Speaker Abstracts (Tentative)
CompMod Transition Workshop
May 14-16, 2007

Ariel Cintrón-Arias
SAMSI and North Carolina State University
Center for Research in Scientific Computation
ariel@samsi.info

"Estimation of Seasonal Reproductive Numbers and Transmission Rates"

The effective reproductive number is a fundamental parameter in the study of transmission dynamics of infectious diseases. This time-dependent parameter $\mathcal{R_t}$ is an ongoing measure of transmission. An analysis of seasonal influenza epidemics follows from the estimation of both $\mathcal{R_t}$ while using longitudinal incidence data in the United States. We also study seasonal epidemics using a differential equation model with periodic-time dependent transmission rates. Observed year to year variability in outbreak incidence can be explained by chaotic behaviour of the solutions of such models or by additional variability in transmission rate parameters in time. To gain insight into possible causes for year to year variability, we investigate the temporal change of the transmission parameter required for a model without chaotic behaviour that is required to reproduce measured data.

Susie Bayarri
susie.bayarri@uv.es
University of Valencia and SAMSI
Department of Statistics and Operations Research

"Some Intriguing Methodological Issues when Statistically Analyzing Computer Model Data"

The special methodological characteristics of (Bayesian) calibration and validation of complex computer models (with, in particular highly confounded parameters), result in very tricky practical MCMC implementations, which need to be addressed with exquisite care. In particular, numerical problems can arise if the statistical model is not entirely satisfactory. We recommend use of an approximation to the full Bayesian approach which treats some components of the overall model separately; we call this approach "modularization". We'll motivate the modular approach and demonstrate its efficacy. Other problems with MCMC appear when full Gibbs sampling is attempted in the expanded space of unknown quantities; we show that "extending the conversation" or explicitly including latent processes can sometimes be dangerous in these problems. Another methodological intriguing aspect in the analysis of computer models is treatment of bias (or model discrepancy). The usual approach of an additive unstructured term is not always satisfactory; an alternative approach (adding an input dependent discrepancy term to some parameters) is presented in Peter Reichert's talk, and is a very promising venue. However, when the computer model is very complex or the code can not be accessed this alternative approach is not directly feasible, but it inspired still another possibility which we are exploring and which involves the model derivatives as part of

the bias term.

Michael Breen
U.S. Environmental Protection Agency
breen.michael@epa.gov

"Systems Biology Models: Challenges and Applications"

Systems biology is a new field that aims at system-level understanding of biological systems. Interests in systems biology has been greatly stimulated by advances in molecular biology that has improved our understanding biological systems by the identification of genes and the function of their products (e.g. proteins, metabolites), which are the components of biological systems. The next challenge is to understand the biological system behaviors from the dynamic interactions of the components revealed by molecular biology. Mathematical systems biology models are necessary to quantitatively describe the dynamic biological processes and the effects of perturbations on biological systems. The development of systems biology models is an iterative process that requires a close collaboration between experimenters and mathematical modelers. At the U.S. Environmental Protection Agency, we are applying a systems biology approach to predict human health and ecological outcomes from chemical exposures. Our research goal is to better understand the dose response behaviors of endocrine active chemicals. Our approach is to develop computational systems biology models that describe the biological perturbations at the biochemical level and integrate information towards higher levels of biological organization. This approach will ultimately enable predictions of dose-response behavior at the organismal level. In this talk, I will discuss the challenges of developing of systems biology models, and the results from a mechanistic mathematical model of steroid biosynthesis to predict the biochemical responses to endocrine active chemicals in fish ovaries.

James Clark
Duke University
Department of Biology
jimclark@duke.edu

Title and abstract: TBD

Pierre Gremaud
North Carolina State University
Department of Mathematics
gremaud@ncsu.edu

"Calibration of a Numerical Model for Cerebral Blood Flow "

The goal of this work is the study of the influence of the topology and material properties of the Circle of Willis on cerebral blood flow. The Circle of Willis is a network of 16 interconnected vessel segments whose structure may change from patient to patient. To this end, it is important to maintain the geometrical complexity of the model to a minimum.

Our approach shares with others the use of cross-sectional averages. This prevents the resolution of the detailed features of the flow but allows for high flexibility of the computational model. Additional features of the models include viscoelastic effects, effects due gravity, various flow profiles and non-Newtonian effects.

The smooth character of the solutions allows the use of high accuracy pseudo-spectral discretization methods. The numerical model is calibrated against velocity flow measurements provided by the group of Dr. Novak (Harvard) through ensemble Kalman filtering. Preliminary results establishing not only the feasibility of the proposed approach but also its efficiency will be presented.

Joint work with K. DeVault, M. Olufsen and G. Vernieres.

Serge Guillas
Georgia Tech
Department of Mathematics
guillas@math.gatech.edu

"Calibration of an Air Quality Model"

We calibrate a regional chemical transport model (RAQAST) for surface ozone. The controllable input variables describe the meteorology. The calibration parameters are the diffusion (corresponding to boundary layer height), nitrogen oxides, anthropogenic VOC, biogenic isoprene, and effects of clouds on photolysis. A set of 100 runs of RAQAST enable us to find the best parameters. Time series and spatial approaches are discussed. We use a combination of Gaussian stochastic processes and MCMC techniques to find the best tuning parameters for forecasting purposes.

Dave Higdon
Los Alamos National Laboratory
Statistics Group
dhigdon@lanl.gov

"Gaussian Process Models for Finding Structure in High Dimensional Spaces"

Gaussian process models have proven to be very useful in modeling computer simulation output. This is because many computer codes are essentially noiseless and respond very smoothly to changes in input settings. In a typical setting, the simulation output is a function of a p-dimensional input vector x. In some applications, p may be as large as 60, however for the application of interest, the output is typically affected by a much smaller subset of these p inputs. After a fixed number of simulations are carried out, a GP model can be used to predict the simulation output at untried settings. When the number of simulations carried out to train the GP model is much over 1000, standard modeling and fitting approaches become computationally burdensome. In this talk, we describe strategies we've found useful for fitting such models (with large p) when the number of simulation runs is large.

Leanna House
Durham University
Department of Mathematical Sciences
leanna.house@durham.ac.uk

"Better Judgments through Multiple Model Ensembles"

Often, several computer models exist that attempt to estimate features of the same physical process. Selecting only the tuned results from the varying models, we combine the outputs into one vector to which we refer as a multiple model ensemble (MME). The goal for this paper is to use the MME sensibly to specify data-informed prior quantities for an uncertainty analysis of an additional model which is not contained within the MME. Specifically, we propose a relatively straightforward method to calculate the prior expectation and variance for the squared discrepancy between the additional model and the underlying true, physical system. The prior quantities are constructed so that they may be incorporated into a formal, computer model uncertainty analysis and/or used to summarise a priori two fundamental sources of error in computer modeling: global and model variation. To develop our data-informed priors, we consider the tuned results within the MME as well as from the model of interest to be second-order exchangeable and apply Bayes linear methods to improve expert judgments.

Authors: Michael Goldstein (1), Leanna House (1), Jonathan Rougier (2)
1. Department of Mathematical Sciences, Durham University, Durham, UK DH1 3LE
2. Department of Mathematics, Bristol University, Bristol, UK BS8 1TW

Cari Kaufman
cgk@samsi.info
SAMSI and NCAR

"State-Space Modelling of Soil Moisture"

Fei Liu
Duke University and SAMSI
Department of Statistics
fei@stat.duke.edu

"Bayesian Functional Data Analysis for Computer Model Validation"

Functional data analysis (FDA) -- inference on curves or functions -- has wide application in statistics. An example of considerable recent interest arises when considering computer models of processes; the output of such models is a function over the space of inputs of the computer model. The output is functional data in many contexts, such as when the output is a function of time, a surface, etc. A nonparametric Bayesian statistics approach, utilizing separable Gaussian Stochastic Process as the prior distribution for functions, is a natural choice for smooth functions in a manageable (time) dimension. However, direct use of separable Gaussian stochastic processes is inadequate for irregular functions, and can be computationally infeasible in high dimensional cases.

In this talk, we will develop and extend several Bayesian FDA approaches for high dimensional irregular functions in the context of computer model validation, tailored to interdisciplinary problems in engineering.

Chunsheng Ma
SAMSI and Wichita State University
Department of Mathematics and Statistics
cma@math.wichita.edu

"Space-Time Downscaling of Regional Ozone Forecasts with Nonseparable Covariance Models"

The Regional Air Quality forecAST (RAQAST) model is a regional chemistry and transport modeling system that provides 48-hour forecast of the concentrations of ozone and its precursors over the United States. Since the grid size is 70 by 70 km, forecasts can not be made at a local level. We use EPA stations from the Georgia and neighborhood area to downscale and improve local forecasts using RAQAST outputs, and analyse the discrepancies between local observations and model outputs in space and time based upon the assumption that the deficiencies of the model can be explained by meteorological variables. A new space-time covariance structure will be presented whose spatial index domain is R^d and temporal index domain is Z. The proposed covariance function is nonseparable, allows positive and negative correlations, possesses certain spatial smoothness, and the unexplained short term deficiencies are autoregressive in time.

Subbagging and tapering approaches will be used to estimate the model parameters.

(This is a joint work with Serge Guillas)

Bruce Pitman
University at Buffalo
Department of Mathematics
pitman@buffalo.edu

Title and abstract: TBD

Samantha Prins
James Madison University
Department of Mathematics and Statistics
prinssc@jmu.edu

"Bayesian Uncertainty Assessment for Multicompartment Deterministic Simulation Models"

I will present work my coauthors (A.E. Raftery & A. Cullen) and I have done on a special case of Bayesian melding to make inference from deterministic models while accounting for uncertainty in the inputs to the model. The method uses all available information,

based on both data and expert knowledge, and extends current methods of 'uncertainty analysis' by updating models using available data. We extend the methodology for use with sequential multicompartment models.

Peter Reichert
EAWAG and SAMSI
reichert@eawag.ch

"Identifying Causes of Bias in Model Predictions with Stochastic Time-Dependent Parameters"

A recently developed technique for identifying continuous-time, time-dependent, stochastic parameters of dynamic models is put into a framework for identifying the causes of bias in model results. The framework is illustrated with a didactical example and applied to a simple hydrological model. In this technique, state estimation of the time-dependent model parameter is combined with the estimation of (constant) model parameters and parameters of the stochastic process underlying the time-dependent parameter. In the framework for identifying and correcting model deficits, this technique is sequentially applied to the model parameters and the degree of bias reduction of model outputs is analyzed for each parameter. In a next step, the identified time dependence of the parameter is analyzed for correlation with external influence factors and model states. If significant relationships between the time-dependent parameter and influence factors or states can be derived, the deterministic model must be improved. Otherwise, or after improving the deterministic model in a first step, the description of uncertainty in model predictions can be improved by considering stochastic model parameters. In the didactical example, the technique identifies the cause of the bias as a deficit in the formulation of the deterministic model. The application of the framework to a simple 8-parameter conceptual hydrological model demonstrates that the parameters (including additional parameters for input modification) have significantly different potential for bias reduction. Only minor dependences of time dependent parameters on internal model states can be found. These are used to slightly improve the deterministic model. However, the major contribution to the residuals of the original model cannot be associated with deterministic relationships. Therefore, a stochastic parameter must be introduced to get a description of the data that does not violate the statistic model assumptions. Because of the high uncertainty in precipitation within the watershed (total amount and spatial distribution), we use this parameter for input uncertainty characterization.

Jonathan Rougier
SAMSI and University of Bristol, UK
J.C.Rougier@bristol.ac.uk

"Atmospheric Modelling: Joint SAMSI and NCAR Projects"

The SAMSI CompMod Program coincides with the NCAR-IMAGe Theme of the Year, Statistics for Numerical Models (http://www.image.ucar.edu/ThemeOfTheYear/), and five joint projects are underway. I outline these, and focus briefly on some of the more unusual and challenging aspects of one particular project: emulating the response of the

NCAR TIEGCM atmospheric model to tidal effects.

Thomas Santner
Ohio State University
Department of Statistics
santner.1@osu.edu

"Optimization of Engineering Computer Models"

One of the key objectives of Engineering Design is product optimization. This talk will describe the use of "improvement" to guide this process when such problems are solved using (calibrated) computer codes.

Elaine Spiller
SAMSI
spiller@math.duke.edu

"A Risk Map for Pyroclastic Flows: Combining Simulations and Data to Predict Rare Events"

We propose an emulator guided sampling method to calculate the probabilities of catastrophic flow events in populated regions. Flow volume is a primary input to the TITAN pyroclastic flow simulator, and we will discuss the use of the Wolpert-Lunagomez flow-volume model in conjunction with TITAN output in calculating these probabilities.

David Steinberg
Tel Aviv University
Department of Statistics and Operations Research
dms@post.tau.ac.il

"Computer Experiments with Complex Output"

Computer simulation of physical processes are a convenient substitute to laboratory experiments or field observation, when such a study is too complex, expensive or time consuming. Most work to date on computer experiments has focused on examples where the output from the computer simulator is scalar or of low dimension. This talk will discuss computer experiments in which the output is much more complex: a dense time trace, a spatially varying function or a temporal-spatial process.

We suggest two-stage methods for modeling and predicting such data, separating the model for time/space dependence from the model that relates the output to the explanatory variables used to define each of the simulator runs. Time dependence can be modeled by fitting known basis functions such as splines or wavelets. We also show how data derived basis functions can be effectively used. Such basis functions are generated by a functional cluster analysis of the data. We also suggest the idea of breaking up the first stage of the modeling into two parts, one focusing on the shape of the output and the other on the scale of the output. Several applications will be used to illustrate these methods: a simulation of response to chemotherapy, which yields the amount of cancer cells in a patient's body in response to different chemotherapy treatment protocols; a circadian rhythm simulation, showing the mRNA production and degradation throughout a sleep-wake cycle; and a geo-dynamic model for simulating ground motion resulting from a strong earthquake.

This is joint work with Sigal Levy.

Gentry White
North Carolina State University and SAMSI
Department of Statistics
white@stat.ncsu.edu

"A Kalman Filter Based Emulator for Engineering Models"

We present here a novel emulator for complex computer models based on the Kalman-Filter solution to a simplified state-space model and a Gaussian Process. The resulting emulator shows excellent promise for a variety of problems, specifically those that are easily characterised as state-space models. Methodological development is shown and results are demonstrated for a simple case. Future development of a more complex example from a hydrological groundwater runoff model is also discussed as well as the estimation of unknown parameters.

Darren Wilkinson
Newcastle University, UK
School of Mathematics & Statistics
wilkinson@samsi.info

"Parameter Estimation for Systems Biology Models"

This talk will consider the problem of the estimation of parameters (especially reaction rate constants) for systems biology models using time-course data. A general framework will be outlined for conducting Bayesian inference for such Markov process models using ideas of sequential and likelihood-free MCMC. Techniques for computationally expensive models using stochastic emulators will also be discussed.

Robert Wolpert
Duke University
Department of Statistics
wolpert@stat.duke.edu

"Pyroclastic Flow Volumes"

The frequency of high-volume pyroclastic flows at active volcano sites appears to fall off with increasing flow volume at a nearly linear rate on a log-log scale. We show how this suggests that the probability distribution for high-volume flows may be Pareto, leading to

the possibility of a non-negligible rate of catastrophically large flows.

We build a Bayesian hierarchical model for several different varieties of data from the Soufri`ere Hills Volcano on the Caribbean island of Montserrat, hoping to make predictions about future events.

This work is collaborative with Simon Lunagomez of Duke University ISDS and has benefited greatly from the active support of Bruce Pitman, Abani Patra, Elaine Spiller, and Eliza Calder of the University of Buffalo.

C. F. Jeff Wu
Georgia Institute of Technology
Industrial and Systems Engineering
jeffwu@isye.gatech.edu

"Some Recent Advances in Data Integration, Modeling and Design Strategies for Computer Experiments"

Standard practice in analyzing data from different types of experiments is to treat data from each type separately. By borrowing strength across multiple sources, an integrated analysis can produce better results. To this end, some Bayesian hierarchical Gaussian process models (BHGP) are proposed. The heterogeneity among different sources is accounted for by performing flexible location and scale adjustments. The approach tends to produce prediction closer to that from the high-accuracy experiment. The Bayesian computations are aided by the use of Markov chain Monte Carlo and Sample Average Approximation algorithms. The proposed method is illustrated with some real examples. The modeling problem also leads to the construction of a novel class of Latin hypercube designs to accommodate experiments at two levels of accuracy. A brief report will also be made on new work in the construction of Gaussian process models for both quantitative and qualitative factors and the associated issues of estimation.

(The reported work is based on joint papers with Professor Zhiguang Qian, U. of Wisconsin, Madison and Professor Huaiqing Wu, Iowa State U.)