

USACM THEMATIC WORKSHOP ON UNCERTAINTY QUANTIFICATION AND DATA- DRIVEN MODELING

FINAL REPORT

May 1, 2017

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND???????

Overview and Purpose of the Workshop

The USACM Thematic Workshop on Uncertainty Quantification and Data-Driven Modeling was held on March 23-24, 2017, in Austin, TX. The organizers of the technical program were James R. Stewart of Sandia National Laboratories and Krishna Garikipati of University of Michigan. The administrative organizer was Ruth Hengst, who serves as Program Coordinator for the USACM. The organization of this workshop was coordinated through the USACM Technical Thrust Area on Uncertainty Quantification and Probabilistic Analysis. The workshop website (<http://uqpm2017.usacm.org>) includes the presentation agenda as well as links to several of the presentation slides (permission to access the presentations was granted by each of those speakers, respectively). Herein, this final report contains the complete workshop program that includes the presentation agenda, the presentation abstracts, and the list of posters.

The purpose of this workshop was to bring together leading experts in uncertainty quantification, statistics, computer science, and computational science to discuss new research ideas in data-driven modeling. With UQ now established as a core area of computational science and engineering, data science is quickly emerging as a critical accompanying technology area for enabling validated, predictive simulations. However, data science is still in its infancy and new ideas are needed, especially in the context of UQ for multi-scale, multi-physics science and engineering applications.

Wide-ranging physical phenomena, from turbulent flows through materials physics, exhibit emergent behavior that can only be understood and predicted by detailed multi-scale models. Existing multi-scale models based on first principles equations, and physical approximations thereof, tend to be either prohibitively expensive or inaccurate. However, recent advances at the intersection of data and computational science have created new opportunities to meet the long-standing challenge of delivering quantitative predictivity in multi-scale computational physics. This has led to a new paradigm of data-driven model development. In addition to classical methods from inverse problems, this new framework brings to bear Bayesian inference, stochastic simulation, reduced order modelling, compressed sensing as well as advances in machine learning. Of particular interest are approaches by which the above and related techniques can enhance the predictivity of large scale computational physics.

Through this workshop, we hoped to highlight new advances in UQ and data-driven modeling that will lead to breakthroughs in our ability to develop predictive models for realistic applications. Areas of interest included, but were not limited to

- Deterministic and stochastic inverse problems
- Data assimilation methodologies, particularly in the context of datasets with a high degree of noise and/or poorly characterized uncertainties
- Machine learning and data mining methods to inform model development
- Inference of models, including parameters and functional forms
- Validation and data-informed predictions for science and engineering applications
- Uncertainty quantification for large-scale, high-dimensional problems

Workshop Participation and DOE Financial Support

A total of 90 people registered for and attended the workshop. The photo below shows the workshop venue at capacity. The registration fee of \$320 covered breakfast, lunch, and coffee breaks on each day, the dinner banquet on the first day, printing costs for the registration materials, and rental of the workshop venue. The workshop format consisted of 20 talks (25 minutes each) organized over six sessions, as well as 34 posters organized into one poster session. The poster session started with a lively one-minute poster introduction by each of the poster presenters.



DOE's Office of Advanced Scientific Computing Research (ASCR) provided financial support for students and postdocs to attend the workshop. A total of 10 postdocs and 27 graduate students, from 19 different (U.S.-based) institutions, received this support; the full list of award recipients is included at the end of this report. The support covered workshop registration, lodging, and airfare for each recipient (with the exception that four of the recipients were local and only required workshop registration expenses). Each of the financial award recipients was required to present a poster at the workshop. We received 43 applications for this financial support, which put us in the difficult position of turning down some very deserving applicants.

Workshop Summary and Path Forward

The final session of the workshop was an open discussion where all participants were given the opportunity to provide feedback and comments. The following major themes and observations, arising from the presentations and the posters, were identified:

- Model selection;
- A divide between the machine learning and computational science & engineering communities;
- Coupling physics modeling with machine learning;
- Discussions of “big data” vs. “little data”;
- Combining information from different sources: high vs. low fidelity;
- Reduced-order modeling – discovering lower dimensional manifolds;
- Need for taxonomy;
- Structure preservation – how does/can data preserve structure in models?
- Sequential data assimilation, e.g., Kalman filtering;
- Data regularization;
- Modeling/predicting rare events;
- UQ benchmarks (both positive and negative experiences were expressed).

Notable missing themes (which could be incorporated into future workshops) included:

- Data wrangling; data characterization; physics identification;
- Model robustness against data;
- Code implementation details and performance;
- Identifiability in Bayesian analysis; e.g., simultaneous parameter uncertainty and model bias (distinguishing the two);
- Data augmentation, identifying gaps in data;
- Decision making – based on results of the data and/or models;
- Data acquisition, experimental design;
- Data science applied to experimental data;
- Artificial intelligence;
- System ID and system theoretic communities.

We note that many of the above topics would need much more thorough discussion; the above list simply captures the open comments from the participants. Regarding the workshop itself, the participants were asked what worked well and what could be improved. The participants liked the one-minute poster introductions and thought that the length of the talks (25 minutes), as well as the number of talks per session, were appropriate. Areas for improvement included having smaller poster sessions (our single session was thought to be too large), providing food/drinks during the poster session, getting a better room (especially for viewing the presentation slides), better grouping of talks within sessions according to themes, and more diversity among speakers (e.g., women, students, postdocs).

Additional ideas that were mentioned for future workshops included the following: Introducing a single new topic and inviting one or two subject matter experts to speak on that topic; highlighting software (especially open source software); sharing of frustrations in applying various UQ and data assimilation methods; hands-on experiences for junior researchers as well as tutorials. It was strongly felt that there are many fruitful directions for possible future

workshops. Finally, it was noted that there is a danger of a small workshop turning into a conference; therefore, it was emphasized that focusing on specific themes would help keep the size down and maintain the value of small-group interactions.

Workshop on Uncertainty Quantification and Data-Driven Modeling

March 23-24, 2017



U.S. Association for Computational Mechanics



Institute for Computational Engineering and Sciences

Workshop on Uncertainty Quantification and Data-Driven Modeling

AT&T Executive Education and Conference Center, Austin, Texas
March 23-24, 2017

Thursday, March 23	
7:30	Registration; Room 104
8:15 - 8:30	Opening Remarks
Session 1	
8:30 – 8:55	Roger Ghanem , University of Southern California <i>Data-driven Sampling and Prediction on Manifolds</i>
8:55 – 9:20	Michael Shields , Johns Hopkins University <i>How Much Data do I Really Need to Conduct Probabilistic UQ?</i>
9:20 – 9:45	Robert Moser , ICES/University of Texas at Austin <i>Validating the Reliability of Predictions based on Unreliable Models</i>
9:45 – 10:15	Break
Session 2	
10:15 – 10:40	Paris Perdikaris , Massachusetts Institute of Technology <i>Data-driven Modeling and Optimization with Probabilistic Multi-fidelity Surrogates</i>
10:40 – 11:05	Rick Archibald , Oak Ridge National Laboratory <i>Sparse Sampling Methods for Large Scale Experimental Data</i>
11:05 – 11:30	Kevin Carlberg , Sandia National Laboratories <i>Reducing Nonlinear Dynamical Systems via Model Reduction and Machine Learning</i>
11:30 – 12:30	Lunch
12:30 – 1:00	Poster Introductions (see list at end of program)
1:00 – 2:25	Poster Session
Session 3	
2:25 – 2:50	Nathan Kutz , University of Washington <i>Data-driven Discovery of Governing Equations in the Engineering and Physical Sciences</i>
2:50 – 3:15	David Stracuzzi , Sandia National Laboratories <i>Uncertainty Quantification for Machine Learning and Statistical Models</i>
3:15 – 3:40	Youssef Marzouk , Massachusetts Institute of Technology <i>Low-dimensional Couplings for Bayesian Inference</i>
3:40 – 4:10	Break
Session 4	
4:10 – 4:35	Daniele Schiavazzi , University of Notre Dame <i>A Generalized Multi-resolution Expansion for Uncertainty Propagation with Application to Cardiovascular Modeling</i>
4:35 – 5:00	Paul Barbone , Boston University <i>Efficiently Computing Covariance of Parameter Estimates in Inverse</i>

	<i>Problems</i>
5:00 – 5:25	Tan Bui-Thanh , ICES/University of Texas at Austin <i>A Randomized Misfit Approach for Data-driven PDE-constrained Bayesian Inverse Problems</i>
6:30 – 7:00	Dinner reception – UT Alumni Center, 2110 San Jacinto Blvd.
7:00 – 9:00	Workshop Dinner - UT Alumni Center, Legends Room
Friday, March 24	
Session 5	
8:15 – 8:40	J. Tinsley Oden , ICES/University of Texas at Austin <i>Selection and Validation of Predictive Models of Tumor Growth and Cancer Therapies</i>
8:40 – 9:05	C. Alberto Figueroa , University of Michigan <i>A Reduced-order Kalman Filtering Approach for Data-driven Parameter Estimation in Arterial Hemodynamics</i>
9:05 – 9:30	Ben Peherstorfer , University of Wisconsin <i>Multifidelity Methods for Uncertainty Propagation and Rare Event Simulation</i>
9:30 – 9:55	Qiqi Wang , Massachusetts Institute of Technology <i>When Does an Accurate Model Lead to a Predictive Simulation?</i>
9:55 – 10:25	Break
Session 6	
10:25 – 10:50	Omar Ghattas , ICES/University of Texas at Austin <i>Scalable Methods for Optimal Control of Systems Governed by PDEs with Random Coefficient Fields</i>
10:50 – 11:15	Jonathan Freund , University of Illinois at Urbana-Champaign <i>Adjoint-based Sensitivity in Turbulent Combustion Simulation</i>
11:15 – 11:40	Mohammad Khalil , Sandia National Laboratories <i>Data-Driven Bayesian Model Selection: Parameter Space Dimension Reduction using Automatic Relevance Determination Priors</i>
11:40 – 12:05	Karthik Duraisamy , University of Michigan <i>A Paradigm for Data-driven Predictive Modeling Using Field Inversion and Machine Learning</i>
12:05 – 1:15	Lunch
1:15 – 2:00	Open Discussion : Key research themes, challenges, opportunities, etc.
2:00	Adjourn

Abstracts of talks follow (in order of presentation)

List of posters (listed alphabetically)

Data-Driven Sampling and Prediction on Manifolds

Roger Ghanem¹ and Christian Soize²

¹University of Southern California, ²Université Paris-Est

With the possibility of interpreting data using increasingly complex models we are often faced with the need to embed the data in an ambient space consistent with the parameterization of these models typically a high-dimensional Euclidean space. Constructing probability measures on these spaces or subsets of them is fairly straightforward once the subsets have been delineated. Quite often fundamental laws associated for example with symmetry of conservation constrain the data to a complex manifold within this ambient space. Acknowledging these constraints serves to focus the scatter in the data around the manifold with significant ramifications to subsequent statistical analysis: the shape of the distributions asymptotic sample properties and the sampling mechanisms would all be affected.

Increasingly more often the exact constraints (hence manifolds) to which the data is subjected are not known either because of unaccounted interaction with other scales or physics (such as in physics-based problems) or because the fundamental governing laws are not yet understood (such as in biological social and economical systems). In such situations manifold discovery is an important step in augmenting statistical analysis with key hidden constraints.

In this talk we present a recent procedure for describing probability measures on diffusion manifolds and sampling from them. The procedure integrates methods from machine learning with statistical estimation functional analysis and white noise calculus to achieve orders of magnitude efficiencies in data requirements for probabilistic characterization and sampling. Examples will be shown from applications from across the sciences and engineering.

How Much Data Do I Really Need to Conduct Probabilistic UQ?

Michael Shields and Jiaxin Zhang

Johns Hopkins University

In purely mathematical terms stochastic systems are expressed in terms of a set of random variables with known probability models. Under certain conditions (e.g. asymptotic normality) the true probability model may indeed be known. More commonly the probability model is assumed or inferred from data. Along these lines a great deal of research has focused on methods for estimating probability distributions – ranging from frequentist parameter estimation to Bayesian model selection. In these methods two common themes arise: 1. In the infinite data limit most of the established methods converge toward a “best” (not necessarily true/correct) solution; 2. In the small data case uncertainty abounds which (among other factors) has inspired separate fields of imprecise probabilities and non-probabilistic approaches. This begs the simple but provocative question posed in the title. In general the answer to this question will be dependent on the sensitivity of the stochastic system to the probability models of its variables and the level of confidence required by the modeler. In this presentation I attempt to provide some insights into this question by presenting an efficient method for propagating uncertain probability distributions quantified through Kullback-Liebler information theoretic multi-model inference with Bayesian parameter estimation. The method retains and propagates the uncertainties associated with both model form and model parameters that stems from lack of data. In this way we establish confidence levels in probabilistic response based on the available data – allowing the user/analyst to make the determination of how much data he/she is willing/able to collect in order to achieve a sufficient level of confidence in statistical response.

Validating the Reliability of Predictions Based on Unreliable Models

Robert Moser

ICES/University of Texas at Austin

One of the most challenging and important applications of computational models of physical systems is to make predictions when no observations of the quantities being predicted are available. This is the usual situation when model results are to be used to support decisions (e.g. design or operations decisions) where predictions are needed precisely because observational data are not available when the decision must be made. Predictions then are essentially extrapolations of available information to the quantities and scenarios of interest. The validation challenge is to assess whether such an extrapolation can be made reliably. The structure of many computational models actually makes this possible. However a critical ingredient is a representation of the uncertainty introduced due to any modeling errors in such a way that the uncertainties can be propagated to the predictions. A primary challenge is formulating a model error representation that respects both what is known about the phenomenon being modeled and the known deficiencies of the model. In this talk we will explore a basis for making reliable extrapolative predictions using imperfect models. We will also discuss the representation of model errors in several physical domains including chemical kinetics, turbulent combustion and contaminant dispersion in porous media.

Data-Driven Modeling and Optimization with Probabilistic Multi-Fidelity Surrogates

Paris Perdikaris¹, George Em Karniadakis²

¹Massachusetts Institute of Technology, ²Brown University

Multi-fidelity stochastic modeling entails the use of variable fidelity methods and models both in physical and probability space. In this new paradigm we target learning from variable information sources as even under-resolved simulations or simplified mathematical models or empirical correlations can be employed to construct an accurate stochastic response surface. The goal of this talk is to provide an introduction to multi-fidelity modeling using probabilistic machine learning and Gaussian processes. In particular we will present a general yet flexible data-driven framework that allows us to simultaneously track both parametric and modeling uncertainties by synergistically combining models of variable fidelity and exploiting their cross-correlation structure. A collection of benchmark problems will demonstrate the robustness of the proposed algorithms with respect to model misspecification (e.g. inaccurate low-fidelity models or noisy measurements) as well as their accuracy and efficiency for a wide range of target applications including uncertainty quantification data-assimilation inverse problems design optimization and beyond.

Sparse Sampling Methods for Large Scale Experimental Data

Rick Archibald

Oak Ridge National Laboratory

This talk will focus on mathematics developed to help with the mathematical challenges face by the Department of Energy (DOE) at the experimental facilities at Oak Ridge National Laboratory (ORNL). This talk will specifically focus on sparse sampling methods for large scale experimental data. Sparse sampling has the ability to provide accurate reconstructions of data and images when only partial information is available for measurement. Sparse sampling methods have demonstrated to be robust to measurement error. These methods have the potential to scale to large computational machines and analysis large volumes of data.

Reducing Nonlinear Dynamical Systems Via Model Reduction and Machine Learning

Kevin Carlberg¹, Harbir Antil², Matthew Barone¹, Sumeet Trehan², Louis Durlofsky²

¹Sandia National Laboratories, ²George Mason University, ³Stanford University

Predictive modeling of complex physical systems often necessitates constructing large-scale nonlinear dynamical-system models that are extremely costly to simulate. This creates a 'computational barrier' that precludes such high-fidelity models from being employed in many-query applications (e.g. uncertainty propagation) that can demand hundreds or thousands of such simulations. This talk describes a two-pronged approach for using available simulation data to significantly reduce the dimensionality and complexity of such large-scale dynamical-system models.

First we propose using simulation data related to the system state and nonlinear functions to construct a projection-based reduced-order model (ROM). Rather than applying the commonly-used Galerkin projection we show that ROMs constructed via least-squares Petrov–Galerkin (LSPG) projection exhibit smaller error bounds time-discrete optimality and greatly improved performance large-scale turbulent-flow problems.

Second we propose using simulation data related to the error between the reduced-order and high-fidelity models to construct a model of the ROM error. This approach applies high-dimensional regression methods from machine learning (e.g. random forests LASSO) to map a large set of inexpensively computed 'error indicators' (i.e. features) produced by the reduced-order model to a prediction of its error in a quantity of interest (i.e. output).

Data-Driven Discovery of Governing Equations in the Engineering and Physical Sciences

Nathan Kutz

University of Washington

We use emerging large-scale time-series data from modern sensors to directly construct in an adaptive manner governing equations even nonlinear dynamics that best model the system measured using sparsity-promoting techniques. Recent innovations also allow for handling multi-scale physics phenomenon and control protocols in an adaptive and robust way. The overall architecture is equation-free in that the dynamics and control protocols are discovered directly from data acquired from sensors. The theory developed is demonstrated on a number of example problems. Ultimately the method can be used to construct adaptive controllers which are capable of obtaining and maintaining optimal states while the machine learning and sparse sensing techniques characterize the system itself for rapid state identification and improved optimization.

Uncertainty Quantification for Machine Learning and Statistical Models

*David Stracuzzi, Maximillian Chen, Matthew Peterson, Michael Darling,
Charles Volume*

Sandia National Laboratories

Charles Darwin once said that "ignorance more frequently begets confidence than does knowledge." This is especially true in data analytics. By necessity we often rely on data of debatable quality relevance and completeness to answer all manner of questions. Moreover our data often come from multiple sources of varying quality and reliability. Under these conditions the easiest way to be certain about results is to avoid looking too closely.

Data-driven uncertainty analysis takes the opposite approach; it looks very closely at the available data and the statistical modeling process to determine both the stability of the learned model and the reliability of its predictions. The talk begins with a discussion of what uncertainty means in a data analytics context where it comes from and why analyzing uncertainty is important. We then summarize applications and preliminary results in two domains: seismic analysis and overhead image analysis. The talk concludes with a discussion of possible long-term implications of the work for data collection data analysis and decision making.

Low-dimensional Couplings for Bayesian Inference

Youssef Marzouk, Alession Spantini, Daniele Bigoni

Massachusetts Institute of Technology

We will discuss how transport maps, i.e. deterministic couplings between probability measures can enable useful new approaches to Bayesian computation. In particular we present a variational approach to Bayesian inference that constructs a deterministic transport between a reference measure and the posterior measure without resorting to MCMC or importance sampling. Independent and unweighted posterior samples can then be obtained by pushing forward reference samples through the map.

Making this approach efficient in high dimensions requires identifying and exploiting low-dimensional structure. We present new results relating the sparsity and decomposability of transports to the conditional independence structure of the target distribution. We also describe conditions common in inverse problems under which transport maps have a particular low-rank or near-identity structure. In general these properties of transports can yield more efficient algorithms. As a particular example we derive new “online” variational algorithms for Bayesian inference in nonlinear and non-Gaussian state-space models with static parameters.

A Generalized Multi-Resolution Expansion for Uncertainty Propagation with Application to Cardiovascular Modeling

Daniele Schiavazzi

University of Notre Dame

Computational models are used in a variety of fields to improve our understanding of complex physical phenomena. Recently the realism of model predictions has been greatly enhanced by transitioning from deterministic to stochastic analysis frameworks where the effects of the intrinsic variability in parameters loads constitutive properties model geometry and other quantities can be more naturally included. A general stochastic system may be characterized by a large number of arbitrarily distributed and correlated random inputs and a limited support response with sharp gradients or event discontinuities. This motivates continued research into novel adaptive algorithms for uncertainty propagation particularly those handling high dimensional arbitrarily distributed random inputs and non-smooth stochastic responses.

In this talk I will discuss a generalization of a previously proposed multi-resolution approach to uncertainty propagation that improves computational efficiency can handle arbitrarily distributed random inputs and non-smooth stochastic responses and naturally facilitates adaptivity i.e. the expansion coefficients encode information on solution refinement. Our approach relies on partitioning the stochastic space into elements subject to binary refinements that are particularly effective in avoiding the exponential increase in the multi-resolution basis cardinality and significantly reduce the regression complexity for moderate to high dimensional random inputs.

The performance of the approach is demonstrated through previously proposed uncertainty propagation benchmarks and stochastic multi-scale finite element simulations in cardiovascular flow. In these latter simulations we show that our approach is able to combine random inputs assimilated by sequentially sampling from their joint posterior distribution and inputs with assumed distributions.

Efficiently Computing Covariance of Parameter Estimates in Inverse Problems

Paul E. Barbone¹, Bryan Chue¹, Quang Ha¹, Assad Oberai²

¹Boston University, ²Rensselaer Polytechnic Institute

Uncertainty quantification in inverse problems presents distinct opportunities and challenges relative to UQ in forward models. A formal solution for the *a posteriori* probability distribution (PPD) for inverse problems is often readily available but is difficult to work with practically. For unimodal posteriors the PPD may be approximately characterized by its mode and covariance about the mode. Computing these quantities can present formidable challenges. In practice finding the MAP estimate may require minimizing a very high dimensional function ($O(10^5 - 10^6)$ parameters) with complex (e.g. discretized nonlinear) PDE constraints. For such a problem merely storing the final covariance may require tens of terabytes. We show that the structure of inverse problems leads to a sparse update on a predictable and *a priori* known covariance operator. This leads to an efficient sparse basis in which to compute and store the covariance. We show how the covariance can be accurately computed with relatively little computation over and above that used in finding the MAP estimate. This method can thus be used to compute the uncertainty in an inverse problem solution relatively efficiently. We describe the approach formally in a relatively general setting and demonstrate it with examples in inverse diffusion and inverse elasticity.

A Randomized Misfit Approach for Data-Driven PDE-Constrained Bayesian Inverse Problems

Tan Bui-Thanh, Elle Le, Aaron Myers

University of Texas at Austin

A randomized misfit approach is presented for the efficient solution of very large-scale PDE-constrained inverse problems with high-dimensional data. The contribution of this work is to offer a theory-based framework for random projections in this inverse problem setting. The stochastic approximation to the misfit is analyzed using random projection theory. The interplay of both Johnson-Lindenstrauss theory and Morozov's discrepancy principle is shown to be essential to the result. The computational cost savings for very large-scale PDE-constrained problems with high-dimensional data is discussed. Numerical verification of theoretical findings are presented for model problems of estimating a distributed parameter in an elliptic partial differential equation. Results with different random projections are presented to demonstrate the viability and accuracy of the proposed approach.

Selection and Validation of Predictive Models of Tumor Growth and Cancer Therapies

J. Tinsley Oden¹, Ernesto Lima¹, Amir Shahmorad¹, Barbara Wohlmuth², Thomas Horger², Thomas Yankeelov¹, David Hormuth¹

¹University of Texas at Austin, ²Technical University of Munich

Stochastic phenomenological models of vascular tumor growth in living tissues based on continuum mixture theory and principles of cancer biology are described. These are used to predict the evolution of brain cancer in laboratory animals. A Bayesian framework is proposed for managing model calibration validation and for prediction. Models of radiation effects on tumor growth are also studied. The OPAL algorithm is implemented to control parameter sensitivities model plausibilities model inadequacy and to the solution of this forward predictions problem. Examples include the applications to predictive methods to model glioma in laboratory rats. Encouraging results in tumor growth predictions are described.

A Reduced-Order Kalman Filtering Approach for Data-driven Parameter Estimation in Arterial Hemodynamics

C. Alberto Figueroa

University of Michigan

A major challenge in constructing three-dimensional patient-specific hemodynamic models is the calibration of boundary condition and material parameters that enable model predictions to match patient data on flow pressure wall motion etc. acquired in the clinic. We have implemented a parameter estimation framework based on reduced-order Kalman filtering [1 , 2] to estimate the Windkessel parameters that characterize the outflow boundary conditions in a 3-D subject-specific fluid-structure interaction model of aortic hemodynamics.

We first demonstrate the suitability of this methodology to recover estimates and covariance metrics of tissue stiffness in an idealized model using synthetic data. Then we estimate outflow boundary condition parameters in a clinical scenario: The estimation algorithm incorporates noninvasive flow data acquired using magnetic resonance imaging and pressure data from applanation tonometry in a healthy human volunteer and successfully produced converged estimates for the subject-specific aorta model. Numerical experimentation suggests that having time-resolved flow measurements in the branches of the model and pressure data at a single location is sufficient for the successful recovery of the boundary condition parameters.

Multifidelity Methods for Uncertainty Propagation and Rare Event Simulation

Benjamin Peherstorfer¹, Karen Willcox², Max Gunzburger³, Boris Kramer²

¹University of Wisconsin ²Massachusetts Institute of Technology

³Florida State University

In many situations across computational science and engineering multiple computational models are available that describe a system of interest. These different models have varying evaluation costs and varying fidelities. Typically a computationally expensive high-fidelity model describes the system with the accuracy required by the current application at hand while lower-fidelity models are less accurate but computationally cheaper than the high-fidelity model. Uncertainty quantification typically requires multiple model solves at many different inputs which often leads to computational demands that exceed available resources if only the high-fidelity model is used. We present multifidelity methods for uncertainty propagation and rare event simulation that leverage low-cost low-fidelity models for speedup and occasionally make recourse to the expensive high-fidelity model to establish unbiased estimators. Our methods combine low-fidelity models of any type including projection-based reduced models data-fit models and response surfaces coarse-grid approximations and simplified-physics models. Our numerical results demonstrate that our multifidelity methods achieve significant speedups while providing unbiased estimators even in the absence of error control for the low-fidelity models.

When Does an Accurate Model Lead to a Predictive Simulation?

Qiqi Wang¹, Alexandre Marke¹, Xiao Heng²

¹Massachusetts Institute of Technology, ²Virginia Tech

Data-based methods have great potential in modeling unclosed term in physics-based simulations. Such data-based methods often minimizes the modeling error on a training set. This error minimization however does not always lead to the most predictive simulation. One factor is the possibility of overfitting which can be addressed by many statistical and machine learning techniques. There is another factor that should be considered when combining data-based modeling and physics-based modeling. This factor is model conditioning which quantifies how much a small error in a model term impacts the solution. We will use analysis and examples to show that modeling conditioning is essential to ensure accurate solution of data-based modeling. We will discuss challenges in constructing well-conditioned data-based modeling and in particular what conditioning means in the context of chaotic fluid flow.

Scalable Methods for Optimal Control of Systems Governed by PDEs with Random Coefficient Fields

*Omar Ghattas¹, Peng Chen¹, Umberto Villa¹, Alen Alexanderian²,
Noemi Petra³, Georg Stadler⁴*

¹ICES University of Texas at Austin ²North Carolina State University ³UC Merced ⁴Courant
Institute NYU

We present a method for optimal control of systems governed by partial differential equations (PDEs) with uncertain parameter fields. We consider an objective function that involves the mean and variance of the control objective a form of risk-averse optimal control. Conventional numerical methods for optimization under uncertainty are prohibitive when applied to this problem; for example sampling the (discretized infinite-dimensional) parameter space to approximate the mean and variance would require solution of an enormous number of PDEs which would have to be done at each optimization iteration. To make the optimal control problem tractable we invoke a quadratic Taylor series approximation of the control objective with respect to the uncertain parameter field. This enables deriving explicit expressions for the mean and variance of the control objective in terms of its gradients and Hessians with respect to the uncertain parameter. The risk-averse optimal control problem is then formulated as a PDE-constrained optimization problem with constraints given by the forward and adjoint PDEs defining these uncertain parameter gradients and Hessians.

The expressions for the mean and variance of the control objective under the quadratic approximation involve the trace of the (preconditioned) Hessian and are thus prohibitive to evaluate for (discretized) infinite-dimensional parameter fields. To overcome this difficulty we employ a randomized eigensolver to extract the dominant eigenvalues of the decaying spectrum. The resulting objective functional can now be readily differentiated using adjoint methods along with eigenvalue sensitivity analysis to obtain its gradient with respect to the controls. Along with the quadratic approximation and truncated spectral decomposition this ensures that the cost of computing the risk-averse objective and its gradient with respect to the control---measured in the number of PDE solves---is independent of the (discretized) parameter and control dimensions leading to an efficient quasi-Newton method for solving the optimal control problem. Finally the quadratic approximation can be employed as a control variate for accurate evaluation of the objective at greatly reduced cost relative to sampling the original objective. We present results of collaborative work with R. Moser and T. Oliver on the optimal control of a turbulent jet flow model with an algebraic turbulence model with uncertain parameter field.

Adjoint-Based Ignition Sensitivity in Turbulent Combustion

Jonathan Freund¹, David Buchta¹, Jesse Capecelatro²

¹University of Illinois at Urbana-Champaign, ²University of Michigan

The talk will cover ongoing efforts using adjoint-based sensitivity in conjunction with large-scale simulations of flow turbulence. A current target of these efforts is adjoint-based sensitivity calculations for large-scale turbulent combustion simulations with the goal of identifying quantifying and reducing prediction uncertainties. It is demonstrated on a non-premixed shear layer a reacting jet-in-crossflow and ignition in decaying turbulence. We distinguish sensitivities between a detailed and a global one-step hydrogen-air mechanism. The primary model system is the ignition of a turbulent jet by a laser-induced optical breakdown (LIB). Ignition defined by a space-time integral of temperature is most sensitive to the modeled plasma kernel geometry and its energy deposited on the gas phase. Thus combining the adjoint-based sensitivity with the LIB's aleatoric interspersed plasma kernels these parameters dominate the propagated output uncertainty which is local to the inputs. The present combustion sensitivity studies are a component of a multi-scale multi-physics combustion application being investigated in the Illinois Center for Exascale Simulation of Plasma-coupled Combustion (XPACC). Similar examples from aeroacoustic flow control will also be included in the discussion.

Data-Driven Bayesian Model Selection: Parameter Space Dimension Reduction using Automatic Relevance Determination Priors

Mohammad Khalil

Sandia National Laboratories

Bayes' theorem provides parameter estimates that blend prior knowledge of the system parameters with indirect observational data. Bayesian model selection utilizes such estimates in comparing the suitability of many plausible models using the so-called model evidence the probability that randomly selected parameters from the prior would generate the observed data. There are various approaches to prescribe the prior distribution depending on the level of knowledge of the modeler. Popular priors include diffuse priors Jeffrey's priors conjugate priors and informative priors. The choice of prior distribution and associated parameters that parametrize such priors (called hyper-parameters) has a major impact on any Bayesian estimation procedure and subsequent model selection analysis.

In the context of feature selection automatic relevance determination (ARD) aka sparse Bayesian learning is an effective tool for pruning large numbers of irrelevant features leading to a sparse explanatory subset. It does so by regularizing the Bayesian inference solution space using a parameterized data-dependent prior distribution that effectively prunes away redundant or superfluous features. The hyper-parameter of each ARD prior explicitly represents the relevance of the corresponding model parameter. The hyper-parameters are estimated using the observational data by performing evidence maximization or type-II maximum likelihood. In the context of model selection ARD priors aid in finding the best model nested under the envisioned model. ARD provides a flexible Bayesian platform to find the optimal nested model by eliminating the need to propose candidate nested models and associated prior pdfs. Thus ARD priors effectively reduce the parameter space dimension of the inference procedure based on available observations.

This talk will motivate the use of ARD priors in the context of physics-based Bayesian model selection. Results will be presented for an application to model selection of complex aeroelastic systems modeled by coupled nonlinear stochastic ordinary differential equations using noisy wind-tunnel experimental observations. The experiments consist of a NACA0012 airfoil undergoing limit cycle oscillation in the transitional Reynolds number regime.

A Paradigm for Data-Driven Predictive Modeling using Field Inversion and Machine Learning

Karthik Duraisamy, Anand Singh, Shaowu Pan

University of Michigan

A data-driven modeling paradigm based on field inversion and machine learning (FIML) is proposed to aid the quantification of model discrepancy in a predictive setting. This work uses inverse modeling to obtain corrective spatio-temporally distributed functional terms offering a route to directly address model-form errors. Once the inference has been performed over a number of problems that are representative of the deficient physics machine learning techniques are used to reconstruct the model corrections in terms of features that appear in the closure model. When these reconstructed functional forms are embedded in existing PDE solvers improved predictions and uncertainty bounds are obtained. The inversion/learning/uncertainty propagation framework is cast in a Bayesian setup allowing for consistent uncertainty quantification. Results are presented in fluid flow and materials problems.

Posters

Ali Akhavan Safaei	A Bayesian Framework for Rheology Model Combination and UQ in Simulation of Geophysical Mass Flows.
Watheq Al-Mudhafar	Boosted Regression-based Proxy Optimization of CO2 Enhanced Oil Recovery in Unconventional Reservoirs
Negin Alemazkoor	A Near-Optimal Sampling Strategy for Sparse Recovery of Polynomial Chaos Expansions
Alessandro Alla	A Certified Model Reduction Approach for Robust Optimal Control with PDE Constraints
Ahmed Attia	Cluster Sampling Algorithms for Non-Gaussian Data Assimilation
Olalekan Babaniyi	Inversion Techniques from Transient Data in the Ultrasound Elastography Field
Ramin Bostanabad	Data-Driven Modeling and UQ in Multiscale Materials
Peng Chen	Taylor Approximation and Variance Reduction for PDE-constrained Optimal Control under Infinite-dimensional Uncertainty
Oriol Colomes	Mesh Discretization Error and Uncertainty Quantification: A Variational Multiscale Approach
M. Chase Dwelle	Flooding Dynamics with Spectral Projection and Bayesian Compressive Sensing
Wouter Edeling	A Return to Eddy Viscosity UQ framework
Gianluca Geraci	Multilevel/Multifidelity Strategies for Forward UQ
Alex Gorodetsky	Low-rank Approximation for Gaussian Filtering
Quang Ha	Krylov Subspace Method for Hessian Approximation - Conjugate Gradient in Infinite Dimensional Hilbert Space
Eric Hall	Uncertainty Quantification for Generalized Langevin Dynamics
Joey Hart	Efficient Computation of Sobol' Indices for Stochastic Models
Jeffrey Hokanson	Data-Driven Polynomial Ridge Approximation Using Variable Projection
Ming Jiang	A Supervised Learning Framework for Arbitrary Lagrangian-Eulerian Simulations
Soo Kyung Kim	Machine Learning Approach of Climate Event Prediction
Ellen Le/Aaron Myers/ Vishwas Rao	Scalable Methods for Large-Scale Data-driven Bayesian Inverse Problems
Jing Lin	Bayesian Learning for High-Dimensional Chaotic Dynamical Systems with Application to Two-Dimensional Turbulent Bottom Gravity Currents
Erin Linebarger	Physically Constrained Ensemble Kalman Filter
Sudipto Mandal	Analyzing Constitutive Parameters In Multi-Scale Material Models Using Canonical Correlation Analysis
Damon McDougall	QUESO: A Parallel C++ Library for Uncertainty Quantification
Joseph Slagel	Stochastic Newton and Quasi-Newton Methods for Large Linear Least-squares Problems
David Sondak	Calibration of a Stochastic Model Inadequacy Representation for Chemical Kinetics
Gray Thomas	Identifying Bounded Uncertainty Models with Convex Optimization

Gregory Teichert	Data-interactive Computational Materials Physics: Studies of Precipitate Morphology by Combination of Experiment Nonlinear Elasticity and Machine Learning
Panagiotis Tsilifis	Polynomial Chaos Basis Adaptation Schemes for Uncertainty Propagation and Design Optimization on a SCRAMJET Engine
Umberto Villa	Integrating Data with Complex Predictive Models under Uncertainty: An Extensible Software Framework for Large-Scale Bayesian Inversion
Tsuyoshi Wakamatsu	Information Contents of a 4dvar Analysis Based on a Reduced Order Model Approach
Jianxun Wang	Physics-Informed Data-Driven Approach for Reducing Model Discrepancies Due to Unresolved Physics
Sichao Wu	GENEUS: a Computational Framework for Uncertainty Quantification Sensitivity Analysis and Experimental Design for Network-based Models
Benjamin Zhang	Measure Transport Methods for Rare Event Simulation
Jiaxin Zhang	Data-Driven Probabilistic Calibration of Material Models from Small Datasets and its Influence on Structural Response

Directions to Alumni Center

Turn left outside the AT&T Center, head north, walk to 21st Street. Turn right on 21st Street and walk until you get to San Jacinto. Turn left on San Jacinto until you reach the Alumni Center, which is directly across from the stadium. The dinner will be held in the Legends Room.

Travel Award Recipients (*Postdocs are denoted by *; others are PhD students*)

Joey Hart	North Carolina State University
Jianxum Wang	Virginia Tech
M. Chase Dwelle	University of Michigan
Jospeh Tanner Slagel	Virginia Tech
Negin Alemazkoor	University of Illinois
Eric Hall*	University of Massachusetts Amherst
Jiaxin Zhang	Johns Hopkins University
Watheq Al-Mudhafar*	Louisiana State University
Gray Thomas	The University of Texas at Austin / NASA JSC
Erin Linebarger	University of Utah
Alla Alessandro*	Florida State University
Ramin Bostanabad	Northwestern University
Aaron Myers	University of Texas at Austin
Panagiotis Tsilifis*	University of Southern California
Ali Akhavan Safael	University at Buffalo
Jing Lin	Massachusetts Institute of Technology
Wouter Edeling*	Stanford University
Vishwas Hebbur Venkata Subba Rao*	University of Texas at Austin
Jeffrey M. Hokanson*	Colorado School of Mines
Sudipto Mandal	Carnegie Mellon University
Gregory Teichert	University of Michigan
Ahmed Attia*	Statistical and Applied Mathematical Science Institute
Oriol Colomes*	Duke University

Olalekan Babaniyi*	Duke university
Benjamin Zhang	Massachusetts Institute of Technology
Sichao Wu	Virginia Tech
Ellen Le	University of Texas at Austin