Dakota Sensitivity Analysis and Uncertainty Quantification, with Examples
Dakota Uncertainty Quantification (UQ)

- UQ goals and examples
- Select Dakota examples for UQ:
  - Monte Carlo sampling
  - Local and global reliability
  - Polynomial chaos expansions / stochastic collocation
  - Mixed aleatory-epistemic approaches
  - Probabilistic design

- Dakota primarily focuses on forward propagation
  - Secondarily on estimating parameter uncertainty given data
  - Not on processing experimental data to calculate uncertainties
Drivers for Dakota UQ

Current Dakota research and development largely focuses on efficient UQ for large-scale engineering analyses.

DOE in general, ASC V&V in particular, are:

- Responding to shift from test-based to modeling and simulation-based design and certification
- Demanding risk-informed decision-making using credible M&S:
  - Predictive simulations: verified, validated for application domain of interest
  - Quantified margins and uncertainties: random variability effect is understood, best estimate with uncertainty prediction for decision-making
Why Perform Uncertainty Quantification?

- What? Determine variability, distributions, statistics of code outputs, given uncertainty in input factors
- Why? Assess likelihood of typical or extreme outcomes. Given input uncertainty...
  - Determine mean or median performance of a system
  - Assess variability in model response
  - Find probability of reaching failure/success criteria (reliability metrics)
  - Assess range/intervals of possible outcomes
- Assess how close uncertainty-endowed code predictions are to
  - Experimental data (validation, is model sufficient *for the intended application*)
  - Performance expectations or limits (quantification of margins and uncertainties; QMU)
Many Potential Uncertainties in Simulation and Validation

- physics/science parameters
- statistical variation, inherent randomness
- model form / accuracy
- material properties
- manufacturing quality
- operating environment, interference
- initial, boundary conditions; forcing
- geometry / structure / connectivity
- experimental error (measurement error, measurement bias)
- numerical accuracy (mesh, solvers); approximation error
- human reliability, subjective judgment, linguistic imprecision

The effect of these on model outputs should be integral to an analyst’s deliverable: best estimate PLUS uncertainty!
Forward Parametric Uncertainty Quantification

- Identify and characterize uncertain variables (may not be normal, uniform)
- *Forward propagate: quantifying the effect that (potentially correlated) uncertain (nondeterministic) input variables have on model output:*

**Input Variables** \( u \) (physics parameters, geometry, initial and boundary conditions) → **Computational Model** → **Variable Performance Measures** \( f(u) \)

**Uncertainties on inputs**
- Parameterized distributions: normal, uniform, gumbel, etc.
- Means, standard deviations
- PDF, CDF from data
- Intervals
- Belief structures

**Uncertainties on outputs**
- Means, standard deviations
- Probabilities
- Reliabilities
- PDF, CDF
- Intervals
- Belief, plausibility
Example: Thermal Uncertainty Quantification

- Device subject to heating (experiment or computational simulation)
- Uncertainty in composition/environment (thermal conductivity, density, boundary), parameterized by $u_1, \ldots, u_N$
- Response temperature $f(u) = T(u_1, \ldots, u_N)$ calculated by heat transfer code

**Final Temperature Values**

Given distributions of $u_1, \ldots, u_N$, UQ methods calculate statistical info on outputs:

- Mean($T$), StdDev($T$), Probability($T \geq T_{\text{critical}}$)
- Probability distribution of temperatures
- Correlations (trends) and sensitivity of temperature
Example: Uncertainty in Boiling Rate in Nuclear Reactor Core

<table>
<thead>
<tr>
<th>Method</th>
<th>ME_nnz</th>
<th>ME_meannz</th>
<th>ME_max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Dev</td>
<td>Mean</td>
</tr>
<tr>
<td>LHS (40)</td>
<td>651.225</td>
<td>297.039</td>
<td>127.836</td>
</tr>
<tr>
<td>LHS (400)</td>
<td>647.33</td>
<td>286.146</td>
<td>127.796</td>
</tr>
<tr>
<td>LHS (4000)</td>
<td>688.261</td>
<td>292.687</td>
<td>129.175</td>
</tr>
<tr>
<td>PCE (Θ(2))</td>
<td>687.875</td>
<td>288.140</td>
<td>129.151</td>
</tr>
<tr>
<td>PCE (Θ(3))</td>
<td>688.083</td>
<td>292.974</td>
<td>129.231</td>
</tr>
<tr>
<td>PCE (Θ(4))</td>
<td>688.099</td>
<td>292.808</td>
<td>129.213</td>
</tr>
</tbody>
</table>

Mean and standard deviation of key metrics

Anisotropic uncertainty distribution in boiling rate throughout quarter core model

Normally distributed inputs need not give rise to normal outputs...
Three Core Dakota UQ Methods

- **Sampling** (Monte Carlo, Latin hypercube): robust, easy to understand, slow to converge / resolve statistics

- **Reliability**: good at calculating probability of a particular behavior or failure / tail statistics; efficient, some methods are only local

- **Stochastic Expansions** (PCE/SC global approximations): efficient tailored surrogates, statistics often derived analytically, *far more efficient than sampling for reasonably smooth functions*
Given distributions of $u_1, \ldots, u_N$, sampling-based methods calculate sample statistics, e.g., on temperature $T(u_1, \ldots, u_N)$:

- **Sample mean**
  \[
  \bar{T} = \frac{1}{N} \sum_{i=1}^{N} T(u^i)
  \]

- **Sample variance**
  \[
  T_{\sigma^2} = \frac{1}{N} \sum_{i=1}^{N} [T(u^i) - \bar{T}]^2
  \]

- **Full PDF (probabilities)**

**Output Distributions**

- Measure 1
- Measure 2

**Model**

- Monte Carlo sampling, Quasi-Monte Carlo
- Centroidal Voroni Tessalation (CVT)
- Latin hypercube (stratified) sampling: better convergence; stability across replicates

Robust, but slow convergence: $O(N^{-1/2})$, independent of dimension *(in theory)*

**Final Temperature Values**

<table>
<thead>
<tr>
<th>Temperature [deg C]</th>
<th>% in Bin</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>36</td>
<td>0.5</td>
</tr>
<tr>
<td>42</td>
<td>1</td>
</tr>
<tr>
<td>48</td>
<td>1.5</td>
</tr>
<tr>
<td>54</td>
<td>2</td>
</tr>
<tr>
<td>60</td>
<td>2.5</td>
</tr>
<tr>
<td>66</td>
<td>3</td>
</tr>
<tr>
<td>72</td>
<td>3.5</td>
</tr>
<tr>
<td>78</td>
<td>4</td>
</tr>
<tr>
<td>84</td>
<td>4.5</td>
</tr>
<tr>
<td>90</td>
<td>5</td>
</tr>
</tbody>
</table>

**N samples**

- $u_1$
- $u_2$
- $u_3$
Example:
Cantilever Beam UQ with Sampling

- Dakota study with LHS
- Determine mean system response, variability, margin to failure given
  - Density $P \sim \text{Normal}(500, 30)$
  - Young’s modulus $E \sim \text{Normal}(2.9e7, 2.e6)$
  - Horizontal load $X \sim \text{Normal}(50, 3)$
  - Vertical load $Y \sim \text{Normal}(100, 6)$
  - *(Dakota supports a wide range of distribution types)*
- Hold width and thickness at 1.0, L at 5.
- Compute with respect to thresholds with probability_levels or response_levels
  - What is the probability(stress < 10000)?
  - What is the probability(mass < 1.5)?
  - What is the probability(displacement < 0.002)?
Example:
Cantilever Beam UQ with Sampling

- Dakota study with LHS
- Determine mean system response, variability, margin to failure given
  - Density $P \sim \text{Normal}(500, 30)$
  - Young’s modulus $E \sim \text{Normal}(2.9e7, 2.e6)$
  - Horizontal load $X \sim \text{Normal}(50, 3)$
  - Vertical load $Y \sim \text{Normal}(100, 6)$
  - (Dakota supports a wide range of distribution types)
- Hold width and thickness at 1.0, L at 5.
- Compute with respect to thresholds with probability_levels or response_levels
  - What is the probability(stress < 10000)? $\sim 0.9$ for uniform, $0.99$ for normal
  - What is the probability(mass < 1.5)? $\sim 0.6$ for uniform, $0.8$ for normal
  - What is the probability(displacement < 0.002)? $\sim 0.6$ for uniform, $0.7$ for normal
Dakota Input:
LHS Sampling for Cantilever Beam

method
  sampling
  sample_type lhs
  samples = 100
  seed = 3845

  num_probability_levels = 17 17 17
  probability_levels =
    .001 .01 .05 .1 .15 .2 .3 .4 .5 .6 .7 .8 .85 .9 .95 .99 .999
    .001 .01 .05 .1 .15 .2 .3 .4 .5 .6 .7 .8 .85 .9 .95 .99 .999
    .001 .01 .05 .1 .15 .2 .3 .4 .5 .6 .7 .8 .85 .9 .95 .99 .999
  cumulative distribution

variables
  active uncertain
  continuous_design = 3
    upper_bounds = 1.2 1.2 6.0
    lower_bounds = 0.8 0.8 4.0
    descriptors    "w"     "t"     "L"

  uniform_uncertain = 4
    upper_bounds = 600. 35.E+6 60. 120.
    lower_bounds = 400. 23.E+6 40. 80.
    descriptors    'p'   'E'   'X'  'Y'

  ... responses
  response_functions = 3
    descriptors = 'mass' 'stress' 'displacement'
  no_gradients no_hessians
Dakota Output:
LHS Sampling for Cantilever Beam

- Moments and confidence intervals

<table>
<thead>
<tr>
<th>Level mappings for each response function:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative Distribution Function (CDF) for mass:</td>
</tr>
<tr>
<td>Response Level</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>1.1683297300e+00</td>
</tr>
<tr>
<td>1.1683297300e+00</td>
</tr>
<tr>
<td>1.2951111800e+00</td>
</tr>
<tr>
<td>1.3316578300e+00</td>
</tr>
<tr>
<td>1.3559746900e+00</td>
</tr>
<tr>
<td>1.3734105800e+00</td>
</tr>
<tr>
<td>1.4003385200e+00</td>
</tr>
<tr>
<td>1.4245467700e+00</td>
</tr>
</tbody>
</table>

Statistics based on 100 samples:

<table>
<thead>
<tr>
<th>Moment-based statistics for each response function:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean              Std Dev              Skewness            Kurtosis</td>
</tr>
<tr>
<td>mass  1.4460475709e+00   8.8239262134e-02   -1.605107447e-01  2.5955294928e-01</td>
</tr>
<tr>
<td>stress 8.9986343326e+04   4.0344159128e+03   6.4230716871e-02  1.0335094626e-01</td>
</tr>
<tr>
<td>displacement 1.9378806350e-03   1.6660999428e-04   5.5574418567e-01  5.8860476955e-01</td>
</tr>
</tbody>
</table>

95% confidence intervals for each response function:

<table>
<thead>
<tr>
<th>LowerCI_Mean</th>
<th>UpperCI_Mean</th>
<th>LowerCI_StdDev</th>
<th>UpperCI_StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>mass  1.4285389869e+00   1.4635561549e+00   7.7474676187e-02  1.0250536737e-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stress 8.9185827682e+04   9.0786858970e+04   3.5422447886e+03  4.6866811355e+03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>displacement 1.9048215975e-03   1.9709396725e-03   1.4628471549e-04  1.9354670764e-04</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- CDF (and PDF) data

CDF plotted in Matlab
Challenge: Calculating Potentially Small Probability of Failure

- Given uncertainty in materials, geometry, and environment, how to determine likelihood of failure: \( \text{Probability}(T \geq T_{\text{critical}}) \)?
- Perform 10,000 LHS samples and count how many exceed threshold; (better) perform adaptive importance sampling

**Mean value:** make a linearity (and possibly normality) assumption and project; great for many parameters with efficient derivatives!

\[
\mu_T = T(\mu_u) \\
\sigma_T = \sum_i \sum_j \text{Cov}_u(i, j) \frac{dg}{d\mu_i}(\mu_u) \frac{dg}{d\mu_j}(\mu_u)
\]

**Reliability:** directly determine input variables which give rise to failure behaviors by solving an optimization problem for a most probable point (MPP) of failure

\[
\text{minimize } u^T u \\
\text{subject to } T(u) = T_{\text{critical}}
\]
Perform optimization in uncertain variable space to determine Most Probable Point (of response or failure occurring) for \( G(u) = T(u) \).

**Reliability Index Approach (RIA)**

\[
\text{minimize} \quad u^T u \\
\text{subject to} \quad G(u) = \bar{z}
\]

_All the usual nonlinear optimization tricks apply…_
Efficient Global Reliability Analysis Using Gaussian Process Surrogate + MMAIS

- Efficient global optimization (EGO)-like approach to solve optimization problem
- Expected feasibility function: balance exploration with local search near failure boundary to refine the GP
- Cost competitive with best local MPP search methods, yet better probability of failure estimates; addresses nonlinear and multimodal challenges

Gaussian process model (level curves) of reliability limit state with 10 samples vs. 28 samples.
Generalized Polynomial Chaos Expansions (PCE)

Approximate response with Galerkin projection using multivariate orthogonal polynomial basis functions defined over standard random variables

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

$$R(\xi) \approx f(u)$$

- Intrusive or non-intrusive
- Wiener-Askey Generalized PCE: optimal basis selection leads to exponential convergence of statistics

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Density function</th>
<th>Polynomial</th>
<th>Weight function</th>
<th>Support range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$</td>
<td>Hermite $H_{\xi}(x)$</td>
<td>$e^{-\frac{x^2}{2}}$</td>
<td>$[-\infty, \infty]$</td>
</tr>
<tr>
<td>Uniform</td>
<td>$\frac{1}{2}$</td>
<td>Legendre $P_n(x)$</td>
<td>1</td>
<td>$[-1, 1]$</td>
</tr>
<tr>
<td>Beta</td>
<td>$\frac{(1-x)^{\alpha}(1+x)^{\beta}}{2^{\alpha+\beta+1}B(\alpha+1,\beta+1)}$</td>
<td>Jacobi $P_{\alpha,\beta}(x)$</td>
<td>$(1-x)^{\alpha}(1+x)^{\beta}$</td>
<td>$[-1, 1]$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$e^{-x}$</td>
<td>Laguerre $L_n(x)$</td>
<td>$e^{-x}$</td>
<td>$[0, \infty]$</td>
</tr>
<tr>
<td>Gamma</td>
<td>$\frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)}$</td>
<td>Generalized Laguerre $L_{n}^{(\alpha)}(x)$</td>
<td>$x^\alpha e^{-x}$</td>
<td>$[0, \infty]$</td>
</tr>
</tbody>
</table>

- Can also numerically generate basis orthogonal to empirical data (PDF/histogram)
Sample Designs to Form Polynomial Chaos or Stochastic Collocation Expansions

Random sampling: PCE

Expectation (sampling):
- Sample w/i distribution of $x$
- Compute expected value of product of $R$ and each $Y_j$

Linear regression ("point collocation"):

$$
\Psi \alpha = R
$$

Tensor-product quadrature: PCE/SC

Tensor product of 1-D integration rules, e.g., Gaussian quadrature

Smolyak Sparse Grid: PCE/SC

Cubature: PCE

Stroud and extensions (Xiu, Cools):
optimal multidimensional integration integration rules
Adaptive PCE/SC: Emphasize Key Dimensions

- Judicious choice of new simulation runs
- Uniform p-refinement
  - Stabilize 2-norm of covariance
- Adaptive p-refinement
  - Estimate main effects/VBD to guide
- h-adaptive: identify important regions and address discontinuities
- h/p-adaptive: p for performance; h for robustness

Anisotropic index sets

Anisotropic Gauss-Hermite
Changes for Reliability, PCE

<table>
<thead>
<tr>
<th>Changes for Reliability, PCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>method,</strong></td>
</tr>
<tr>
<td>local_reliability</td>
</tr>
<tr>
<td>num_probability_levels = 17 17 17</td>
</tr>
<tr>
<td>probability_levels =</td>
</tr>
<tr>
<td>.001 .01 .05 .1 .15 .2 .3 .4 .5 .6 .7 .8</td>
</tr>
<tr>
<td>.85 .9 .95 .99 .999</td>
</tr>
<tr>
<td>.001 .01 .05 .1 .15 .2 .3 .4 .5 .6 .7 .8</td>
</tr>
<tr>
<td>.85 .9 .95 .99 .999</td>
</tr>
<tr>
<td>.001 .01 .05 .1 .15 .2 .3 .4 .5 .6 .7 .8</td>
</tr>
<tr>
<td>.85 .9 .95 .99 .999</td>
</tr>
<tr>
<td>cumulative distribution</td>
</tr>
<tr>
<td>responses</td>
</tr>
<tr>
<td>response_functions = 3</td>
</tr>
<tr>
<td>descriptors = 'mass' 'stress' 'displacement'</td>
</tr>
<tr>
<td>numerical_gradients</td>
</tr>
<tr>
<td>method_source dakota</td>
</tr>
<tr>
<td>interval_type central</td>
</tr>
<tr>
<td>fd_gradient_step_size = 0.0001</td>
</tr>
<tr>
<td>no_hessians</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changes for Reliability, PCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>method,</strong></td>
</tr>
<tr>
<td>polynomial_chaos</td>
</tr>
<tr>
<td>sparse_grid_level = 2</td>
</tr>
<tr>
<td>sample_type lhs</td>
</tr>
<tr>
<td>samples = 10000</td>
</tr>
<tr>
<td>seed = 8572</td>
</tr>
<tr>
<td>num_probability_levels = 17 17 17</td>
</tr>
<tr>
<td>probability_levels =</td>
</tr>
<tr>
<td>.001 .01 .05 .1 .15 .2 .3 .4 .5 .6 .7 .8</td>
</tr>
<tr>
<td>.85 .9 .95 .99 .999</td>
</tr>
<tr>
<td>.001 .01 .05 .1 .15 .2 .3 .4 .5 .6 .7 .8</td>
</tr>
<tr>
<td>.85 .9 .95 .99 .999</td>
</tr>
<tr>
<td>.001 .01 .05 .1 .15 .2 .3 .4 .5 .6 .7 .8</td>
</tr>
<tr>
<td>.85 .9 .95 .99 .999</td>
</tr>
<tr>
<td>cumulative distribution</td>
</tr>
</tbody>
</table>
Uncertainty Quantification Research in Dakota: New algorithms bridge robustness/efficiency gap

<table>
<thead>
<tr>
<th></th>
<th>Production</th>
<th>New</th>
<th>Under dev.</th>
<th>Planned</th>
<th>Collabs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sampling</strong></td>
<td>Latin Hypercube, Monte Carlo</td>
<td>Importance, Incremental</td>
<td></td>
<td>Bootstrap, Jackknife</td>
<td>FSU</td>
</tr>
<tr>
<td><strong>Stochastic expansion</strong></td>
<td><strong>Adv. Deployment</strong></td>
<td>PCE and SC with uniform &amp; dimension-adaptive p-/h-refinement</td>
<td>Local adapt refinement, gradient-enhanced, compr sens</td>
<td>Discrete rv, orthogonal least interp.</td>
<td>Stanford, Purdue</td>
</tr>
<tr>
<td><strong>Other probabilistic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cornell, Maryland</td>
</tr>
<tr>
<td><strong>Epistemic</strong></td>
<td>Interval-valued/ Second-order prob. (nested sampling)</td>
<td>Opt-based interval estimation, Dempster-Shafer</td>
<td>Bayesian, discrete/model form</td>
<td>Imprecise probability</td>
<td>LANL, UT Austin</td>
</tr>
<tr>
<td><strong>Metrics &amp; Global SA</strong></td>
<td>Importance factors, Partial correlations</td>
<td>Main effects, Variance-based decomposition</td>
<td></td>
<td>Stepwise regression</td>
<td>LANL</td>
</tr>
</tbody>
</table>
Aleatory/Epistemic UQ: Nested (“Second-order”) Approaches

- Propagate over epistemic and aleatory uncertainty, e.g., UQ with bounds on the mean of a normal distribution (hyper-parameters)
- Typical in regulatory analyses (e.g., NRC, WIPP)
- Outer loop: epistemic (interval) variables, inner loop UQ over aleatory (probability) variables; *potentially costly, not conservative*
- If treating epistemic as uniform, do not analyze probabilistically!

50 outer loop samples: 50 aleatory CDF traces

50 outer loop samples: 50 aleatory CDF traces

\[ m \in [L, U] \]

\[ u \sim N(m, \sigma) \]

“Envelope” of CDF traces represents response epistemic uncertainty

bound probability or bound response
Dakota Mixed UQ with Nested Model

- Two models, each with a different set of variables
- Outer method operates on nested model
- Inner method operates on simulation model

```python
method
    id_method = 'EPISTEMIC'
    model_pointer = 'EPIST_M'
    sampling sample_type lhs
    samples = 5 seed = 12347

model,
    id_model = 'EPIST_M'
    nested
        variables_pointer = 'EPIST_V'
        sub_method_pointer = 'ALEATORY'
        responses_pointer = 'EPIST_R'
        primary_variable_mapping = 'X' 'Y'
        secondary_variable_mapping = 'mean' 'mean'
        primary_response_mapping = 1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0
                                            0.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0
                                            0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0

variables,
    id_variables = 'EPIST_V'
    interval_uncertain = 2
    num_intervals = 1 1
    interval_probabilities = 1.0 1.0
    upper_bounds = 600. 1200.
    lower_bounds = 400. 800.

responses,
    id_responses = 'EPIST_R'
    response_functions = 3
    descriptors = 'mean_mass' '95th_perc_stress' '95th_perc_disp'
    no_gradients no_hessians
```
Example Output: Intervals on Statistics

<<<< Iterator nond_sampling completed.
<<<< Function evaluation summary (ALEAT_I): 971 total (971 new, 0 duplicate)

Statistics based on 50 samples:

Min and Max values for each response function:
- mean_Wt: Min = 9.5209117200e+00, Max = 9.5209117200e+00
- ccdf_beta_s: Min = 1.8001336086e+00, Max = 4.0744019409e+00
- ccdf_beta_d: Min = 1.9403177486e+00, Max = 3.7628144053e+00

Simple Correlation Matrix between input and output:

```
<table>
<thead>
<tr>
<th></th>
<th>mean_wt</th>
<th>ccdf_beta_s</th>
<th>ccdf_beta_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>X_mean</td>
<td>9.40220e-16</td>
<td>-6.38145e-01</td>
<td>-9.14016e-01</td>
</tr>
<tr>
<td>Y_mean</td>
<td>1.38778e-15</td>
<td>-7.93481e-01</td>
<td>-4.39133e-01</td>
</tr>
</tbody>
</table>
```
Interval Estimation Approach (Probability Bounds Analysis)

- Propagate intervals through simulation code
- Outer loop: determine interval on statistics, e.g., mean, variance
  - global optimization problem: find max/min of statistic of interest, given bound constrained interval variables
  - use EGO to solve 2 optimization problems with essentially one Gaussian process surrogate
- Inner loop: Use sampling, PCE, etc., to determine the CDFs or moments with respect to the aleatory variables

\[
\begin{align*}
\min_{u_E} & \quad f_{\text{STAT}}(u_A \mid u_E) \\
u_{\text{LB}} & \leq u_E \leq u_{\text{UB}} \\
u_A & \sim F(u_A \mid u_E)
\end{align*}
\]

\[
\begin{align*}
\max_{u_E} & \quad f_{\text{STAT}}(u_A \mid u_E) \\
u_{\text{LB}} & \leq u_E \leq u_{\text{UB}} \\
u_A & \sim F(u_A \mid u_E)
\end{align*}
\]
Interval Analysis can be Tractable for Large-Scale Apps

Converge to more conservative bounds with 10—100x less evaluations
Model Form UQ in Fluid/Structure Interactions

Discrete model choices for same physics:

- A clear hierarchy of fidelity (low to high)
- An ensemble of models that are all credible (lacking a clear preference structure)
  - With data: Bayesian model selection
  - Without data: epistemic model form uncertainty propagation

Combination:

- Potential flow
- Vortex lattice

- SA-RANS → KE-RANS-NBC → KE-RANS-DBC

- SA-RANS
  - Smagorinsky-LES
  - Germano-LES
  - DNS

wind turbine applications

Horizontal Axis Wind Turbine
Vertical Axis Wind Turbine
Multifidelity UQ using Stochastic Expansions

- High-fidelity simulations (e.g., RANS, LES) can be prohibitive for use in UQ
- Low fidelity “design” codes often exist that are predictive of basic trends
- Can we leverage LF codes w/i HF UQ in a rigorous manner? → global approxs. of model discrepancy

\[ f_{hi}(\xi) = \sum_{j=1}^{N_{lo}} f_{lo}(\xi_j) L_j(\xi) + \sum_{j=1}^{N_{hi}} \Delta f(\xi_j) L_j(\xi) \]

\[ R_{hi}(\xi) = e^{-0.05\xi^2} \cos 0.5\xi - 0.5 e^{-0.02(\xi-5)^2} \]

\[ R_{lo}(\xi) = e^{-0.05\xi^2} \cos 0.5\xi, \]

\[ N_{lo} \gg N_{hi} \]

Low fidelity: CACTUS: Code for Axial and Crossflow TUrbine Simulation

High fidelity: DG formulation for LES Full Computational Fluid Dynamics/Fluid-Structure Interaction
Uncertainty Quantification
not Addressed Here

- Efficient epistemic UQ [Dakota]
- Fuzzy sets (Zadeh)
- Imprecise Probability (Walley)
- Dempster-Shafer Theory of Evidence (Klir, Oberkampf, Ferson) [Dakota]
- Possibility theory (Joslyn)
- Probability bounds analysis (p-boxes)
- Info-gap analysis (Ben-Haim)

- Bayesian model calibration / inference via MCMC [Dakota]
- Other Bayesian approaches: Bayesian belief networks, Bayesian updating, Robust Bayes, etc.
- Scenario evaluation

(Some available in [Dakota])
# Dakota UQ: Summary, Relevant Methods

- **What?** Understand code output uncertainty / variability
- **Why?** Risk-informed decisions with variability, possible outcomes
- **How?** What Dakota methods are relevant?

<table>
<thead>
<tr>
<th>character</th>
<th>method class</th>
<th>problem character</th>
<th>variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>aleatory</td>
<td>probabilistic sampling</td>
<td>nonsmooth, multimodal, modest cost, # variables</td>
<td>Monte Carlo, LHS, importance</td>
</tr>
<tr>
<td>local reliability</td>
<td>smooth, unimodal, more variables, failure modes</td>
<td>mean value and MPP, FORM/SORM,</td>
<td></td>
</tr>
<tr>
<td>global reliability</td>
<td>nonsmooth, multimodal, low dimensional</td>
<td>EGRA</td>
<td></td>
</tr>
<tr>
<td>stochastic expansions</td>
<td>nonsmooth, multimodal, low dimension</td>
<td>polynomial chaos, stochastic collocation</td>
<td></td>
</tr>
<tr>
<td>epistemic</td>
<td>interval estimation</td>
<td>simple intervals</td>
<td>global/local optim, sampling</td>
</tr>
<tr>
<td>evidence theory</td>
<td>belief structures</td>
<td></td>
<td>global/local evidence</td>
</tr>
<tr>
<td>both</td>
<td>nested UQ</td>
<td>mixed aleatory / epistemic</td>
<td>nested</td>
</tr>
</tbody>
</table>

- See Dakota Usage Guidelines in User’s Manual
- Analyze tabular output with third-party statistics packages
UQ References


- Dakota User’s Manual: Uncertainty Quantification Capabilities
- Dakota Theory Manual
- Corresponding Reference Manual sections