Model Calibration under Uncertainty: Matching Distribution Information

Laura P. Swiler, Brian M. Adams, and Michael S. Eldred
September 11, 2008

AIAA Multidisciplinary Analysis and Optimization Conference
Victoria, Canada

*Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

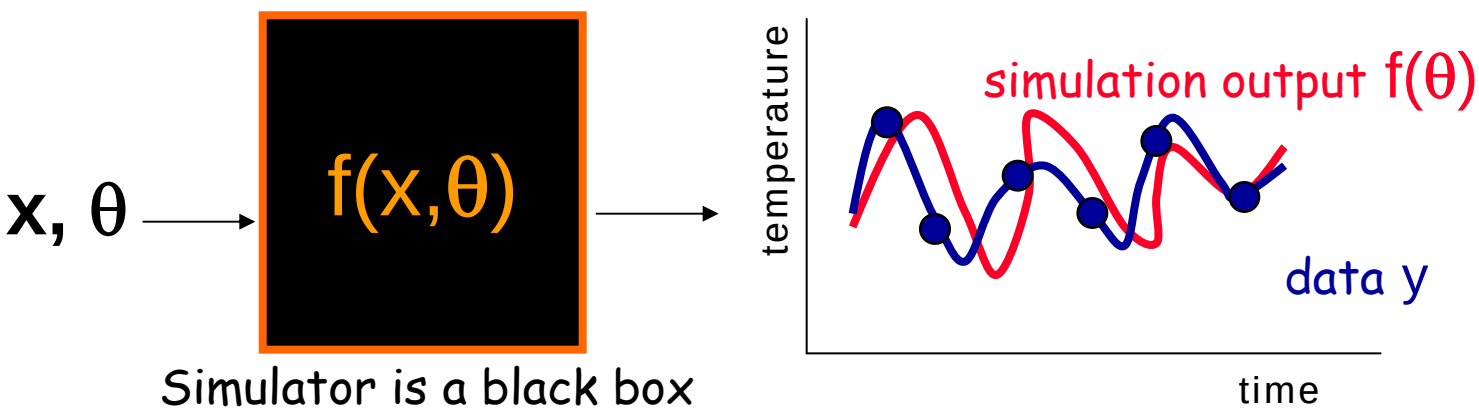
What is traditional calibration?

Calibration: Adjust model parameters (θ) to maximize agreement with a set of experimental data (AKA parameter estimation, parameter identification, nonlinear least-squares)

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} [(y_i(\boldsymbol{x}_i) - f(\boldsymbol{x}_i;\boldsymbol{\theta})]^2 = \sum_{i=1}^{n} [R_i(\boldsymbol{\theta})]^2$$

Experimental data

Simulation output that depends on x



Simulator is a black box

Calibration under Uncertainty

Calibration: Determine *optimal* parameter settings for parameters of computer models

Problem: How to determine "optimal"?

- Standard approach: find the parameters that minimize the sum squared difference between the model-computed data and the actual experimental data.
- This approach usually allows for some error in the experimental data through a Gaussian error term, e.g. $N(0,\sigma^2)$, but the error is assumed independently and identically distributed at each data point
- Instead: Best distribution fit. Match means, means and variances, or multiple percentile values from the distributions of model results and experimental data.

Calibration under Uncertainty:
Best Pointwise Fit → Best Distribution Fit

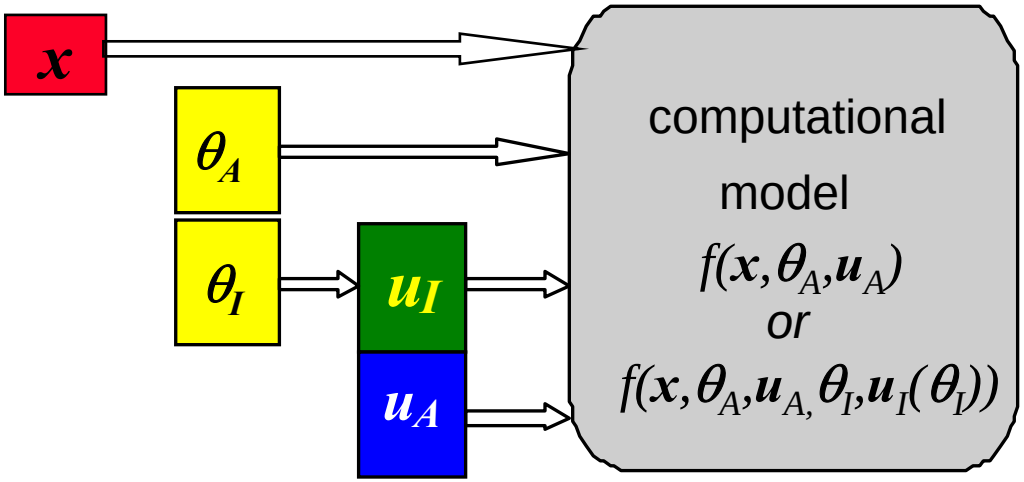
Sandia National Laboratories

Why do this?

- Often our experimental data is of poor quality, has very few data points, or may not be directly relevant to the particular simulation model under investigation → all we want to do is match moments
- Frequently modelers want to get a ballpark estimate based on bounds matching: I have had people say to me "if I can get my model upper and lower bound estimates to match an upper/lower bound on the data, I will be happy"
- Might have several sets of experiments and want to match the mean or moments of each
- Our model may have some uncertain variables (some of which may be "calibrated" (e.g. pick the mean) and others not) and we want to incorporate this uncertainty in the response
- The output of the simulation has variability due to uncertainty in some of the input parameters to the model

Sandia National Laboratories

Form of computational model



$\boldsymbol{x}$ = model inputs (not calibrated: scenario or case specific inputs)

$y(\boldsymbol{x})$ = experimental outputs: $\boldsymbol{x}$ and y are observable but we assume $\mathbf{u}$ is not

$\boldsymbol{\theta}_\mathbf{A}$ = parameters to be calibrated entering directly into the model

$\boldsymbol{\theta}_\mathbf{I}$ = parameters to be calibrated which influence the characterization of $\mathbf{U}$

$\mathbf{u}_\mathbf{I}$ = stochastic inputs which depend on a calibrated parameter

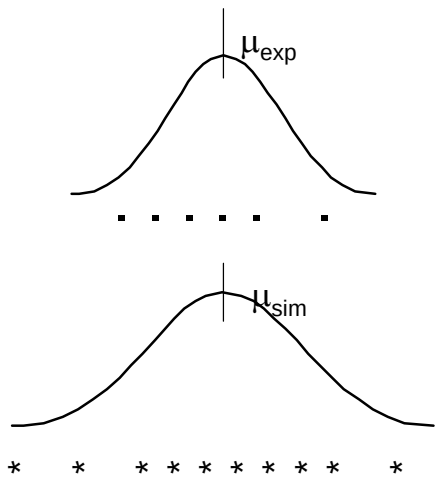
$\mathbf{u}_\mathbf{A}$ = stochastic inputs which do not depend on a calibrated parameter

$f(\boldsymbol{x}_i, \theta, \mathbf{u}_\mathbf{j})$ = a model result based on inputs $\boldsymbol{x}_i$, θ, and

a realization $\mathbf{u}_\mathbf{j}$ of the uncertain variables

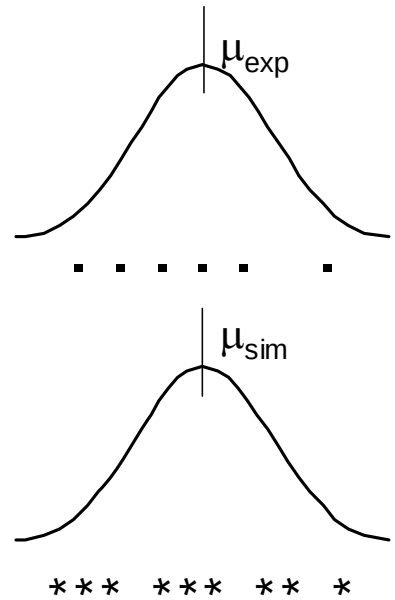
Examples

Matching means



$$S(\boldsymbol{\theta}) = (\mu_{\exp} - \mu_{sim})^2 = \left(\frac{\sum_{i=1}^{n} y_i}{n} - \frac{\sum_{j=1}^{m} f(\mathbf{x};\boldsymbol{\theta};\boldsymbol{u}_j)}{m} \right)^2$$

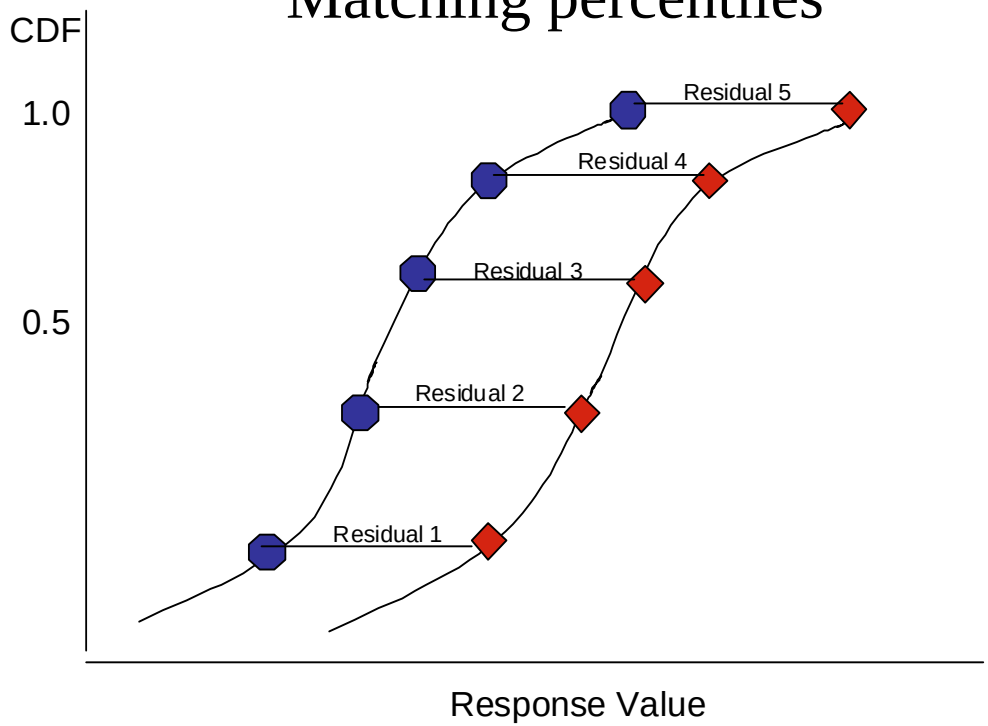
Matching means and variances



$$S(\boldsymbol{\theta}) = (\mu_{\exp} - \mu_{sim})^2 + (\sigma_{\exp} - \sigma_{sim})^2$$

Examples (cont'd)

Matching percentiles

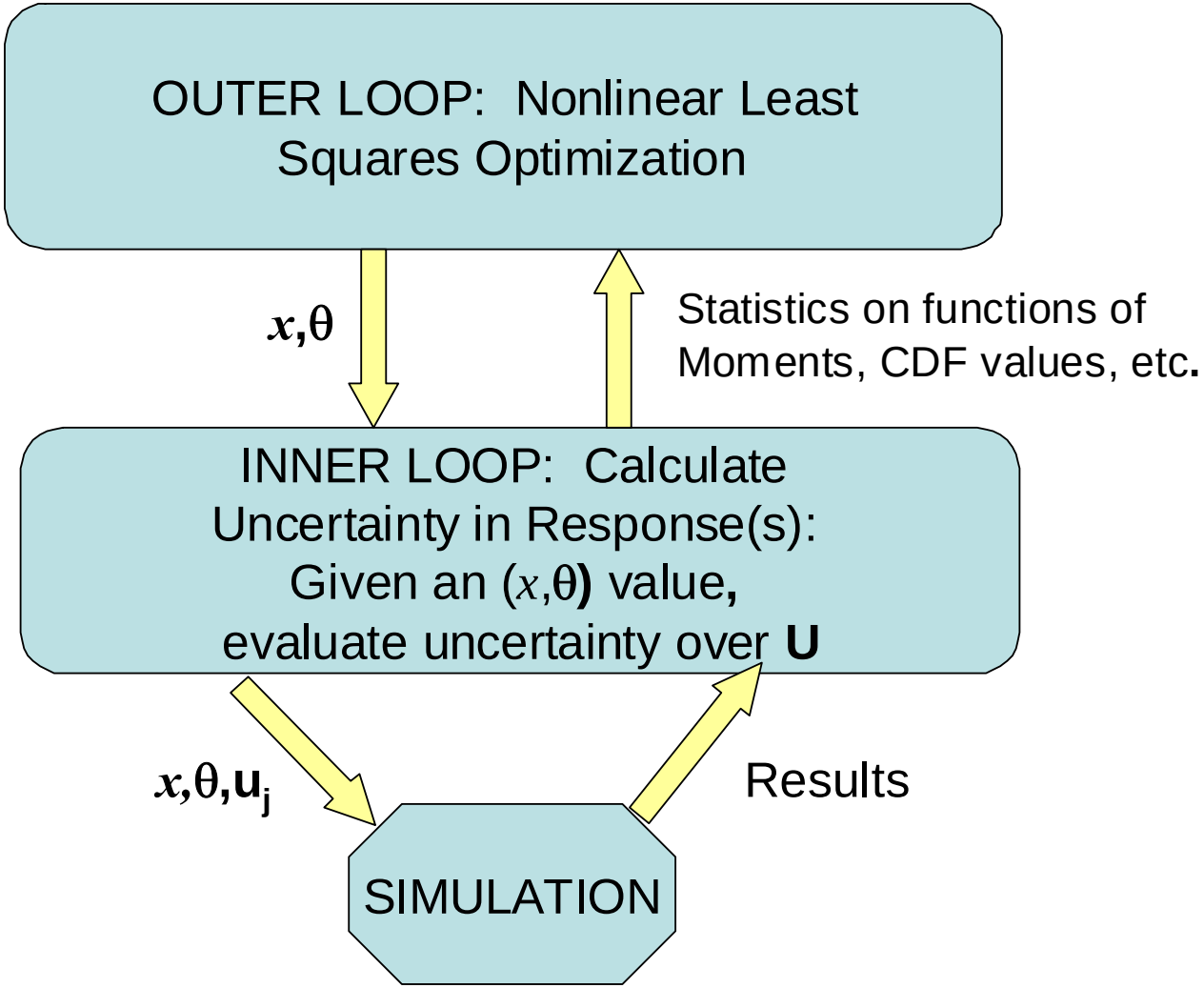


$$S(\boldsymbol{\theta}) = \sum_{k=1}^{K} (CDF_{\exp(k)} - CDF_{sim(k)})^2 = \sum_{k=1}^{K} \left(\left[Y_k : \Pr(y \le Y_k) = p_k \right] - \left[f_k : \Pr(f(\mathbf{x};\boldsymbol{\theta};\boldsymbol{U}) \le f_k) = p_k \right] \right)^2$$

or

$$S(\boldsymbol{\theta}) = \sum_{k=1}^{K} (CDF_{\exp(k)} - CDF_{sim(k)})^2 = \sum_{k=1}^{K} \left(\Pr(y \le Y_k) - \Pr(f(\mathbf{x};\boldsymbol{\theta};\boldsymbol{U}) \le Y_k) \right)^2$$

Solution Approach



Solution Approaches: Inner Loop

- **Uncertainty quantification can be calculated using:**
  - **Sampling**
    - **We used Latin Hypercube sampling (LHS), a stratified sampling method which results in a good sample spread over the input domain.**
  - **Reliability Methods**
    - **Transform the UQ problem into an optimization problem, finding the most probable point (MPP).**
    - **Requires finite difference or analytic gradients of the output variable with respect to the uncertain variables. Finding each point in a cumulative distribution function requires performing a separate optimization run. We used the mean value method or the advanced mean value (AMV) with first and second order Taylor series expansions.**
  - **Stochastic Expansions**
    - **Polynomial Chaos: Approximate a stochastic function in terms of finite-dimensional series expansions, often based on Hermite polynomials which are functions of the Gaussian random variable.**
    - **Stochastic collocation: Lagrange polynomial interpolants are used instead of orthogonal polynomials.**
    - **Used 3rd order quadrature to sample the points upon which the expansion is based.**

Sandia National Laboratories

Solution Approaches: Outer Loop

- The preceding formulations become nonlinear least squares (NLLS) optimization under uncertainty problems
- NLLS algorithms exploit the structure of the objective function. By assuming that the residuals are close to zero near the solution, the Hessian matrix of second derivatives of $S(\theta)$ can be approximated using only first derivatives of the residual terms.
- For the NLLS algorithm, we considered 3 approaches:
  - Gauss-Newton: approximates the Hessian as $J^T J$, where J is the Jacobian matrix.
  - NLSSOL: uses a sequential quadratic programming formulation (SQP)
  - NL2SOL uses a trust-region method (and thus can be viewed as a generalization of the Levenberg-Marquardt algorithm)
- The approximation of gradients can be problematic: these approaches rely on accurate results from finite differencing

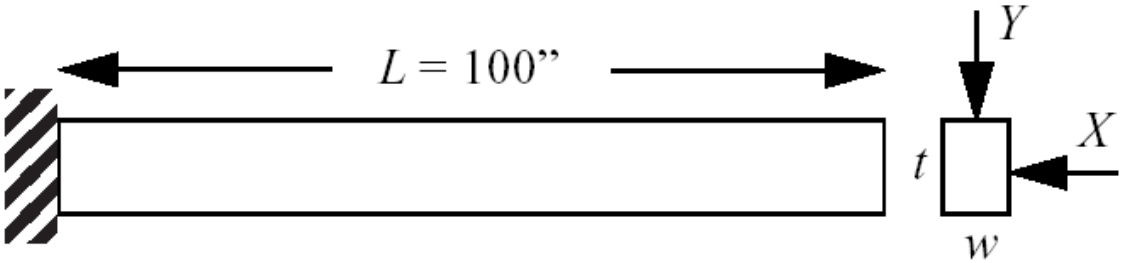
Sandia National Laboratories

Case Studies

- Used problems in the reliability literature (cantilever beam, steel column)
- Used DAKOTA software for implementation
  - Allows a nested approach, NLLS outer loop, UQ inner loop
  - Use of 3 NLLS methods, 3 UQ methods
  - Option to utilize analytic or generate finite difference gradients
    - Derivatives may be needed for both optimization (w.r.t. calibration parameters), or UQ (w.r.t uncertain variables).
  - Design variable insertion: the capability to calibrate parameters which specify or govern distributions (such as means, variances, and bounds).
  - In some cases, we have the ability to compute analytic derivatives of the statistics of interest with respect to the calibration parameters, without resorting to additional evaluations of the computational model.

Sandia National Laboratories

Cantilever Beam



$$Stress = \frac{600}{wt^2}Y + \frac{600}{w^2 t}X$$

$$Displacement \quad D = \frac{4L^3}{Ewt}\sqrt{\left(\frac{Y}{t^2}\right)^2 + \left(\frac{X}{w^2}\right)^2}$$

- **We are interested in obtaining calibrated estimates for the width and thickness of the beam (*w* and *t*) which result in the simulated displacement “matching” the experimental displacement in a variety of ways**
- **Simulated 20 experimental data points by holding the design variables fixed at w*=2.5 in. and t* = 2.5 in., and sampling over the uncertain variables. The mean displacement of this particular sample is 3.993 inches.**

Cantilever Beam

	Variable	Description	Nominal Value
constant	L	length	100 inches
design $\boldsymbol{\theta}$	w	width	4 inches
	t	thickness	4 inches
uncertain $\boldsymbol{u}$	X	horizontal load	500 lbs
	Y	load	1000 lbs
	R	yield stress	40000 psi
	E	elastic modulus	2.9E7 psi

- **The displacement of the nominal beam given the properties in Table 1 above (with *w* = *t* = 4 inches) is 0.605 inches.**
- **Thus, we expect the calibration to result in smaller values for *w* and *t* to match with the greater displacement observed in the simulated experiment.**

Sandia National Laboratories

Cantilever Beam: Matching Mean

NLLS Method (Outer Loop)	UQ Method (Inner Loop)	$\hat{t}$	$\hat{w}$	Residual Norm Value	LHS-verified Res Norm	Num Fn Evaluations
NL2SOL	sampling (20)	2.116	2.795	1.03E-10	4.52E-03	840
Gauss Newton	sampling (20)	3.237	2.224	1.00E-06	4.76E-04	1700
NLSSOL	sampling (20)	2.437	2.53	2.22E-15	2.78E-03	2120
NL2SOL	mean value, numerical deriv.	2.11	2.791	7.33E-11	3.48E-02	42
Gauss Newton	mean value, numerical deriv.	3.243	2.22	1.19E-06	1.51E-02	85
NLSSOL	mean value, numerical deriv.	2.428	2.53	3.11E-15	2.77E-02	106
NL2SOL	mean value, analytic deriv.	2.11	2.791	0	3.48E-02	20
NL2SOL	PCE	2.117	2.795	4.44E-16	4.76E-06	270
NL2SOL	stochastic collocation	2.117	2.795	4.44E-16	7.19E-07	270

- **Several combinations of (t,w) achieve a mean displacement very close to 3.993**
- **Non-uniqueness or lack of identifiability of parameter values is a common issue in inverse problems → Regularization approaches may be used to address this**
- **Tradeoff between accuracy and efficiency: Mean value methods use fewer function evaluations, not as accurate**
- **The NL2SOL algorithm consistently performs well. The remaining results are presented using NL2SOL.**

Sandia National Laboratories

Cantilever Beam: Matching Mean and Std. Dev.

UQ Method (Inner Loop)	$\hat{t}$	$\hat{w}$	Residual Norm values (mean, std.dev)	LHS-verified Residual Norm (mean, std.dev)	Num Fn Evaluations
sampling (20)	2.552	2.469	6.13E-03, 5.24E-02	3.86E-3, 5.30E-03	1640
mean value, numerical deriv.	2.02	2.91	1.48E-8, 1.92E-10	3.54E-2, 8.38E-02	49
PCE	2.328	2.602	4.69E-12, 1.15E-12	9.48E-6, 2.95E-04	351
stochastic collocation	2.328	2.602	4.76E-12, 1.16E-12	9.48E-6, 2.88E-04	351
PCE specifying reliability indices	2.328	2.603	5.94E-5, 3.68E-05	5.00E-5, 3.31E-04	324
stochastic collocation specifying reliability indices	2.328	2.603	5.95E-5, 3.68E-05	5.00E-5, 3.25E-04	324

- **Again, several combinations of (t,w) achieve a mean displacement very close to 3.993 and a standard deviation close to 0.414**
- **Matching mean and standard deviation requires more function evaluations than matching mean only**
- **How to weight the mean vs. standard deviation in a consistent way?**
  - **Use of reliability indices**
  - **Weighted least squares**

Sandia National Laboratories

Reliability Indices for Implicit Weighting

- **A reliability index can be thought of as the number of standard deviations a particular output value z* is from the mean.**

$$\beta = \frac{z^* - \bar{z}_{\exp}}{\bar{\sigma}_{\exp}}$$

- **Instead of explicitly matching the mean and standard deviation, we re-formulate the problem to match several reliability indices corresponding to experimental data.**
- **For this particular problem, we took 5 experimental data points corresponding to the 5th, 25th, 50th, 75th, and 95th percentile values.**
- **We then optimized so that the simulation reliability indices would approximately correspond to the experimental reliability indices.**
- **Doing this implicitly matches the simulation mean and standard deviation to the experimental mean and standard deviation, because the simulation betas are calculated as:**

$$\beta_{sim} = \frac{z^*_{sim} - \bar{z}_{sim}}{\bar{\sigma}_{sim}}$$

- **This approach may be more appropriate than arbitrarily choosing a weighting scheme such as 2:1 for mean: standard deviation.**

Sandia National Laboratories

Cantilever Beam: Matching Percentiles

UQ Method (Inner Loop)	$\hat{t}$	$\hat{w}$	Residual Norm values (one row per 0.05, 0.25, 0.5, 0.75, and 0.95 percentile)	LHS-verified residual norm values of percentiles	Num Fn Evaluations
sampling (20)	2	2.958	5th: 0.1582 25th: 0.1213 50th: 0.02687 75th: 0.0832 95th: 0.2452 TOTAL: 0.3279	5th: 0.2298 25th:0.0211 50th: 0.0980 75th: 0.0123 95th: 0.1826 TOTAL: 0.3104	1440
sampling (100)	2.444	2.534	5th: 0.0889 25th: 0.0489 50th: 0.0712 75th: 0.1333 95th: 0.0886 TOTAL: 0.2024	5th: 0.0924 25th: 0.0298 50th: 0.0885 75th: 0.1002 95th: 0.0140 TOTAL: 0.1658	6200
MPP x-Taylor mean (AMV)	4.3	2.008	5th: 0.0799 25th: 0.0658 50th: 0.1231 75th: 0.0760 95th: 0.0315 TOTAL: 0.1807	5th: 0.1232 25th: 0.0373 50th: 0.1180 75th: 0.0363 95th: 0.0866 TOTAL: 0.1983	385
MPP x-Taylor mpp (AMV+)	4.305	2.007	5th: 0.0799 25th: 0.0659 50th: 0.1231 75th: 0.0760 95th: 0.0315 TOTAL: 0.1807	5th: 0.1233 25th: 0.0373 50th: 0.1179 75th: 0.0363 95th: 0.0866 TOTAL: 0.1983	550
PCE, specifying percentiles	2.395	2.56	5th: 0.0903 25th: 0.0046 50th: 0.1063 75th: 0.0769 95th: 0.0077 TOTAL: 0.1660	5th: 0.0826 25th: 0.0444 50th: 0.1078 75th: 0.0817 95th: 0.0093 TOTAL: 0.1648	1809
stochastic collocation, specifying percentiles	2.408	2.552	5th: 0.0895 25th: 0.0475 50th: 0.1065 75th: 0.0776 95th: 0.0071 TOTAL: 0.1663	5th: 0.0827 25th: 0.0445 50th: 0.1079 75th: 0.0818 95th: 0.0075 TOTAL: 0.1649	2349

Sandia National Laboratories

Cantilever Beam: Matching Percentiles

- Matching percentiles requires more function evaluations than matching first and second moments
- Some parameter values result in better matches to the middle percentiles and are very inaccurate in the tail
- There is a tradeoff in matching the various percentiles: some can be matched closely at the expense of others. Could employ weighting to obtain better matches at the percentiles of interest.
- UQ performed with 20 sample LHS suffered from problems resolving tail probabilities.
- Increasing the sample size from 20 to 100 greatly increases the accuracy of the result, but at the cost of significantly more function evaluations.
- Adding second order information to the reliability method does not improve the accuracy of the solution and requires more function evaluations.
- The stochastic expansion methods perform well.

Sandia National Laboratories

Steel Problem: Design Variable Insertion

Variable	symbol	reference value	unit
mean of flange breadth	b	200.00	mm
mean of flange thickness	d	17.50	mm
mean of height of steel profile	h	100.00	mm

variable	symbol	distribution	mean/standard deviation	unit
yield stress	F_s	lognormal	400/35	MPa
dead weight load	P_1	normal	500000/50000	N
variable load	P_2	Gumbel	600000/90000	N
variable load	P_3	Gumbel	600000/90000	N
flange breadth	B	lognormal	b/3	mm
flange thickness	D	lognormal	d/2	mm
height of profile	H	lognormal	h/5	mm
initial deflection	F_0	normal	30/10	mm
Young's modulus	E	Weibull	21000/4200	MPa

,

$$f(\boldsymbol{\theta}, \boldsymbol{u}(\boldsymbol{\theta})) = F_s - P\left(\frac{1}{A_s} + \frac{F_0}{M_s} \cdot \frac{E_b}{E_b - P}\right) \qquad P = (P_1 + P_2 + P_3)$$

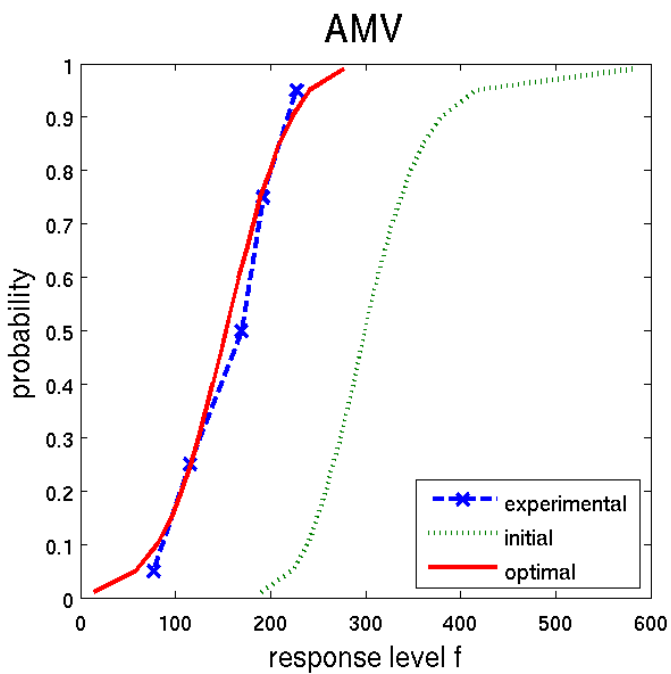
$$A_S = 2BD, M_S = BDH, M_i = 0.5BDH^2, \text{and } E_b = \frac{\pi^2 E M_i}{s^2}$$

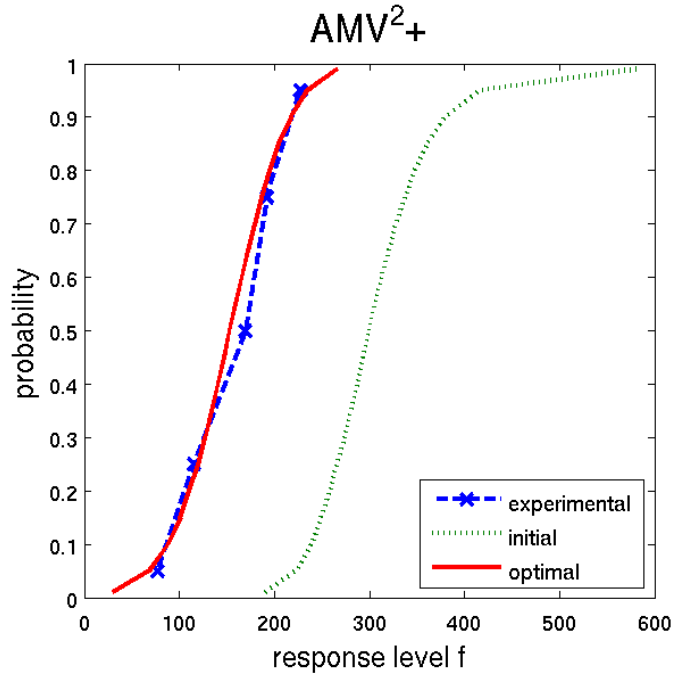
Sandia National Laboratories

Steel Column Problem

UQ algorithm	*b* *	*d* *	*h* *	solver residual norm	LHS-verified residual	function evals
mean value	200	17.07	100	19.32	20.87	91
u-space AMV	228.7	14.42	220.4	24.07	28.77	145
u-space AMV+	200	17.03	100	18.63	21.45	399

- We cannot always precisely recover the optimal parameter *b* since the parameters *B* and *D* are somewhat confounded in the response function *f*, and there are only 5 percentiles on the CDF to match.





Summary

- Presented several formulations to estimate model parameters which result in the “best distribution fit” between experimental and simulation data
- Our results show reliability methods or stochastic expansion methods are more effective than sampling in the propagation of uncertainty in the inner loop, due in part to availability of analytic derivatives of statistics for these methods.
- NL2SOL generally was more efficient than NLSSOL or Gauss Newton on the problems in our case studies.
- The optimal model parameters obtained when matching statistical metrics may be non-unique, a common situation in inverse problems.
- Best distribution fit problems may require regularization approaches to address the lack of identifiability in parameter estimation.
- Finding optimal parameters to match percentiles may result in some percentiles being closely matched and others (such as tail probabilities) not being as well matched. Weighting the residual terms of interest can help address this.
- Future work:
  - Treat the uncertainty as epistemic instead of aleatory, and matching interval bounds.
  - Further investigate stochastic response approximation such as PCE for use in calibration problems
  - Use on Sandia problems

Sandia National Laboratories