SAND2014-20703C

Overview of Selected DOE/NNSA Predictive Science Initiatives: the Predictive Science Academic Alliance Program and the DAKOTA Project

M. S. Eldred[*], L. P. Swiler[†], B. M. Adams[‡], and J. D. Jakeman[§]
Sandia National Laboratories[¶], Albuquerque, NM 87185

This paper supports a special session on "Frontiers of Uncertainty Management for Complex Aerospace Systems" with the intent of summarizing two aspects of the DOE/NNSA Accelerated Strategic Computing (ASC) program, each of which is focused on predictive science using complex simulation models. The first aspect is academic outreach, as enabled by the Predictive Science Academic Alliance Program (PSAAP). The second aspect is the Dakota project at Sandia National Laboratories, which develops and deploys uncertainty quantification capabilities focused on high fidelity modeling and simulation on large-scale parallel computers.

I. Introduction and context

This paper supports a special session on "Frontiers of Uncertainty Management for Complex Aerospace Systems" with the intent of summarizing two aspects of the DOE/NNSA Accelerated Strategic Computing (ASC) program, each of which is focused on predictive science using complex simulation models.

The first aspect is academic outreach, as realized by a sequence of academic alliance programs that seek to partner with universities to develop the future talent needed to develop predictive computational models at extreme scale. The current incarnation of this outreach program is the Predictive Science Academic Alliance Program (PSAAP), which has recently initiated its second cycle: PSAAP II.

The second aspect of this overview is the Dakota project at Sandia National Laboratories, which develops and deploys uncertainty quantification capabilities in support of a variety of DOE mission areas. We will highlight recent investments in the scalability of uncertainty quantification (UQ) capabilities, both in terms of problem size and in terms of efficient utilization of large-scale parallel computers.

II. PSAAP

The Accelerated Strategic Computing (ASC) program funded by DOE/NNSA has a history of university outreach through the ASAP and PSAAP programs. The initial cycles of this outreach program (ASAP, 1997-2007) focused primarily on high fidelity modeling and simulation on the most advanced parallel architectures available at that time. The predictive science emphasis began in 2008 with the initial PSAAP program (2008-2013), and continues with new PSAAP II programs initiating in 2014. Lead institutions in PSAAP II include Florida, Illinois, Notre Dame, Stanford, Texas A&M, and Utah. The centers are divided into Single Discipline Centers (Florida, Notre Dame, Texas A&M) and larger Multidisciplinary Simulation Centers (Illinois, Stanford, Utah). Each center is unified around a motivating application:[1]

[*]Distinguished Member of Technical Staff, Optimization and Uncertainty Quantification Department, AIAA Associate Fellow.
[†]Distinguished Member of Technical Staff, Optimization and Uncertainty Quantification Department, AIAA Senior Member.
[‡]Principal Member of Technical Staff, Optimization and Uncertainty Quantification Department.
[§]Foreign National Interim Technical Staff, Optimization and Uncertainty Quantification Department.
[¶]Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

- *Compressible multiphase turbulence (CMT) in explosive-driven particle-laden flows (Florida).* This center will advance the field of CMT through rigorous first-principle multiscale modeling, advance large-scale predictive simulation science on present and near-future platforms, and advance a co-design strategy that combines exascale emulation with a novel energy-constrained numerical approach.
- *Exascale Simulation of Plasma-coupled Combustion (Illinois).* This center will use large-scale predictive simulation to advance the science and technology of plasmas to initiate and control turbulent combustion. The target application is a gas-phase fuel jet.
- *Shock Wave-processing of Advanced Reactive Materials (Notre Dame).* This center will develop controlled microstructures in order to design novel materials with unique properties. It will predict shock conditions that generate high temperatures and pressures under which new materials can be synthesized.
- *Predictive Simulations of Particle-laden Turbulence in a Radiation Environment (Stanford).* This center blends efforts in computer science, UQ, and computational engineering to tackle multiphysics simulation for harvesting solar energy. Fine particles suspended within a fluid absorb sunlight and directly transfer heat throughout the fluid volume, potentially increasing energy absorption efficiency in solar-thermal systems.
- *Exascale Radiation Transport (Texas A&M).* This center will focus on thermal radiation transport (TRT), which is essential for multiphysics modeling of high energy density physics (HEDP) experiments at NNSA facilities such as the National Ignition Facility. The center will avoid the complexity of multiphysics HEDP experiments by performing surrogate neutron transport experiments, exploiting the strong analogy between thermal radiation transport and neutron transport.
- *High Efficiency Electric Power Generation with Carbon Capture (Utah).* This center will use exascale UQ-predictive simulation science to rapidly design and deploy a new technology for secure electric energy generation, namely, a high efficiency advanced ultrasupercritical oxy-coal power boiler. Key physics components include large eddy simulations, multiphase flow, particle combustion, and radiation.

Each of these programs shares a focus on predicting complex phenomena through research investments in physical modeling and simulation, verification and validation, and computer science for next-generation extreme-scale computing platforms (i.e., exascale co-design). In the presentation, these three aspects are specifically highlighted for each center.

III. DAKOTA

The DAKOTA project began as a laboratory-directed research and development project in 1994, focused on developing a reusable toolkit for design optimization. It has since broadened into other related areas, and now provides a variety of algorithms for:

- optimization with gradient-, nongradient-, and surrogate-based methods;
- uncertainty quantification with sampling, reliability, stochastic expansion, and epistemic methods;
- deterministic and stochastic parameter estimation with nonlinear least squares and Bayesian inference methods; and
- sensitivity/variance analysis with design of experiments and parameter study methods.

These capabilities may be used on their own or as components within advanced "meta-iteration" strategies such as hybrid optimization, surrogate-based optimization, mixed integer nonlinear programming, optimization under uncertainty, or mixed aleatory-epistemic UQ.

DAKOTA has been a core component of the ASC Verification and Validation program since its inception (around 2000). Researchers at Sandia, Los Alamos, and Lawrence Livermore National Laboratories regularly deploy DAKOTA's UQ capabilities in support of mission requirements for assessing quantified margins and uncertainties (QMU) within the stockpile stewardship program. In this paper, we focus on the core foundational algorithms for UQ and then highlight several areas of active research that seek to address key challenges in deploying UQ to DOE mission areas.

A. Uncertainty Quantification

Within the area of uncertainty quantification algorithms, Dakota provides a variety of sampling, reliability, stochastic expansion, and epistemic methods that support differing analysis goals. The following sections provide short overviews for each algorithm class. For more detailed descriptions, refer to [2, 3].

1. Sampling methods

Arguably the most commonly used approach for propagating uncertainty is random sampling. Starting from assumed distributions on the uncertain input values, this method samples randomly from within these distributions, runs the model at the input sample values, and collects the corresponding ensemble of simulation results. The ensemble of output values is analyzed to determine characteristics of the response, e.g. maximum and minimum response values, response moments (mean, standard deviation, skewness, kurtosis), probability density and cumulative distribution function values, etc.

Random sampling Pure Monte Carlo (MC) methods simply generate samples from the particular distribution specified using a random number generator. There is no structure to the samples: one simply "picks a random draw" each time, and the particular realization does not influence samples that come before or after it. Random sampling has the advantage that the accuracy and computational burden is independent of the number of uncertain parameters. It may be preferred in high dimensions and when a sufficiently large number of samples are affordable. Pure MC sampling produces unbiased estimates for the means, variances, and percentiles of the outputs. A disadvantage of random sampling is that one must have a large number of samples to get accurate estimate of the output statistics. The accuracy of the mean estimate obtained from a set of random samples displays $1/\sqrt{N}$ convergence, meaning that on average one needs to quadruple the number of sample points to halve the error. To resolve small probabilities regarding predicted responses, one may be required to take hundreds of thousands or more sample points to obtain a desired accuracy with random sampling.

Latin Hypercube Sampling A good alternative to pure random sampling is Latin Hypercube Sampling (LHS).[4–6] LHS is a stratified random sampling method where the distribution is divided into strata or bins. Each stratum is chosen to be equally probable, so that the strata are of equal length for uniform distributions but of unequal length for normal distributions (shorter strata near the distribution centers and longer strata near the tails). If one wants to create a total sample of size N using LHS, an individual sample value is chosen from each of the N equally probable strata for each input variable. This stratification achieves sampling across the entirety of each distribution and eliminates some of the clustering of sample points often seen in pure random sampling.

For multidimensional sampling, LHS also serves to achieve a good "mixing" of sample values from different input variables. For example, one would not want the sample in strata 1 from input A to be paired with the sample in strata 1 from input B. Instead, the pairing of the strata is performed in such a way to generate multi-dimensional samples that are "well-mixed" or randomized. Pairing algorithms have been designed to achieve this.[6] An important point is that these pairing algorithms allow the generation of independent inputs but also will allow the generation of samples which honor a user-specified correlation structure.

Finally, LHS is more efficient than pure Monte Carlo in the sense that it requires fewer samples to achieve the same accuracy in statistics (standard error of the computed mean, for example). LHS gives an estimator for a function mean that has lower variance than MC for any function having finite second moment.[7,8] Further, the convergence behavior of LHS improves if the function is additively separable, meaning it can be decomposed into additive functions of the individual input parameters.

2. Reliability methods

Reliability methods are probabilistic approaches that compute approximate response statistics (moments and probabilities) based on specified uncertain variable distributions. These methods are often more efficient at computing statistics in the tails of the response distributions (events with low probability) than sampling based approaches.

The methods all answer the fundamental question: "Given a set of uncertain input variables, $\mathbf{X}$, and a scalar response function, g, what is the probability that the response function is below or above a certain level, $\bar{z}$?" The former can be written as $P[g(\mathbf{X}) \leq \bar{z}] = F_g(\bar{z})$ where $F_g(\bar{z})$ is the cumulative distribution

function (CDF) of the uncertain response $g(\mathbf{X})$ over a set of response levels. The latter can be written as $P[g(\mathbf{X}) > \bar{z}]$ and defines the complementary cumulative distribution function (CCDF). This probability calculation involves a multi-dimensional integral over an irregularly shaped domain of interest. The reliability methods all involve the transformation of the user-specified uncertain variables, $\mathbf{X}$, with probability density function, $p(x_1, x_2)$, which can be non-normal and correlated, to a space of independent Gaussian random variables, $\mathbf{u}$, possessing a mean value of zero and unit variance (i.e., standard normal variables). The Nataf transformation,[9] which is identical to the Rosenblatt transformation[10] in the case of independent random variables, is used in Dakota to accomplish this mapping. In the transformed space, probability contours are circular in nature allowing the multi-dimensional integrals to be approximated by simple functions of a single parameter, β, called the reliability index. β is the minimum Euclidean distance from the origin in the transformed space to the response surface and is known as the most probable point (MPP) of failure[a].

Within the class of reliability methods, there are local and global algorithms. The most well-known are local methods, which locate a single MPP and then utilize an approximation centered around this point. In contrast, global methods generate approximations for the full random variable space and can find multiple MPPs if they exist. In both cases, a primary strength of the methods lies in the fact that the computational expense is generally unrelated to the probability level: the cost of evaluating a probability in the far tails is no more than that of evaluating near the means.

Local Reliability Methods The Dakota Theory Manual[3] provides the algorithmic details for the local reliability methods as well as references to related research activities. Local methods include first- and second-order versions of the Mean Value method (MVFOSM and MVSOSM) and a variety of most probable point (MPP) search methods, including the Advanced Mean Value method (AMV and AMV^2), the iterated Advanced Mean Value method (AMV+ and AMV^2+), the Two-point Adaptive Nonlinearity Approximation method (TANA-3), and the traditional First Order and Second Order Reliability Methods (FORM and SORM).[11] The MPP search methods may be used in forward (Reliability Index Approach (RIA)) or inverse (Performance Measure Approach (PMA)) modes, as dictated by the type of level mappings. Each of the MPP search techniques solve local optimization problems in order to locate the MPP, which is then used as the point about which approximate probabilities are integrated (using first- or second-order integrations in combination with refinements based on importance sampling). Given settings for limit state approximation, approximation order, integration approach, and MPP-search type, the number of algorithmic combinations is significant. Table 1 provides a succinct mapping for some of these combinations to common method names from the reliability literature, where blue indicates the most well-known combinations and gray indicates other supported combinations.

Table 1. Mapping from Dakota options to standard reliability methods.

	Order of approximation and integration	
MPP search	First order	Second order
none	MVFOSM	MVSOSM
x_taylor_mean	AMV	AMV^2
u_taylor_mean	u-space AMV	u-space AMV^2
x_taylor_mpp	AMV+	AMV^2+
u_taylor_mpp	u-space AMV+	u-space AMV^2+
x_two_point	TANA	
u_two_point	u-space TANA	
no_approx	FORM	SORM

Global Reliability Methods Global reliability methods are designed to handle nonsmooth, multi-modal, and highly nonlinear failure surfaces by creating global approximations and adaptively refining the approximation in the vicinity of a particular response threshold. Three variants are described below: EGRA, GPAIS, and POFDarts.

[a]this nomenclature is due to the origin of these methods within the disciplines of structural safety and reliability; however, the methodology is equally applicable for computation of probabilities unrelated to failure.

Efficient Global Reliability Analysis (EGRA)[12] has its roots in efficient global optimization (EGO).[13,14] The main idea in EGO-type optimization methods is that a global approximation is made of the underlying function. This approximation, which is a Gaussian process model, is used to guide the search by finding points which maximize the expected improvement function (EIF). The EIF is used to select the location at which a new training point should be added to the Gaussian process model by maximizing the amount of improvement in the objective function that can be expected by adding that point. A point could be expected to produce an improvement in the objective function if its predicted value is better than the current best solution, or if the uncertainty in its prediction is such that the probability of it producing a better solution is high. Because the uncertainty is higher in regions of the design space with fewer observations, this provides a balance between exploiting areas of the design space that predict good solutions, and exploring areas where more information is needed.

Gaussian Process Adaptive Importance Sampling (GPAIS)[15] starts with an initial set of LHS samples and adds samples one at a time, with the goal of adaptively improving the estimate of the ideal importance density during the process. The approach uses a mixture of component densities. An iterative process is used to construct the sequence of improving component densities. At each iteration, a Gaussian process (GP) surrogate is used to help identify areas in the space where failure is likely to occur. The GPs are not used to directly calculate the failure probability; they are only used to approximate the importance density. Thus, the Gaussian process adaptive importance sampling algorithm overcomes limitations involving using a potentially inaccurate surrogate model directly in importance sampling calculations.

Probability of Failure Darts (POFDarts)[16] is a novel method for estimating the probability of failure based on random sphere-packing. Random spheres are sampled from the domain with the constraint that each new sphere center has to be outside prior disks. The radius of each sphere is chosen such that the entire sphere lies either in the failure or the non-failure region. This radius depends of the function evaluation at the disk center, the failure threshold, and an estimate of the function gradient at the disk center. After exhausting the specified sampling budget, which is the number of spheres per failure threshold, the domain is decomposed into two regions. These regions correspond to failure and non-failure, each represented by the union of the spheres of each type. The volume of the union of failure spheres gives a lower bound on the required estimate of the probability of failure, while the volume of the union of the non-failure spheres subtracted from the volume of the domain gives an upper estimate. After the spheres are constructed, a surrogate model is constructed and then sampled extensively to estimate the probability of failure for each threshold. The surrogate model can either be a Gaussian process model or a Voronoi piecewise surrogate.

3. Stochastic expansion methods

Stochastic expansion approaches capture the functional relationship between a set of output response metrics and a set of input random variables. In Dakota, we focus on two stochastic expansion methods: Polynomial Chaos Expansion (PCE) and Stochastic Collocation (SC). Polynomial Chaos Expansion is based on a multidimensional orthogonal polynomial approximation and stochastic collocation is based on a multidimensional interpolation polynomial approximation, both formed in terms of standardized random variables. A distinguishing feature of these two methodologies is that the final solution is expressed as a functional mapping, and not merely as a set of statistics as is the case for many nondeterministic methodologies.

One advantage of stochastic expansion methods is their convergence rate: for smooth functions (i.e., analytic, infinitely-differentiable) in L^2 (i.e., possessing finite variance), exponential convergence rates can be obtained under order refinement for integrated statistical quantities of interest such as mean, variance, and probability. Another advantage of stochastic expansion methods is that the moments of the expansion (e.g. mean or variance of the response) can be written analytically, along with analytic formulations of the derivatives of these moments with respect to the uncertain variables. This nice feature can be exploited in design optimization under uncertainty problems or epistemic uncertainty problems.[17] Related to these analytic variance calculations is the capability to perform variance-based decomposition by computing Sobol' indices for main, interaction, and total effects. These indices provide a powerful capability for global sensitivity analysis, where influential input parameters can be identified and rank ordered. Finally, approaches for automated refinement include uniform and adaptive approaches, where the latter includes dimension-adaptive refinement (anisotropic dimension weighting computed from total Sobol' indices or spectral coefficient decay rates) or generalized sparse grid refinement (greedy refinement that evolves the frontier of a sparse grid in an unstructured and goal-oriented manner).

Polynomial Chaos Expansion In PCE, the output response is modeled as a function of the input random variables using a basis of multivariate orthogonal polynomials:

$$R(\boldsymbol{\xi}) \cong \sum_{j=0}^{P} \alpha_j \Psi_j(\boldsymbol{\xi}) \tag{1}$$

where each univariate basis polynomial ψ_{jk} is tailored to the distribution type of the random variable ξ_k. Dakota implements the generalized PCE approach using the Wiener-Askey scheme,[18] in which Hermite, Legendre, Laguerre, Jacobi, and generalized Laguerre orthogonal polynomials are used for modeling the effect of continuous random variables described by normal, uniform, exponential, beta, and gamma probability distributions, respectively. These orthogonal polynomial selections are optimal for these distribution types since the inner product weighting function corresponds to the probability density functions for these continuous distributions. This Wiener-Askey set is then augmented with numerically-generated orthogonal polynomials for non-Askey probability density functions (e.g., truncated normal, Weibull, histogram, etc.). Note that when independent standard random variables are used (or computed through transformation), the variable expansions are uncoupled, allowing the polynomial orthogonality properties to be applied on a per-dimension basis. This allows one to mix and match the polynomial basis used for each variable without interference with the spectral projection scheme for the response.

In non-intrusive PCE, simulations are used as black boxes and the calculation of chaos expansion coefficients for response metrics of interest is based on a set of simulation response evaluations. To calculate these response PCE coefficients, two primary classes of approaches are used: spectral projection and regression. The spectral projection approach projects the response against each basis function using inner products and employs the polynomial orthogonality properties to extract each coefficient. Each inner product involves a multidimensional integral over the support range of the weighting function, which can be evaluated numerically using sampling, tensor-product quadrature, Smolyak sparse grid,[19] or cubature[20] approaches. The regression approach finds a set of PCE coefficients which best match a set of response values obtained from either a design of computer experiments ("point collocation"[21]) or from sub-sampling a set of tensor Gauss points ("probabilistic collocation"[22]). Various methods can be used to solve the resulting linear system, including least squares methods for over-determined systems and compressed sensing methods for under-determined systems. Details of these methods are documented in the Linear regression section of the Dakota Theory Manual.[3]

Stochastic collocation Stochastic collocation (SC) is closely related to PCE, but replaces projection and integration of spectral coefficients with interpolation. An SC interpolant takes the form

$$R(\boldsymbol{\xi}) \cong \sum_{j=1}^{N_p} r_j \boldsymbol{L}_j(\boldsymbol{\xi}) \tag{2}$$

where N_p is the number of unique collocation points in the multidimensional grid. Since the expansion coefficients are just the response values at the sampled points, the core of the method is formation of the multidimensional interpolation polynomials used as the basis. This basis can be formulated to be nodal or hierarchical and either value-based (Lagrange interpolation polynomials) or gradient-enhanced (Hermite interpolation polynomials). Point sets are generally defined using either tensor or sparse grid constructions, where the latter lead to sparse interpolants formed from linear combinations of low-order tensor interpolants. Relative to PCE, SC can be a good selection when either local error estimates (formed from hierarchical surpluses) or explicit interpolation of derivatives (using Hermite interpolants) is important.

4. Epistemic methods

Uncertainty quantification is often used as part of the risk assessment of performance, reliability, and safety of engineered systems. Increasingly, uncertainty is separated into two categories for analysis purposes: aleatory and epistemic uncertainty.[23,24] Aleatory uncertainty is also referred to as variability, irreducible or inherent uncertainty, or uncertainty due to chance. Examples of aleatory uncertainty include the height of individuals in a population, or the temperature in a processing environment. Aleatory uncertainty is typically modeled with probability distributions. In contrast, epistemic uncertainty refers to lack of knowledge or lack of information about a particular aspect of the simulation model, including the system and environment being

modeled. An increase in knowledge or information relating to epistemic uncertainty will lead to a reduction in the predicted uncertainty of the system response or performance. Epistemic uncertainty is referred to as subjective, reducible, or lack of knowledge uncertainty. Examples of epistemic uncertainty include little or no experimental data for a fixed but unknown physical parameter, incomplete understanding of complex physical phenomena, uncertainty about the correct model form to use, etc.

There are many approaches that have been developed to model epistemic uncertainty, including fuzzy set theory, possibility theory, and evidence theory. We have three options for modeling epistemic uncertainties in Dakota: interval analysis, evidence theory, or subjective probability. In the case of subjective probability, the same probabilistic methods for sampling, reliability, or stochastic expansion may be used, albeit with a different subjective interpretation of the statistical results. We describe the interval analysis and evidence theory capabilities in the following sections.

Interval Methods for Epistemic Analysis In interval analysis, one assumes that nothing is known about an epistemic uncertain variable except that its value lies somewhere within an interval. In this situation, it is *not* assumed that the value has a uniform probability of occuring within the interval. Instead, the interpretation is that any value within the interval is a possible value or a potential realization of that variable. In interval analysis, the uncertainty quantification problem is one of determining the resulting bounds on the output (defining the output interval) given interval bounds on the inputs. Correspondingly, any output response that falls within the output interval is a possible output with no frequency information assigned to it.

Interval analysis can be performed using either global or local methods. In the global approach, one uses either a global optimizer or a sampling method to assess the bounds. The optimization-based approach employs efficient global optimization (EGO), which relies on adaptive refinement of a Gaussian process model and reuses data for the related minimization and maximization solves. In the case of discrete epistemic parameters (such as an enumeration of epistemic model form), mixed integer optimization approaches (evolutionary algorithm or surrogate-based global optimization) are used to compute bounds. If the problem is not expected to contain multiple extrema, then one can use local gradient-based optimization methods (sequential quadratic programming or nonlinear interior point methods) to calculate epistemic bounds.

Dempster-Shafer Theory of Evidence Evidence theory, also known as Dempster-Shafer theory or the theory of random sets,[23] is a generalization of probability theory. In evidence theory, there are two complementary measures of uncertainty, belief and plausibility, that together define lower and upper bounds, respectively, on probabilities. In evidence theory, it is not possible to specify one probability value; rather, there is a range of values that is consistent with the evidence.

In Dempster-Shafer evidence theory, the uncertain input variables are modeled as sets of intervals with a basic probability assignment (BPA) for each interval, indicating how likely it is that the uncertain input falls within the interval. The BPAs for a particular uncertain input variable must sum to one and may be overlapping, contiguous, or have gaps. Epistemic analysis is then performed using either global or local methods, using the same algorithm approaches described previously for interval analysis. The primary difference from interval analysis is the number of solves that must be performed, as each unique input BPA bound defines new cells requiring separate minimum and maximum response values. This ensemble of cell minima and maxima are used to define cumulative distribution functions on belief and plausibility.

B. Key research challenges

Within many DOE NNSA, Office of Science, and Energy programs, UQ activities are faced with the daunting challenge of performing moderate-to-high dimensional UQ with very high fidelity models subject to severe resource constraints. For example, UQ for climate predictions requires uncertainty propagations over tens to hundreds of random climate parameters while being restricted to a small number of black-box Community Earth System Model (CESM) simulations. These challenges may be further compounded when faced with a mixture of aleatory and epistemic uncertainties (indicating the need for nested iteration to maintain separation), when requiring evaluation of the probability of rare events (requiring accuracy in distribution tails), and/or when faced with nonsmoothness in response QoI (precluding the use of fast-converging spectral methods using global basis polynomials). In these extreme cases, our most efficient and scalable resource-constrained UQ methodologies may be insufficient on their own, requiring additional deployment of dimension reduction, problem decomposition, and multifidelity modeling strategies to render the problem tractable.

To address these challenges, research and development activites are focusing on extending the scalability of uncertainty quantification (UQ) capabilities, both in terms of problem size and in terms of effective utilization of large-scale parallel computers. Selected R&D highlights include compressed sensing with adaptive basis selection in high dimensions, multifidelity methods for uncertainty quantification, Bayesian inference applied to random field models (Karhunen-Loeve expansions), and tailored methods for mixed aleatory-epistemic UQ including epistemic model form.

1. *Compressed sensing in high dimensions*

A highly effective technique for resource-constrained UQ in high dimensions involves the use of compressed sensing approaches for recovering a sparse polynomial chaos expansion given sparse sampling of the random parameter space. Starting from an under-determined system of equations:

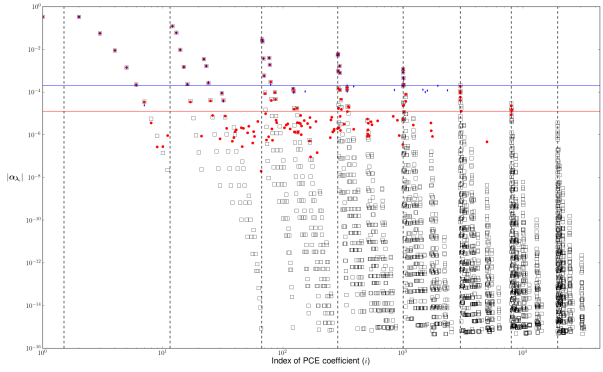
$$\boldsymbol{\Psi\alpha} = \boldsymbol{R} \tag{3}$$

where we wish to recover the most important terms α using a basis set Ψ of cardinality M that is (much) larger than the number of observations N ($M >> N$). Since the under-determined system has infinitely many potential solutions, we apply an ℓ_1 regularization to seek an accurate solution with minimum terms:

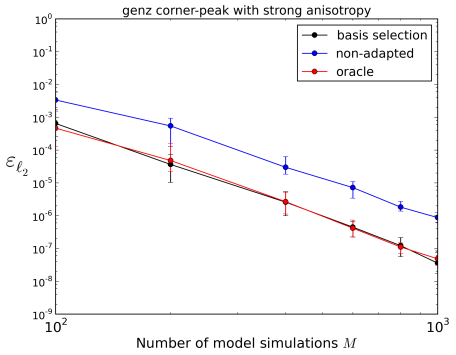
$$\boldsymbol{\alpha} = \arg\min \|\boldsymbol{\alpha}\|_{\ell_1} \quad \text{such that} \quad \|\boldsymbol{\Psi\alpha} - \boldsymbol{R}\|_{\ell_2} \leq \varepsilon \tag{4}$$

In high dimensions (for example, 80 dimensions), the size of a typical total-order candidate basis grows very rapidly (e.g., with orders 2, 3, and 4, M = 3321, 91881, and 1929501, respectively), such that the highest potential polynomial order that can be considered within the candidate basis becomes restricted, both due to computational tractability of the regression solve and due to numerical conditioning (mutual coherence increases with basis size and impedes the ability to discern between basis terms). To address this challenge, we have developed an adaptive selection procedure that evolves the candidate basis in high dimensions.[25]

Figure 1(a) shows the PCE coefficients for the anisotropic Genz corner peak test problem, with comparison of the recovered compressed sensing solutions using the traditional and adapted candidate basis sets. Figure 1(b) compares the convergence rates, where it is evident that selecting an effective candidate basis results in a much more accurate recovery (roughly an order of magnitude lower ℓ_2 error in this example using the same simulation data).



(a) PCE coefficients: full (black), total-order recovery (blue), adapted recovery (red).



(b) Error convergence using total-order (blue) and adapted (black) basis sets compared to oracle (red).

Figure 1. Compressed sensing results for anisotropic Genz corner peak test problem.

2. *Multifidelity UQ*

In a multifidelity UQ approach employing stochastic expansions, we seek to utilize a predictive low-fidelity model to our advantage in reducing the number of high-fidelity model evaluations required to compute high-fidelity statistics to a particular precision. When a low-fidelity model captures useful trends of the

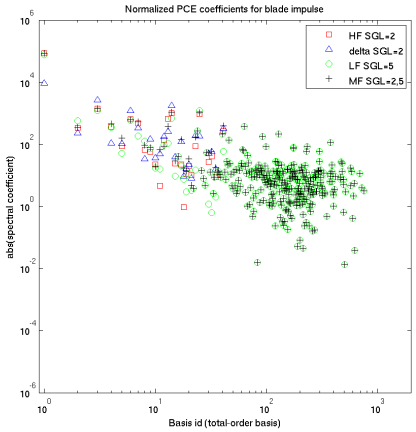
high-fidelity model, then the model discrepancy (the difference between high and low fidelity model results for additive discrepancy or the ratio of high to low results for multiplicative discrepancy) may have lower complexity, lower variance, or greater sparsity, requiring less computational effort to resolve its functional form than that required for the original high-fidelity model.[26] In the case of polynomial chaos expansions, this reduction is computational effort can often be linked to a more rapid decay in the coefficient spectrum of the model discrepancy relative to the decay of the high fidelity coefficient spectrum.

An example multifidelity expansion for a two-model hierarchy is:

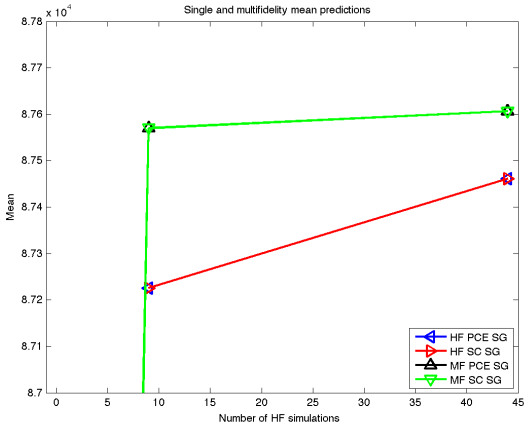
$$\hat{R_{hi}}(\boldsymbol{\xi}) \cong \sum_{j=1}^{N_{lo}} r_{lo}(\boldsymbol{\xi}_j)\boldsymbol{L}_j(\boldsymbol{\xi}) + \sum_{j=1}^{N_{hi}} \Delta r(\boldsymbol{\xi}_j)\boldsymbol{L}_j(\boldsymbol{\xi}) \tag{5}$$

where we employ stochastic collocation with an additive discrepancy Δr and $N_{lo} >> N_{hi}$. In this case, we simply sum the contributions from the low fidelity and discrepancy expansions and define moments and other statistics from this combined expansion. Other variations employing polynomial chaos, multiplicative discrepancies, and additional model instances are similarly straightforward to define.

Figure 2 shows a production result for multifidelity UQ of a vertical axis wind turbine (VAWT), using low fidelity potential flow models and high-fidelity large-eddy simulation. This is a challenging case with nonsmooth response QoI and no known reference solution; yet the multifidelity approach still appears to result in faster convergence to an asymptote in the mean statistic.



(a) Multifidelity coefficient spectrum: multifidelity coefficients (black) are composed from low fidelity (green) and discrepancy (blue) to extend spectrum from high fidelity coefficients (red).



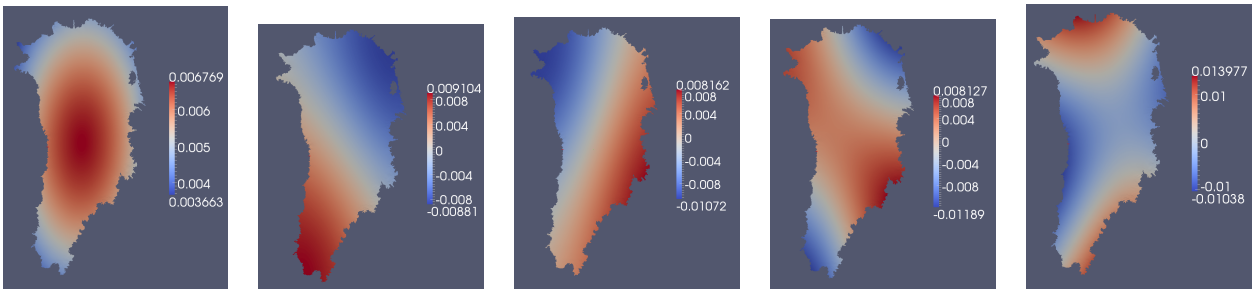
(b) Apparent convergence in mean of blade impulse for multifidelity PCE/SC compared to single fidelity PCE/SC.

Figure 2. Multifidelity stochastic expansions for wind turbine response to an uncertain gust: high fidelity LES uses sparse grid level 2; low fidelity potential flow uses sparse grid level 5.

3. Random field inference

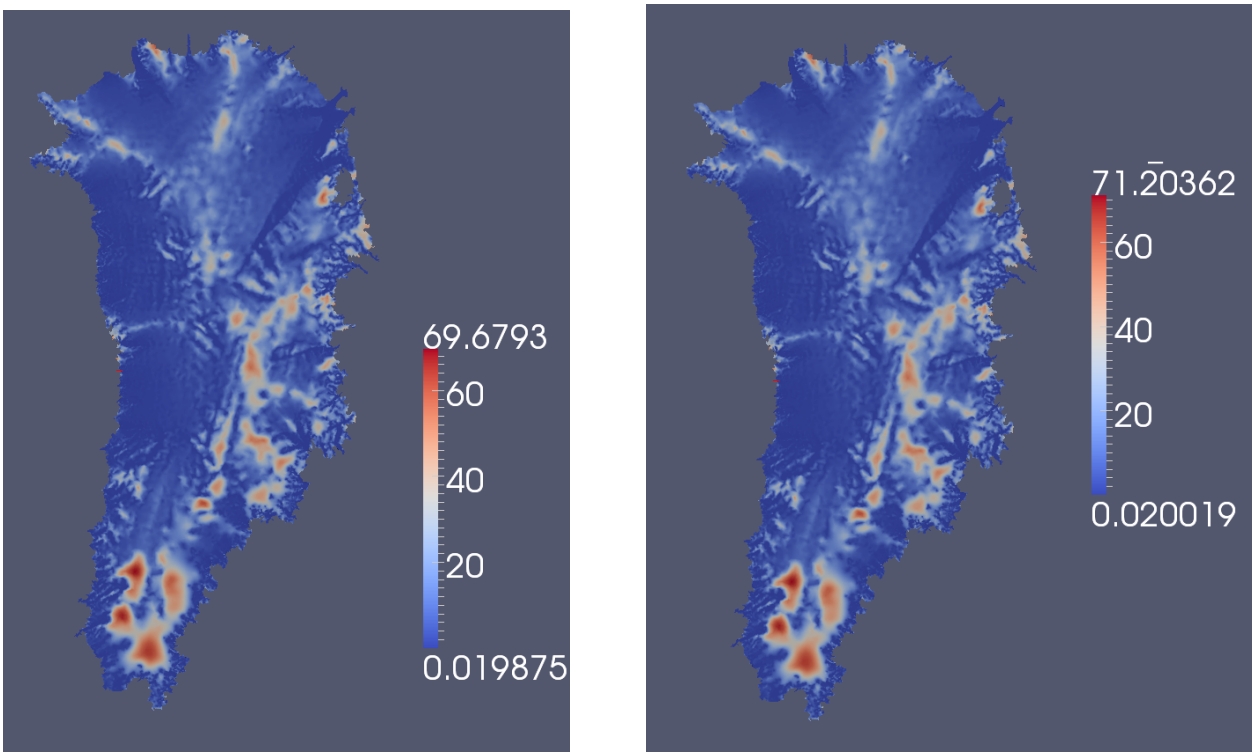
The Scientific Discovery through Advanced Computing (SciDAC) program funded by the DOE Office of Science supports the QUEST institute for uncertainty quantification and the PISCEES application partnership for ice-sheet evolution due to climate forcing. A key challenge for these two programs is the development of capabilities to infer random field characteristics from observational data, in this case inferring the characteristics of the basal sliding interface between the ice sheet and the underlying Greenland bed topography based on ice sheet surface velocity observations. To solve this large-scale statistical inverse problem in a restricted set of dimensions, a Karhunen-Loève expansion (KLE) model is used to define a manageable set of spatial modes for the basal sliding field. Bayesian inference is then performed to update prior distributions for the random coefficients of these spatial modes using the available observational data.

Figure 3 shows the spatial KLE modes for Greenland computed from an assumed exponential form for the spatial covariance, where these modes define excursions from a deterministic reference solution. Figure 4 compares a synthetic truth solution to the maximum a posteriori (MAP) solution recovered from Bayesian



(a) Mode 1. (b) Mode 2. (c) Mode 3. (d) Mode 4. (e) Mode 5.

Figure 3. KLE modes for basal sliding field in Greenland.



(a) Synthetic truth solution for basal field. (b) MAP solution from Bayesian inversion.

Figure 4. Result of Bayesian inversion for random coefficients of spatial KLE modes.

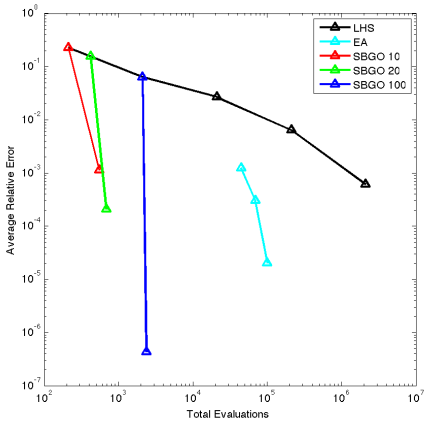
inference. Ongoing work is focused on defining more realistic spatial modes that can better reflect known topographic features.

4. Mixed Aleatory-Epistemic UQ

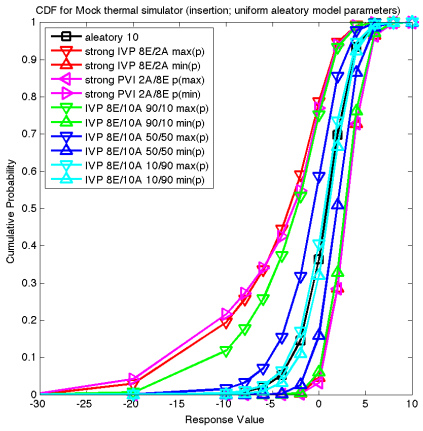
Mixed UQ approaches employ a nested iterative approach when an outer loop UQ is commonly linked to epistemic uncertainties and the inner loop UQ is commonly linked to aleatory uncertainties. The outer level generates sets of realizations of the epistemic parameters, and each set of these epistemic parameters in used within a separate inner loop probabilistic analysis over the aleatory random variables. In this manner, ensembles of aleatory statistics are generated, one set for each realization of the epistemic parameters. In Dakota, we support three approaches for mixed UQ: interval-valued probability (IVP), Dempster-Shafer theory of evidence (DSTE), and second-order probability (SOP). These three approaches support a range of assumptions about the nature of the epistemic variables in the outer loop: they are simple intervals in IVP, more finely-resolved BPA structures in DSTE, and subjective probability distributions in SOP. This

provides a spectrum of assumed epistemic structure, from strongest assumptions in SOP to weakest in IVP.

Recent work has occurred along two fronts: (1) inclusion of discrete model form as epistemic uncertainty, and (2) demonstration of the effects of different assumptions about the separation between aleatory and epistemic sources. Figure 5(a) demonstrates algorithms for the former, where efficient mixed-integer nonlinear programming (MINLP) solvers have been demonstrated for IVP including discrete model form parameters.[27] Convergence of surrogate-based global optimization (SBGO) can be seen to be much more rapid than for the reference Latin hypercube sampling-based approach. Figure 5(b) demonstrates the effect of different assumptions about the separation of aleatory and epistemic assumptions. In particular, it is often the case that model parameters have both inherent variability and lack of knowledge, where a rigorous separation would require a dual parameterization, e.g., epistemic parameters modeling the lack of knowledge about the distribution parameters of an aleatory uncertainty. Neglecting this subtlety and mischaracterizing irreducible variability as reducible epistemic uncertainty (strong assumption cases) is shown to provide conservative bounds so long as irreducible tails are not being truncated. This behavior results from interval optimization (ℓ_∞ metrics) computing more extreme values than statistical integration (ℓ_2 metrics).



(a) MINLP algorithms for mixed UQ including model form: surrogate-based global optimization outperforms Latin hypercube and evolutionary algorithm approaches.



(b) Comparison of different aleatory-epistemic separation assumptions: strong assumption cases bound more rigorous dual parameterization cases.

Figure 5. Demonstration of mixed UQ algorithm capabilities.

IV. Summary Remarks

This paper has described two different aspects of the DOE/NNSA Accelerated Strategic Computing (ASC) program. While these aspects are not directly related, they both support the DOE vision of performing predictive science using high-fidelity simulation models on leadership class computing architectures, one through university outreach and the other through production deployment of scalable UQ algorithms.

Acknowledgments

The overview of the ASC PSAAP II program is provided on behalf of Robert Voigt (Krell Institute) and Lucille Gentry (DOE/NNSA).

References

[1] "Predictive Science Academic Alliance Program II," *2014 Stewardship Science Academic Programs Annual*, 2014, pp. 23–27.

[2] Adams, B. M., Bauman, L. E., Bohnhoff, W. J., Dalbey, K. R., Eddy, J. P., Ebeida, M. S., Eldred, M. S., Hough, P. D., Hu, K. T., Jakeman, J. D., Swiler, L. P., and Vigil, D. M., "Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 5.4 Users Manual," Tech. Rep. SAND2010-2183, Sandia National Laboratories, Albuquerque, NM, Updated Nov. 2013, Available online from

http://dakota.sandia.gov/documentation.html.

[3]Adams, B. M., Bauman, L. E., Bohnhoff, W. J., Dalbey, K. R., Eddy, J. P., Ebeida, M. S., Eldred, M. S., Hough, P. D., Hu, K. T., Jakeman, J. D., Swiler, L. P., and Vigil, D. M., "Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 5.4 Theory Manual," Tech. Rep. SAND2011-9106, Sandia National Laboratories, Albuquerque, NM, Updated Nov. 2013, Available online from http://dakota.sandia.gov/documentation.html.

[4]Swiler, L. P. and Wyss, G. D., "A User's Guide to Sandia's Latin Hypercube Sampling Software: LHS UNIX Library and Standalone Version," Tech. Rep. SAND04-2439, Sandia National Laboratories, Albuquerque, NM, July 2004.

[5]McKay, M. D., Beckman, R. J., and Conover, W. J., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, Vol. 21, No. 2, 1979, pp. 239–245.

[6]Iman, R. L. and Conover, W. J., "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables," *Communications in Statistics: Simulation and Computation*, Vol. B11, No. 3, 1982, pp. 311–334.

[7]Stein, M., "Large sample properties of simulations using latin hypercube sampling," *Technometrics*, Vol. 29, No. 2, 1987, pp. 143–151.

[8]Owen, A., "A central limit theorem for latin hypercube sampling," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 54, No. 2, 1992, pp. 541–551.

[9]Der Kiureghian, A. and Liu, P. L., "Structural Reliability Under Incomplete Information," *J. Eng. Mech., ASCE*, Vol. 112, No. EM-1, 1986, pp. 85–104.

[10]Rosenblatt, M., "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, Vol. 23, No. 3, 1952, pp. 470–472.

[11]Haldar, A. and Mahadevan, S., *Probability, Reliability, and Statistical Methods in Engineering Design*, Wiley, New York, 2000.

[12]Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M., "Efficient Global Reliability Analysis for Nonlinear Implicit Performance Functions," *AIAA Journal*, Vol. 46, No. 10, 2008, pp. 2459–2468.

[13]Jones, D., Schonlau, M., and Welch, W., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, 1998, pp. 455–492.

[14]Huang, D., Allen, T. T., Notz, W. I., and Zeng, N., "Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models," *Journal of Global Optimization*, Vol. 34, 2006, pp. 441–466.

[15]Dalbey, K. and Swiler, L. P., "Gaussian Process Adaptive Importance Sampling," *International Journal for Uncertainty Quantification*, Vol. 4, No. 2, 2014, pp. 133–149.

[16]Ebeida, M., Mitchell, S., Swiler, L., and Romero, V., "POF-Darts: Geometric Adaptive Sampling for Probability of Failure," *SIAM Journal of Uncertainty Quantification, submitted Aug. 2014*, 2014.

[17]Eldred, M. S., Swiler, L. P., and Tang, G., "Mixed Aleatory-Epistemic Uncertainty Quantification with Stochastic Expansions and Optimization-Based Interval Estimation," *Reliability Engineering and System Safety*, Vol. 96, No. 9, 2011, pp. 1092–1113.

[18]Xiu, D. and Karniadakis, G. M., "The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations," *SIAM J. Sci. Comput.*, Vol. 24, No. 2, 2002, pp. 619–644.

[19]Smolyak, S., "Quadrature and interpolation formulas for tensor products of certain classes of functions," *Dokl. Akad. Nauk SSSR*, Vol. 4, 1963, pp. 240–243.

[20]Stroud, A., *Approximate Calculation of Multiple Integrals*, Prentice Hall, 1971.

[21]Walters, R. W., "Towards Stochastic Fluid Mechanics via Polynomial Chaos," *Proceedings of the 41st AIAA Aerospace Sciences Meeting and Exhibit*, No. AIAA-2003-0413, Reno, NV, January 6–9, 2003.

[22]Tatang, M., *Direct incorporation of uncertainty in chemical and environmental engineering systems*, Ph.D. thesis, MIT, 1995.

[23]Oberkampf, W. L. and Helton, J. C., "Evidence Theory for Engineering Applications," Tech. Rep. SAND2003-3559P, Sandia National Laboratories, Albuquerque, NM, 2003.

[24]Helton, J. C., Johnson, J. D., Oberkampf, W. L., and Storlie, C. B., "A Sampling-based Computational Strategy for the Representation of Epistemic Uncertainty in Model Predictions with Evidence Theory," *Comp. Methods in Applied Mechanics and Engineering*, Vol. 196, 2007, pp. 3980–3998.

[25]Jakeman, J., Eldred, M., and Sargsyan, K., "Enhancing ℓ_1-minimization estimates of polynomial chaos expansions using basis selection," *Journal of Computational Physics*, to appear.

[26]Ng, L. W. T. and Eldred, M. S., "Multifidelity Uncertainty Quantification Using Nonintrusive Polynomial Chaos and Stochastic Collocation," *Proceedings of the 14th AIAA Non-Deterministic Approaches Conference*, No. AIAA-2012-1852, Honolulu, HI, April 23-26, 2012.

[27]Eldred, M. S. and Wildey, T. M., "Propagation of Model Form Uncertainty for Thermal Hydraulics using RANS Turbulence Models in Drekar," Tech. Rep. SAND2012-584, Sandia National Laboratories, Albuquerque, NM, 2012.