An Overview of an Approximation approach to Validation

Roger Ghanem
University of Southern California
Los Angeles, CA

Sandia CSRI Workshop on Mathematical Methods for Validation, New Mexico, August 14-16, 2007.

Outline

- Introduction and Objectives
- Representation of Information
  - Stochastic Representations
  - Representing the missing information
- Propagation of Information:
  - Stochastic Galerkin Projections
- Validation
- Conclusions

Introduction

- Objectives:
  - Determine certifiable confidence in model-based predictions:
    - Certifiable = amenable to analysis
    - Accept the possibility that certain statements, given available resources, cannot be certified.
  - Compute actions to increase confidence in model predictions: change the information available to the prediction.
    - More experimental/field data, more detailed physics, more resolution for numerics…

Two meaningful questions

Given a (physics model, data, and computational resources):

- **compute limits on predictability**

  i.e. which statements about system performance can be certified ?

- **compute resource allocation**

  (data/computing/modeling) to achieve validation

Critical steps:

- **represent** all available evidence as faithfully as possible.
- **propagate** the weight of the evidence to obtain the corresponding weight of evidence in prediction.

Probabilistic models package information in a manner suitable for analysis:

- Adapt this packaging to the needs of our decision-maker
  - **Craft** a mathematical model that is parameterized with respect to the relevant uncertainties.

Error Budget

$$\mathbf{U} = \hat{\mathbf{U}}|_{h,d,p,m} + \underbrace{\epsilon_h|_{p,d,m} + \epsilon_p|_{d,m} + \epsilon_d|_m}$$

limit on predictability, given a model

MUST BE QUANTIFIED !!!!

$\epsilon_h|_{d,m}$: can be reduced through better numerics.

$\epsilon_p|_{d,m}$: can be reduced through better statistics.

$\epsilon_d|_m$: can be reduced through better data.

Cameron-Martin Theorem

Let $x(t)$ be Brownian motion

$\{\alpha_i(t)\}$ is a CONS in $L_2[0,1]$

$\int_0^1 \alpha_p(t)dx(t)$: coordinate of Brownian noise relative to CONS

Polynomials w.r.t. to each of these coordinates:

$$\Phi_{m,p}(x) = H_m\left[\int_0^1 \alpha_p(t)dx(t)\right] \quad m = 0,1,\cdots,; \quad p = 1,2,\cdots$$

Extend over all coordinates:

$$\Psi_{m_1,\cdots,m_p}(x) = \Phi_{m_1,1}(x)\cdots\Phi_{m_p,p}(x)$$

Then

$$\int_C^w \left| F[x] - \sum_{m_1,\cdots,m_N} A_{m_1,\cdots,m_N}\Psi_{m_1,\cdots,m_N}(x)\right|^2 d_w x \to 0 \quad \text{as} \quad N \to \infty$$

Cameron-Martin

IN SIMPLE WORDS:

The polynomial chaos decomposition of any square-integrable functional of the Brownian motion converges in mean-square as N goes to infinity.

For a finite-dimensional representation, the coefficients are functions of the missing dimensions: they are random variables.

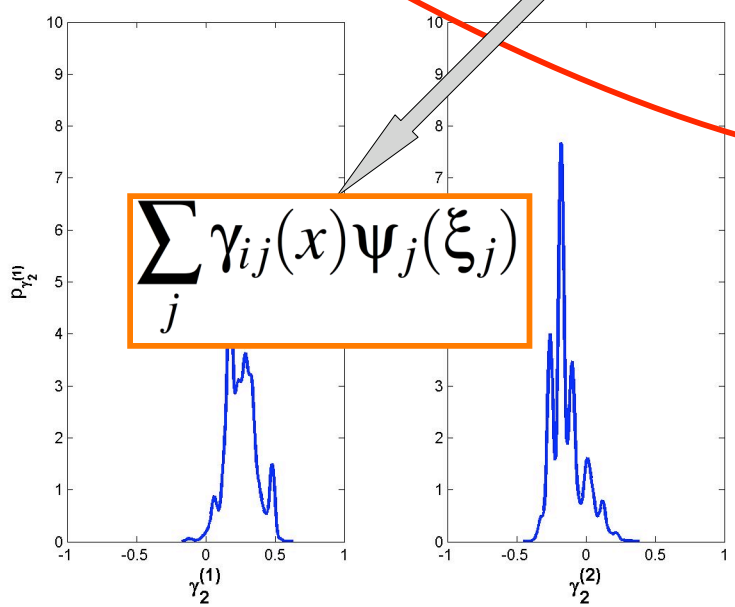
Representation of Uncertainty

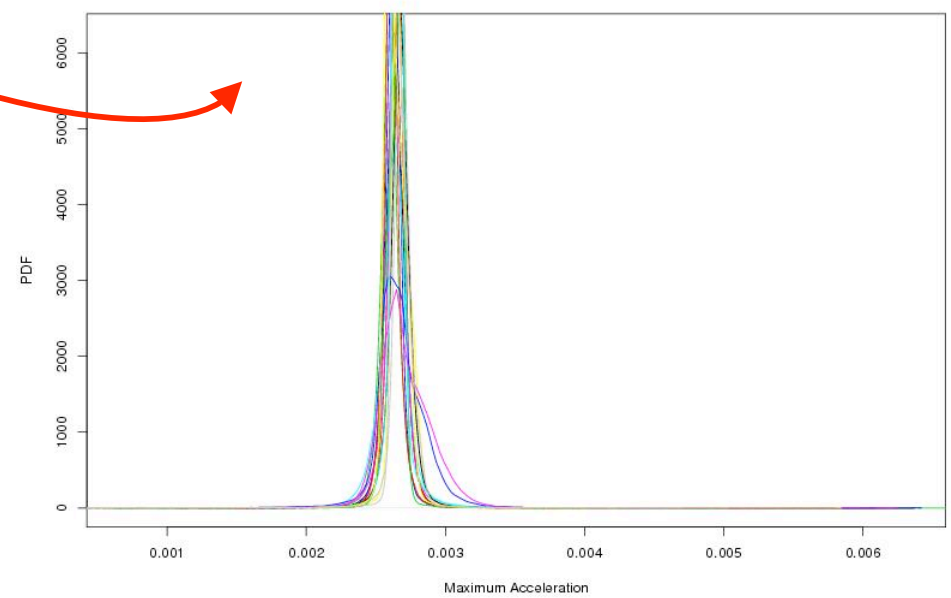
The random quantities are resolved as surfaces in a normalized space:

$$\alpha(\mathrm{x},\theta) = \sum_{i=1}^{\infty} \alpha_i(\mathrm{x}) \underbrace{\Psi_i(\{\underbrace{\boldsymbol{\xi}(\theta)}\})}$$

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= \xi_1 \\
\psi_2 &= \xi_2 \\
\psi_3 &= \xi_1^2 - 1 \\
\psi_4 &= \xi_1\xi_2 \\
\psi_5 &= \xi_2^2 - 1 \\
\ldots
\end{aligned}$$

$$\boxed{\sum_j \gamma_{ij}(x)\psi_j(\xi_j)}$$





Representation of Uncertainty

$$\alpha(x,\theta) = \sum_i \hat{\alpha}_i(x)\psi_i(\underbrace{\xi_1,\cdots,\xi_n}_{\text{Uncertainty in model parameters}},\underbrace{\xi_{n+1},\cdots,\xi_m}_{\text{Uncertainty due to small experimental database}})$$

Dimension of vector ξ reflects complexity of α

Some Common Bases

Infinite-dimensional case:

This is an exercise in stochastic analysis:

- Hermite polynomials: Gaussian measure (Wiener: Homogeneous Chaos)
- Charlier polynomials: Poisson measure (Wiener: Discrete Chaos)
- Very few extensions possible: Friedrichs and Shapiro (Integration of functionals) provide characterization of compatible measures. Segall and Kailath provide an extension to martingales.

Finite-dimensional case: independent variables:

This is an exercise in one-dimensional approximation:

- Askey polynomials:measures from Askey chart (Karniadakis and co-workers)
- Wavelets: Olivier Le-Maitre
- Multiorthogonal polynomials and local bases: Babuska, Tempone et.al

Finite-dimensional case: dependent variables:

This is an exercise in multi-dimensional approximation:

- Arbitrary measures: Soize and Ghanem.

Characterization of Uncertainty

$$\alpha(\mathbf{x}, \theta) = \sum_{i=1}^{\infty} \alpha_i(\mathbf{x}) \Psi_i(\{\boldsymbol{\xi}(\theta)\})$$

- Galerkin Projections
- Maximum Likelihood
- Maximum Entropy
- Bayes Theorem
- Ensemble Kalman Filter

Uncertainty Propagation: Stochastic Projection

Discretized Governing Equation:

$$\mathbf{K}(\mathbf{u}, \kappa(\theta))\mathbf{u}(\theta) \;=\; \mathbf{f}(\theta) \;,$$

Expand Basic Uncertain Property:

$$\kappa(\theta) = \sum_{j=0}^{N} \kappa_j \psi_j(\xi(\theta))$$

Expand Solution:

$$\mathbf{u} = \sum_{j=0}^{N} \mathbf{u}_j \Psi_j(\theta)$$

$\langle u \rangle = u_0$
$R_{uu} = \sum_j u_j u_j^T \langle \psi_j^2 \rangle$
realizations of u can be readily obtained from realizations of ξ

A-Posteriori Error Analysis

A posteriori error analysis of SGS:

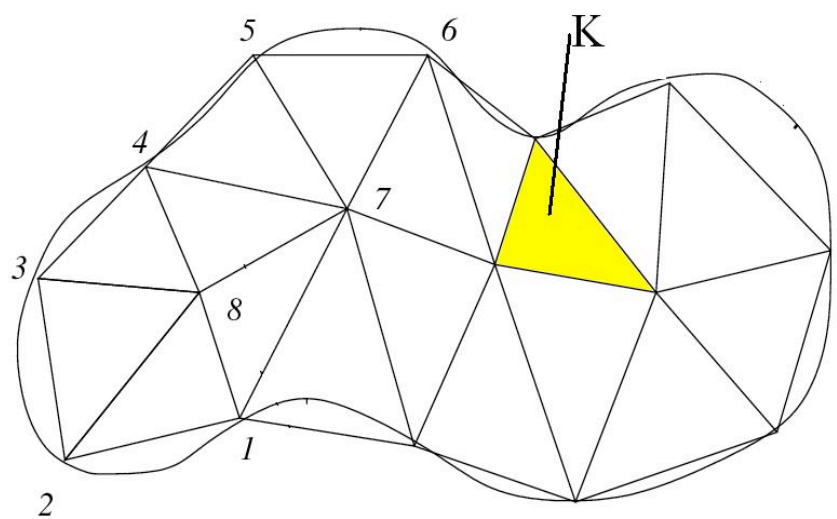
$$\|e\|^2_{H^1_0(\mathcal{D})\otimes L_2(\Omega)} \leq C\left\{\underbrace{\sum_K h_K^2 \|r\|^2_{L_2(K)\otimes L_2(\Omega)}}_{\text{Spatial discretization error}} + \underbrace{p^{-s}\|r\|^2_{L_2(\mathcal{D})\otimes L_2(\Omega)}}_{\text{Random discretization error}}\right\}$$

Where:

$$e = u - u_{h,p}$$

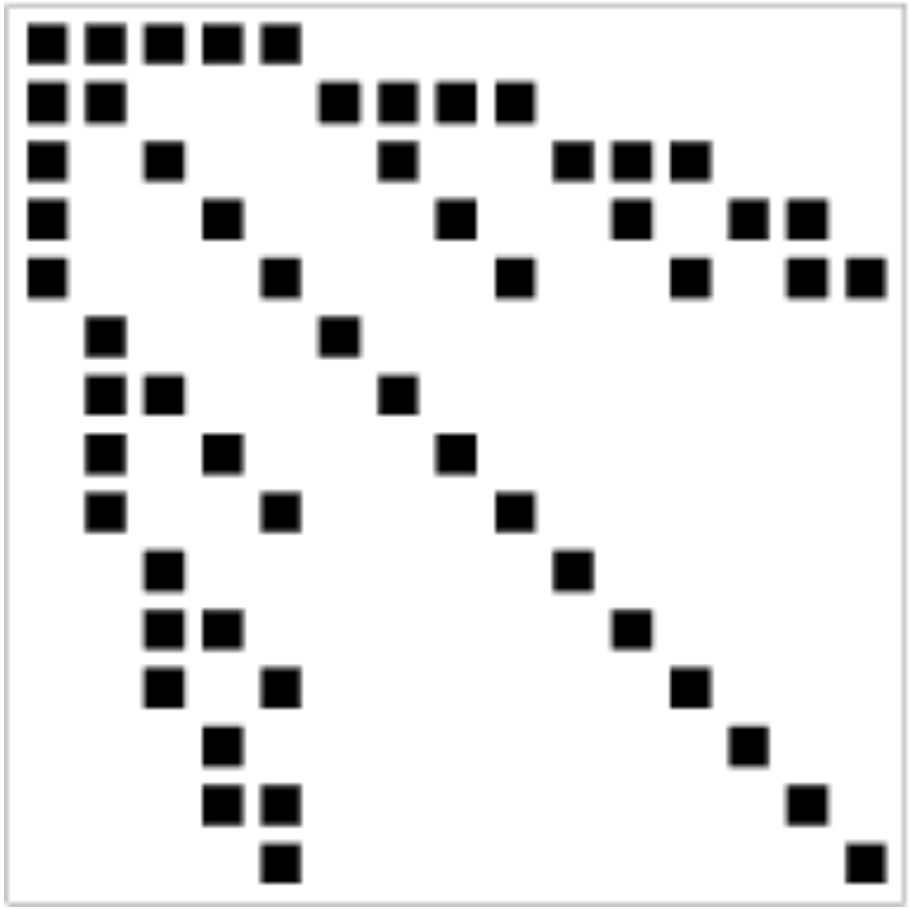
$$r = f + \nabla.\left(a\nabla u_{h,p}\right) \quad \text{in} \quad K$$

C : independent from h_K and

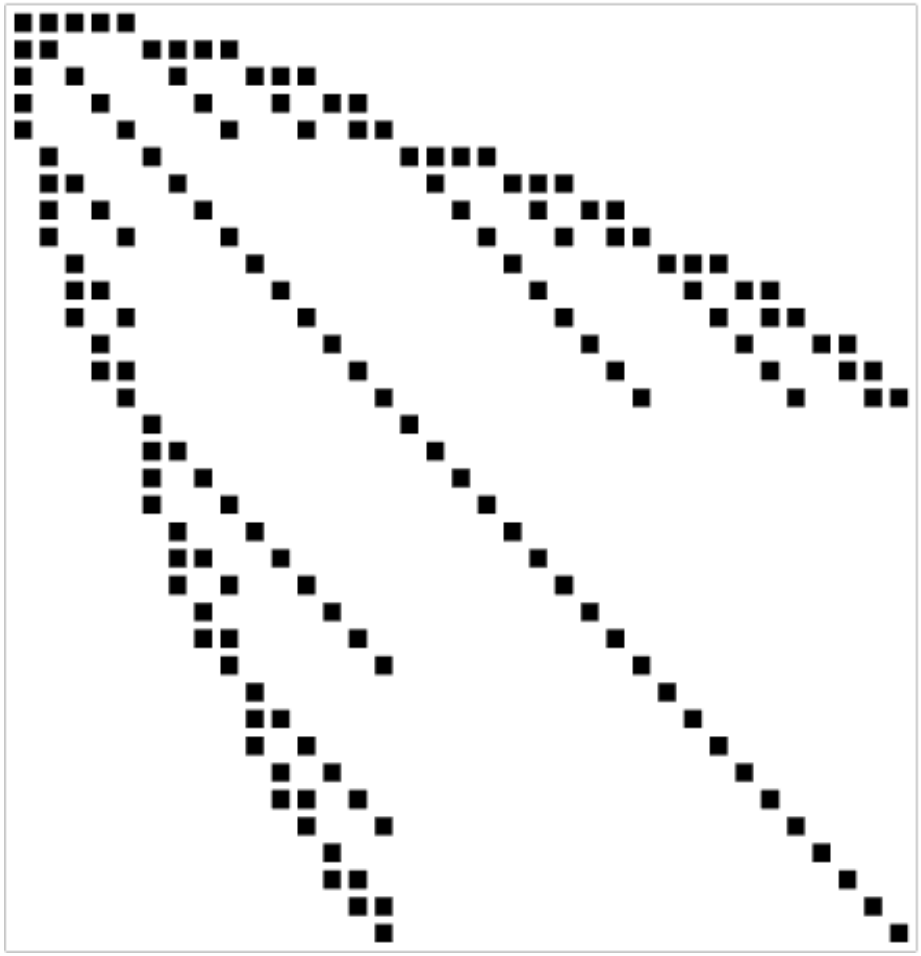


SFEM:Typical system matrix

SFEM:Typical system matrix



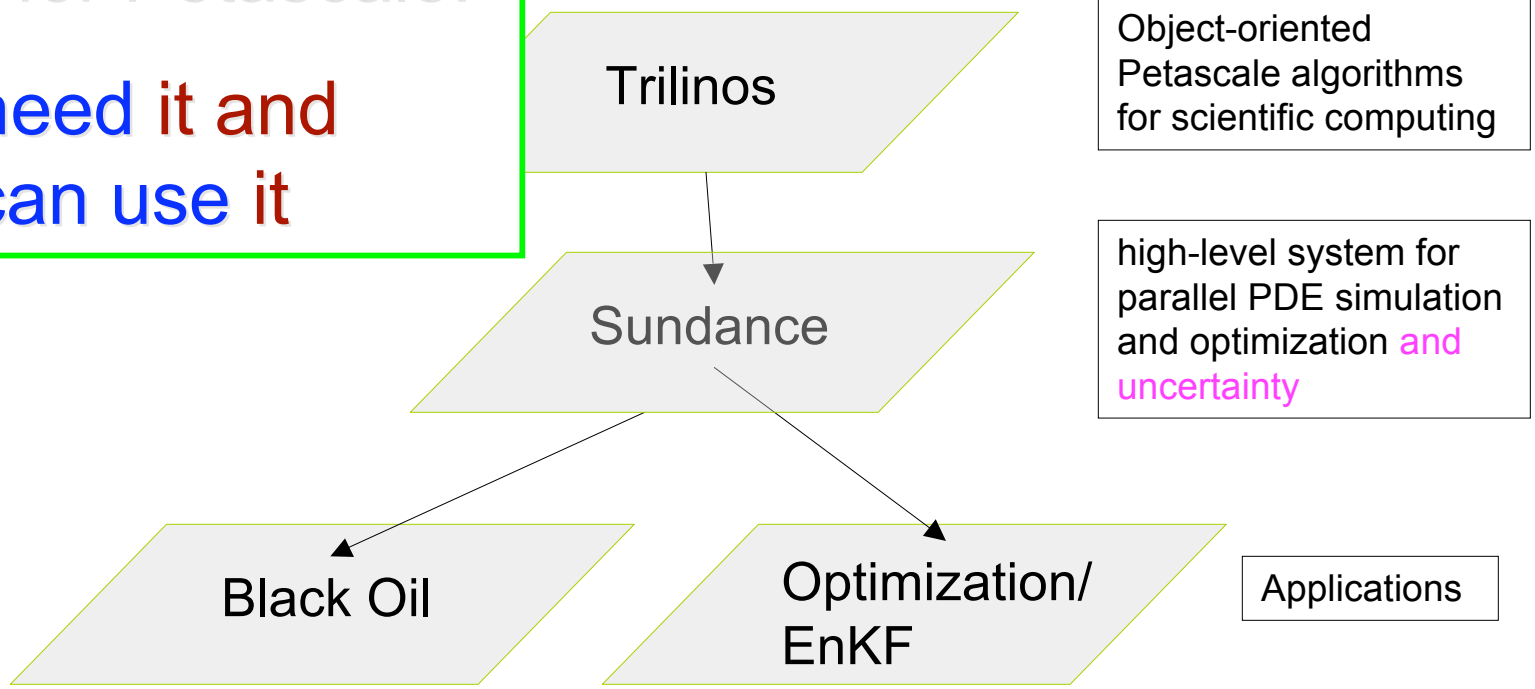
SFEM:Typical system matrix



Example Implementation

Ready for Petascale:

need it and
can use it



Implementation

```
Sundance::init(&argc, &argv);

/* We will do our linear algebra using Epetra */
VectorType<double> vecType = new EpetraVectorType();

/* Create a mesh. It will be of type BasisSimplicialMesh, and will
 * be built using a PartitionedLinedMesher. */
int np = 120;
MeshType meshType = new BasicSimplicialMeshType();
MeshSource mesher = new PartitionedLineMesher(0.0, 1000, np, meshType);
Mesh mesh = mesher.getMesh();

/* Create a cell filter that will identify the maximal cells
 * in the interior of the domain */
CellFilter interior = new MaximalCellFilter();
CellFilter points = new DimensionalCellFilter(0);
CellFilter rightPoint = points.subset(new RightPointTest());
CellFilter leftPoint = points.subset(new LeftPointTest());

/* Create the Spectral Basis */
int ndim = 3;
int order = 4;

SpectralBasis sbasis = new HermiteSpectralBasis(ndim, order);

/* Create unknown and test functions, discretized using first-order
 * Lagrange interpolants */
Expr Se = new UnknownFunction(new Lagrange(2), sbasis, "Se");
Expr Ns = new TestFunction(new Lagrange(2), sbasis, "Ns");
Expr Pw = new UnknownFunction(new Lagrange(2),sbasis, "Pw");
Expr Nn = new TestFunction(new Lagrange(2), sbasis, "Nn");

/* Create differential operator and coordinate functions */
Expr x = new CoordExpr(0);
Expr dx = new Derivative(0);


/* We need a quadrature rule for doing the integrations */
QuadratureFamily quad = new GaussianQuadrature(4);
```

Define mesh

Define boundaries

Define stochastic bases

Define SFEM interpolation

Non-intrusive Implementation

The solution $\mathbf{u}$ is a function of the basic random variables $\mathbf{x}_i$:

$$\mathbf{u} = f(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

where $f(....)$ is a mapping available through an analysis code.

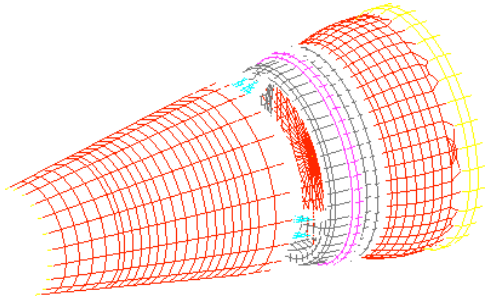
A Polynomial Chaos Decomposition of the solution has the form:

$$\boxed{\mathbf{u} = \sum_i \psi_i \mathbf{u}_i \qquad \mathbf{u}_i = \frac{\langle \mathbf{u}\psi_i \rangle}{||\psi_i||^2}}$$

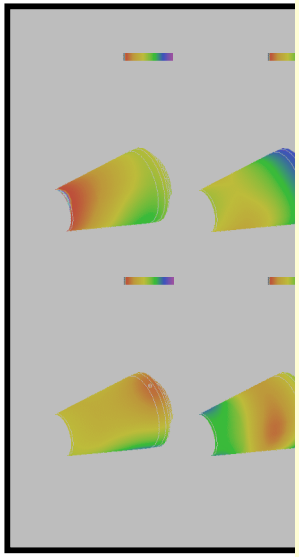
where the mathematical expectation is evaluated through a statistical average over a finite number of samples.

Implementation: Dakota

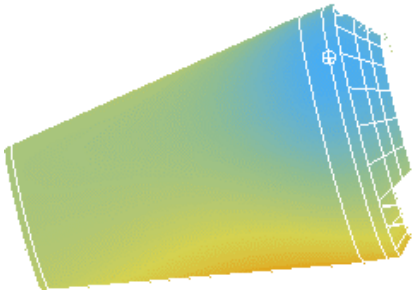
```
( {nond_polynomial_chaos}                                          \
        {expansion_terms = <INTEGER>} |                            \
        {expansion_order = <INTEGER>}                              \
        {expansion_samples  = <INTEGER>} |                         \
        {quadrature_order   = <INTEGER>} |                         \
        {collocation_points = <INTEGER>} |                         \
        {expansion_import_file = <STRING>}                         \
        [samples = <INTEGER>] [seed = <INTEGER>] [fixed_seed]      \
        [ {sample_type} {random} | {lhs} ]                         \
        [ {distribution} {cumulative} | {complementary} ]          \
        [ {response_levels = <LISTof><REAL>}                       \
          [num_response_levels = <LISTof><INTEGER>]                \
          [ {compute} {probabilities} | {reliabilities} ] ]         \
        [ {probability_levels = <LISTof><REAL>}                    \
          [num_probability_levels = <LISTof><INTEGER>] ]           \
        [ {reliability_levels = <LISTof><REAL>}                    \
          [num_reliability_levels = <LISTof><INTEGER>] ] )         \
|                                                                  \
```



Implementation: Salinas



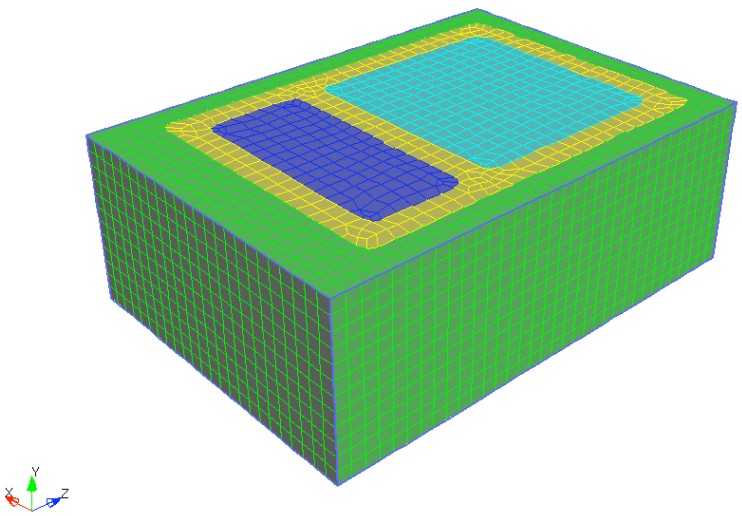
```
Material 1
            name 'Aluminum'
            E 10.0E6
            nu 0.33
            density 253.82e-6
END

Material 3
            s_isotropic
            name 'mat3.dat'
            density 253.82e-6
END

Material 4
            s_isotropic
            name 'mat4.dat'
            density 253.82e-6
END
```



Example Application: W76 Foam study

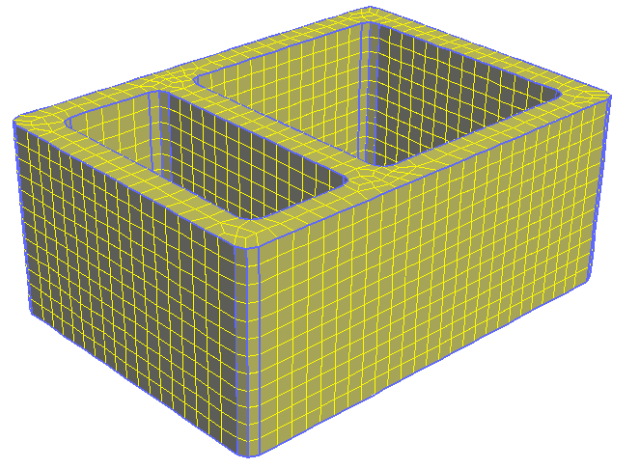
System has 10320 HEX elements.

Stochastic block has 2832 elements.



Built-up structure with shell, foam and devices.

Foam domain.



1. Modeled as non-stationary random field.
2. Accounting for random and structured variations
3. Limited observations are assumed:

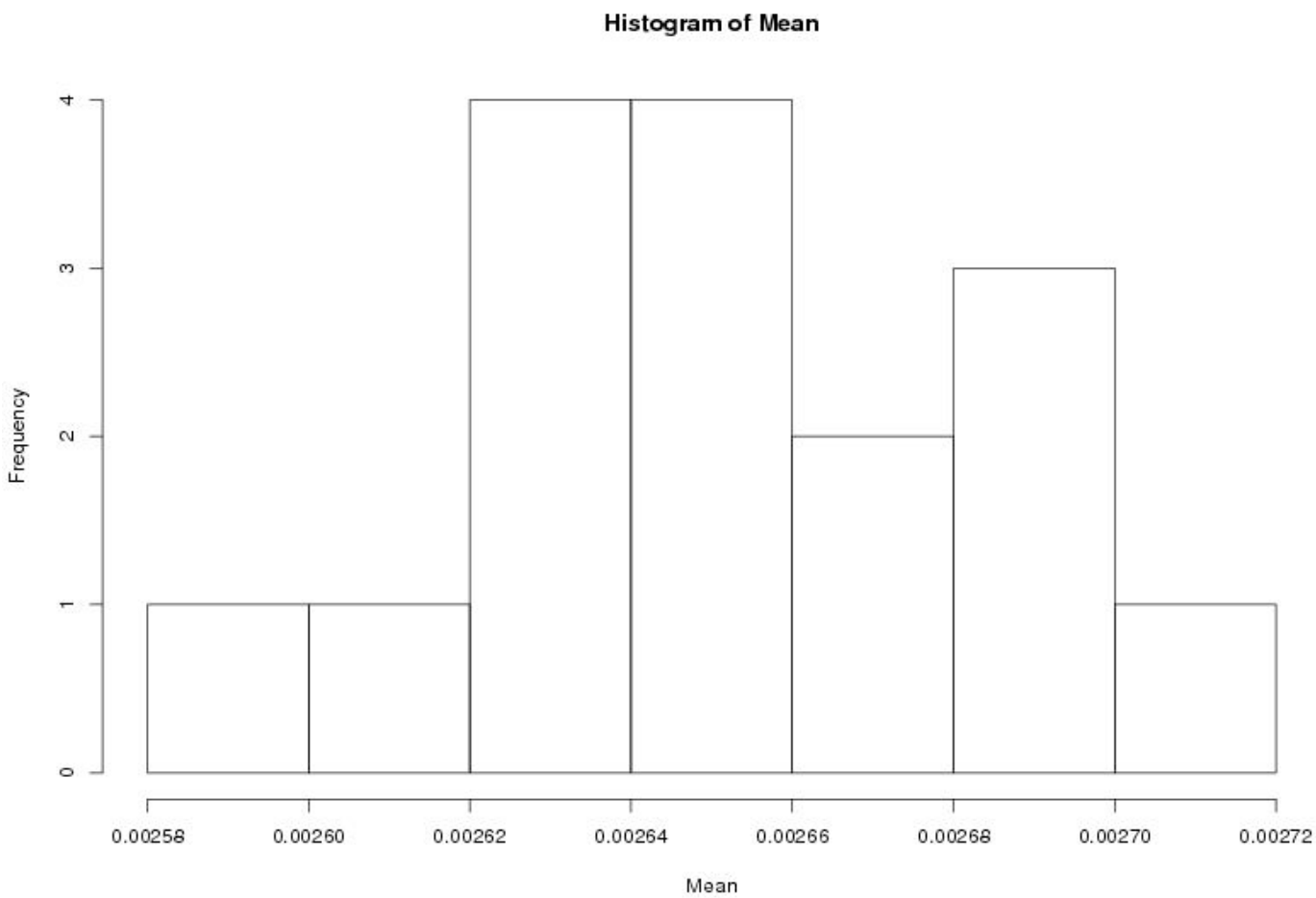
   selected 30 locations on the foam.

   Limited statistical observations: Correlation estimator from small sample size: interval bounds on correlation matrix.

Example Application: W76 Foam study

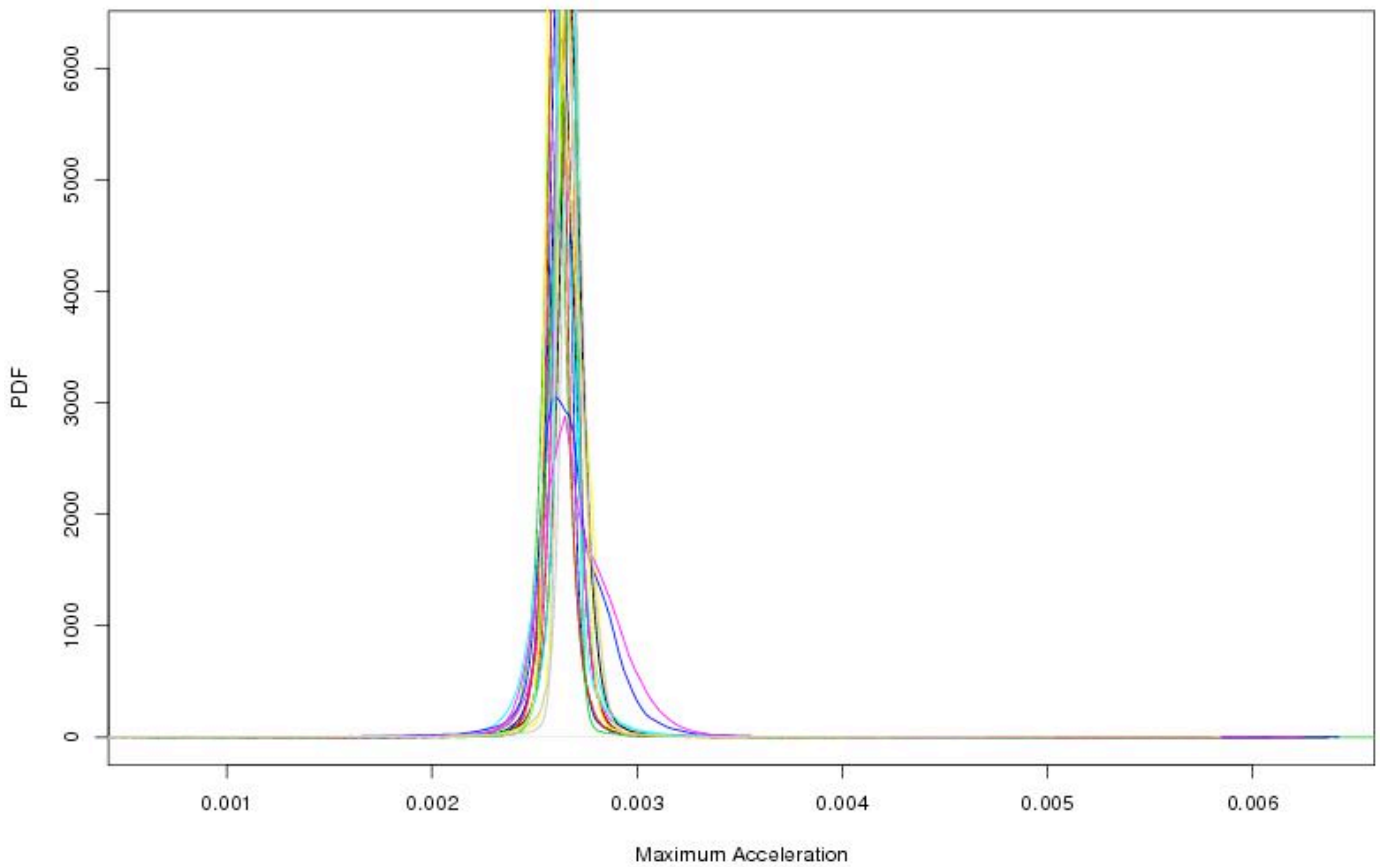
- Polynomial Chaos representation of epistemic information
  - Constrained polynomial chaos construction
- Radial Basis function consistent spatial interpolation
- Cubature integration in high-dimensions

Foam study
Statistics of maximum acceleration



Histogram of average of maximum acceleration

Foam study
Statistics of maximum acceleration



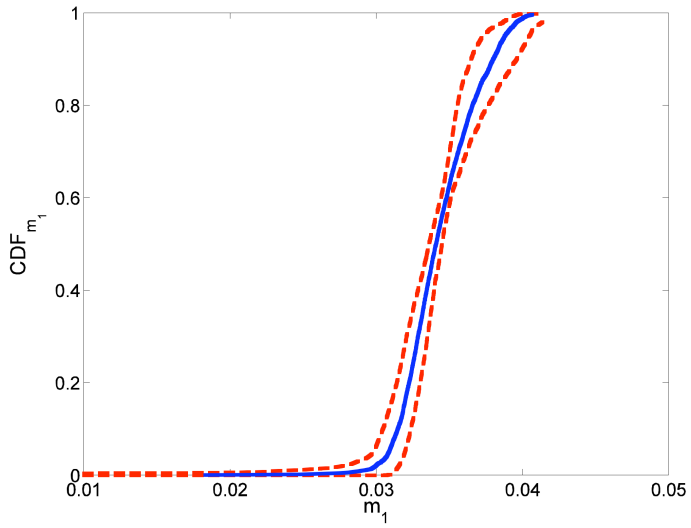
Plots of density functions of the maximum acceleration

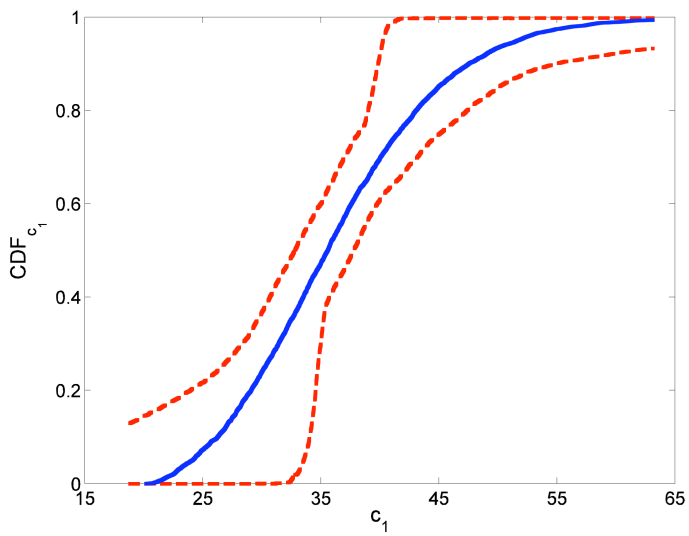
Validation Example

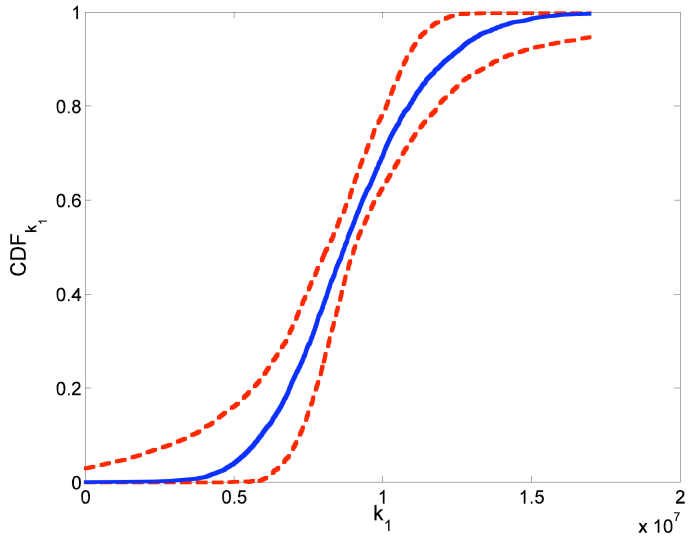
The Sandia validation challenge problem:

- Given a few samples from a related experiment
- Given a model (map between input/output)
- Decide whether the model is valid to predict a certain quantity of interest
- If the model is valid, use it to decide on certification of a design

CDF of calibrated stochastic parameters (3 out of 9 shown)







Estimate

%95 probability box

Remarks:

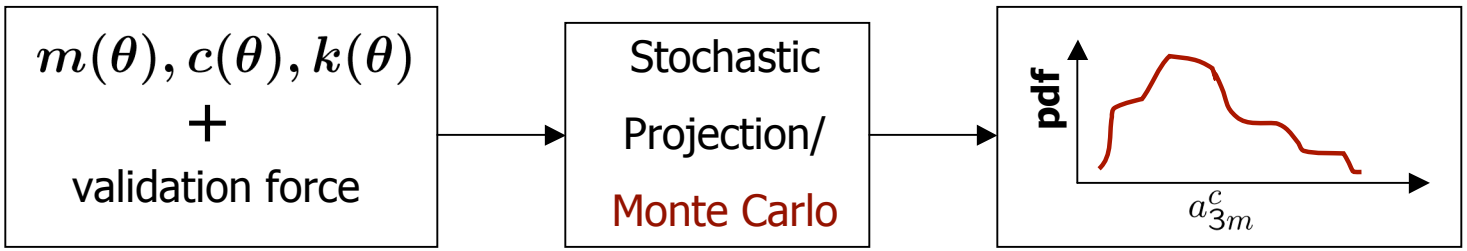
- Confidence intervals are due to finite sample size.

Validation path: hypothesis test

System Response Quantity (SRQ):

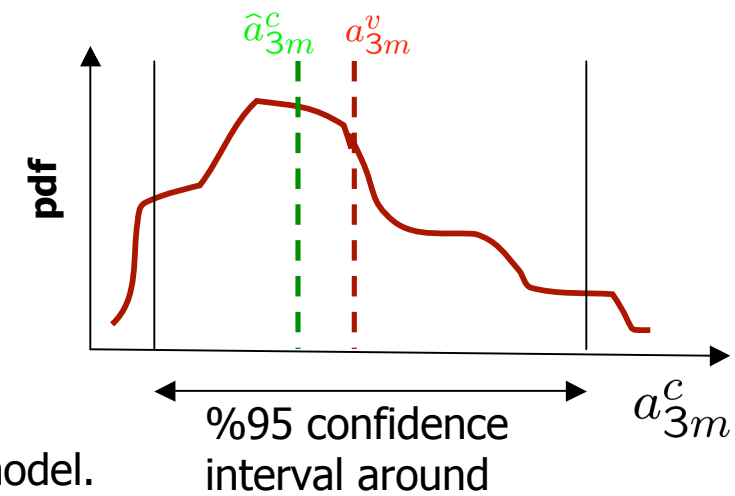
Maximum acceleration of the top mass = $\mathbf{a_{3m}}$

Propagation using calibrated stochastic linear model:



Equivalent hypothesis test:

$$H_0 : \hat{a}^c_{3m} = a^v_{3m}$$



Remark:

Parameters are calibrated under H_0 .

$\hat{a}^c_{3m}$ = mean of predicted a^c_{3m} from linear model.

Subsystem validation outcome

Calibration Based On Excitation Level	Validation Excitation Level	Hypothesis
Low	Low	**Accepted**
	Medium	**Accepted**
	High	**Accepted**
Medium	Low	**Accepted**
	Medium	**Accepted**
	High	**Accepted**
High	Low	**Accepted**
	Medium	**Accepted**
	High	**Accepted**

Prediction on target application

$$P_{am} := \text{Prob}\left\{\max_{t>0} |a(t)| > 1.8(10^4) in/sec^2\right\} < 10^{-2}$$

Calibration Based On Excitation Level	Sample Mean of P_{am}	Sample Variance of P_{am}
Low	0.0835	0.000830
Medium	0.0662	0.001500
High	0.1269	0.004300

Remark: Based on only 25 samples.

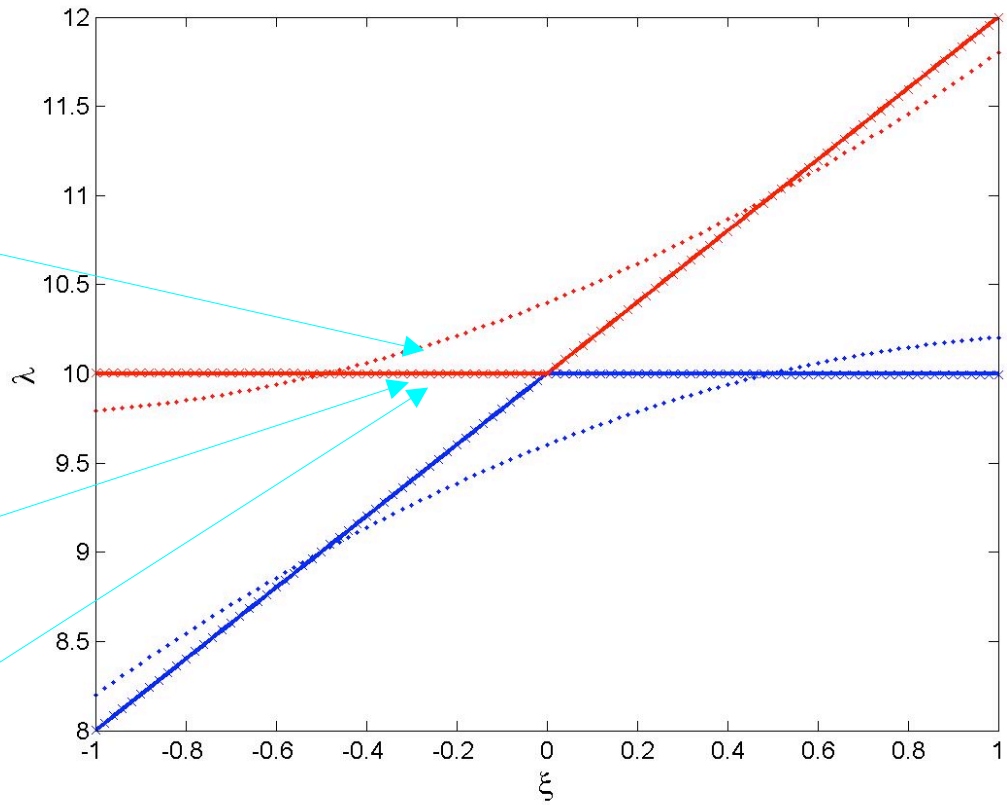
Efficiency Issues: Basis Enrichment

$$\lambda = \sum_{i=0}^{P} \lambda_i \psi_i + \lambda_{P+1} |\xi|$$

Solution with 3rd order chaos

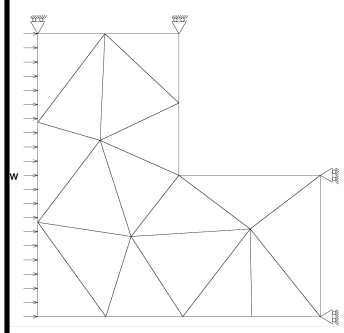
Solution with 3rd order chaos and enrichment

Exact solution



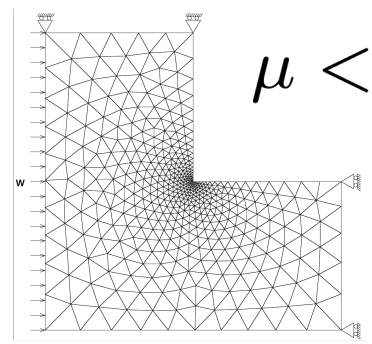
Efficiency Issues: Reduced Order Models

A Two-Scale Analysis



$+ \{\psi_i\}_{i=1}^{P}$

$$\eta_i(\omega) = \eta_i(\{\psi_i\}_{i=1}^{P}) \quad \forall i$$

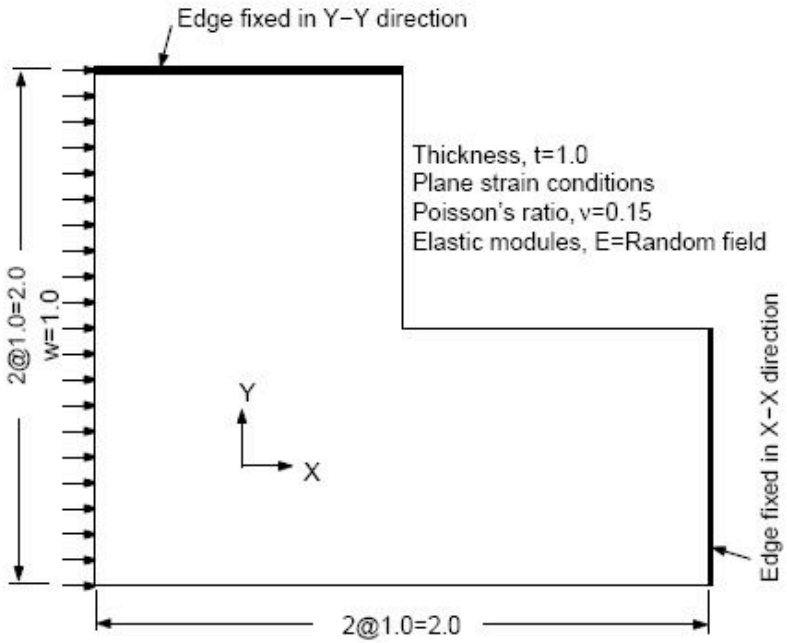


$\mu << P$

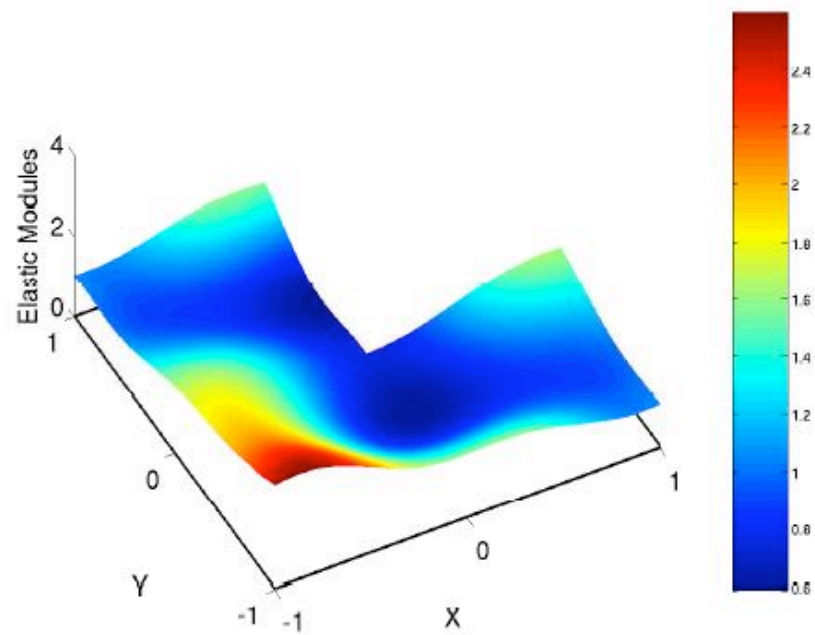
$+ \{\eta_i\}_{i=1}^{\mu}$

Coarse Analysis | **Algebraic operations** | **Fine Analysis**

Model Reduction: Example



(a)

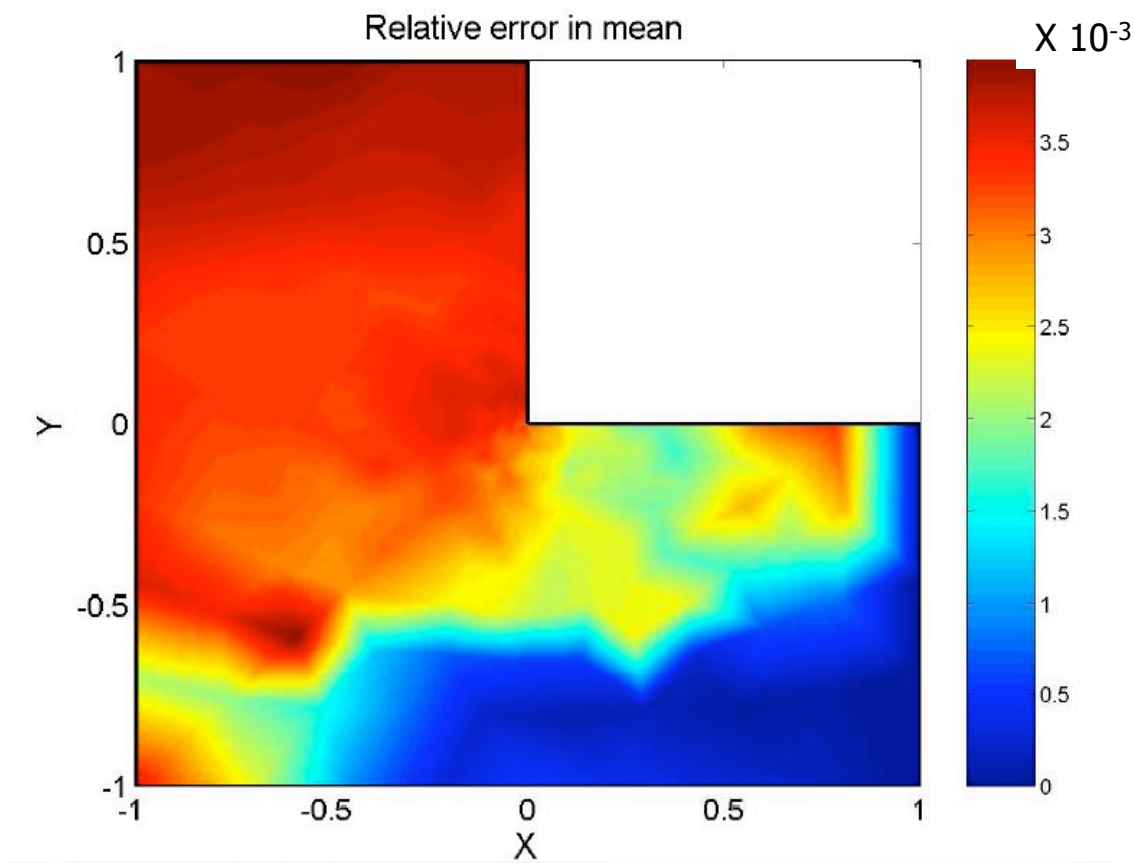


(b)

2D solid mechanics problem. (a) Geometry of the problem. (b) A sample realization of the modules of elasticity E.

Model Reduction: Example

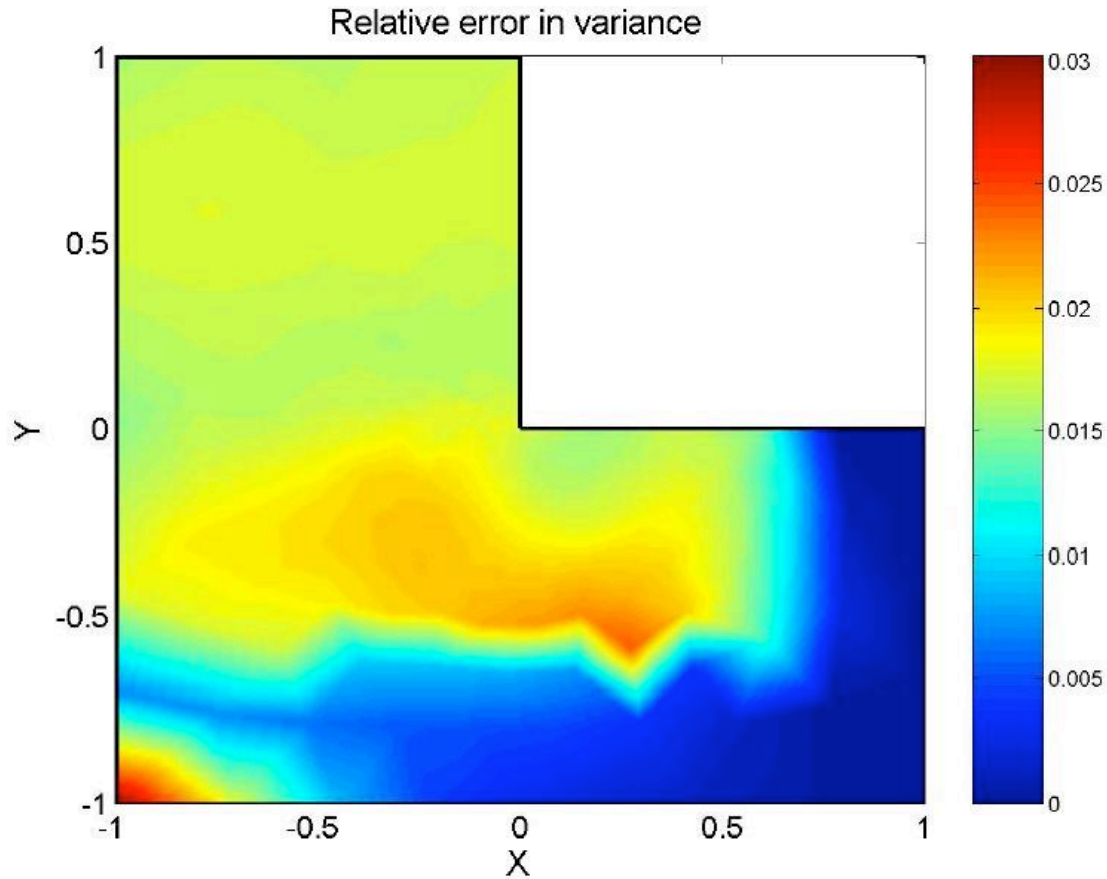
Relative Error in Response Mean



$p = 3; q = 8$ $\quad COV\,(u) = \%60 - \%65$

Model Reduction: Example

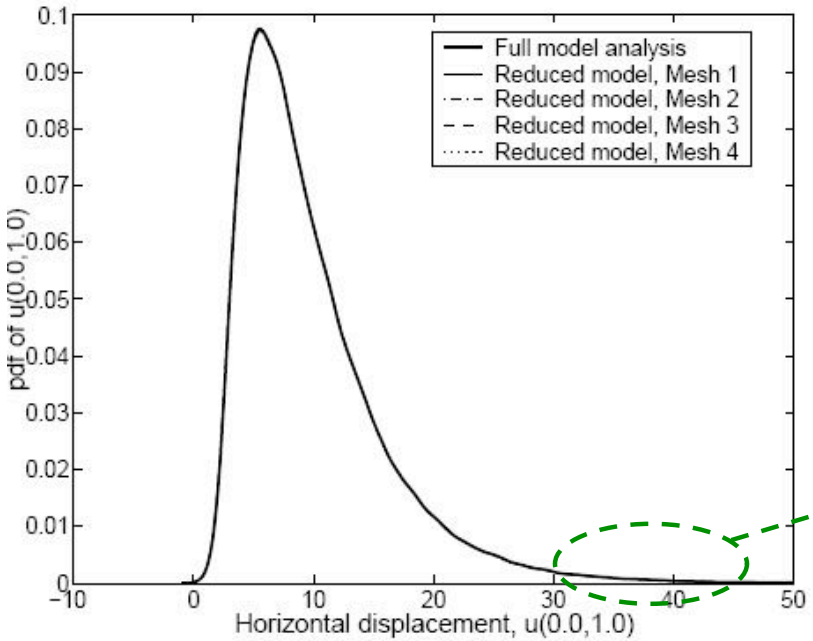
Relative Error in Response Variance

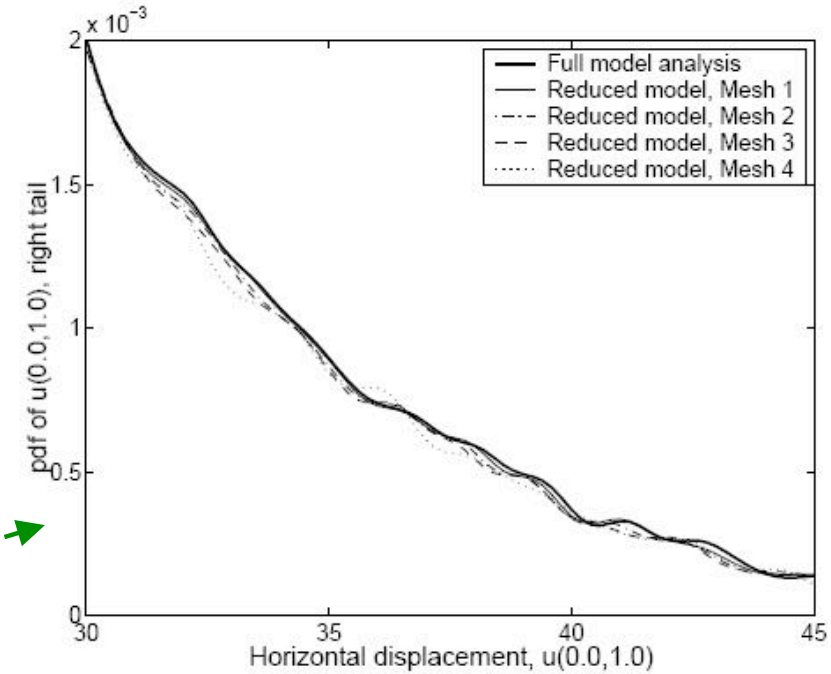


$p = 3; q = 8$ $\quad COV\,(u) = \%60 - \%65$

Model Reduction: Example

pdf of the Horizontal Displacement At (x,y)=(0,1)





$$\mu + 1 = 5 << P + 1 = 165$$

$$COV\,(u) = \%60 - \%65$$

Some Remaining Challenges

- Theory:
  - Chaos expansions under “epistemic” constraints.
  - More on stochastic (and data) error estimation.

- Efficiency:
  - Develop adapted linear and nonlinear solvers that take advantage of the algebraic structure of resulting equations.
  - Petascaliziation.
  - Adaptation of order of PCE over space.

- Implementation:
  - Integrate with black-box computational physics engines.
  - Adaptation to full library of equations of state and constitutive equations.
  - Matrix-free implementations.
  - Couple with sampling-based techniques.
  - Challenges associated with large-scale complex computational systems.

Conclusion

$$P(E) \geq 0$$
$$P(\Omega) = 1$$
$$P(A \cup B) = P(A) + P(B)$$

- Very flexible framework for formulating rational statements and developing suitable models.

- Provides a complete extension for deterministic science that permits the interpretation of experimental evidence and the synthesis of information.