SAND2014-0289C

Uncertainty Quantification Methods for Model Calibration, Validation, and Risk Analysis

C. Safta[1],K. Sargsyan[1],H.N. Najm[1],K. Chowdhary[1], B.Debusschere[1],L.P. Swiler[2],M.S. Eldred[2]

[1]Sandia National Laboratories
Livermore, CA, USA

[2]Sandia National Laboratories
Albuquerque, NM, USA

16th AIAA Non-Deterministic Approaches Conference
National Harbor, MD
January 17, 2014

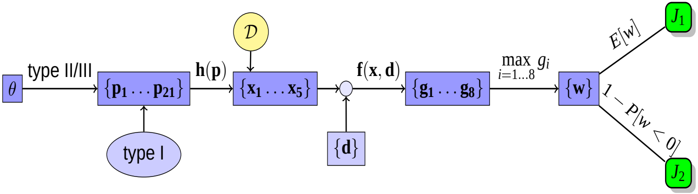
Acknowledgement

- Support for this work was provided through Scientific Discovery through Advanced Computing (SciDAC) program funded by U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research.

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.

Outline

Model Hierarchy and Notations



- $\theta = \{\theta_{1,1}, \ldots, \theta_{1,8}, \theta_{2,1}, \ldots, \theta_{2,7}, \ldots\}$ - set of epistemic parameters
- $\theta \to p$ - Uniform, Gaussian, Beta distributions (for type II/III parameters)
- $p \to x$ is computationally cheap
- $x \to g$ is computationally expensive

Problem A - Model Calibration

Employ a Bayesian framework to characterize the epistemic variables for the input parameters of submodel h_1

$$\underbrace{p(\theta_1|\mathcal{D})}_{\text{Posterior}} = \underbrace{L_{\mathcal{D}}(\theta_1)}_{\text{Likelihood}}\underbrace{p(\theta_1)}_{\text{Prior}}/\underbrace{p(\mathcal{D})}_{\text{Evidence}}$$

- uniform priors for all parameters with bounds specified by the problem setup
- Data $\mathcal{D}$ consists of either 25 or 50 independent samples of x_1 values.
- Likelihood $L_{\mathcal{D}}(\theta_1) = \prod_{i=1}^{N_d} p(x_{1,i}|\theta_1)$

The posterior probability is sampled using a Markov Chain Monte Carlo algorithm

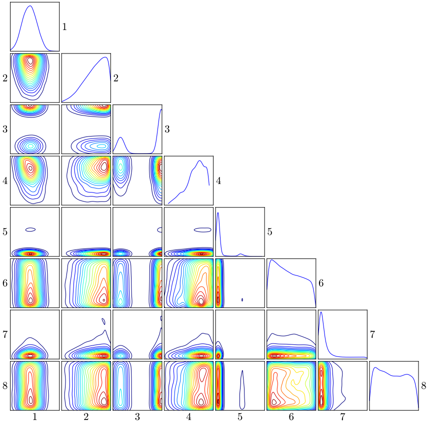
Likelihood Construction

Input: $\theta_1 = \{\theta_{1,1}, \ldots, \theta_{1,8}\}$, Algorithm knobs: N_{spl}, N_{bin}
Output: $L_{\mathcal{D}}(\theta_1)$

Generate N_{spl} sets of $(p_1, \ldots, p_5)$:
foreach $k = 1, \ldots, N_{spl}$ **do**
    $p_{1,k} \sim Beta(\alpha(\theta_{1,1}, \theta_{1,2}), \beta(\theta_{1,1}, \theta_{1,2}))$
    $p_{2,k} = \theta_{1,3}$
    $p_{3,k} \sim U[0, 1]$
    $(p_{4,k}, p_{5,k}) \sim N[\mu(\theta_{1,4}, \theta_{1,5}), \Sigma(\theta_{1,6}, \theta_{1,7}, \theta_{1,8})]$
end
$\{x_1|_{\theta_1}\} = \{h_1(\mathbf{p_k}), k = 1, \ldots, N_{spl}\}$
Evaluate $p(x_{1,i})$, $i = 1, \ldots, N_d$ by binning the $\{x_1|\theta_1\}$ ensemble
Evaluate $L_{\mathcal{D}}(\theta_1) = \prod_{i=1}^{N_d} p(x_{1,i}|\theta_1)$

Posterior Distributions-1D and 2D marginals

While computationally expensive, the Bayesian framework allows the computation of full posterior distributions based on available data

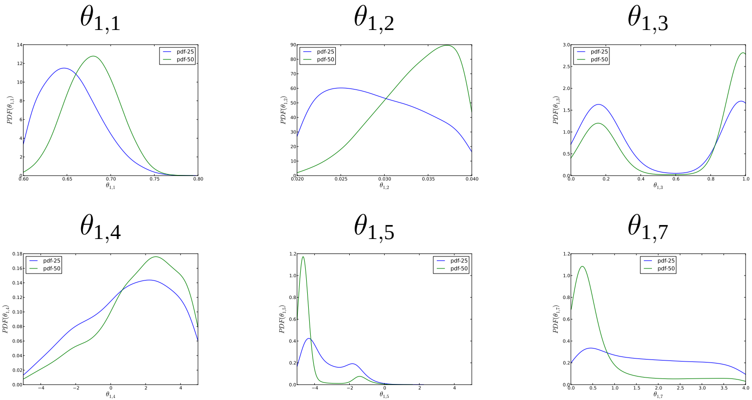


Results based on 50 x_1 values

$\theta_{1,1}$	$E[p_1]$
$\theta_{1,2}$	$V[p_1]$
$\theta_{1,3}$	p_3
$\theta_{1,4}$	$E[p_4]$
$\theta_{1,5}$	$E[p_5]$
$\theta_{1,6}$	$V[p_4]$
$\theta_{1,7}$	$V[p_5]$
$\theta_{1,8}$	$\rho[p_4, p_5]$

1D Posterior Marginals Based on 25 and 50 x_1 Values

Generally, more data points lead to sharper PDF's for the parameters that are informed by the data



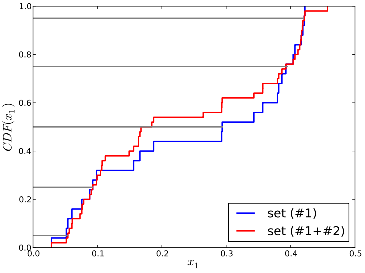
Approximate Bayesian Computations

Full calibration is expensive - each likelihood evaluation requires 10^5 h_1 model evaluation for approximately $10^5 - 10^6$ MCMC samples.

$$L_{\mathcal{D}}(\theta) \propto \frac{1}{\epsilon} K\left(\frac{\|s(\mathcal{D}^*(\theta)) - s(\mathcal{D})\|}{\epsilon}\right)$$

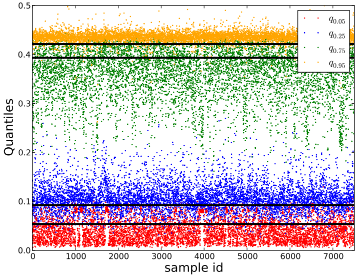
$$\|s^* - s\| = \sqrt{\sum_{i=1}^{N_s} (s_i^* - s_i)^2}$$

- s - summary statistics based on provided data
- s^* - summary statistics based on simulated data

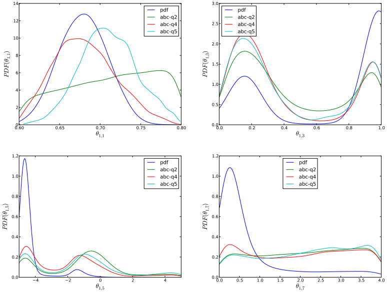


Approximate Bayesian Computations

Comparison of summary statistics based on ABC computations



Comparison of 1D marginal PDFs



The ABC methodology is about two orders of magnitude less expensive compared to the full Bayesian approach.

Problem B - Sensitivity Analysis

Employ variance-based Global Sensitivity Analysis (GSA) to understand the effects of the epistemic variables θ and the aleatoric parameters p on the intermediate variables x and the Quantities of Interest (QoI) J_1 and J_2.

- Use Sobol indices to understand the connectivities between p and x and between θ and x, and address question B1
- Employ a cut-HDMR approach to answer B2 and B3 about J_1 and J_2.

Sobol Indices

First order sensitivity:

$$S_i = \frac{\mathrm{Var}[\mathbb{E}(h(\mathbf{p})|p_i)]}{\mathrm{Var}[h(\mathbf{p})]},$$

Total order sensitivity:

$$S_i^T = \frac{\mathbb{E}[\mathrm{Var}(h(\mathbf{p})|p_{-i})]}{\mathrm{Var}[h(\mathbf{p})]}$$

Joint sensitivity:

$$S_{ij} = \frac{\mathrm{Var}[\mathbb{E}(h(\mathbf{p})|p_i, p_j)]}{\mathrm{Var}[h(\mathbf{p})]} - S_i - S_j,$$

Sobol Indices $p \to x = h(p)$

h_1

	S_i	S_i^T
$\mathbf{p_1}$	**95.80**	**97.20**
p_2	0.20	0.01
p_3	0	0
p_4	3.00	0
$\mathbf{p_5}$	**3.40**	**4.00**

h_2

	S_i	S_{ij}			
		p_7	p_8	p_9	p_{10}
$\mathbf{p_6}$	**7.57**	**82.50**	0.22	0.08	0.01
$\mathbf{p_7}$	**6.66**	-	0.28	0.14	0.01
p_8	0.27	-	-	0.01	0
p_9	0.11	-	-	-	0
p_{10}	0	-	-	-	-

h_3

	S_i	S_{ij}		
		p_{12}	p_{14}	p_{15}
p_{11}	**1.07**	0.88	0.03	0
$\mathbf{p_{12}}$	**92.53**	-	**2.97**	0.19
$\mathbf{p_{14}}$	**2.22**	-	-	0
p_{15}	0.20	-	-	-

h_4

	S_i	S_{ij}		
		p_{17}	p_{18}	p_{20}
$\mathbf{p_{16}}$	**42.76**	**28.02**	**1.63**	0.17
$\mathbf{p_{17}}$	**11.17**	-	0.40	0.03
$\mathbf{p_{18}}$	**14.42**	-	-	0.05
p_{20}	0.36	-	-	-

Some parameters have negligible effect on select sub-model variances

Sobol Indices $\theta \rightarrow A_{p-box}$

"Area" of a p-box is defined as

$$\int_{\min_i(x_i)}^{\max_i(x_i)} \left(\max_{\theta_\mathbf{i}} F(x_i|\theta_\mathbf{i}) - \min_{\theta_\mathbf{i}} F(x_i|\theta_\mathbf{i})\right) dx_i$$

h_1:	p_1				p_2		p_5			
	$\theta_{1,1}$		$\theta_{1,2}$		$\theta_{1,3}$		$\theta_{1,5}$		$\theta_{1,7}$	
	0.67	**0.42**	**0.35**	**0.33**	**0.33**	**0.32**	**0.47**	**0.56**	**0.38**	**0.35**

h_2:	p_6		p_7				p_8			
	$\theta_{2,1}$		$\theta_{2,2}$		$\theta_{2,3}$		$\theta_{2,4}$		$\theta_{2,5}$	
	0.27	**0.52**	**0.23**	0.05	0.04	0.04	0.03	0.03	0.03	0.03

h_3:	p_{12}		p_{13}				p_{14}				p_{15}			
	$\theta_{\mathbf{3},\mathbf{1}}$		$\theta_{3,2}$		$\theta_{3,3}$		$\theta_{3,4}$		$\theta_{3,5}$		$\theta_{3,6}$		$\theta_{3,7}$	
	0.95	**0.11**	0	0	0	0	0	0	0	0	0	0	0	0

h_4:	p_{16}		p_{17}				p_{18}			
	$\theta_{\mathbf{4},\mathbf{1}}$		$\theta_{4,2}$		$\theta_{4,3}$		$\theta_{4,4}$		$\theta_{4,5}$	
	0.25	**0.83**	0.02	0.01	0.01	0.01	0.03	0.01	0.01	0.01

High-Dimensional Model Representation (HDMR) to Examine the Relative Impact of θ on J_1 and J_2

The model output is expressed in terms of a hierarchy of functions that account for the interaction between model parameters

$$M(\theta) = f_0 + \sum_{i=1}^{n} f_i(\theta_i) + \sum_{1 \le i < j \le n} f_{ij}(\theta_i, \theta_j) + \dots$$

Resort to a "cut-HDMR" approach to compute the component functions f:

$$\begin{aligned}
f_0 &= M(\theta^0) \\
f_i(\theta_i) &= M(\theta_i, \theta^0_{-i}) - f_0 \\
f_{ij}(\theta_i, \theta_j) &= M(\theta_i, \theta_j, \theta^0_{-ij}) - f_i(\theta_i) - f_j(\theta_j) - f_0 \\
&\dots
\end{aligned}$$

HDMR Component Functions approximated via Polynomial Chaos Expansions

We adopt a Polynomial Chaos (PC) representation for the HDMR component functions. For univariate components:

$$f_i(\theta_i) \approx \sum_{k=0}^{K} c_k^i \Psi_k(\xi_i(\theta_i))$$

while for bi-variate components the PC approximations are given by

$$f_{ij}(\theta_i, \theta_j) \approx \sum_{\substack{0 \le k_1, k_2 \\ k_1+k_2 \le K}} c_{k_1,k_2}^{ij} \Psi_{k_1}(\xi_i(\theta_i)) \Psi_{k_2}(\xi_j(\theta_j))$$

Since θ have bounded support, we choose a Legendre-Uniform basis with $\xi_i \sim U[-1, 1]$ and Ψ_k the Legendre polynomial of order k.

PC Expansion Coefficients Computed via Galerkin Projection

Using the orthogonality of basis terms:

$$c_k^i = \frac{\langle (M(\theta_i, \theta_{-i}^0) - M_0)\Psi_k \rangle}{\langle \Psi_k^2 \rangle}$$

- For univariate coefficients, 5 model M evaluations are necessary for a $3rd$ order expansion
- Each model evaluation requires about 5×10^4 $g(x)$ evaluations to obtain converged values for $M = J_1$ and about 10^4 evaluations for converged $M = J_2$ values.

Rank $p(\theta)$ based on Fractional Contribution to the Total Variance of J_1 and J_2

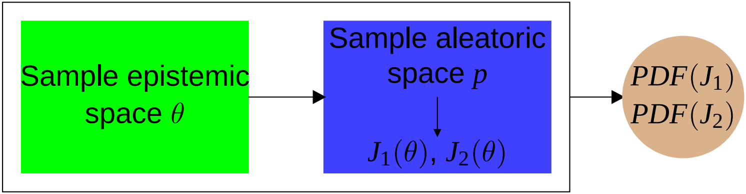
Fractional contribution of each input parameter to the total variance of the model:

$$V_i = \sum_{k=1}^{K} (c_k^i)^2 \langle \Psi_k^2 \rangle$$

Rank	J_1		J_2	
	p	$V_p/\max(V_p)$	p	$V_p/\max(V_p)$
1	p_{21}	1	p_1	1
2	p_1	0.17	p_{12}	0.26
3	p_5	0.02	p_5	0.14
4	p_7	0.02	p_{14}	2×10^{-3}

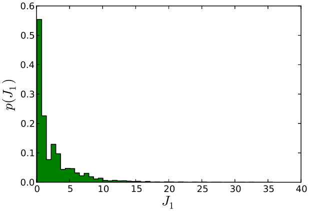
Problem C - Uncertainty Propagation

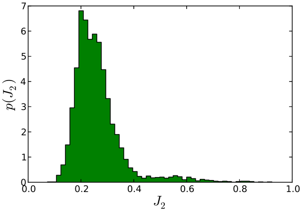
Employ a nested sampling approach to quantify the probability distributions for the Quantities of Interest (QoIs) J_1 and J_2



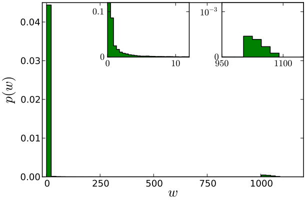
- $N_\theta = 5,000$ and $N_p = 1,000$ samples in the outer and inner loop, respectively

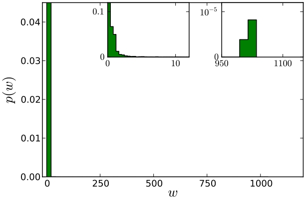
Histograms of J_1 and J_2 based on Original Epistemic Information



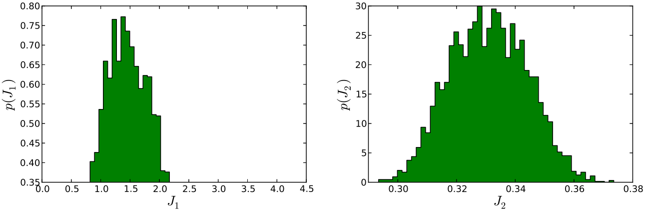


Histogram of aleatoric samples at epistemic realizations corresponding to smallest and highest J_1 values:





Histograms of J_1 and J_2 based on Improved Epistemic Information



- Results based on $N_\theta = 5,000$ and $N_p = 10,000$ samples in the outer and inner loop, respectively

Problem E - Robust Design

Extend the nested sampling approach to determine the optimal values for design variables d

Find optimal values of $d \in \mathfrak{R}^{14}$ resulting in minimal upper bound of J_1

Find bounds on J_1 given epistemic parameters $\theta \in \mathfrak{R}^{30}$

Determine J_1 given $p \in \mathfrak{R}^{21}$

Problem E - Algorithm Choices in Dakota

- Design Optimization
  - Tailored approach: exploit well-behaved functionals, use gradient-based optimization or adaptive GP surrogate-based optimization (e.g. efficient global optimization, EGO).
  - Fallback approach: genetic algorithm or pattern search.
- Epistemic Uncertainty
  - Tailored approach: Adaptive Gaussian-Process based interval optimization methods (variant of EGO). Can also use gradient-based optimization to determine bounds.
  - Fallback approach: LHS.
- Aleatory Uncertainty
  - Tailored approach: exploit solution smoothness of response, use Polynomial Chaos Expansions with Compressed Sensing to reduce dimensionality.
  - Fallback approach: LHS sampling. Reason: aggregation of 8 performance metrics and selecting the max results in nonlinear behavior and possible discontinuities

Summary

- Employed probabilistic methodologies to handle inverse and forward UQ studies for the models posed by the NASA LaRC UQ challenge.
- The calibration exercise is performed in a Bayesian framework. Posterior distributions are compared with results based on Approximate Bayesian Computation concepts.
- Variance-based Global Sensitivity Analysis for the effects the epistemic and aleatoric inputs have on models h and on the statistics J_1 and J_2.
- Double-nested sampling loop for forward UQ. Methodology challenged by the computational expense and the multi-modal behavior in the computation of QoIs.
- Proof-of-concept triple-nested loop for the design optimization exercise.