Statistical Perspectives on Uncertainty Quantification

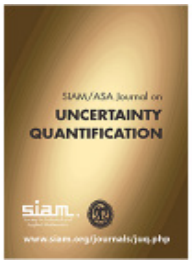
May 29 – 30, 2017
Georgia Institute of Technology
Atlanta, GA

Georgia Tech H. Milton Stewart School of Industrial & Systems Engineering
College of Engineering

Georgia Tech College of Engineering

jmp Statistical Discovery™ From SAS.

P&G



siam

NISS

Table of Contents

Sponsors

Statistical Perspectives of Uncertainty Quantification (SPUQ) 2017 gratefully acknowledges generous support from the following sponsors (in alphabetical order):

- College of Engineering, Georgia Tech
- H. Milton Stewart School of Industrial & Systems Engineering, Georgia Tech
- JMP (SAS Institute)
- National Institute of Statistical Sciences (NISS)
- Procter & Gamble
- SIAM / ASA Journal on Uncertainty Quantification (JUQ)
- Society for Industrial and Applied Mathematics (SIAM)

Organizing Committee

V. Roshan Joseph (Co-Chair)	Georgia Tech
David Higdon (Co-Chair)	Virginia Tech
Matthew Plumlee	University of Michigan
Peter Qian	University of Wisconsin
Benjamin Haaland	Georgia Tech

Local support: The committee would like to thank the ISyE staff Harry Sharp and Scott Jacobson and students Simon Mak, Arvind Krishna, Chih-Li Sung, Fang Cao, Wenjia Wang, and Li-Hsiang Lin for their help.

General Information

Sessions – All sessions will be held in Salon VI at the Georgia Tech Hotel & Conference Center. Each session consists of three 30-minute talks, centered around a specific topic on the uncertainty quantification of complex computational models.

Poster Presentation – Poster presentations are held on Monday from 17:30 – 19:00 in Salon V & VI at the Georgia Tech Hotel & Conference Center. Easels and poster boards are provided at the venue. Please set-up your posters during the break time 15:30-16:00, so that we can start the presentations on-time.

Meals – For all attendees, continental breakfast (near the general session in the second level) and lunch (conference dining room in the first level) is provided on Monday and Tuesday at the Georgia Tech Hotel & Conference Center. A banquet dinner is scheduled on Monday from 19:00 – 21:00 in Conference Room A.

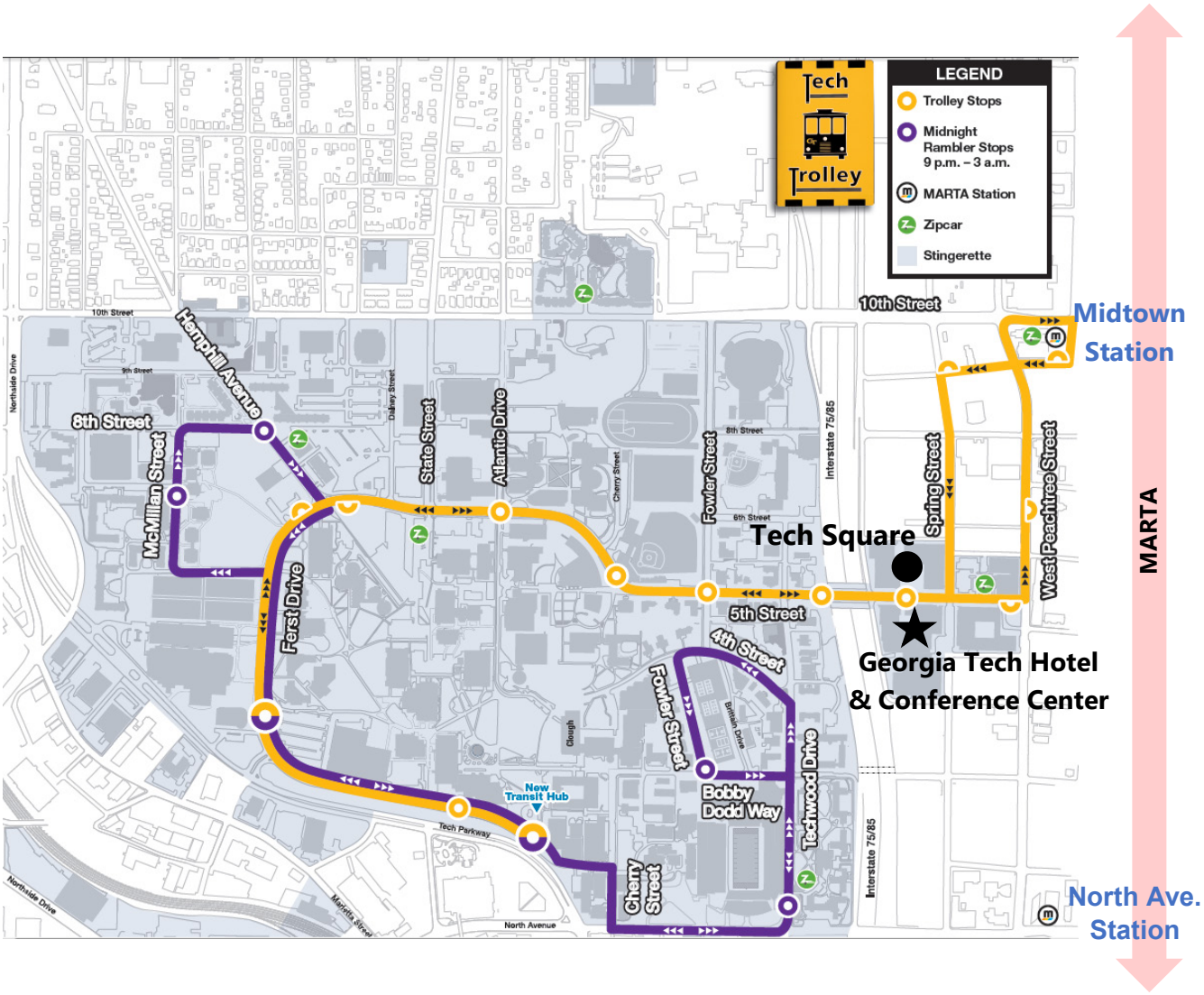
Parking

Parking is provided for attendees at the Georgia Tech Hotel & Conference Center. After parking your vehicle in the hotel parking lot, please ask the front desk volunteers for a parking pass. This pass can be redeemed at the parking booth on your way out.

Ground Transportation & Local Information

Transportation – For attendees who are not driving, there are two main modes of transportation around the Georgia Tech campus: the Tech Trolley and the MARTA rail system. The map on the following page how to get to the Trolley and the MARTA from the Georgia Tech Hotel & Conference Center (marked by a star):

- **Tech Trolley –** The Trolley is a complementary (free) shuttle around the Georgia Tech campus, and provides transportation to the MARTA Midtown Station to the hotel. The Trolley runs from 5:45am to 10:15pm along the route indicated on the map. It is accessible at a frequency of around 10 minutes during the day, and around half an hour during early morning (before 7:15 am) and late evening (post 7:30 pm). To take the Trolley, exit the Georgia Tech hotel onto 5th Street, and walk a few steps left. The trolley stop to Midtown Station should be right outside Tin Drum.
- **MARTA –** The MARTA, Atlanta's rapid transit system, can take you directly to the airport from the conference. The nearest MARTA station (North Ave. Station) is an 8-minute walk away from the Georgia Tech Hotel; the Midtown Station is a few minutes ride on the Tech Trolley. To get to the airport, simply take the southbound train; the final station is the Airport Station.
- **Traveling at night –** While the surrounding neighborhood is safe during the day, this may not be the case at night. We strongly suggest calling an Uber / Lyft / taxi instead of walking at night.



Dining options – There are a variety of restaurants around the Tech Square area which you might enjoy after the conference:

- **Within Tech Square (see map)**: Subway, Moe's Southwest Grill, Tin Drum Asian Kitchen, Umma's House Restaurant, Ray's New York Pizza, Starbucks and Waffle House. Most restaurants close at 9 pm, but Waffle House is open 24/7.
- **Midtown:** Several restaurants and pubs are located in the midtown area, which is less than a 15-minute walk away from the Georgia Tech Hotel. Popular choices include Five Guys, Café Agora, Ecco, Fado Irish Pub and Hudson Grille.

Program

Monday, May 29, 2017

Time	Session	Event
07:30 – 08:30		Continental Breakfast
08:30 – 09:00		**Welcome and Opening Remarks** • **Edwin Romeijn** (*H. Milton Stewart School Chair and Professor, Georgia Tech*) • **C. F. Jeff Wu** (Coca-Cola Chair in Engineering Statistics and Professor, *Georgia Tech*)
09:00 – 10:30	Session 1	**Computer Experiments: Early Work and Recent Developments** Organizer: Peter Qian • **Michael Stein** – University of Chicago *Climate Model Emulation and Future Climate Simulation* • **Art Owen** – Stanford University *On Shapley value for measuring importance of dependent inputs* • **Tom Santner** – Ohio State University *Challenges in Designing and Using Simulator Experiments in Biomechanics and Biomaterials Research*
10:30 – 11:00		Coffee Break
11:00 – 12:30	Session 2	**Calibration of Models under Uncertainty** Organizer: Matthew Plumlee • **Daniel Apley** – Northwestern University *An Empirical Adjustment of the Uncertainty Quantification in Gaussian Process Modeling* • **Oksana Chkrebtii** – Ohio State University *Advances in discretization uncertainty quantification for differential equation models* • **Rui Tuo** – Chinese Academy of Sciences *Adjustments to Computer Models via Projected Kernel Calibration*
12:30 – 14:00		Lunch

14:00 – 15:30	Session 3	**Applied Math, Engineering and Computation** Organizer: Dave Higdon • **Paul Constantine** – Colorado School of Mines *Active Subspaces: Emerging Ideas for Dimension Reduction in Computational Science and Engineering Models* • **Akil Narayan** – University of Utah *Sampling strategies for efficient approximations* • **Clayton Webster** – ORNL *Sparse polynomial approximation for high-dimensional uncertainty quantification*
15:30-16:00		Coffee Break
16:00-17:30	Session 4	**UQ in Action** Organizer: Ben Haaland • **Hendrik Hamann** – IBM *Big data gets physical* • **Johannes Hoetzer** – KIT *Application of data-science approaches for the analysis of directional solidified ternary alloys* • **Bill Myers** – P&G *Robust Parameter Design using Computer Experiments*
17:30-19:00		Poster Presentation
19:00-21:00		Banquet Dinner

Tuesday, May 30, 2017

Time	Session	Event
07:30 – 08:30		Continental Breakfast
08:30 – 10:00	Session 5	**UQ in Emulation, Simulation, and Calibration** Organizer: Roshan Joseph • **Jim Berger** – Duke University *On Fitting Gaussian Process Emulators* • **Bani Mallick** – Texas A&M University *Bayesian variational approaches for inverse problems* • **Barry Nelson** – Northwestern University. *Reducing Simulation Input-Model Risk via Input Model Averaging*
10:00 – 10:30		Coffee Break
10:30 – 12:00	Session 6	**UQ and Model-based Applications** Organizer: Dave Higdon • **Marc Genton** – KAUST *Statistics-Based Compression of Global Wind Fields* • **Sez Atamturktur** – Clemson University *Model Calibration, Validation, and Uncertainty Quantification in Scientific Computing: State-Aware Calibration of Computer Models* • **Salman Habib** – ANL *The Universe as a Statistical Inverse Problem*
12:00 – 13:30		Lunch
13:30 – 15:00	Session 7	**UQ for Complex Problems** Organizer: Peter Qian • **Derek Bingham** – Simon Fraser University *Bayesian model calibration for generalized linear models: An application in radiation transport* • **Antony Overstall** – University of Southampton *Bayesian Optimal Design for Ordinary Differential Equation Models* • **Peter Marcy** – LANL *Bayesian Gaussian Process Models on Spaces of Sufficient Dimension Reduction*
15:00-15:30		Coffee Break

15:30-17:00	Session 8	**Engineering Applications** Organizer: Roshan Joseph • **Simon Mak** – Georgia Tech *Physics-based modeling of large-eddy simulations for rocket injectors* • **Matthew Plumlee** – University of Michigan *Bayesian calibration of inexact computer models* • **Peter Qian** – University of Wisconsin *Invariance-Preserving Emulation for Computer Models, with Application to Structural Energy Prediction*
-- END --		

Abstracts

Monday, May 29, 2017

Session 1: Computer Experiments: Early Work and Recent Developments

Time: 9:00 am - 10:30 am

Organizer & Chair: Peter Qian, University of Wisconsin

Climate Model Emulation and Future Climate Simulation

Michael Stein
University of Chicago
stein@galton.uchicago.edu

This talk will discuss some general issues in the emulation of global climate models. In particular, although most such models are deterministic, their sensitivity to initial conditions makes them effectively stochastic. Furthermore, the temporal structure of the output means that even one run of a climate model can produce a lot of information on the relationship between inputs and outputs. These characteristics can lead to simple and effective emulators of some aspects of climate model output. I will also touch on approaches to simulating future climate under changing forcings, which, because climate models only capture some aspects of the actual climate accurately, is a very different problem than climate model emulation.

On Shapley value for measuring importance of dependent inputs

Art Owen
Stanford University
owen@stanford.edu

Variable importance in UQ is commonly defined by Sobol' indices or other measures derived from a functional ANOVA. When the inputs are dependent, then it becomes difficult to define an ANOVA. The Stone-Hooker solution is available, but it requires strong assumptions and can give negative importances. Song, Nelson and Staum (2016) proposed Shapley value for dependent inputs building on earlier work of the speaker on Shapley value for independent inputs. This talk investigates Shapley value further for the dependent variable case, working out some special cases. For instance, the domain of (x1,x2) may be so restricted that x2 cannot change at all for a given x1. That rules out pick-freeze approaches, but Shapley value still goes through. Shapley value also makes logically reasonable choices for cases where there are bijections between two variables or when one transforms the input space. This is joint work with Clementine Prieur, of the University of Grenoble.

Challenges in Designing and Using Simulator Experiments in Biomechanics and Biomaterials Research

Tom Santner

Ohio State University

santner.1@osu.edu

The goal of this talk is to describe several applications in the design of Biomechanical prosthetic systems for the treatment of diseased or damaged joints and the development of new biomaterials which can be thought of as the next generation of such treatments. Research in this area is often conducted using cadaver components in laboratory settings. Finite Element and other simulator models are used to describe the physical experimental results. Standard statistical methodology for combining simulator and physical experiments is used for interpolating or extrapolating the calibrated simulator to other environmental or other control variables settings. The limitations of this type of application are described.

Session 2: Calibration of Models under Uncertainty

Time: 11:00 am - 12:30 pm

Organizer & Chair: Matthew Plumlee, University of Michigan

An Empirical Adjustment of the Uncertainty Quantification in Gaussian Process Modeling

Daniel Apley

Northwestern University

apley@northwestern.edu

Gaussian process (GP) models have emerged as the standard surrogate models for deterministic computer response surfaces, due in part to their built-in mechanism for providing uncertainty quantification (UQ) in the form of prediction intervals on the response, in addition to the response prediction itself. However, their covariance parameter estimation methods tend to favor giving good response prediction, at the expense of poor UQ. We develop a post-processing method that takes a fitted GP model and its training data, and then empirically adjusts the built-in GP UQ so that it is in better agreement with the actual uncertainty in the response predictions. We demonstrate that this substantially improves the accuracy of the UQ, especially for nonstationary response surfaces that have more complex behavior in some input regions.

Advances in discretization uncertainty quantification for differential equation models

Oksana Chkrebtii

The Ohio State University

oksana@stat.osu.edu

When models are defined implicitly by systems of differential equations without a closed form solution, small local errors in finite-dimensional solution approximations can propagate into large deviations from the true underlying state trajectory. Inference for such models relies on a likelihood approximation constructed around a numerical solution, which underestimates posterior uncertainty. This talk will introduce and discuss progress in a new formalism for modeling and propagating discretization uncertainty through the Bayesian inferential framework, allowing exact inference and uncertainty quantification for discretized differential equation models.

Adjustments to Computer Models via Projected Kernel Calibration

Rui Tuo

Chinese Academy of Sciences

tuorui@gatech.edu

Identification of model parameters in computer simulations is an important topic in computer experiments. We propose a new method, called the projected kernel calibration, to estimate these model parameters. The proposed method is proven to be asymptotic normal and semi-parametric efficient. As a frequentist method, the proposed method is as efficient as the L2 calibration method proposed by Tuo and Wu (2015). On the other hand, the proposed method has a natural Bayesian version, which the L2 method does not have. This Bayesian version allows user to calculate the credible region of the calibration parameters without using a large sample approximation. We also show that, the inconsistency problem of the calibration method proposed by Kennedy and O'Hagan (2001) can be rectified by a simple modification of the kernel matrix.

Session 3: Applied Math, Engineering and Computation

Time: 14:00 pm - 15:30 pm

Organizer & Chair: Dave Higdon, Virginia Tech

Active Subspaces: Emerging Ideas for Dimension Reduction in Computational Science and Engineering Models

Paul Constantine

Colorado School of Mines

pconstan@mines.edu

Scientists and engineers use computer simulations to study relationships between a physical model's input parameters and its output predictions. However, thorough parameter studies---e.g., constructing response surfaces, optimizing, or averaging---are challenging, if not impossible, when the simulation is expensive and the model has several inputs. To enable parameter studies in these cases, the engineer may

attempt to reduce the dimension of the model's input parameter space. Active subspaces are part of an emerging set of subspace-based dimension reduction tools that identify important directions in the input parameter space. I will (i) describe computational methods for discovering a model's active subspaces, (ii) propose strategies for exploiting the reduced dimension to enable otherwise infeasible parameter studies, and (iii) review results from several science and engineering applications. For more information, visit activesubspaces.org Toward the goals of the workshop, I will compare and contrast deterministic approximation and statistical regression perspectives in both the problem formulation and the algorithm analysis. The difference in perspectives leads to dramatically different interpretations of the algorithm results---even if portions of the algorithms are identical.

Sampling strategies for efficient approximations

Akil Narayan
University of Utah
akil@sci.utah.edu

One main goal of parametric uncertainty quantification (UQ) is to understand how system states or responses depend on input parameters. Under some simplifications, this is essentially the problem of approximating a multivariate function whose inputs are the parameters. In practice, only point evaluations of the function can be collected. When parameters are high-dimensional, this poses a serious issue: Naïve strategies for resolving dependence on a high-dimensional parameter require an exponentially increasing number of samples. However, if the function has some structure (dimensional anisotropy, low-order interactions, sparse representations, etc.), then one can circumvent this exponential complexity. One of the main ingredients in such structure-exploiting approaches is a sampling strategy. Some modern methods have used importance-sampling-like procedures where non-standard sampling designs, randomized or deterministic, are devised in order to obtain optimal stability and convergence properties. We will give a brief overview of such designs in the context of polynomial approximation, which is a standard approximation strategy for parametric UQ problems.

Sparse polynomial approximation for high-dimensional uncertainty quantification

Clayton Webster
University of Tennessee and Oak Ridge National Lab
akil@sci.utah.edu

In this talk, we present an overview of sparse polynomial approximation, including interpolation, projection, discrete least squares, and compression techniques, applicable to a wide class of parameterized PDEs with both deterministic and stochastic inputs. Such methods construct an index set that corresponds to the "best s-terms," based on sharp estimates of the polynomial coefficients. In particular, we consider several cases of d-dimensional affine and non-affine coefficients, and prove analytic dependence of the PDE solution map in a polydisc or polyellipse of the multi-dimensional complex plane respectively. Several types of isotropic and anisotropic (weighted) multi-index sets are explored, and rigorous proofs reveal sharp asymptotic error estimates in which we achieve sub-exponential convergence rates with respect to the total number of degrees of freedom. Through several theoretical examples, we explicitly derive the rate constant and use the resulting sharp bounds to illustrate the effectiveness of our approach, as well as compare our rates of convergence with current published

results. Finally, computational evidence complements the theory and shows the advantage of our generalized methodology compared to previously developed estimates.

Session 4: UQ in Action

Time: 16:00 pm - 17:30 pm

Organizer & Chair: Ben Haaland, Georgia Tech

Big data gets physical

Hendrik Hamann

IBM T.J. Watson Research Center

hendrikh@us.ibm.com

While in the past most information on the internet was generated by humans or computers, with the emergence of the Internet of Things, vast amount of data is now being created by sensors from devices, machines etc., which are placed in the physical world. Here we present a series of example applications enabled by such sensor data and what we call "Physical Analytics", which provides the underlying intelligence for smarter applications using a combination of physical and statistical models. The smarter solutions, which are being presented in this talk, range from active energy management and optimization, environmental sensing and controls, precision agriculture to renewable energy forecasting. All these different applications have been built using a single platform, which is comprised of a set of "configurable" technologies components including ultra-low power sensing and communication, big data management technologies, numerical modeling for physical systems, machine learning based physical model blending, and physical analytics based automation and control. We will also discuss the uncertainly of the solutions presented.

Application of data-science approaches for the analysis of directional solidified ternary alloys

Johannes Hoetzer

University of Applied Sciences Karlsruhe

johannes.hoetzer@kit.edu

For the development of improved and new materials the understanding of the relationship between the process conditions, the chemical composition and the microstructure is of high interest. Depending on the evolving microstructure, different macroscopic material properties arise. One class of materials, which are excellently suited to study these relationships are directionally solidified ternary eutectic alloys, due to their wide range of arising microstructures. To study these materials, massive parallel large-scale phase-field simulations are applied. For the analysis of the resulting complex three-dimensional microstructures, different statistical methods are used. Applying principal component analysis (PCA) based on two-point correlations the necessary domain size is determined to gain structures unaffected by the boundary conditions and statistical volume elements. With this PCA method, also a quantitative agreement with experimental micrographs is revealed. To analyze the complex spatial arrangement of the solidified phases in simulations and experiments, a graph based approach is applied. This method allows to analyze the microstructure rearrangement mechanisms during the solidification, like nucleating, splitting, merging and overgrowing of the phases.

Robust Parameter Design using Computer Experiments

Bill Myers

The Procter & Gamble Company

myers.wr@pg.com

Computer experiments provide a competitive advantage in industry where fast and cost effective product development is critical. In many industrial applications computer experiments are replacing physical experiments because the physical creation and testing of prototypes is very prohibitive in terms of time and cost. Applying robust parameter design in computer simulation experiments has become important. This talk discusses new experimental design methods for computer simulations involving noise factors. Noise factors are difficult to control during normal process conditions or during product use. Most existing methods focus on uniformly distributing the points in the design space, which are not suitable for noise factors, because they usually follow non-uniform distributions such as normal distribution. This would suggest placing more points with in regions with high probability mass. However, noise factors also tend to have a smooth relationship with the response and therefore, placing more points towards the trails of the distribution is also useful for accurately estimating the relationship. These opposing effects make the experimental design methodology a challenging problem. We propose several possible solutions to this problem and demonstrate their advantages using simulated examples and a real industry example involving a manufacturing packing line.

Poster Presentation

Time: 17:30 pm - 19:00 pm

Nonparametric Functional Calibration of Computer Models

Andrew Brown

Clemson University

ab7@clemson.edu

Standard methods in computer model calibration treat the calibration parameters as constant throughout the domain of control inputs. In many applications, systematic variation may cause the best values for the calibration parameters to change across different settings. When not accounted for in the code, this variation can make the computer model inadequate. We propose a framework for modeling the calibration parameters as functions of the control inputs to account for a computer model's incomplete system representation in this regard, while simultaneously allowing for possible constraints imposed by prior expert opinion. We demonstrate how inappropriate modeling assumptions can mislead a researcher into thinking a calibrated model is in need of an empirical discrepancy term when it is only needed to allow for a functional dependence of the calibration parameters on the inputs. We apply our approach to plastic deformation of a visco-plastic self-consistent material in which the critical resolved shear stress is known to vary with temperature.

Modeling Precipitation Extremes using Log-Histospline

Whitney Huang

Purdue University

huang251@purdue.edu

One of the commonly used approaches to modeling univariate extremes is the peaks-over-threshold (POT) method. The POT method models exceedances over a (sufficiently high/low) threshold as a generalized Pareto distribution (GPD). To apply this method, a threshold has to be chosen and the estimates might be sensitive to the chosen threshold. Here we propose an alternative, the "Log-Histospline", to explore modeling the tail behavior and the remainder of the density in one step using the full range of the data. Log-Histospline applies smoothing spline on a finely binned histogram of the log transformed data to estimate its log density. By construction, we are able to preserve the polynomial tail behavior, a feature commonly observed in daily rainfall data. The Log-Histospline can be extended to the spatial setting by treating the marginal (log) density at each location as spatially indexed functional data, and perform a dimension reduction and spatial smoothing. We illustrate the proposed method by analyzing precipitation data from both regional climate model output (North American Regional Climate Change and Assessment Program (NARCCAP)) and weather stations in China.

Coupling computer models through linking their Gaussian process emulators

Ksenia Kyzyurova

Duke University/Statistical Science

ksenia.kyzyurova@gmail.com

Gaussian processes, together with an objective Bayesian implementation of the processes, have become a common tool for emulating (approximating) complex computer models. Sometimes more than one computer model needs to be utilized for the predictive goal. For instance, to model the true danger of a volcano pyroclastic flow, one might need to combine the flow model (which can produce the flow size and force at a location) with a computer model that provides an assessment of structural damage, for a given flow size and force. Direct coupling of computer models is often difficult for computational and logistical reasons. We propose coupling two computer models by linking independently developed Gaussian process emulators (GaSPs) of these models. The linked emulator results in a smaller epistemic uncertainty than a direct GaSP of the coupled computer model would have (if such a model were available). This feature is illustrated via simulations. The application of the methodology to complex computer models is demonstrated as well.

A Class of Nonseparable and Nonstationary Covariance Functions for Multi-output Gaussian Process

Pulong Ma

University of Cincinnati

mapn@mail.uc.edu

Gaussian process is widely used as a surrogate to emulate many computationally expensive simulators (computer models) in uncertainty quantification field due to its good theoretic properties and simplicity. The use of a separable covariance function in Gaussian process emulator is computationally convenient, which, however, ignores the interaction among input dimensions and performs very poor when quantifying uncertainties for computer model outputs. To allow efficient computation as well as quantify

uncertainty correctly, we build a multi-output Gaussian process with its covariance function coming from two parts, one of which is constructed from (modified) predictive process, and the other of which has separable form with any covariance function for each input subspace. The resulting covariance function for the multi-output Gaussian process emulator is nonseparable, nonstationary, and allows efficient computation for very large datasets. Bayesian inference including parameter estimation and prediction are derived for the proposed multi-output Gaussian process emulator, and its performance is demonstrated with simulation examples and real data analysis.

Global Optimization of Expensive Functions Using Adaptive RBF-Based Surrogate Model Via Uncertainty Quantification

Ray-Bing Chen

National Cheng Kung University

rbchen@mail.ncku.edu.tw

Global optimization of expensive functions has important applications in physical and computer experiments. It is a challenging problem to develop efficient optimization scheme, because each function evaluation is costly and the derivative information of the function is usually not available. We propose a novel global optimization framework using adaptive Radial Basis Functions (RBF) based surrogate model via uncertainty quantification. The framework consists of two iteration steps. It first employs an RBF-based Bayesian surrogate model to approximate the true function, where the parameters of the RBFs can be adaptively estimated and updated each time a new point is explored. Then it utilizes a model-guided selection criterion to identify a new point from a candidate set for function evaluation. The selection criterion incorporates the expected improvement (EI) of function prediction and its uncertainties. We conduct numerical studies with standard test functions to illustrate the performance of the proposed method.

A Bayesian Approach to Model Inadequacy

Brad Marvin

The University of Texas at Austin

brad_marvin@utexas.edu

In this poster we cast the model inadequacy problem into a Bayesian inverse problem. We postulate that the discrepancy between the PDE and true (physical) solutions can be described by a Gaussian process. Following a Bayesian approach, we assume a Gaussian prior measure on the discrepancy and incorporate observation data to form a posterior mea- sure. The posterior mean can be used to make predictions about the true state and quantities of interest while the posterior covariance allows us to quantify the uncertainty in those predictions. We provide a rigorous construction and analysis of the proposed approach in the infinite dimensional setting for linear PDEs. The method is tested on an elliptic problem in one, two, and three spatial dimensions solved using the finite element method. Ongoing work will extend the approach to hyperbolic PDEs and nonlinear PDEs.

Stochastic agent based model calibration and uncertainty quantification

Arindam Fadikar

Biocomplexity Institute of Virginia Tech

afadikar@vt.edu

Dynamic network simulation models can evolve a very large system of agents over a network that changes with time. Such models are often used to simulate epidemics or transportation, typically producing random trajectories, even when model parameters and initial conditions are identical. This introduces a number of challenges in designing ensembles of model runs for sensitivity analysis and computer model calibration. Gaussian process meta modelling combining observations and experimental data is a popular and well-known technique for uncertainty quantification task, but it requires significant amount of computing power when model dimension increases; it also has difficulty with random output. The present work focuses on a novel model calibration technique based on a sequential monte-carlo approach, to an epidemic simulator, seeking to forecast the epidemic's behavior by geographic region. Based on a simple and interpretable importance sampling approach, this method sequentially selects plausible model parameter settings from a pool of design points, enabling exploration of the posterior and direct description of prediction uncertainty.

Exploiting Ridge Structure to Solve Chance-Constrained Design Under Uncertainty Problems

Jeffrey Hokanson

Colorado School of Mines

hokanson@mines.edu

A chance constrained design under uncertainty problem seeks to minimize an objective function over a set of design variables subject to the constraint that failure functions of both design and random variables are below some threshold with a specified probability. This optimization problem poses significant difficulty because in the absence of exploitable structure evaluating the chance constraints requires thousands of model evaluations for each candidate design. Here we introduce an approach that removes this requirement provided that failure functions can be well approximated by ridge functions – functions that act on a low-dimensional subspace of the input parameters. When this is the case, the ridge structure makes the random variables act as an additive perturbation to the low-dimensional input of the ridge function. This permits the chance constraints to be replaced by a constraint on the design variables alone, removing the stochastic component. Without the random variables, solving this simplified optimization problem is inexpensive. We demonstrate the effectiveness of this technique on both a simple test problem from the literature and a jet nozzle design application.

Analysis of dimension reduction in Gaussian process regression

Minyong Lee

Stanford University

minyong@stanford.edu

We analyze the quality of dimension reduction using projected Gaussian processes. We quantify the proportional variance of the Gaussian process explained by the dimensionally reduced space. This helps us choose the dimension of the dimensionally reduced space.

Hidden Parameter Hypothesis Testing

Zachary del Rosario
Stanford University
zdr@stanford.edu

Dimensional Analysis is a powerful tool that arises from the fundamental principle of dimensional homogeneity. Its power lies in its generality; from no more than a list of physical inputs, one can derive the governing dimensionless parameters for a Quantity of Interest. However, the correct identification of relevant input quantities is not obvious. We present a technique for doing data-driven identification of hidden parameters -- crucial parameters which affect a Quantity of Interest, but are missing from the analysis. This technique relies on recent developments in Dimensional Analysis, posing the Buckingham Pi theorem in ridge function form, and uses bootstrap hypothesis testing to test the null hypothesis of no hidden parameters. We present our hidden parameter formalism and illustrative results.

Inverse Regression for Ridge Recovery

Andrew Glaws
Colorado School of Mines
aglaws@mines.edu

We investigate the application of sufficient dimension reduction (SDR) to a deterministic function of several variables. In this context, SDR provides a framework for ridge recovery. A scalar function of m inputs is referred to as a ridge function when it may be expressed as a function of n < m linear combinations of the original inputs. The ridge directions contain the weights of these linear combinations. Ridge recovery methods attempt to discover the ridge subspace spanned by the ridge directions. Many ridge recovery methods, such as active subspaces, require the gradient of the function relative to the original inputs. Such gradient information may be unavailable, and methods for approximating the gradient (e.g., finite differences) might be prohibitively expensive. Sufficient dimension reduction (SDR) provides a theoretical framework for dimension reduction in statistical regression. In contrast to deterministic functions, regression problems contain noise in the relationship between the predictors and response—i.e., multiple values of y are possible for a single x. SDR defines a subspace of the predictors in which no loss of statistical information occurs to the conditional random variable y|x. This is accomplished by enforcing conditional independence of the response and predictors given the reduced predictors. We prove that, given the deterministic relationship y = f(x), the dimension reduction subspaces from SDR are equivalent to the ridge subspaces from ridge recovery. We study two inverse regression methods for SDR—sliced inverse regression (SIR) and sliced average variance estimation (SAVE)—that use predictor/response pairs to search for dimension reduction subspaces. These methods require no gradient information, which makes them promising candidates for gradient-free approaches to ridge recovery. We show that, from the proper perspective, these algorithms may be used to approximate ridge subspaces and provide convergence results. The methods are implemented on a physically-motivated test problem to illustrate their usefulness.

Practical heteroskedastic Gaussian process modeling for large simulation experiments

Mickael Binois

University of Chicago Booth School of Business

mbinois@chicagobooth.edu

We present a unified view of likelihood based Gaussian progress regression for simulation experiments exhibiting input-dependent noise. Replication plays an important role in that context, however previous methods leveraging replicates have either ignored the computational savings that come from such design, or have short-cut full likelihood-based inference to remain tractable. Starting with homoskedastic processes, we show how multiple applications of a well-known Woodbury identity facilitate inference for all parameters under the likelihood (without approximation), bypassing the typical full-data sized calculations. We then borrow a latent-variable idea from machine learning to address heteroskedasticity, adapting it to work within the same thrifty inferential framework, thereby simultaneously leveraging the computational and statistical efficiency of designs with replication. The result is an inferential scheme that can be characterized as single objective function, complete with closed form derivatives, for rapid library-based optimization. Illustrations are provided, including real-world simulation experiments from manufacturing and the management of epidemics.

Establishing Design Continuity in Pareto Fronts with Active Subspaces

Zachary Grey

Colorado School of Mines

zgrey334@gmail.com

Multi-objective shape optimization problems involve parameterizing a shape and modifying the shape to minimize a set of objectives. However, as in the case of transonic flow regimes, the computational cost of approximating a partial differential equation to compute quantities of interest is prohibitive. Response surfaces and surrogate modeling methods enable computationally efficient approximations of quantities of interest. These surrogate methods benefit from dimension reduction of the parameter space. A reduced dimension active subspace enables dimension reduction of the parameter space based on a particular quantity of interest. Changing parameters in the active subspace change the quantities more, on average, than changing parameters orthogonally to the active subspace. The active subspace also characterizes sensitivity information for each quantity of interest. We quantify the active subspaces present in transonic lift and drag coefficients and relate the subspaces to physical characteristics of the shape parameterization. Using the resulting subspaces, we characterize continuous shape perturbations over a Pareto frontier.

Highly efficient spatiotemporal flow dynamics prediction via data-driven analysis and LES-based surrogate model

Yu-Hung Chang

Georgia Institute of Technology

ychang99@gatech.edu

This interdisciplinary study, which combines machine-learning techniques, statistical methodology, and physics analysis, demonstrates a paradigm of design strategy for swirl injector design in new generation. Large eddy simulation has been widely used to simulate flow physics and combustion characteristics inside rocket engines for decades; however, it consumes great amount of time and resources, which is

impractical for design purposes. Our previous study proposed Common-grid Proper Orthogonal Decomposition (CPOD), which uses a physics-based surrogate model to predict mean flow mechanism efficiently with statistical analysis for design exploration in a broad geometry range. As the successor for CPOD, Kernel-smoothed POD (KSPOD) is proposed this time. KSPOD is trained by high-fidelity simulation datasets to predict spatiotemporal flow dynamics of a swirl injector via utilizing Kriging based weight function from design matrix. The implemented machine learning concept enhances the model's natural applicability in emulation provides an improvement over POD. This model performs well for the analytical estimation of the performance measures such as liquid film thickness and spreading angle. It also well captures shapes of flowfield; especially the subtle details of traveling vortex structures. Moreover, power spectrum densities based on predicted flowfields POD coefficients reveals the dominant frequencies as same as the frequencies from simulation POD coefficients. To contrast, the turnaround computation time for evaluating new design points is reduced significantly compared with other algorithms and over 42,000 times faster than simulation.

Bayesian Uncertainty Quantification of Biological Networks

Karen Larson

Brown University

karen_larson@brown.edu

We will present a Bayesian uncertainty quantification tool that is applicable to many dynamic processes in networks. We use transitional Monte Carlo Markov Chains (TMCMC) to populate the posterior distribution for the model parameters. The method is highly parallelizable and can be used on computationally complex systems. While formulating the posterior distributions, the TMCMC algorithm finds the evidence for the model conditioned on the measured data. Thus, it is also relevant to Bayesian model selection. The application we will look at is one-dimensional blood flow in arterial networks. This is also related to other work we have done on shape optimization for lipid bilayer membranes by finding the angle of the lipid tails to take on a specific shape and investigating phase transitions of epidemic spread on networks. This work is done jointly with Anastasios Matzavinos, Petros Koumoutsakos, Costas Papadimitriou, and Panagiotis Hadjidoukas. This research was supported in part by NSF CDS&E-MSS 1521266 and NSF CAREER 1552903.

Bayesian Inference of Material Constitutive Properties from Instrumented Spherical Indentation Experiments using Kriging Finite Element Model Surrogates

Patxi Fernandez-Zelaia

Georgia Institute of Technology

pfz3@gatech.edu

Instrumented indentation testing enables the characterization of mechanical behavior in materials across multiple length scales. The heterogeneous deformation fields beneath the indenter however make difficult the inference of intrinsic constitutive properties. Therefore, as constitutive properties cannot be directly inferred from experiments, model-based inverse methods must be employed to estimate the unknown properties. Finite element analysis of the indentation process may be used to accurately emulate and study physical experiments yet are prohibitively costly for use in inverse analysis. Therefore we propose an inverse strategy which employees a kriging surrogate model, a class of Gaussian Process models, trained from expensive finite element model evaluations, to solve the inverse problem. The inverse solution is obtained using a Bayesian framework and Markov Chain Monte Carlo sampling strategy to infer the

posterior probability densities of the underlying underlying constitutive properties. A sequential design strategy based on Bayesian decision theory is utilized to intelligently augment the surrogate model near potential solutions.

Calibrating material properties with functional output using Bayesian model calibration

Lauren Hund

Sandia National Laboratories

lbhund@sandia.gov

In experiments conducted on the Z-machine at Sandia National Laboratories, dynamic material properties cannot be analyzed using traditional analytic methods, necessitating solving an inverse problem. Bayesian model calibration is a statistical framework for solving an inverse problem to estimate parameters input into a computational model in the presence of multiple uncertainties. Disentangling input parameter uncertainty and model misspecification is often poorly identified problem. When using computational models for physical parameter estimation, the issue of parameter identifiability must be carefully considered to obtain accurate and precise estimates of physical parameters. Additionally, in dynamic material properties applications, the experimental output is a function, velocity over time. While we can sample an arbitrarily large number of points from the measured velocity, these curves only contain a finite amount of information about the calibration parameters. In this report, we propose modifications to the Bayesian model calibration framework to simplify and improve the estimation of physical parameters with functional outputs. Specifically, we propose scaling the likelihood function by an effective sample size rather than modeling the discrepancy function; and modularizing input nuisance parameters with weakly identified parameters. We evaluate the performance of these proposed methods using a statistical simulation study and then apply these methods to estimate parameters of the tantalum equation of state. We conclude that these proposed methods can provide simple, fast, and statistically valid alternatives to the full Bayesian model calibration procedure; and that these methods can be used to estimate parameters of the equation of state for tantalum.

Controlling Sources of Inaccuracy in Stochastic Kriging

Wenjia Wang

Georgia Institute of Technology

wenjiawang@gatech.edu

Computer experiments have become ubiquitous in science and engineering. For expensive or time-consuming stochastic simulations, stochastic kriging (Ankenman, et al. (2010)) is commonly used to generate predictions. Here, we decompose error in stochastic kriging predictions into nominal, numeric, parameter estimation and parameter estimation numeric components and provide means to control each in terms of properties of the underlying experimental design. The design properties implied for each source of error are weakly conflicting and several broad principles are proposed. In brief, the space-filling properties "small fill distance" and "large separation distance" should balance with replication at unique input configurations, with number of replications depending on the relative magnitudes of the stochastic and process variability, while non-stationarity implies more input density in more active regions and regression functions imply a push towards balancing with traditional design properties. This work can be applied to the deterministic case proposed in Haaland et al. (2014). A few examples are presented to illustrate the results.

Asymptotics of Bayesian Risk Formulations For Data-driven Stochastic Optimization

Di Wu

Georgia Institute of Technology

dwu80@gatech.edu

In stochastic simulation, input distributions are the distributions (of stochastic uncertainty) fed into a simulation model to drive the simulation process. Since input distributions are usually estimated from finite data, the model is subject to input uncertainty caused by finite-sample error. As a result, optimizing the model may yield solutions that perform poorly under the true input distributions. In this paper, we consider a case where input distributions have known parametric forms with unknown parameters. To hedge against the risk of input uncertainty, we optimize the "risk formulations", which is to minimize the risk measures of mean output with respect to the unknown parameters' posterior distribution. Of our particular interest are four risk measures: mean, mean-variance, value-at-risk and conditional value-at-risk. We establish the consistency of objective functions and optimal solutions, as well as the asymptotic normality of objective functions and optimal values. Moreover, our analysis reveals the hidden intuition of risk formulations: the risk formulations can be approximately viewed as a weighted sum of posterior mean performance and the (squared) half-width of the true performance's confidence interval. As is shown by our numerical examples, optimizing risk formulations leads to an explicit tradeoff between average performance and the risk of actual performance.

A Generalized Gaussian Process Model for Computer Experiments with Binary Time Series

Chih-Li Sung

Georgia Institute of Technology

csung33@gatech.edu

Conventional analysis for computer experiments is based on Gaussian process (GP) models. Non-Gaussian observations such as binary responses are common in some computer experiments, but the extensions of GP models to these cases have received scant attention in the literature. Motivated by the analysis of a class of cell adhesion experiments, we introduce a generalized Gaussian process model for binary responses, which shares some common features with standard GP models. In addition, the proposed model incorporates a flexible mean function that can capture different types of time series structures. Asymptotic properties of the estimators are derived and their performance is examined via a simulation study. An optimal predictor and its predictive distribution are constructed based on the proposed model. The methodology is applied to study two different cell adhesion mechanisms, which were conducted by computer simulations. The fitted models reveal important biological differences between the two mechanisms in repeated bindings, which cannot be directly observed experimentally.

Effective Model Calibration via Sensible Variable Identification and Adjustment, with Application to Composite Fuselage Simulation

Xiaowei Yue

Georgia Institute of Technology

xwy@gatech.edu

Estimation of model parameters of computer simulations, also known as calibration, is an important topic in many engineering applications. In this paper, we consider the calibration of computer model

parameters with the help of engineering design knowledge. We introduce the concept of sensible (calibration) variables. Sensible variables are model parameters which are sensitive in the engineering modeling, and whose optimal values are different from the pre-specified design values. We propose an effective calibration method to identify and adjust the sensible variables with limited physical experimental data. A numerical study and a composite fuselage simulation example show the effectiveness and efficiency of the proposed method.

Interleaved lattice-based minimax distance designs

Xu He

Chinese Academy of Sciences, Beijing

hexu@amss.ac.cn

We propose a new method to construct minimax distance designs, which are useful for computer experiments. To circumvent computational difficulties, we consider designs with interleaved lattice structure, a newly defined class of lattice that has repeated or alternated layers based on any single dimension.

Quantifying Uncertainty in Random Forest Predictions

Tim Coleman

University of Pittsburgh

tsc35@pitt.edu

Random forests and related supervised learning ensembles remain popular and robust tools in the scientific community due to their straightforward implementation and high predictive accuracy in a variety of settings. This work introduces a modification to these traditionally bootstrap-based procedures by instead incorporating a subsampling scheme during model construction. This careful adjustment allows for a precise structure on final predictions, which we demonstrate can be treated as U-statistic extensions allowing for a formal means by which predictive uncertainty can be precisely quantified. Specifically, predictions at fixed test locations can be shown to be asymptotically normal allowing for point-wise confidence (prediction) intervals as well as formal hypothesis testing procedures to compare nested forest models and ultimately assess variable importance. Importantly, these procedures can be fully parameterized and the relevant parameters estimated within the ensemble itself, allowing for such inferential results without resorting to additional layers of bootstrapping or further resampling. We apply these procedures to tree swallow (Tachycineta bicolor) migration data sourced from the e-Bird Citizen Science project, which tracks sightings of migratory birds throughout North America, and demonstrate the significance of 'maximum daily temperature' in predicting tree swallow abundance during fall migration.

Robust Uncertainty Quantification for Deterministic Interpolation

Joon Jin Song
Baylor University
Joon_Song@baylor.edu

Deterministic spatial interpolators, like inverse distance weighting (IDW) and regression-based inverse distance weighting (RIDW), offer flexible and fast spatial prediction, but fail to provide estimate of uncertainty. In addition, real-world data often contain outliers that lead to unexpected impact on spatial prediction and uncertainty quantification. In this study, we propose a robust and flexible stochastic uncertainty attachment scheme for spatial deterministic interpolators. The proposed method is applied to radar rainfall estimation and compared with some existing methods.

Model Calibration with Censored Data

Fang Cao
Georgia Institute of Technology
fcao6@gatech.edu

The purpose of model calibration is to make the model predictions closer to reality. The classical Kennedy-O'Hagan approach is widely used for model calibration, which can account for the inadequacy of the computer model while simultaneously estimating the unknown calibration parameters. In many applications, the phenomenon of censoring occurs when the exact outcome of the physical experiment is not observed, but is only known to fall within a certain region. In such cases, the Kennedy-O'Hagan approach cannot be used directly, and we propose a method to incorporate the censoring information when performing model calibration. The method is applied to study the compression phenomenon of liquid inside a bottle. The results show significant improvement over the traditional calibration methods, especially when the number of censored observations is large.

Tuesday, May 30, 2017

Session 5: UQ in Emulation, Simulation, and Calibration

Time: 8:30 am - 10:00 am

Organizer & Chair: V. Roshan Joseph, Georgia Tech

On Fitting Gaussian Process Emulators

Jim Berger

Duke University

berger@stat.duke.edu

We discuss estimation of the correlation parameters of a Gaussian Stochastic Process (GaSP), in the context of emulation (approximation) of computer models. Often only a moderate number of observations (computer model runs) are available and the likelihood function for the correlation parameters will then frequently be problematical, in that the likelihood for some correlation parameters will be maximized at 0 or at infinity. This makes maximum likelihood estimation or full Bayesian analysis over the correlation parameters problematical (e.g., constant priors do not work, resulting in improper posterior distributions). Reference priors for the correlation parameters are available and do result in good posterior distributions for the correlation parameters; alas, the reference priors are computationally challenging for full Bayesian analysis. Reference priors can, however, be efficiently utilized for posterior mode estimation of the correlation parameters (actually, the marginal posterior mode with mean and variance parameters integrated out). The choice of parameterization for the posterior mode estimation is, however, critical, and a seemingly optimal choice will be discussed. The resulting methodology is implemented in a new R package, RobustGaSP.

Bayesian variational approaches for inverse problems

Bani Mallick

Texas A&M University

bmallick@stat.tamu.edu

We consider a Bayesian approach to inverse problems with complex error structure. A hierarchical Bayesian model is developed in this inverse problem setup. The Bayesian approach contains a natural mechanism for regularization in the form of a prior distribution. Different regularized prior distributions have been used to strongly induce sparseness. We propose a variational type algorithm by minimizing the Kullback-Liebler divergence between the true posterior and a separable approximated one. The proposed method is illustrated on several two dimensional linear and nonlinear inverse problems.

Reducing Simulation Input-Model Risk via Input Model Averaging

Barry Nelson

Northwestern University

nelsonb@northwestern.edu

Input uncertainty is an aspect of simulation model risk that arises when the driving input distributions are derived or "fit" to real-world, historical data. Examples of input processes are arrivals and service times in queueing simulations, bed occupancy times in hospital simulations, and times to machine failure in manufacturing simulations. While there has been significant progress on quantifying and hedging against input uncertainty, there has been no direct attempt to reduce it. In this paper we show that frequentist model averaging can be a provably effective way to create input models that better represent the true, unknown input distributions, thereby reducing model risk. Input model averaging builds from standard input modeling practice, and requires no change in how the simulation is executed nor any follow-up experiments. We provide theoretical and empirical support for our approach. This is joint work with Alan T K Wan and Shuqin Fan of City University of Hong Kong, and Xinyu Zhang of the Chinese Academy of Sciences.

Session 6: UQ and Model-based Applications

Time: 10:30 am - 12:00 pm

Organizer: Dave Higdon, Virginia Tech

Chair: Ying Hung, Rutgers University

Statistics-Based Compression of Global Wind Fields

Marc Genton

King Abdullah University of Science and Technology

marc.genton@kaust.edu.sa

Wind has the potential to make a significant contribution to future energy resources; however, the task of locating the sources of this renewable energy on a global scale with climate models, along with the associated uncertainty, is hampered by the storage challenges associated with the extremely large amounts of computer output. Various data compression techniques can be used to mitigate this problem, but traditional algorithms deliver relatively small compression rates by focusing on individual simulations. We propose a statistical model that aims at reproducing the data-generating mechanism of an ensemble of runs by providing a stochastic approximation of global annual wind data and compressing all the scientific information in the estimated statistical parameters. We introduce an evolutionary spectrum approach with spatially varying parameters based on large-scale geographical descriptors such as altitude to better account for different regimes across the Earth's orography. We consider a multi-step conditional likelihood approach to estimate the parameters that explicitly accounts for nonstationary features while

also balancing memory storage and distributed computation, and we apply the proposed model to more than 18 million points on yearly global wind speed. The proposed model achieves compression rates that are orders of magnitude higher than those achieved by traditional algorithms on yearly- averaged variables, and once the statistical model is fitted, decompressed runs can be almost instantaneously generated to better assess wind speed uncertainty due to internal variability. This is based on joint work with Jaehong Jeong, Stefano Castruccio, and Paola Crippa.

Model Calibration, Validation, and Uncertainty Quantification in Scientific Computing: State-Aware Calibration of Computer Models

Sez Atamturktur

Clemson University

sez@clemson.edu

Computational models developed to predict mechanical behavior of engineering systems define a representation of physics behavior linking the state variables (loads, temperature, etc.) to relevant output responses (deformation, stresses, etc.) with the goal of simulating the behavior of the system under various operational conditions. Representation of physics behavior requires additional input parameters to be defined that reflect the properties of the engineering system. Model calibration is the area concerned with determining the appropriate values of these input parameters. Standard methods in computer model calibration treat the calibration parameters as constant throughout the domain of control inputs. In many applications, however, systematic variation not accounted for in the computer code may cause the best values for the calibration parameters to change between different experimental settings. Problems arise when such relationships are unforeseen or insufficiently understood. Often, such behavior tends to be ignored so that an overly simplistic model is used. This problem is exacerbated as the complexity of the system increases and as the available knowledge of the associated mechanistic behavior becomes increasingly insufficient. In this talk, I will present a framework for modeling the calibration parameters as systematic functions of the control inputs to account for a computer model's incomplete system representation while simultaneously allowing for possible constraints imposed by prior expert opinion. We refer to this approach as State-Aware Calibration. In our proposed methodology, experiments are used to infer important relationships between model input parameters and state variables. Hence, State-Aware Calibration provides the capability to gain insight concerning the underlying and mechanistically-relevant physical processes absent from physics-based models. The major benefit of state-aware calibration is that it can guide model developers by pinpointing the most influential, mechanistically-relevant processes that are absent from the models, thereby offering a definitive guide to the prioritization of future model developments as well as the efficient allocation of limited resources. Through state-aware calibration, we expect to further the engineer's knowledge of the physics principles governing the system behavior of interest, as well as to significantly reduce the remaining systematic bias between model predictions and experimental observations. We demonstrate the feasibility and performance of our approach through simulation as well as application to a study of plastic deformation of a visco-plastic self-consistent material (5182 aluminum alloy) in which the critical resolved shear stress parameter is known to vary with temperature.

The Universe as a Statistical Inverse Problem

Salman Habib

Argonne National Laboratory

habib@anl.gov

Modern cosmology has made rapid strides in the last two decades. The advent of multi-wavelength wide and deep surveys of the sky has allowed measurements of a large number of summary statistics and cross-correlations with small statistical errors and these are expected to become even smaller in the near future. The challenge is to convert these measurements into robust scientific inferences that can shed some light on a number of cosmic puzzles, such as the cause of cosmic acceleration, the nature of dark energy, and the origin of primordial fluctuations that have seeded the current large-scale structure of the Universe. Additionally, cosmological observations can provide measurements of fundamental parameters, such as the sum of neutrino masses. I will discuss some of the general approaches and challenges in the area of cosmological inference, and then focus on emulation as a forward modeling strategy with which to address these challenges.

Session 7: UQ for Complex Problems

Time: 13:30 pm -15:00 pm

Organizer: Peter Qian, University of Wisconsin
Chair: Andrew Brown, Clemson University

Bayesian model calibration for generalized linear models: An application in radiation transport

Derek Bingham

Simon Fraser University

dbingham@stat.sfu.ca

Using computer models to explore physical systems is commonplace in scientific applications. Statistical computer model calibration uses outputs from a simulator and field data to estimate unknown parameters that govern the computer model and to build a predictive model for the system. The conventional approach to model calibration assumes that the observations are continuous outcomes. In many applications this is not the case. In this talk, new methodology that allows the conventional calibration framework to be embedded into different Bayesian hierarchical models that reflect the structure of the observation error and scientific understanding of the generating processes is presented. The proposed methodology was motivated by an application in modeling photon counts at the Center for Exascale Radiation Transport at Texas A&M University. There, high performance computing is used for simulating the flow of neutrons through graphite and a requirement to characterize the distribution of impurities within the graphite.

Bayesian Optimal Design for Ordinary Differential Equation Models

Antony Overstall

University of Southampton

A.M.Overstall@soton.ac.uk

Bayesian optimal design is considered for physical models derived from the (intractable) solution to a system of ordinary differential equations (ODEs). Bayesian optimal design requires the minimisation of the expectation (over all unknown and unobserved quantities) of an appropriately chosen loss function. This can be non-trivial due to 1) the high dimensionality of the design space; and 2) the intractability of the expected loss. In this case, a further complication arises from the intractability of the solution to the system of ODEs. We propose a strategy that employs a modification of the continuous coordinate exchange algorithm where a statistical emulator is employed to approximate the expected loss function, and a probabilistic solution to the system of ODEs. The strategy is demonstrated on several illustrative examples from the biological sciences.

Bayesian Gaussian Process Models on Spaces of Sufficient Dimension Reduction

Peter Marcy

Los Alamos National Laboratory

pmarcy@lanl.gov

It is often known that important directions within the input space of Gaussian process (GP) regression models do not align with the original coordinate directions. It might be the case that changes along unknown linear combinations of the inputs effect the biggest changes in the response variables. There are two goals associated with this GP model: (1) the usual goal of in-fill prediction, i.e. "emulation", and (2) inferring a sufficient dimension reduction space, along with uncertainty of the linear combinations. In this talk I present a Bayesian formulation of the problem and describe the challenges to inference. I then describe a general Metropolis-Hastings algorithm for exploring the posterior, which is known to lie on a Stiefel manifold. The methodology is illustrated using data from a complex computer model.

Session 8: Engineering Applications

Time: 15:30 pm -17:00 pm

Organizer: V. Roshan Joseph, Georgia Tech
Chair: Lulu Kang, Illinois Institute of Technology

Physics-based modeling of large-eddy simulations for rocket injectors

Simon Mak

Georgia Institute of Technology

smak6@gatech.edu

In the quest for advanced propulsion systems, a new design methodology is needed which combines engineering physics, computer simulations and statistical modeling. There are two main challenges for

such a task: the simulation of high-fidelity flow data is computationally expensive, and the analysis and modeling of this data requires both physical insights and statistical tools. This talk is divided into three short parts, each addressing these challenges from a different perspective. First, a surrogate model is presented for efficient prediction of turbulent flows in swirl 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 turn-around times, and also reveals interesting flow physics which can guide further investigations. Next, a flame transfer function model is proposed which quantifies the effect of flow field variables on unsteady heat release in a rocket injector. Such a model is useful not only for analyzing the stability of an injector design, but also identifies key physical mechanisms which contribute to combustion instability. This talk concludes with some ongoing work on calibration and forward UQ propagation in the context of turbulent flow simulation.

Bayesian calibration of inexact computer models

Matthew Plumlee

University of Michigan

mplumlee@umich.edu

Bayesian calibration is used to study computer models in the presence of both a calibration parameter and model bias. The parameter in the predominant methodology is left undefined. This results in an issue where the posterior of the parameter is sub-optimally broad. There have been no generally accepted alternatives to date. This talk proposes and studies a relatively straightforward fix for Bayesian calibration where the prior distribution on the bias is made orthogonal to the gradient of the computer model. Problems associated with Bayesian calibration are shown to be mitigated through analytic results in addition to examples.

Invariance-Preserving Emulation for Computer Models, with Application to Structural Energy Prediction

Peter Qian

University of Wisconsin-Madison

peter.qian@wisc.edu

Computer models with invariance properties appear frequently in materials science, physics, biology and other fields. These properties are consequences of dependency on structural geometry, and cannot be accommodated by standard emulation methods. We propose a new statistical framework for building emulators to preserve invariance. This framework uses a weighted complete graph to represent the geometry and introduces a new class of function, called the relabeling symmetric functions, associated with the graph. We establish a characterization theorem of the relabeling symmetric functions, and propose a nonparametric kernel method for estimating such functions. The effectiveness of the proposed method is illustrated by several examples from materials science.

Notes