samsi
NSF·Duke·NCSU·UNC

MUMS: Transition and SPUQ Workshops
May 14-17, 2019

SPEAKER TITLES/ABSTRACTS

Taylor Asher
University of North Carolina

"Surge Hazards WG Issues and Efforts"

Coastal flood hazards are amongst the deadliest and most costly natural disasters on the planet. Their underlying processes are, in some regards, in an advanced state of knowledge. Yet, the scale and variety of both causes and effects leave open many challenging questions. And our advanced state of knowledge has failed to realize a reduction in deaths or damages.

In this talk, I will address the underlying physical processes of coastal flooding and knowledge gaps, with an eye toward current research and the overarching issue of turning knowledge to action.

Evan Baker
University of Exeter

"Stochastic Simulators: Issues, Methods, Unresolved Questions"

Stochastic computer models, wherein reruns of the code with the exact same inputs does not yield the exact same output, are becoming increasingly commonplace. Effective statistical analysis of such output can be more challenging and more crucial than the statistical analysis of deterministic computer models. Even so, stochastic simulation is currently subject to less statistical research focus.

This talk will outline a review we have been working on, in which we aim to spur additional research on the topic â€“ introducing the objectives; outlining what statistical models currently exist; discussing how one can efficiently use such models to answer key questions about a stochastic computer model, and explaining what challenges currently still exist.

Pierre Barbillon
AgroParis Tech

"Embedding a Discrepancy in the Computer Model"

In spite of the increasing complexity of computer models, they may still fail to perfectly represent real systems. Many works advocate for modeling the reality as the computer model corrected by an additive discrepancy function. When the computer model is derived from ordinary differential equations (ODE), another approach consists in embedding the discrepancy model inside the computer model through a stochastic relaxation which turns the ODE into Stochastic Differential Equations (SDE). We will explore this idea on two computer models: a failure model and mass-spring-damper system.

"Variable Selection in the Discrepancy Function Associated with a Simulator"

A statistical model which links field experiments with a simulator usually embeds a discrepancy function. The discrepancy function models the systematic gap between the simulator and the real system. Analyzing the discrepancy should help to understand to what extent the simulator is reliable. In particular, determining that some variables are active or inert in the discrepancy function is of major interest since it indicates which variables are correctly modeled or not by the simulator. Therefore, this could give some leads to improve the simulator and help to determine if extrapolation is safe or not with respect to a specific input.

The discrepancy function is modeled as a Gaussian process which is parametrized as in Linkletter et al. (2006). This parametrization provides a simple distinction between active and inert variables. The variable selection is performed through a model selection where the models in competition differ on the prior distribution considered for the parameter associated with the variables in the Gaussian process.

We resort to computations of Bayes factors by using a bridge sampling to perform the model selection. Contrasted synthetic examples are considered to support the proposed technique.

Joint work with Rui Paulo (Universidade de Lisboa) and Anabel Forte (Universitat de Valencia)

Nikolas Bravo
North Carolina State University

"Parameter-Dependent Surrogate Model Development and Control Design for PZT Bimorph Actuators Employed for Micro-Air Vehicles"

In the talk, we discuss the development of the homogenized energy model (HEM) and the surrogate model dynamical mode composition for a PZT bimorph actuator used for micro-air vehicles including Robobee. HEM quantifies the nonlinear, hysteretic, and rate-dependent behavior inherent to PZT in highly dynamic operating regimes. Due to computation complexity of HEM, we must develop a surrogate model. The surrogate model must be parameter and control dependent to be able to perform inverse problems or uncertainty quantification in different driving regimes. In the literature, DMD can be adapted to handle different control inputs. We will discuss using interpolation over the parameters to adapt the DMD to include parameter dependence. Finally, we will discuss control design using the surrogate model and quantifying the uncertainty in the controls.

Won Chang
University of Cincinnati

"Emulation for Large Storm Surge Simulation Ensemble"

As a building block for our storm surge emulator we are currently developing a statistical surrogate model that can predict the spatial patterns of simulated storm surge for any given hurricane locations and properties. This problem poses nontrivial statistical challenges due to the semi-continuous marginal distribution of the surge patterns and some severe nostationarity in the input-output relationship. Our method combines a new emulation method for semi-continuous data and treed Gaussian process approach to mitigate these challenges. A preliminary application of our approach to storm surge patterns at selected locations in Southwestern Florida shows that the method is highly promising for emulating the complicated storm surge patterns from simulation. Some future research plan on optimal ensemble design will be also discussed.

Oksana Chkrebtii
Ohio State University

"Adaptive Step-Size Selection for State-Space Probabilistic Differential Equation Solvers"

When models are defined implicitly by systems of differential equations with no closed-form solution, small local errors in finite-dimensional solution approximations can propagate into deviations from the true underlying model trajectory. Some recent perspectives in quantifying this uncertainty are based on Bayesian probability modeling: a prior is defined over the unknown solution and updated by conditioning on interrogations of the forward model. Improvement in accuracy via grid refinement must be considered in order for such Bayesian numerical methods to compete with state of the art numerical techniques. We apply principles of Bayesian statistical design to develop an adaptive probabilistic method to sequentially select time-steps for state-space probabilistic ODE solvers. We investigate the behaviour of local error under the adaptive scheme which underlies numerical variable step-size methods. Numerical experiments are used to illustrate the performance of such adaptive schemes, showing improved accuracy in terms of global error over uniform designs when small step lengths are considered.

Simon Cotter
University of Manchester

"Competing Complexities in Bayesian Inverse Problems: Models and Distributions"

There are many sources of complexity when considering Bayesian inverse problems for systems in applied mathematics. The primary sources of complexity lie in two broad areas; the mechanistic model of the system that is being analysed, and the structure of the distributions of the parameters that arise. In this talk we will look briefly at two problems which each demonstrate one of these two areas. In the first we will consider an inverse problem for thermodynamical properties of materials from laser flash experiments. In this work we aim to reduce the complexity of the Bayesian inversion by using a stochastic Galerkin approximation of the observation operator in order to make full posterior characterisation via MCMC feasible. In the second we will consider an ensemble importance sampling method which allows for efficient sampling of posterior distributions with complex structure, for example where the posterior density is closely concentrated around a lower dimensional manifold in parameter space.

Aaron Danielson
Simon Fraser University

"Some Strategies to Quantify Uncertainty for Extrapolation in Physical Systems"

Although we often told not to do it, statistical scientists frequently predict the value of outcome measures of physical systems at input points far the observed data. Since predictions are made in new regions of the input space, a statistical theory cannot dictate optimal rules for measures of uncertainty associated with extrapolation. This talk presents several solutions based on simple principles. The solutions are illustrated via the analysis of data generated by dropping spheres of varying radii and masses from different heights. Some of the techniques apply to more complex physical systems. The efficacy of these techniques is demonstrated using data (experimental and simulated) of the level of complexity physical scientist frequently face. Scientists should tailor these techniques to fit the needs of a particular application.

Arindam Fadikar
Virginia Tech University

"Clustering Based Gaussian Process Emulation and Calibration of a Stochastic Agent Based Model"

Gaussian process (GP) model is an effective tool for emulating complex computer simulations. Heterogeneous gaussian process (Binois et al, 2017) has been shown to be superior in the presence of input dependent noise as in the case for any stochastic computer simulation. However, all GP models impose a gaussian variability assumption in the emulator. In this talk, we propose a new approach based on heterogeneous GP and a clustering based technique to emulate and hence calibrate a stochastic agent based simulation. The basic idea is to relax the normality assumption by borrowing the standard gaussian mixture model and coupling that with a traditional GP. The study is motivated by with an example taken from the 2015 Ebola challenge workshop which simulated an Ebola epidemic to evaluate methodology.

Michael Grosskopf
Los Alamos National Laboratory

"Structural Model Discrepancy in Nuclear Energy Density Functional Simulators"

Nuclear Energy Density Functional (EDF) theory has been invaluable in understanding the properties of nuclei and their impact on applications from stellar phenomena to energy applications. Numerical implementations of EDF models are able to simulate the properties of elements across the nuclide table, but involve approximations to the underlying physics. The impact of missing physics has systematic effects that are important for calibration, prediction, and extrapolation using the physics model. This talk will present work on calibration of the EDF models and analysis of a structured discrepancy model with the goal of improving the ability to extrapolate to new nuclei properties beyond the observed and further understanding of the ways in which the simulator can be improved.

Mengyang Gu
Johns Hopkins University

"A Review of Model Calibration Methods with an Application by Fusing Multiple Sources of Data from the Eruption of the Kilauea Volcano in 2018"

Model calibration or data inversion involves using experimental or field data to estimate the unknown parameters in a mathematical model. In the first part of the talk, I will present a review of a few approaches for model calibration or data inversion with the focus on model discrepancy and measurement bias. A few state-of-art methods, such as modeling the discrepancy by the Gaussian stochastic process (GaSP) or scaled Gaussian stochastic processes (S-GaSP), L2 calibration, Least squares (LS) calibration and orthogonal Gaussian process calibration, will be introduced. The connection and difference between these methods will be discussed. In the second part of talk, I will discuss our ongoing works on calibrating a geophysical model by integrating the different types of the field data, such as the interferometric synthetic aperture radar satellite (InSAR) interferograms, GPS data, velocities of tilt and lava lake from the Kilauea Volcano during the eruption in 2018. This task is complicated by the discrepancy between the model and reality different sample sizes and possible bias in field data. We introduce the scaled Gaussian stochastic process (S-GaSP), a new stochastic process to model the discrepancy function in calibration for the identifiability issue between the calibrated mathematical model and the discrepancy function. We also compare a few approaches to model the measurement bias in the data. A feasible way to fuse the field data from multiple sources will then be discussed. The calibration models are implemented in the "RobustCalibration" R Package on CRAN. The scientific goal of this work is to use data in May 2018 during the earthquake and the eruption of the Kilauea Volcano to resolve the location, volume, and pressure change in the Halema'uma'u Reservoir, as well as relating the results to the inferences from the past caldera collapses.

Jan Hannig
University of North Carolina

"Are Reported Likelihood Ratios Well Calibrated?"

Many computer programs and software systems used in the interpretation of forensic evidence have as their output Bayes factors also commonly referred to as likelihood ratios. For example, it is not unusual to see it reported that the DNA recovered at the crime scene is a million times more likely under the assumption that the defendant is a contributor to the crime stain than under the assumption that the defendant is not a contributor. In this talk we summarize existing approaches for examining the validity of likelihood ratio systems and discuss a new statistical methodology, based on generalized fiducial inference, for empirically examining the validity of such likelihood ratio assessments. Using data from ta number of sources, such as glass, paint and DNA evidence, we illustrate our approach by examining LR values calculated using standard approaches in forensic literature.

Joint work with Hari Iyer at National Institute of Standards and Technology

Whitney Haung
University of Victoria

"Some Thoughts on Estimating Input Distribution of Storm Surge Simulations"

While accurately accessing the storm surge risk is crucial for coastal regions, the limited observational records of storms prevents traditional data--based statistical approaches. The common practice involves running hydrodynamics simulation where each run simulates the storm surge response with respect to a hurricane scenario, which is parameterized by a set of storm characteristics. One of the key components for surge risk assessment is the estimation of the distribution of the storm characteristics (input). In this talk I will present our working group's effort on addressing this critical but often overlooked issue.

Zhuoqiong He
University of Missouri

"Bayesian CUSP Catastrophe Model for Sudden Changes"

The cusp catastrophe model uses a discontinuous nonlinear function to predict sudden changes. Due to the complexity of the discontinuous nonlinear relationship, there are some issues in fitting the statistical cusp regression model, i.e., gradient-based optimization methods no longer work. We have developed a Bayesian method for the cusp regression model and used the posterior mean to obtain estimates of the parameters. The partial swarm optimization algorithm is used to speed up the convergence of the Markov chain Monte Carlo algorithm. The simulation study shows that the Bayesian method yields a better estimate than both the maximal likelihood estimation and the traditional stochastic differential equations method under the Maxwell convention.

Authors: Jiayi Hou[1], Ranadeep Daw[1], Cong He[1], Dinggen Chen[2], Dongchu Sun[1]
[1] University of Missouri, [2] University of North Carolina

Radu Herbei
Ohio State University

"Estimating Ocean Circulation Structure: Deterministic and Stochastic Simulators"

The state of the world oceans is a crucial factor in the understanding of our climate. However, ocean circulation cannot be measured directly, it has to be inferred from indirect observations of various tracers (oxygen, salinity, temperature, etc.) or Lagrangian data. In this talk we consider the inverse problem of inferring ocean circulation structure (water velocities, diffusion coefficients) based on tracer concentration measurements. We propose a Bayesian solution based on a quasi-horizontal flow model connecting water velocities and diffusion coefficients to tracer concentrations. The forward problem of solving the advection-diffusion equations does not have a closed form solution. In order to evaluate the likelihood function, we consider two strategies. The first one is based on a deterministic multi-grid PDE solver. The second approach is based on a stochastic solver derived using the Feynman-Kac probabilistic representation of a PDE solution. In each case an adequate Markov chain Monte Carlo sampler is used to explore the posterior distribution for the quantities of interest.

Ying Hung
Rutgers University

"Computer Experiments with Binary Time Series and Applications to Cell Biology: modeling, estimation and calibration"

Computer experiments have become ubiquitous in various applications from rocket injector designs to weather forecast. Although extensive research has been devoted in the literature, computer experiments with binary time-series outputs have received scant attention. Motivated by the analysis of a class of cell adhesion experiments, we introduce a new emulator as well as a new calibration framework for binary time-series outputs. More importantly, we provide their theoretical properties to ensure the estimation performance in an asymptotic setting. The application to the cell adhesion experiments illustrates that the proposed emulator and calibration framework not only provide an efficient alternative for the computer simulation, but also reveal important insight on the underlying adhesion mechanism, which cannot be directly observed through existing methods.

Roshan Joseph
Georgia Tech University

"Transformation and Additivity in Gaussian Process"

We discuss the problem of approximating a deterministic function using Gaussian Processes (GP). The role of response transformation in GP modeling is not well understood. We argue that transformations can be used for making the deterministic function approximately additive, which can then be easily estimated using an additive GP. We call such a GP a Transformed Additive Gaussian (TAG) process. To capture possible interactions which are unaccounted for in an additive model, we propose an extension of TAG process called Transformed Approximately Additive Gaussian (TAAG) process. We develop efficient techniques for fitting a TAAG process. In fact, we show that it can be fitted to high-dimensional and big data much more efficiently than the usual GP. Furthermore, we show that the use of TAAG process leads to better estimation, interpretation, visualization, and prediction.

Authors: Li-Hsiang Lin and V. Roshan Joseph (Georgia Institute of Technology, Atlanta)

Emily Kang
University of Cincinnati

"Simulation Experiments and Uncertainty Quantification in Remote Sensing"

Remote sensing data sets produced by NASA and other space agencies are the result of complex algorithms that infer geophysical state from observed radiances using retrieval algorithms. Simulation experiments are important tools to design new observing systems, to evaluate new data assimilation algorithms, to calibrate parameters, and to study uncertainty propagation. This talk will discuss opportunities and challenges involved in such experiments and to advance statistical methodology. Examples will be given of modeling and simulating carbon dioxide and development of a stochastic simulator of the physical forward model in the retrieval algorithm for NASA's Orbiting Carbon Observatory-2.

Lulu Kang
Illinois Institute of Technology

"Gaussian Process Model Assisted Active Learning of Physical Laws"

PDE and ODE are commonly used to describe a wide variety of physical phenomena such as sound, heat, diffusion, electrostatics, electrodynamics, fluid dynamics, elasticity, and quantum mechanics, etc. Discovering the governing equations from noisy data is an essential challenge in many areas of science and engineering, which is critical to understand the physical phenomena and predict the future behaviors of the dynamical system. However, in many cases, data are costly or time-consuming to obtain. Therefore, it is desirable to use the smallest possible amount of data to learn the PDE or ODE systems with a user-specified level of accuracy. To achieve this goal, this paper provides an active learning method to estimate the underlying PDE and ODE models. We propose an adaptive design criterion combining the D-optimality and the maximin space-filling criterion, with the weights automatically decided by the data. Different from the typical active learning methods for statistical models, the proposed design criterion involves both the function values and the different orders of derivatives, which are unobserved at the potential design points before data collection. Thus, we use the Gaussian process model to make predictions for these values at the potential design points. After the data collection, variable-selection-based regression methods are used to estimate the PDE or ODE models. The proposed active learning approach is entirely data-driven and requires no tuning parameters. Through three case studies of the commonly used PDE and ODE models, we show the proposed approach outperforms the current literature in terms of model accuracy and data economy.

Earl Lawrence
Los Alamos National Laboratory

"Some Pieces of Exascale Uncertainty Quantification"

Exascale computing is coming to the Department of Energy. The new prefix brings some new challenges. Raw computing power will overwhelm the I/O capability of these machines. This is a particular difficulty for the practice of statistical uncertainty quantification (UQ) which relies on ensembles of simulations, each of which will necessarily be only partially available. This talk will present two separate piece of work that may be useful parts of the exascale UQ workflow. The first part of the talk will cover in situ analysis of simulations in order to summarize simulations as they are running. The second part of the talk will cover statistical emulation with incomplete data.

Gang Li
University of North Carolina

"Deep Fiducial Inference and Approximate Fiducial Computation"

Since the mid-2000s, there has been a resurrection of interest in modern modifications of fiducial inference. To date, the main computational tool in extracting a Generalized Fiducial Distribution is MCMC and similar methods. In this paper, we propose an alternative way of computing a Generalized Fiducial Distribution that could be used in complex situations. Under generalized fiducial inference framework, we first design the Approximate Fiducial Computation (AFC) algorithm to generate approximate generalized fiducial samples, on which we can form inference without knowing the closed form of the fiducial density. To overcome the difficulty when the inverse function or the marginal fiducial density is intractable, we further design the Fiducial Autoencoder (FAE) to approximate the inverse function. AFC is then used to get generalized fiducial samples. The universal approximation theorem in neural networks provides theoretical guarantees for the approximation performance, and our simulations further validate our method.

Cong Lin
ECNU

"Bayesian Smoothing Spline with a Generalized Constraint Operator"

Smoothing spline is one of the most popular curve-fitting methods, for the reason of its flexibility and a natural Bayesian interpretation. However, the extrapolation of smoothing spline in data missing case is subject to a high variation and may produce meaningless results, meantime, people can usually obtain prior information about the underlying function, such as, mean or reasonable range of the initial or boundary values. Our intention is to incorporate such prior information to improve the extrapolation performance by extending the minimization problem of classical spline model, this new model is a generalization to classical smoothing spline by introducing constraint operator, we solved the minimizer via reproducing kernel Hilbert space theory and Green's function. Similar to Wahba (1978), we also proposed a stochastic process prior on the functions of objective function space, the main difference to Wahba's prior is that our prior with respect to the polynomial space is proper rather than diffuse. We would have two tuning parameters instead of single tuning parameter, to perform fully Bayesian analysis, independent Pareto priors and bivariate

Pareto prior on these two tuning parameters are assumed separately, simulation result will also be presented in our talk.

Devon Lin
Queens University

"A Sequential Design Approach for Calibrating a Dynamic Population Growth Model"

Abstract: A comprehensive understanding of the population growth of a variety of pests is often crucial for efficient crop management. Our motivating application comes from calibrating a two-delay blowfly (TDB) model which is used to simulate the population growth of Panonychusulmi (Koch) or European red mites that infest on apple leaves and diminish the yield. We focus on the inverse problem, that is, to estimate the set of parameters/inputs of the TDB model that produces the computer model output matching the field observation as closely as possible. The time series nature of both the field observation and the TDB outputs makes the inverse problem significantly more challenging than in the scalar valued simulator case.

In spirit, we follow the popular sequential design framework of computer experiments. However, due to the time-series response, a singular value decomposition based Gaussian process model is used for the surrogate model, and subsequently, a new expected improvement criterion is developed for choosing the follow-up points. We also propose a new criterion for extracting the optimal inverse solution from the final surrogate. Three simulated examples and the real-life TDB calibration problem have been used to demonstrate higher accuracy of the proposed approach as compared to popular existing techniques.

Pulong Ma
SAMSI

"An Emulator Approach for Quantifying the Risk Due to Storm Surge"

Complex computer models of real-world processes (or simulators) are an essential ingredient to carry out uncertainty quantification in science and engineer. In coastal emergency risks assessment, storm surge is one of the most severe natural disasters that can lead to significant flooding in coastal areas and severe damages to the life and property from a hurricane. Quantifying the risk due to storm surge requires large-scale numerical simulations of hurricanes from storm surge modeling systems. A crucial need is the development of an emulator to facilitate risk quantification. This talk will present a statistical surrogate model to tackle this problem and show some initial results.

Simon Mak
Georgia Tech University

"cmenet: a new method for bi-level variable selection of conditional main effects"

This talk introduces a novel variable selection method for conditional main effects (CMEs), which capture the conditional effect of a factor given a fixed level of another factor. CMEs represent interpretable, domain-specific phenomena for a wide range of applications in the physical and social sciences. The key challenge is in incorporating the grouped structure of CMEs within the variable selection procedure itself. We propose a new method, cmenet, which employs two principles (CME coupling and CME reduction) to effectively navigate the selection algorithm. In simulated tests where CMEs are indeed present, cmenet provides improved performance over standard variable selection and interaction analysis methods. Applied to real-world gene association studies, cmenet not only gives more parsimonious models and improved predictive performance over existing methods, but also reveals important insights on gene activation behavior which can guide further experiments.

Sue Minkoff
University of Texas

"Parameter Subset Selection for Coupled Flow and Deformation Modeling"

Many subsurface reservoirs compact or subside due to production-induced pressure changes. Numerical simulation of this compaction process is important for predicting and preventing well-failure in deforming hydrocarbon reservoirs. However, development of sophisticated numerical simulators for coupled fluid flow and mechanical deformation modeling requires a considerable manpower investment. This development time can be shortened by loosely coupling pre-existing flow and deformation codes via an interface. Pressure changes after set of flow time steps can be used to load mechanical deformation. The mechanics code then solves for displacement and strain. Changes in strain can lead to porosity changes which are then sent back to the flow simulator for the next set of time steps. Loose two-way coupling even for a simple flow and deformation model leads to interesting questions regarding development of coupled emulators. Parameter subset selection can be used to determine influential parameters for the coupled simulator which may lead to fast reduced-order models or emulators.

Max Morris
Iowa State University

"Sensitivity Analysis of Computer Models: A Statistical Perspective"

I will address what some have come to refer to as global sensitivity analysis, based on the expression a model's inputs as random. Current usage sometimes refers to uncertainty analysis as the characterization of the marginal distribution of outputs resulting from random inputs, while sensitivity analysis focuses on characterization of conditional distributions deemed to be indicators of the importance of individual inputs or groups of them (e.g. the Sobol' indices). A ``statistical perspective'' on this activity comes (I believe) from the premise that incomplete information results in uncertainty that is different from that associated with the input distribution, but that must be accounted for in a careful analysis. In this context, ``incomplete information'' may mean a sample of inputs used in place of an input distribution, and/or a surrogate/emulator used in place of the computer models of interest. The presentation will not be technically detailed, and the primary intent will be to explain one statistician's viewpoint to members of other disciplines within the mathematical sciences.

Rebecca Morrison
MIT

"Representing Model Inadequacy in Reduced Models of Interacting Systems"

In many applications of interacting systems, we are only interested in the dynamic behavior of a subset of all possible active species. For example, this is true in combustion models (many transient chemical species are not of interest in a given reaction) and in epidemiological models (only certain critical populations are truly consequential). Thus it is common to use greatly reduced models, in which only the interactions among the species of interest are retained. However, reduction introduces a model error, or inadequacy, which typically is not well characterized. In this talk, I explore the use of an embedded and statistically calibrated inadequacy operator to represent model error. The operator is constrained by available physical information and embedded within the differential equations of the model. Both the reduced and operator models are also designed to respect what I call *species correspondence, i.e.,* that the model outputs of the reduced and operator models correspond to the species of the original model. This design of an augmented, yet physically realistic, model is intended to allow for reliable predictions under extrapolative conditions---in, for example, time or scenario parameters.

Rui Paulo
Universidade de Lisboa

"Model Selection in the Context of Computer Models"

We approach the screening problem - i.e. detecting which inputs of a computer model significantly impact the output - from a formal Bayesian model selection point of view. That is, we place a Gaussian process prior on the computer model and consider the 2^p models that result from assuming that each of the subsets of the p inputs affect the response. The goal is to obtain the posterior probabilities of each of these models. In this talk, we focus on the specification of objective priors on the model-specific parameters and on convenient ways to compute the associated marginal likelihoods. These two problems that normally are seen as unrelated, have challenging connections since the priors proposed in the literature are specifically designed to have posterior modes in the boundary of the parameter space, hence precluding the application of approximate integration techniques based on e.g. Laplace approximations. We explore several ways of circumventing this difficulty, comparing different methodologies with synthetic examples taken from the literature.

Joint work with Gonzalo Garcia-Donato (Universidad de Castilla-La Mancha)

Matthew Plumlee
Northwestern University

"Emulation for Forecasting Storm Surge"

Emulators often viewed as replacements of a computer simulation model to perpetually supplant the computationally expensive simulation. This is not the case when the simulated phenomenon of interest is a singular event. This talk will outline an attempt to create an emulation strategy for forecasting the resulting surge after a unique hurricane. Michael from 2018 was used as a testbed for this investigation.

Laura Schulz
George Mason University

"Practical Bayesian Optimization for Transportation Simulators"

Simulators play a major role in analyzing multi-modal transportation networks. As their complexity increases, optimization becomes an increasingly challenging task. Current calibration procedures often rely on heuristics, rules of thumb and sometimes on brute-force search. Alternatively, we provide a statistical method which combines a distributed, Gaussian Process Bayesian optimization method with dimensionality reduction techniques and structural improvement. We then demonstrate our framework on the problem of calibrating a multi-modal transportation network of city of Bloomington, Illinois. Our framework is sample efficient and supported by theoretical analysis and an empirical study. We demonstrate on the problem of calibrating a multi-modal transportation network of city of Bloomington, Illinois. Finally, we discuss directions for further research.

Vadim Sokolov
George Mason University

"Practical Bayesian Optimization for Agent Based Transportation Simulators"

Simulators play a major role in analyzing multi-modal transportation networks. As complexity of simulators increases, development of calibration procedures is becoming an increasingly challenging task. Current calibration procedures often rely on heuristics, rules of thumb and sometimes on brute-force search. In this talk we consider an statistical framework for calibration that relies on Bayesian optimization. Bayesian optimization treats the simulator as a sample from a Gaussian process (GP). Tractability and sample efficiency of Gaussian processes enable computationally efficient algorithms for calibration problems. We show how the choice of prior and inference algorithm effect the outcome of our optimization procedure. We develop dimensionality reduction techniques that allow for our optimization techniques to be applicable for real-life problems. We develop a distributed, Gaussian Process Bayesian regression and active learning models. We demonstrate those to calibrate ground transportation simulation models.

Chengyuan Song
ECNU

"Bayesian Analysis for one-way MANOVA and a 3-Level Hierarchical Model"

We study a new class of commutative shrinkage priors for two covariance components in multivariate one-way ANOVA models and Normal hierarchical models, respectively. Multivariate one-way ANOVA model is of substantial importance in contemporary statistical theory and application. It could be used to data fusion for analyzing data from different resources. One primary interest is to estimate unknown overall mean and two covariance components (matrices). The usual MLE and moment estimators for some covariance component may not exist. For two covariance components, a new class of commutative priors is also proposed, which is a conjugate class. Propriety and moment existence are derived for both the prior and their posterior. Simulation and real data analysis show the advantages of commutative priors. In addition, we also study the commutative shrinkage priors in normal hierarchical models. Normal hierarchical models are quite important for Bayesian analysis, yet the choice of objective priors to use for hyperparameters is often in a casual fashion. It is common to use a constant prior for higher level variances or covariances, but the constant prior is much too diffuse, requiring twice as many observations to obtain posterior propriety as is logically needed (see Berger et al., 2018), especially when the dimension is high. Using formal priors from non-hierarchical models, such as the Jeffreys-rule or reference prior approach, result in improper posterior distributions if they are used at higher levels of a hierarchical model (Fatti, 1982ï¼› Sun et al., 2001). Berger et al. (1996, 2005, 2018) approached the question of choice of hyperpriors in a 2-level normal hierarchical model from the frequentist notion of admissibility of resulting estimators. We study the commutative shrinkage priors for use in 3-level normal hierarchical models. Our work is based on considerations of posterior propriety, admissibility, ease of implementation (including computational considerations), and performance. This work is joint with Dongchu Sun, James O. Berger, Zhuoqiong He, and William R. Bell.

Paul Speckman
University of Missouri

"Data Fusion for Correlated, Shape-Restricted Curves with Varying Support"

Data fusion methods are implemented to estimate multiple bond yield curves. A bond yield curve describes the relationship between the yield of a particular class of bonds (expressed as an annual interest rate) to the time to maturity for all bonds traded on a given day. These curves are theoretically smooth with possible shape restrictions (e.g., monotonic). The curves for different rating classes are theoretically ordered and may have different support. Application is made to data from the Chinese bond market, where data may be sparse. Thus it is imperative that estimates combine information across different classes and multiple days. The methods discussed here are based in part on interaction smoothing splines to estimate curves for multiple classes. A weighted least squares approach is used to fuse estimates across multiple days. Fully automatic estimation has proved elusive, and provision for expert opinion must be incorporated.

Dongchu Sun
University of Missouri

"Bayesian Model Selection for a Linear Model with Grouped Covariates"

Model selection for normal linear regression models with grouped covariates is considered under a class of Zellner's (1986) g-priors. The marginal likelihood function is derived under the proposed priors, and a simple closed form expression is given assuming the commutativity of the projection matrices from the design matrices. As illustration, the marginal likelihood8 functions of the balanced m-way ANOVA models, either solely with main effects or with all interaction effects, are calculated using the closed form expression. The performance of the proposed priors in model comparison problems is demonstrated by simulation studies on two way ANOVA models and by two real data studies.

Matthias HY Tan
City University of Hong Kong

"Opening Up the Black Box: Gaussian Process Modeling Using Information from Partial Differential Equation Models"

Gaussian process (GP) emulators of computer models are typically constructed based purely on data from a computer experiment using a standard stationary GP prior with product Matérn or Gaussian correlation function. This often ignores valuable engineering and mathematical knowledge about the behavior of the computer model. In this talk, I will present my research on the use of known behavior/properties of partial differential equation models solved numerically by computer codes to improve construction of GP emulators for this type of computer models.

William Welch
University of British Columbia

"Dimensional Analysis in Computer Experiments"

Dimensional analysis pays attention to the units of measurement when modeling scientific and engineering systems. It goes back at least a century (Buckingham, 1914) but has recently caught the attention of statisticians for statistical modeling, particularly in design of experiments (e.g., Shen, Davis, Lin, and Nachtsheim, 2014; Shen, Lin, and Chang 2017; Shen and Lin, 2018). Advantages have been demonstrated for both traditional physical experiments and computer experiments. The basic idea is to analyze in terms of more fundamental dimensionless quantities derived from the original variables, and possibly design for them too. With the "right" variables, prediction accuracy will hopefully improve. While these goals make much sense for scientific applications and statistical modeling, implementation of dimensional analysis is far from straightforward; choosing the derived quantities is particularly problematic. Empirical approaches to finding "good" derived variables in computer experiments will be described, based on the ongoing PhD thesis of G. Alexi RodrÃ-guez-Arelis. Tentatively, an application to storm surge in collaboration with Whitney Huang will be used for illustration too.

Brian Williams
Los Alamos National Laboratories

"Gradient-Free Construction of Active Subspaces for Dimension Reduction"

Recent developments in the field of reduced order modeling - and in particular, active subspace construction - have made it possible to efficiently approximate complex models by constructing low-order response surfaces based upon a small subspace of the original high dimensional parameter space. These methods rely upon the fact that the response tends to vary more prominently in a few dominant directions defined by linear combinations of the original inputs, allowing for a rotation of the coordinate axis and a consequent transformation of the parameters. In this talk, we discuss a gradient free active subspace algorithm that is feasible for high dimensional parameter spaces where finite-difference techniques are impractical. We illustrate an initialized gradient-free active subspace algorithm for a neutronics example implemented with SCALE6.1.

David Woods
University of Southampton

"Design of Experiments for the Calibration of Computational Models"

Computational modelling now underpins much research in the sciences and engineering, allowing in silico investigations and predictions of complex systems. Reliable and accurate computational modelling often relies on the calibration of the model using physical data, usually collected via designed experiments. These data are used to tune, or estimate, unknown model parameters and, perhaps, to learn the discrepancy between the computational model and reality.

In this talk, we present some new methods for the optimal design of physical experiments for this calibration problem. A Bayesian approach is adopted, with a Gaussian process prior assumed for the output from the computational model. New decision-theoretic optimal designs are sought for this problem, using novel methods for the numerical approximation of the expected utility of a design. The results are motivated by, and demonstrated on, problems from science and technology.

Authors: David Woods, Yiolanda Englezou (University of Southampton, UK) Tim Waite (University of Manchester, UK)

Jeff Wu
Georgia Tech University

“Navier-Stokes, Spatial-temporal Kriging and Combustion Stability: a prominent example of physics-based analytics**”**

Most “learning” in big data is driven by the data alone. Some people may believe this is sufficient because of the sheer data size. If the physical world is involved, this approach is often insufficient. In this talk I will give a recent study to illustrate how physics and data are used jointly to learn about the “truth” of the physical world. It also serves as an example of physics-based analytics, which in itself has many forms and meanings. In an attempt to understand the turbulence behavior of an injector, a new design methodology is needed which combines engineering physics, computer simulations and statistical modeling. There are two key challenges: the simulation of high-fidelity spatial-temporal flows (using the Navier-Stokes equations) is computationally expensive, and the analysis and modeling of this data requires physical insights and statistical tools. A surrogate model is presented for efficient flow prediction in injectors with varying geometries, devices commonly used in many engineering applications. The novelty lies in incorporating properties of the fluid flow as simplifying model assumptions, which allows for quick emulation in practical turnaround times, and also reveals interesting flow