate models would be a novelty. Furthermore, the simulators need to be run at specific combinations of inputs to get the most information out of the runs. Questions concerning design of experiments for high performance computing are: (1) how to take make use of the fact that failed runs in a design need to be taken into account in a sequential manner; (2) how to design an experiment that thoroughly explores the parameter space but also optimize the structure of the cluster (e.g. cores in nodes that share memory); (3) how to design sequential experiments efficiently using machine learning algorithms to select points, e.g. advanced Learning (Gramacy and Lee, 2009). The group coded novel approaches to the design of computer experiments that allow for the joint optimization of fast computational resources and enough statistically relevant information. The ground-breaking idea, which to our knowledge was not previously implemented on simulators, is to replace some costly (in terms of computation time) parts of the simulator by an emulator, as against emulating the entire model as a whole. For instance, one idea is to explore how to use emulators to represent some of the ocean processes to reduce computational time (a full embedding of an emulator of parts of the model is too challenging to address). The emulator of some outputs of the ocean was built by one member of the team (Tokmakian). This requires a lot of expertise and needs to be tried on toy examples first.

The working group put together a proposal to the European PASCAL network `http://www.pascal-network.org`. **This proposal has been fully funded** and will allow the group to not only continue its work through meetings both in the UK and the US but also allow the pursuit of the scientific part of the project. Indeed, this is an open-source project based on open softwares: Python, R, CESM. The supercomputer Legion will be freely available for our project up to the CPU time constraints in the Mathematical Sciences consortium to test ideas, and with a specific budget within the funded proposal for ring-fenced computer time (around 10,000 CPU hours). This specific additional budget will enable the completion of large scale efforts with no waiting time in the queue, and the possibility of utilizing Legion beyond current restricted practice for a Grand Challenge. For that goal, the group is able to use the Legion supercomputer at UCL. The group plans a training session on Message Passing Interface (MPI) to be able to fully implement our techniques in large clusters with in Python, C++, Fortran and R.

3.10 Climate: statistics of extremes

Overall leader: Dan Cooley (Colorado State)

The statistics of extremes working group formed primarily out of interest in extremes expressed by participants in both the Climate Change Workshop and the UQ Methodology Workshop. The working group had 34 enrolled members. Due to the group's large size and diverse extremes-related interests, the working group formed a number of subgroups each of which had a particular focus.

3.10.1 Change-points in Extremes

Leader: Debbie Dupuis (HEC Montreal)

Members: Ying Sun (SAMSI postdoc), Alex Chen (SAMSI postdoc), Judy Wang (NCSU), Fabrizio Ruggeri (CNR-IMATI Milano)

The group considered a random process X and examined whether the distribution of X changed at some point. Although the change in the distribution can be exhibited in many different ways, much of the literature looks for changes in location of scale when carrying out a change-point analysis. The objective of this working group was to look for changes in the extremes of the distribution. More specifically, the first goal was to do a complete literature review of change-points in extremes. Second, the group classified all suggested approaches according to ease of implementation and statistical optimality. Their principal motivation was the analysis of global climate model data and this means that only fast detection methods can be considered. Their third objective was thus to propose a change-point detection method that offered some compromise between implementation (computational speed) and statistical optimality (e.g., power of detection), assessing what is lost along the way. A possible outcome could be a survey type paper or an application-oriented paper, more tightly focused on historical and/or climate model temperature data, for example.

3.10.2 Spatial Extremes

Leader: Dan Cooley (SAMSI visiting scholar, Colorado State)

Members: Elizabeth Mannshardt (Ohio State), Rob Erhardt (PhD student, UNC), Jessi Cisewski (PhD student, UNC), Bernard Omolo (University of South Carolina Upstate), Soyoung Jeon (PhD student, UNC), Ying Sun (SAMSI postdoc).

The initial focus of this group was to write a review paper on spatial extremes (Cooley *et al.*, 2012). The full subgroup took on this task as a way for everyone to learn the current state-of-the-art in modeling/describing spatial extreme data and to learn what are open questions in the field. The remainder of the group's emphasis was to conduct new theoretical or methodological research in spatial extremes. Topics that were discussed included downscaling of gridded climate model output to point-referenced data, appropriate models for climate model data of high spatial dimension, and appropriate representations and measures for spatial dependence .

3.10.3 Hurricanes

Leader: Sneh Gulati (Florida International University)

Members: Elizabeth Mannshardt (Ohio State), Harold Brooks (NOAA), Soyoung Jeon (PhD Student, UNC), Judy Wang (NCSU), Mi Zhou (PhD Student, NCSU), Alex Jarman (London School of Economics).

The hurricane group had about six active members and four additional members who were interested but could attend every meeting. The first working meeting of the group was held on October 28, 2011 where Sneh Gulati presented her work on Uncertainty Analysis and Probable Maximum Loss Calculations in the Public Hurricane Loss Model. The following week, Judy Wang presented her work on Quantal Analysis. Other topics considered by the group were: study of a public hurricane database (HURDAT) to see if hurricane activity is increasing, and better methods of estimating extreme hurricane events.

3.10.4 UQ in Geotechnical Failures

Leader: Eugene Morgan (SAMSI New Researcher Fellow, Tufts University)

Members: Ilenia Battiato (Max Planck Institute), Fabrizio Ruggeri (CNR-IMATI Milano), John Jakeman (SAMSI New Researcher Fellow, Purdue)

This sub-working group focused on sediment stability, both at large (regional) and small (grain size) scales. The group was interested in quantifying uncertainty associated with modeling when and where failure will occur. At a large scale, the interestxlies in mapping landslide hazard (probability of slope failure), while at a small scale the interest is in how hydrodynamic forces affect the stability of granulate material. These applications incorporate a variety of fields, including reliability-based techniques, spatial statistics, and extreme value theory. The group's work primarily consisted of formalizing the regional landslide hazard mapping problem as a Bayesian hierarchical model that treats slope stability as a Gaussian process. This subgroup also participated in the Geosciences working group meetings.

3.10.5 Simulation of Rare Events Crossover Group

Leader: Shankar Bhamidi (UNC)

Members: Dan Cooley (SAMSI visitor, Colorado State) and others

A subgroup of members was interested in exploring the intersection between statistical theory and methods for extremes and simulation methods for rare events. Group members attended both the statistics of extremes working group meetings and the simulation of rare events group meetings. Bhamidi gave presentations in both groups regarding a possible project involving social networks, graph theory, and possibly extremal dependence. The last few years have witnessed an explosion in the amount of information collected on social networks such as Faceboook, Twitter and the blogosphere. One of the central questions in this area is an understanding of models of diffusion of information through such networks as well as reconstruction of networks of influence through partial observations. The group aimed to develop and use models for extremal dependence between times of spread of information between different vertices in the network to reconstruct such networks (Network Tomography). They also attempted to develop comprehensive statistical methodology with regards to change point detection in social networks with power law degree distribution via a study of the evolution of the extremes of the distribution. The group tried to validate these techniques using the data presently being collected by UNC Chapel Hill.

3.10.6 Issues in Extremal Dependence

Leader: Robert Wolpert (Duke)

Members: Dan Cooley (SAMSI visitor, Colorado State), Jianyu Wang (PhD Student, Duke), Grant Weller (SAMSI visitor, PhD Student, Colorado State), Soyoung Jeon (PhD Student, UNC), Richard Smith (SAMSI, UNC)

This group explored current issues in extremal dependence. In particular, the group explored the current state-of-the-art in describing and modeling asymptotic independence as well as current dependence metrics in both the asymptotic dependent and independent cases. The group looked at the hidden regular variation work of Ledford and Tawn as that of Resnick. The group also explored the work of Nolan on the connection between α-stable and max-stable distributions. Some of our subgroup's Goals were to:

- Build testbed multivariate extreme examples for which inference is tractable, and considering their theoretical properties
- Think harder about the "in-between" range for dependent but extremally independent random variables (e.g. correlated Gaussians), hoping to extend the work of Ledford & Tawn about dependence within joint tail regions, of Resnick on hidden regular variation, etc.
- Explore various measures of dependence in the literature; relate them to spectral measures H(dw) for asymptotically dependent random variables, and seek to relate and extend them.
- Think harder about threshold models for multivariate extremes

3.10.7 Working papers and publications

1. D. Cooley, J. Cisewski, R. Erhardt, S. Jeon, E. Mannshardt, B. Omolo and Y. Sun, *A Survey of Spatial Extremes: Measuring Spatial Dependence and Modeling Spatial Effects*, RevStat, 2012, 10, pp. 135–165.

2. G.B. Weller and D. Cooley, *An Alternative Characterization of Hidden Regular Variation in Joint Tail Modeling*, submitted.

3. G.B. Weller and D. Cooley, *A Model for Extremes on a Regular Lattice and Inference based Hidden Regular Variation*, in preparation.

4. E. Morgan, F. Ruggeri, *Bayesian hierarchical modeling for landslide susceptibility mapping*, in preparation.

5. F. Ruggeri, E. Spiller, R. Wolpert, *Models for Pyroclastic Flow volumes and directions, based on data and experience from Sufriere Hills Volcano on Monteserrat*, in preparation.

6. Erhardt, R. and Smith, R.L., Approximate Bayesian computing for spatial extremes. *Computational Statistics and Data Analysis* **56**, 1468–1481, 2012.

7. S. Jeon and R.L. Smith, Dependence structure of spatial extremes using a threshold approach. *Extremes*, under revision.

3.11 Engineering: nuclear energy

Leader: Don Estep (Colorado State), Dongbin Xiu (Purdue)

3.11.1 Participants

Anter El Azab (Purdue), Peter Polyakov (Wyoming), Michael Pernice (Idaho National Laboratory), Troy Butler (UT Austin), Will Newton (CSU), Simon Tavener (CSU), Yushu Yang (SAMSI, Purdue), Russ Johnson (Wyoming), Hany Abdel-Khalik (NCSU), several grad students from NE/NCSU.

3.11.2 Research activities and results

The NE working group was focused on developing methods for analyzing multiphysics, multiscale models used in Nuclear Engineering, primarily of the physical properties of new fuel types under typical stresses found in a reactor, though the group also explored other multiscale systems involved in nuclear reactors. Nuclear fuels exist under extreme environmental conditions, e.g. pressure, stress, heat, that cause changes in the microstructure of the material, e.g. growth of voids and cracks, which then affect its physical structure and energy producing properties. It is too expensive to develop new fuels or new reactors using a purely experimental approach. So, computational simulation is centrally important to obtain further progress in nuclear engineering. However, the construction, analysis, simulation and UQ (which is absolutely required in the NE industry) for the complex multiscale models involved in NE are currently at a relatively undeveloped stage. The NE working group conducted some activities intended to reveal common mathematical problems across various NE application domains. It also focused on some particular research projects motivated by the development of new fuel types. In this working group, emphasis was on algorithmic development and construction of C++ code that can be imported into engineering application codes that are used to simulate nuclear fuels. Testing of new algorithms has to be carried out before investing time into the analysis.

1. Mesoscale models of microscale structure in nuclear fuels. Currently, there is a great deal of interest in developing mesoscale models based on the Cahn-Hilliard equations constructed through a phase field model for describing the microscale development of cracks and voids in nuclear fuel material. Part of the working group concentrated on the development of the models themselves, and how these models might be validated by experiment and existing data bases. Another part of the group focused on the use of polynomial chaos techniques for representing the solutions of the stochastic models that result when physical values are used in the parameters of the model. Since the Cahn-Hilliard and Allen-Cahn equations are nonlinear, time evolution problems, theory and application of PC techniques is currently lacking. We also explored the construction of Allen-Cahn equation models to mimic the key behavior of the Cahn-Hilliard-based models, but with reduced mathematical complexity. Cahn-Hilliard models present a number of technical challenges stemming from the fourth order operator in space.

2. Polynomial chaos representation of stochastic output of ODEs. The working group developed and analyzed polynomial chaos techniques for representing the behavior of stochastic solutions of stochastic evolution ODEs. Most of the literature on PC techniques involves stationary problems, e.g. elliptic equations.

3. Coupling of mesoscale and macroscale models. There are a number of ad hoc techniques that are used to couple microscale and mesoscale model outputs into engineering or macroscale equations throughout multiscale engineering applications. A very popular technique, which is currently used in modeling of nuclear fuels, is to evaluate the mesoscale model at each quadrature point in a finite element discretization of the macroscale equations. Unfortunately, this is extremely problematic from a mathematical point of view as it imparts arbitrary roughness to the behavior of the parameter values on the macroscale. The problem is that this amounts to sampling different parameter surfaces at each quadrature point. The working group pursued the construction of an intermediate scale representation of the mesoscale model that is used to transfer an entire stochastic parameter surface constructed by evaluation of multiple instances of the mesoscale model.

4. Stochastic representation of fine scale quantities in multiphysics simulations. In many cases of multiscale multiphysics systems, the fine scale (micro or meso-scale) information is presented as a stochastic quantity, e.g. as a random field, which is populated using some combination of data and simulations of a microscale model. This is somewhat curious from a mathematical point of view because the microscale model is often deterministic. The motivation is likely some combination of intuition that the microscale behavior has a stochastic nature and an ad hoc attempt to deal with the fact that the microscale behavior is woefully underresolved (because of computational limitations). The multiphysics working group developed an innovative approach to create a stochastic representation of a deterministic microscale model that uses stochasticity to handle underresolution in a systematic and quantifiable way. The probability distributions in the stochastic representation reflect the underlying behavior of the multiscale model. Test simulations are extremely promising. The group investigated the theoretical analysis of this approach. A successful conclusion would provide a fully rigorous explanation of this widely used ad hoc engineering technique.

3.11.3 Working papers and publications

1. Peter Polyakov, Keith Lenth, Victor Ginting, Russ Johnson, Polynomial chaos representations of stochastic Cahn-Hillard equations, preprint

2. Yushu Yang, Akil Narayan, Dongbin Xiu, On numerical strategy to decouple stochastic system of equations, preprint

3. C++ Code for coupling mesocale and macroscale models through an intermediate scale for representation of stochastic parameters computed using an iterative approach

4. C++ code changes to the Idaho National Laboratory codes MOOSE and BISON used for multiscale simulations of nuclear fuel behavior

5. C++ code for PC representation of stochastic Cahn-Hilliard models

3.12 Engineering: renewable energy

Leader: Ralph Smith (NCSU)

3.12.1 Participants

Nate Burch (SAMSI), Will Cousins (NCSU), Ryan Elmore (NREL), Mami Tonoe (NCSU)

3.12.2 Research activities and results

This working group focused on the investigation of nonintrusive techniques to construct reduced-order models when the parameter dimensionality is high (e.g., greater than 50). The focus on nonintrusive methods was motivated by the fact that in many cases, legacy codes or executable files constitute the sole source for implementation. HDMR methods are a class of techniques that have proven highly effective for model reduction when high-order (e.g., higher than 2nd order) parameter interactions are negligible as is often the case for high-dimensional problems with smooth dynamics. The two primary classes of techniques are ANOVA-HDMR and CUT-HDMR. The former is equivalent to certain stochastic collocation techniques and hence it relies on sparse grid collocation points for high dimensional parameter spaces. CUT-HDMR utilizes function evaluations at only a few points so it is highly efficient to implement but its accuracy is also highly dependent on the regularity of the underlying function.

Based on research initiated during the program, the group constructed CUT-HDMR techniques for neutron transport equations arising in the analysis of light water nuclear reactors. This was investigated in the context of the DOE Consortium for Advanced Simulation of Light-Water Reactors (CASL). Initial investigations demonstrated a significant improvement in efficiency while retaining reasonable accuracy. The group extended these techniques to provide models that can be calibrated using physical data for use in simulation packages for nuclear reactor design.

3.12.3 Working papers and publications

1. Z. Hu, R.C. Smith, J. Willert, C.T. Kelley, J. Hite and H.S. Abdel-Khalik, "High-dimensional model reduction techniques for the neutron transport equation," Chapter in the September, 2012, DOE CASL Milestone report.

2. Z. Hu, R.C. Smith, J. Willert and C.T. Kelley, "High-dimensional model reduction (HDMR) techniques for the neutron transport equation," paper in progress.