SAND2018-13929C

Real-Space Model Validation and Predictor-Corrector Extrapolation applied to the Sandia Cantilever Beam End-to-End UQ Problem[1]

Vicente Romero

Sandia National Laboratories,[2] Albuquerque, NM

Abstract

This paper describes and demonstrates the Real Space (RS) model validation approach and the Predictor-Corrector (PC) approach to extrapolative prediction given model bias information from RS validation assessments against experimental data. The RS validation method quantifies model prediction bias of selected output scalar quantities of engineering interest (QOIs) in terms of directional bias error and any uncertainty thereof. Information in this form facilitates potential bias correction of predicted QOIs. The PC extrapolation approach maps a QOI-specific bias correction and related uncertainty into perturbation of one or more model parameters selected for most robust extrapolation of that QOI's bias correction to prediction conditions away from the validation conditions. Such corrections are QOI dependent and not legitimate corrections or fixes to the physics model itself, so extrapolation of the bias correction to the prediction conditions is not expected to be perfect. Therefore, PC extrapolation employs both the perturbed and un-perturbed models to estimate upper and lower bounds to the QOI correction that are scaled with extrapolation distance as measured by magnitude of change of the predicted QOI. An optional factor of safety on the uncertainty estimate for the predicted QOI also scales with the extrapolation. The RS-PC methodology is illustrated on a cantilever beam end-to-end uncertainty quantification (UQ) problem. Complementary "Discrete-Direct" model calibration and simple and effective sparse-data UQ methods feed into the RS and PC methods and round out a pragmatic and versatile systems approach to end-to-end UQ.

I. Introduction

Methodologies for modeling and prediction in the presence of uncertainty are being actively researched and formulated by the modeling and simulation community. Comprehensive and detailed frameworks for end-to-end uncertainty quantification (E2E UQ) that spans experiment design, experimental uncertainty and data UQ processing, model calibration and validation, and extrapolative prediction under uncertainty, are still elusive. Over 100 references considered in [1]

---

[1] Sandia National Laboratories document SAND2018- (unlimited release). This paper is a work of the United States Government and is not subject to copyright protection in the U.S.

[2]Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

- [7] reveal various lines of thinking and progress in these areas. This paper concentrates on the data UQ and model validation and extrapolative prediction elements of end-to-end UQ, while keeping in mind their necessary connectivity to the up-stream UQ elements.

Model validation is still a fast-developing field in engineering and science. For example, institutions [8] – [13] cite the following common definition of model validation (within a few slight variations): *the process of determining the degree to which a computer model is an accurate representation of the real world from the perspective of an intended use of the model*. But as pointed out in [4] and [5] there are important differences in how the definition is interpreted by the institutions and how it might be implemented in practice. Indeed, a wide variety of associated viewpoints and methodologies exist in the literature. Besides those already mentioned, references such as [8] – [30] explore many important philosophical and implementation issues in model validation, and survey various paradigms and tactical approaches for performing model validation for application problems. Some of these references go on to develop a specific approach and implementation framework.

With both philosophical and practical considerations in mind, a "Real Space" (RS) approach to model validation has been developed and applied at Sandia over the past decade. The RS approach adopts some elements and constructs from the literature (sometimes adding needed refinement) and adds pivotal new elements and constructs. The approach reflects pragmatism, versatility, and capabilities derived from the author's experience with many industrial-scale validation and calibration applications in the following areas: device thermal response and failure [2], [34]; foam thermal pyrolysis and vaporization [31], [32]; computational fluid dynamics fire combustion and object heating [23]; propellant fire combustion and object heating (draft report not publicly available); sealing glass solid mechanics constitutive model [33]; device thermal-structural response and failure [28], [30]; structural dynamics model of arming and fusing assembly (report not publicly available); radiation-damaged electronics behavior models (reports not publicly available).

These applications have involved calibration and/or validation of models of stochastic phenomena and/or systems with small random variations from unit to unit and test to test. Significant tactical UQ difficulties addressed include appropriate representation, interpretation, propagation, and aggregation of heterogeneous sources and types of experimental uncertainty: probabilistic, interval, aleatory, epistemic, traveling, non-traveling, continuous, discrete, scalar and functional. Additional challenges included probability distributions of uncertain form and/or shape, and dealing with data sparseness both in terms of only one or a few imperfectly replicated tests available/affordable to address physical system variability, and limited spatial sensor data to approximately reconstruct spatial field boundary conditions. Challenges of solution verification to estimate model discretization related solution uncertainty were also involved, as was added uncertainty from using response-surface surrogate models to reduce the number of physics model simulations for uncertainty propagation.

The RS methodology also implements approaches for dealing with the following strategic difficulties.

- Model accuracy characterization—deciding on the formulation or metric for characterizing the discrepancy between model and experiment results.
- Model adequacy characterization—deciding on the threshold or criterion for model adequacy (acceptable agreement with reality).
- Dealing with extrapolation of model validation information/results/ products—deciding how to use information from the validation assessments to best qualify or improve model predictions at conditions different from the validation setting. (This is not part of model validation, but is an essential consideration in the formulation of a workable validation paradigm.)

These items are considered to be strongly interdependent. It is reasoned that only by considering them together can a comprehensive, workable, and relevant model validation framework be formulated to support end-to-end UQ in realistic industrial-scale problems.

The RS validation method quantifies model prediction bias of selected output quantities of engineering interest (QOIs) in terms of directional bias error and any uncertainty thereof. Information in this form facilitates potential extrapolation of any model-form related bias corrections to those QOIs when predicted at other conditions beyond the validation setting. The "Predictor-Corrector" (PC) extrapolation approach explained and demonstrated in this paper maps a QOI-specific bias correction and related uncertainty into perturbation of one or more model parameters selected for most robust extrapolation of that QOI's bias correction to the new prediction conditions. Such corrections are QOI dependent and not legitimate corrections or fixes to the physics model itself, so extrapolation of the bias correction to the prediction conditions is not expected to be perfect. Therefore, PC extrapolation employs both the perturbed and un-perturbed models to estimate upper and lower bounds to the QOI correction that are scaled with extrapolation distance as measured by magnitude of change of the predicted QOI. An optional factor of safety on the uncertainty estimate for the predicted QOI also scales with the extrapolation. Complementary "Discrete-Direct" model calibration [7] and simple and effective sparse-data UQ methods ([35], [36]) feed into the RS and PC methods and round out a pragmatic and versatile systems approach to end-to-end UQ.

Many important challenges encountered in real problems are featured in a Sandia-developed Cantilever Beam End-to-End UQ problem [37]. The problem emphasizes difficult paradigm and strategy issues while being computationally trivial so that approaches and methodologies can be focused on. The next section presents some model validation and extrapolative prediction related aspects of the Beam E2E problem. Section 3 summarizes the RS model validation approach and its application to the Beam problem. Section 4 summarizes the PC approach to extrapolative prediction, and its application to the Beam problem. Section 5 provides some closing remarks.

II. Cantilever Beam Problem Set-Up

This section summarizes the pertinent set-up information for the validation and extrapolation demonstrations in the next two sections. From Part C of the Cantilever Beam E2E UQ problem statement [37], a large population of cantilever beams is considered where the rectangular beams have dimensions given in Table 1.

Table 1 – Dimensions of Beams in Validation Experiments (negligible dimensional variability and measurement errors)

L	W	H
2.20	0.09292	0.18580

Only two replicate tests can be afforded to get an indication of response variability in a large full population of beams (asymptotically ∞) that will be put into service in the tested loading configuration, but at a moderately higher temperature. Accordingly, two tests are performed with vertical downward loading on the end of the cantilever beams as shown in Figure 1. The beams have zero deflection (D=0) and zero slope (dD/dx = 0) where they horizontally protrude from a rigid vertical wall. Beam height is measured upwards from the bottom of the beam as shown. Beam width is measured perpendicular to the height and length directions. The beam is made of a homogeneous isotropic material that has strength parameter E which is suspected to be a function of temperature, E(T). The physical and modeled beams are spatially uniform in temperature. The validation tests take place at a temperature Tval = 60C.

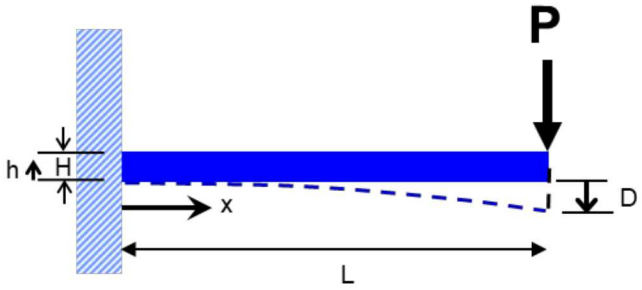


Figure 1 – Cantilever Beams in validation and prediction scenarios

Table 2 presents the measured deflections and end loads in the validation experiments. Small load control variations occur about the experimental target of Po = 7.5E5.

The deflection measurements include possible errors, where

$$\text{measurement error} = \text{measured value} - \text{true value}. \quad (1)$$

The problem statement explains that potential "systematic" error associated with beam deflection measurement is expected to lie within the following range, with "high" confidence.

$$U[\text{defl\_err\_sys}] = I[-2\%, 0\%] \text{ of measurement value} \quad (2)$$

Systematic errors and associated uncertainties are perfectly or fully correlated from test to test, per the problem statement. U[·] in Eqn. 2 signifies that the quantity within the brackets is uncertain. I[·] signifies an interval type uncertainty over the range specified in the brackets.

Table 2 – Beam deflection and end-load measurements, which include possible random and systematic measurement errors as described in the text.

	deflection D (includes possible systematic and random meas. errors in these results)	**end load P (includes possible systematic and random meas. errors in these results)**
ValTest B.1	0.3880	7.769E5
ValTest B.2	0.3840	7.390E5

The deflection measurements in Table 1 further include a component of measurement error that varies randomly from test to test. The problem statement gives the following probabilistic uncertainty information for these.

U[defl_err_rand] = Normal(mean=0%, stdev=0.5%) of deflection measurement (3)

Similarly, potential random and systematic measurement errors and uncertainties in the measured load data in Table 2 are:

U[load_err_rand] = Normal(mean = 0%, stdev=1%) of load measurement. (4)

U[load_err_sys] = I[-2%, +2%] of load measurement (5)

An ordinary differential equation (ODE) for beam deflection derived from a balance of forces and moments in the classical beam problem we are considering is ([38]):

$$\frac{d^2}{dx^2}\left(\frac{1}{12}WH^3E\frac{d^2}{dx^2}D(x)\right) = q(x) = P\delta(x-L)\,. \quad (6)$$

Here x is a horizontal coordinate that starts at the wall (x=0) and runs along the length of the beam to its free end at x=L as indicated in Figure 1. E is the beam's modulus of elasticity, an effective stiffness/strength property of the material. The model is written for beams with isotropic and spatially uniform modulus E. The generalized loading case involves a distributed load $q(x)$ on the beam. The point load P in Figure 1 is represented by $q(x)$ being a delta function $\delta(x-L)\cdot$P that mathematically recovers the point load P at x=L (see [38]).

Equation 6 together with the relevant geometry and material property values and boundary conditions constitute the model for beam deflection behavior. An analytic solution to the governing equations and parameter variables of the ODE model is ([38]):

$$D = 4PL^3/(EWH^3). \quad (7)$$

To add a source of error/uncertainty resembling discretization-related solution error to the problem, the following is specified in [37]. Consider the hypothetical situation where solutions of the model ODE are performed computationally with a discretized finite-element model. Let the discretization effects in the calculations performed to validate the model have tip deflection solutions that are biased to a smaller computed deflection than a mesh-converged model would predict. Let the bias be 3% such that instead of working with the mesh-converged solution Eqn. 7 the analyst obtains solution results from

$$D = 0.97 * 4PL^3/(EWH^3). \quad (8)$$

Note that the magnitude of discretization related error, which is the deflection result from Eqn. 8 minus the exact solution Eqn. 7, varies over the uncertainty ranges (uncertainty space) of the model input variables P, L, E, W, H. Let a solution verification analysis at a mean point[3] in this uncertainty space provide an estimate that the asymptotic mesh-converged tip deflection is greater than computed deflections from Eqn. 8 with the following range of uncertainty, to a proclaimed high degree of belief.

$$\text{mesh-converged deflection} = [102\%, 105\%] \text{ of Eqn. 8 working-mesh deflection} \quad (9)$$

The material strength property (modulus E) varies randomly from beam to beam in the population. The probability density function (PDF) governing the material property variation is unknown to the analyst. Accordingly, a parameter calibration/estimation activity is set up in Part B of the E2E problem. For parameter estimation data, four test beams from the same batch of material and nominally 10% shorter than the validation beams are tested with a target end-load magnitude of Po = 7.5E5. The estimation of modulus E variation from the test data and a calibration inversion of the solution equation (7) of the ODE model is somewhat muddied by the following uncertainties.

The four beams have slightly different length, width, and height dimensions. The dimensions are machined from three different types of machines. Control accuracy variations are independent among the three machines. Dimension measurement errors/uncertainties are also independent for each dimension. The measurement errors/uncertainties have random and systematic components specified in [37]. The point end-loads are slightly different in each test as listed in [37], varying about the target testing load Po=750,000 by significantly more than explainable by possible random and systematic measurement errors defined in [37]. The measured deflections for each beam are also subject to random and systematic measurement errors defined in [37]. The calibration tests take place at a temperature Tcal = 20C. The physical and modeled beams are spatially uniform in temperature.

A novel "Discrete-Direct" model calibration and uncertainty propagation approach demonstrated in [7] addresses all the uncertainties mentioned, including accounting for the fact that only four samples of the material property random variability are available with which to estimate or bound

[3] The analyst is asked to comment on whether the supplied discretization bias information is sufficient and how it can best be used in view of the fact that discretization bias is a function of location in the uncertainty space of the inputs P, L, E, W, H. The analyst is asked whether discretization refinement studies should be run at alternative locations in the uncertainty space and where these should be and how the results would be used. Some strategies in this general area are demonstrated in [39] on some application problems. The present paper ignores the Beam problem's discretization-bias variation over the uncertainty space.

the material variability over the whole population. This is accomplished by performing a separate calibration to each experiment, yielding a value E_i for the ith beam tested (i = 1 to 4), or rather, a set of possible values E_i,k for each beam according to 1000 possible values/realizations (k = 1 to 1000) of beam i's true experimental dimensions, load magnitude, and deflection within the prescribed experimental errors/uncertainties. Discretization related solution error for the ODE model is also specified in [37] and adds a source of systematic error/uncertainty for the four beam calibrations (as might error associated with the use of response-surface surrogate models if employed for the calibrations), but these are ignored in [7] and therefore in the present calibration results. Sample values of E_i,k from [7] are presented in Figure 2. The calibrations are accomplished with standard EXCEL spreadsheet functions, given the compact expression Eqn. 8 for beam deflection and the relative simplicity of the DD methodology even when accounting for all the said sources and types of uncertainty in the calibrations.

	A	B	C	D	E	F
1	Data Set 2					
2			Calibration results			
3						
4		realization k	Test/Beam 1 uncer. realizations of E from calibration to experimental data	Test/Beam 2 uncer. realizations of E from calibration to experimental data	Test/Beam 3 uncer. realizations of E from calibration to experimental data	Test/Beam 4 uncer. realizations of E from calibration to experimental data
5						
6		1	1.928E+11	1.971E+11	2.070E+11	2.125E+11
7		2	1.945E+11	2.075E+11	2.135E+11	2.157E+11
1004		999	1.893E+11	2.014E+11	2.062E+11	2.154E+11
1005		1000	1.801E+11	1.896E+11	1.961E+11	2.028E+11
1006						
1007						
1008	stats	mean =	1.923E+11	2.024E+11	2.090E+11	2.148E+11
1009	from	max =	2.064E+11	2.188E+11	2.263E+11	2.336E+11
1010	1000	min =	1.762E+11	1.877E+11	1.949E+11	1.993E+11
1011	samples	stdev =	4.971E+09	5.192E+09	5.453E+09	5.706E+09

Figure 2. Realizations of possible effective material strength calibration parameter values $E_{i,k}$ for the four beams tested.

The DD calibration and uncertainty propagation approach straightforwardly accommodates many sources and types of heterogeneous uncertainties as previously mentioned. The methodology straightforwardly accommodates problems with many calibration parameters or just one as in the Beam problem. The approach appears to have several advantages over Bayesian and other

calibration approaches for capturing and utilizing the information obtained from the typically small number of experiments in model calibration situations. In particular, the DD methodology better preserves the fundamental information from the experimental data in a way that enables model predictions to be more directly traced back to the supporting experimental data. The approach is also presently more viable for calibration involving sparse realizations of random function data (e.g. stress-strain curves) and random field data. The DD methodology is conceptually simpler than Bayesian calibration approaches and is straightforward to implement.

The validation assessment in the next section uses the models, parameter values, and uncertainties presented in this section to predict the variability of beam deflections in an asymptotically large population under the loading specified.

III. Real-Space Model Validation Assessment Methodology and Results

In this section, predicted variability of beam deflections in an asymptotically large population under the specified loading are compared against estimated quantities from two replicate validation tests involving two beams drawn at random from the population. The Real Space validation methodology used for the model validation assessment is introduced next.

III.A. Introduction to the Real Space (RS) Model Validation Approach

The high-level paradigm that underlies the Real Space model validation approach is summarized as follows.

- Presume the model is imperfect but potentially useful for predictions and analysis informed by the model validation activity.
- Quantify the prediction error for engineering-relevant QOIs and to a resolution level allowed by the magnitude of uncertainty in the validation experiments, simulations, and UQ procedures.
- Assess whether predicted QOIs bound or envelope reality in a useful way for an anticipated model use, where "reality" is inferred from well-designed and targeted experiments and inference procedures appropriate for the anticipated model use/purpose and for the nature of credibility statements sought about model predictions.

The RS method quantifies model prediction bias of selected output scalar or derived scalar quantities of engineering interest (QOIs, see bullets below) in terms of directional bias error and any uncertainty thereof. Information in this form facilitates potential extrapolation of any model-form related prediction-bias corrections to those QOIs when predicted at other conditions beyond the validation setting (e.g. section IV).

Examples of scalar QOIs admissible for RS validation and PC extrapolation are:

- velocity, pressure, strain, stress, temperature, etc. at a given time and/or spatial location;
- statistics such as mean, variance, and selected percentiles of the above quantities if the model validation activity involves multiple replicate experiments and associated modeling involving random variability of the tested systems and/or test conditions and/or measurement errors on measured inputs and outputs of the tests. (This bullet applies to this paper's Beam problem.)

The last bullet contrasts with some model validation methods and metrics that compare whole distributions of experimental and model-predicted response variability (aleatory variability) when stochastically varying systems are involved. These approaches arrive at a global measure of difference/discrepancy/disagreement between the experimental and predicted response distributions. However, such global discrepancy metrics (like Kolmogorov–Smirnov "distance" between two distribution curves) yield numerical values that seem to lack a high degree of validation interpretability, decision-making utility, or usability for engineering purposes. Non-uniqueness further contributes to a degree of arbitrariness of numerical values from such metrics.

The RS methodology is presently configured for point evaluation of model prediction error "point validation") from comparing experimental and simulation results of engineering importance at the same specified input "scenario" conditions (here, at the same end-load magnitude Po=750,000). The larger paradigm is to point-validate a model at one or more relevant points in the scenario space and map a bias correction into the model at each validation point and then predict at the new scenario conditions, weighting the prediction results according to how close the validation scenario points are to the prediction conditions. The set of predictions then define a scale of model-form related uncertainty at the new conditions. This may be preferable to approaches that would attempt to validate the model over multiple points (sets of conditions) simultaneously. Much more research on validation-to-extrapolation methodology must be conducted to assess this. (An example calibration analogue for the latter validation approach is the Bayesian approach illustrated in [27] involving multiple calibration scenario points.)

The RS uncertainty accounting and comparison system transforms all experimental and simulation *input* uncertainties into equivalent effects on *output* response uncertainty and then combines these with any existing output uncertainties. Treatment of the uncertainties is done according to their natures—probabilistic, interval, epistemic, aleatory, correlated, uncorrelated, traveling, and non-traveling. Population statistics such as response mean, variance, and percentiles of response quantities are compared when significant stochastic experimental and simulated behavior exists in the system of interest, as in the present validation problem. Tail percentiles of the distributions are preferred here because they reflect uncertainties in both the mean and variance of the distributions. Moreover, comparing tail percentiles of experimental and predicted response is particularly useful for validating models anticipated to be used in the analysis of performance and safety margins, which is one objective in the beam problem.

Figure 3 depicts uncertainty regarding the true PDF of aleatory uncertainty or variability of experimental response (i.e., variability of deflections in the population of beams in the current problem). Several factors cause this uncertainty, including experimental control and measurement uncertainties, and sparse (only a few) test replicates that sample the variability of the physical systems. The depicted uncertainty in the model-predicted PDF of response variability is, in the current problem, due to propagation of the material strength variability and uncertainty information residing in the calibration values of E in Figure 2. (Figure 3 is a generic figure for illustrative purposes. Its aleatory-epistemic uncertainty representation of the simulation results applies when the epistemic uncertainty in the simulation problem is defined and propagated parametrically, having a systematic character. If a random component of epistemic uncertainty exists in the simulation problem, as it does in the Beam problem through uncorrelated measurement

error/uncertainty contributions to the discretely represented calibration parameter realizations in Figure 2, then the simulation aleatory-epistemic uncertainty is more properly represented in the manner at left in Figure 3 for the experimental uncertainty. This will become apparent in sections III.B and III.C.)

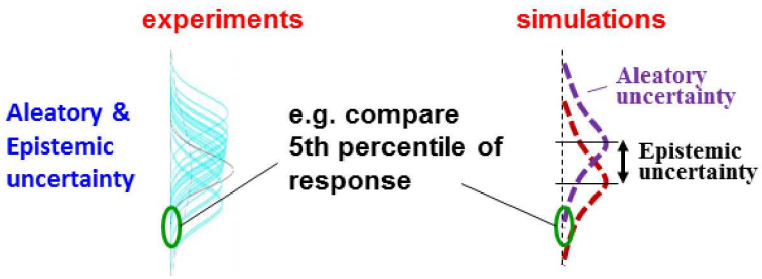


Figure 3. Real Space comparison approach for experimental and simulation results and uncertainties when stochastic behavior exists in the system and response of interest (as in this paper's Beam problem).

III.B. Processing and Roll Up of Experimental Uncertainties for Model Validation Comparisons

Here we demonstrate the specific manner in which the experimental data is processed for Real Space validation comparison to the simulation results. An "apples-to-apples" comparison basis must be established between predicted and experimentally derived percentiles of deflection. Consequently, we *normalize* the experimental data to the following reference conditions input to the simulations in section III.C.

- nominal service load of Po = 7.5e5 point end-load downward
- beam dimensions specified in Table 1

We also account for inference uncertainty on estimated percentiles of response from small numbers of replicate tests.

The problem statement says the beam dimensions in Table 1 are tightly controlled experimentally such that they exist for both test beams. However, if the beams are loaded the same, their deflections will differ according to material strength variations among the beams. This unit-unit stochastic variability effect is to be assessed by replicate testing of two beams randomly drawn from the larger population. Reflective of many real testing situations, the target load of Po=7.5e5 applied in the validation simulations is not exactly produced in the two "replicate" validation tests, per Table 2. Accordingly, in the following we normalize the experimental displacements in Table 2 to values that are much more consistent with the target load Po used in the simulations. In doing so, we also account for the fact that the table's values of measured load and deflection may differ from the true or actual experimental loads and deflections by the random and systematic measurement errors/uncertainties defined by equations 1 to 5.

We start by writing an identity for the ith test and beam, where i=1 corresponds to beam/test B.1 and i=2 corresponds to beam/test B.2 per the designations in Table 2.

$$D_{beam\_i}(Po) = D_{beam\_i}(P_{i,true}) + [D_{beam\_i}(P_{i,meas}) - D_{beam\_i}(P_{i,true})] + [D_{beam\_i}(Po) - D_{beam\_i}(P_{i,meas})] \quad (10)$$

The 'meas' subscript in this equation signifies the measurement result (which is usually different from the true value) for the quantity being measured.

The term on the left-hand side (LHS) of Eqn. 10 represents the sought deflection of beam i if loaded with the target end-load Po=7.5e5. The quantity in the top row of the equation's right-hand side (RHS) represents the true deflection of beam i under its actual or true downward end-load of magnitude Pi,true in the test. This quantity is equal to the RHS of the following equation via application of a rearranged form of the measurement error equation, Eqn. 1.

$$D_{beam\_i}(P_{i,true}) = D_{beam\_i,meas}(P_{i,true}) - D_{beam\_i,measErr} \quad (11)$$

Whereas the RHS term $D_{beam\_i,meas}(P_{i,true})$ is the fixed nominal measured deflection given in Table 2 for beam i, the measurement-error term $D_{beam\_i,measErr}$ is uncertain per Eqns. 2 and 3. Therefore, also uncertain is the LHS term (which is also the RHS term in the first row of Eqn. 10). These uncertainties will be figured-in later.

The third row of Eqn. 10 represents the difference in beam i's deflection that would exist if the beam is subjected to the target load Po and then to an actual or true load equal to the nominal measurement value stated in Table 2. The model is used to approximate this difference as follows, by simulating at these two loads.

$$D_{beam\_i}(Po) - D_{beam\_i}(P_{i,meas}) \approx D_{beam\_i\_model}(Po) - D_{beam\_i\_model}(P_{i,meas}) \quad (12)$$

The simulation model is used as the best available mechanism for estimating how behaviors would change in reality under these different input conditions. The model does not have to be accurate in an absolute sense, but it must be sufficiently accurate in a relative sense of providing reasonable trend information, such that accounting for this term improves the validation analysis. In practice, uncertainty can be added to the result according to judgment of the physics modelers and VVUQ analysts in the project. No such uncertainty is applied in the following, but it is easy to do so in the related column of the data processing spreadsheet introduced later (Figure 4).

The second row of Eqn. 10 represents the difference in beam i's deflection between its actual and nominal measured (Table 2) experimental loads. The model is used for evaluating this difference as well. However, $D_{beam\_i}(P_{i,true})$ in the second row is uncertain due to measurement-related uncertainty regarding what the true experimental load $P_{i,true}$ is, and this load uncertainty is sampled 1000 or more times in the RS validation-UQ approach (illustrated below). So model approximations are employed to avoid running the (often expensive) physics model at each of the sampled possible true values of the experimental inputs.

The easiest model approximation approach comes from a first-order Taylor Series approximation expanded about the nominal measured values of the uncertain experimental input factors and conditions. For the single uncertainty here, load magnitude $P_{i,true}$:

$$D_{beam\_i}(P_{i,meas}) - D_{beam\_i}(P_{i,true})] \approx \frac{\partial(D_{beam\_i})}{\partial(P)}|_{P_{i,meas}} \bullet (P_{i,meas} - P_{i,true}). \quad (13)$$

The difference in parenthesis at right in the above equation is replaced using a relevant form of the measurement error equation, Eqn. 1.

$$D_{beam\_i}(P_{i,meas}) - D_{beam\_i}(P_{i,true})] \approx \frac{\partial(D_{beam\_i})}{\partial(P)}|_{P_{i,meas}} \bullet P_{i,measErr}. \quad (14)$$

The two model results on the RHS of Eqn. 12 could be used to approximate the derivative in Eqn. 14 by finite differencing (FD). Normally, large-perturbation FD is preferred, with a relevant step size indicative of the extent of the input uncertainty relative to its nominal value. If the model evaluation at Po on the RHS of Eqn. 12 is not suitable in these terms, then an added simulation could be performed at an appropriate load value P* to use with the result $D_{beam\_i\_model}(P_{i,meas})$ from the RHS of Eqn. 12 for FD derivative approximation. (These considerations presume an expensive physics model, which is not the case here, but the consideration of economy here is generally important in real problems.) In the following, an analytic derivative of Eqn. 8, with mesh effect adjustment, is used in Eqn. 14 for expediency.

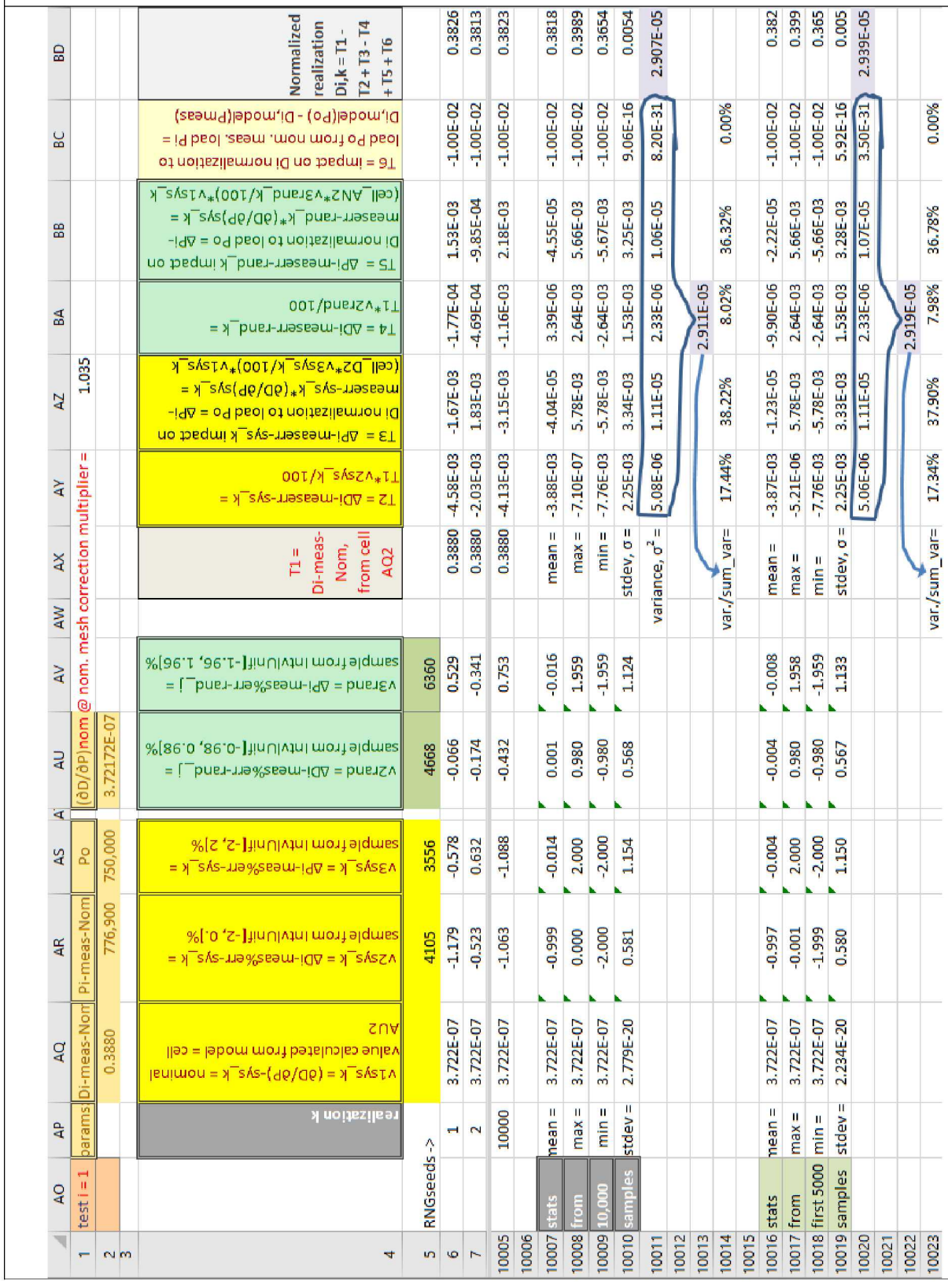
Note that derivatives as in Eqn. 14 can be evaluated from appropriate experimental trend information from testing as in [28]. Real problems usually involve multiple experimental inputs that are significantly uncertain, so a multivariate Taylor Series approximation would be used, as in [28]. Higher-order Taylor Series approximations could be used as well. Indeed, multi-variate linear and quadratic polynomial response-surface approximation analogues to Taylor Series approximations, including dimension- and order- adaptive mixed linear-quadratic response surfaces with estimation of associated error/uncertainty have been used in [34]. Insignificant differences were found between linear and quadratic results, so the simple linear Taylor Series approximation is used in the present analysis.

Eqns. 11, 12, and 14 are substituted into Eqn. 10, imparting an approximate equivalence ≈ between left and right sides of the equation, and giving a general form that applies for any particular realization (signified by added index k in the following) of possible values of the uncertain quantities discussed above.

$$D_{beam\_i,k}(Po) \approx D_{beam\_i,meas}(P_{i,true}) - D_{beam\_i,measErr\_k} + \frac{\partial(D_{beam\_i})}{\partial(P)}|_{P_{i,meas}} \bullet P_{i,measErr\_k} + [D_{beam\_i\_model}(Po) - D_{beam\_i\_model}(P_{i,meas})] \quad (15)$$

The spreadsheet in Figure 4 evaluates this equation for various uncertainty realizations of its k subscripted variables for Beam/test B.1 (i=1). Column AX holds the value of the nominal measured deflection in the test, designated $D_{beam\_i,meas}(P_{i,true})$ in Eqn. 15.

Figure 4. Beam/Test B.1 (i=1) spreadsheet of MC samples of possible random and systematic measurement errors on experimental inputs and outputs + other adjustments to normalize beam displacement to load Po.



Columns AR and AU hold realizations of possible systematic and random displacement measurement error percentages randomly sampled from interval uncertainty ranges Eqns. 2 and 3.[4] These error percentages are converted into displacement error quantities (terms T2 and T4 respectively) in columns AY and BA. These systematic and random error contributions add vectorially (i.e., sign is crucial) into a realization of possible total displacement-measurement related uncertainty designated $D_{\text{beam\_i,measErr\_k}}$ in Eqn. 15. Thus, in column BD, terms T2 and T4 are subtracted from the terms T1, T3, T5, T6 associated with the other RHS members of Eqn. 15.

Similarly, columns AS and AV hold realizations of possible systematic and random load measurement error percentages randomly sampled from interval uncertainty ranges in Eqns. 5 and 4 respectively (see Footnote 4). These error percentages are converted into *displacement* error quantities (terms T3 and T5 respectively) in columns AZ and BB. These systematic and random error contributions add vectorially into a realization of possible total load-measurement related uncertainty equivalent to the middle row in Eqn. 15. Thus, these terms are added to the sum in column BD.

Columns with yellow headers signify that the uncertainties are treated as fully correlated or "systematic" from test to test as asserted in the problem statement and echoed in section II. The sample values in columns AQ – AS are the same in an analogous spreadsheet for Beam/test B.2 (i=2). That spreadsheet's columns AY and AZ have similar values to those in Figure 4's columns AY and AZ, but are slightly different because of each spreadsheet's different base measurement values in column AX that the measurement error percentages in columns AR and AS multiply. Nonetheless, though slightly different values exist between columns AY in the two spreadsheets, the values in the two columns are perfectly correlated. Similar statements apply to column AZ values in the two spreadsheets.

Column AQ holds the value of $\frac{\partial D}{\partial P}$, which is constant in the present treatment but could in general be treated as uncertain. In this case, different realizations (numerical values) would exist in the rows of the column. The present treatment uses the same $\frac{\partial D}{\partial P}$ value for both test beams. In principle these could be different, for example if the beams' dimensions were slightly different or uncertain, then the model solution Eqn. 8 used to approximate the derivate $\frac{\partial D}{\partial P}$ by FD, or the equation's analytic derivative as used here, would give slightly shifted numerical values for different beams. In some real problems the systematic approximation error of the calculated derivative(s) would dominate the unit-to-unit related differences in the calculated derivatives. The approximation uncertainty would then be sampled, with the various realizations in the derivative column (here column AQ) applying across test units, so the column would be colored yellow and have the same set of realizations in each test's spreadsheet.

[4] Interval ranges for the probabilistically described measurement uncertainty distributions in equations (3) and (4) are taken as the central 95% confidence range defined by the mean ± 2 standard deviations of the Normal distributions specified. From the description in [37] of the origins and definition of these measurement uncertainties, this conversion from probability distribution uncertainty treatment to an interval treatment based on a 95% confidence interval draws from the classical measurement uncertainty paper [40].

Columns with green headers have uncertainties treated as uncorrelated or "randomly varying" from test to test. Analogous green columns with the same prescribed uncertainty ranges appear in the spreadsheet for Beam/test B.2, but the sample values will be different because different initial seeds are used for the random-number generator ('RNG' in the spreadsheet).

The summary statistics on the lower half of the figure show that the maximum, minimum, mean, and standard deviation of the samples in each column are not meaningfully different when increasing from 5k to 10k samples, so sufficient sampling exists here.

For brevity in illustrating the methodology going forward, only the first one thousand (1K) results from column BD's possible realizations of adjusted (normalized) deflection for Beam B.1 will be utilized. There are presented in Figure 5 in column BX. The first 1K results for normalized deflection of Beam B.2 are presented in column BY. Each row of data from these two columns represents a pair of plausible deflections of beams B.1 and B.2 under load Po.

Each pair of results is processed according to reliable sparse-data UQ techniques described and tested in [35], [36]. We start by forming 95%_coverage/90%_confidence Tolerance Intervals (95/90 TIs). These give a fairly reliable conservative but not overly conservative bounding estimate of the central 95% range between the actual population's 2.5 and 97.5 percentiles for a large variety of PDF shapes (e.g. 89% of the 144 PDFs tested in [35], [36]). The central 95% range is useful for model validation as discussed above. The TI method is an easy and economical way to obtain such bounding estimates. Their construction is explained next.

For each row or pair of plausible beams' deflections, a mean $\tilde{\mu}$ and standard deviation $\tilde{\sigma}$ are calculated. These are given in columns CA and CB. For a 95/90 TI and two samples, f=18.8 is the value of the factor that multiplies the calculated standard deviation $\tilde{\sigma}$ to create a TI of total length $2f\tilde{\sigma}$. This value of f comes from a look-up table in [41] reproduced from [42]. The interval is centered about the calculated mean $\tilde{\mu}$ of the samples, so the interval's top and bottom ends are defined by $\tilde{\mu} \pm 6.37\tilde{\sigma}$.

The top and bottom ends of 95/90 TIs created for each row of deflections in Figure 5 are listed in columns CD and CI. For the 1000 TIs created, the uncertainty range of the top ends of the TIs (that is, the range of 1000 possibilities below which the 97.5 percentile of deflection lies with ~90% confidence, for an asymptotically large beam population from which the two test beams were drawn) is given by the max and min values in cells CD1009 and CD1010. Cells CI1009 and CI1010 give the uncertainty range for the bottom ends of the 1000 TIs (that is, the range of possibilities above which the 2.5 percentile of deflection lies with ~90% confidence). These min-to-max ranges of the conservatively estimated 2.5 and 97.5 percentiles are plotted at left in Figure 7 in section III.D.

Figure 5. Normalized deflection possibilities of experimental beams B.1 and B.2 if loaded at the reference load Po. Associated 90% confident tolerance interval realizations are presented of central 95% deflection response of an asymptotically large population of similar beams under the reference load Po input to the validation simulations.

	BW	BX	BY	BZ	CA	CB	CC	CD	CE	CF	CG	CI	CJ	CK
1		95/90 TIs for 2 experimental data points												
2														
3								N=2 samples -> 95/90 TI fac			18.8			
4		Normalized deflection for validation Beam1/Test 1	Normalized deflection for validation Beam2/Test 2		μ, N tests	σ, N tests		TI top end, μ + TI_factor*σ				TI bot. end, μ - TI_factor*σ		
5									sorted in ascendin				sorted in ascending	
6	1	0.3826	0.3971		0.3899	0.0103		0.5828	0.3828			0.1969	0.0693	
7	2	0.3813	0.3888		0.3851	0.0053		0.4844	0.3879			0.2857	0.0697	
1004	999	0.3875	0.3963		0.3919	0.0062		0.5087	0.6921			0.2752	0.3902	
1005	1000	0.3927	0.4039		0.3983	0.0080		0.5479	0.7014			0.2487	0.3903	
1006														
1007					μ stats	σ stats		stats on 97.5%ile				stats on 2.5%ile		
1008				mean	0.387	0.0071		0.5206				0.2534		
1009				max	0.401	0.0168		0.7014				0.3903		
1010				min	0.374	0.0000		0.3828				0.0693		

III.C Model Predictions for Validation Comparisons

The variability of the material strength property (modulus E) to be used in the predictions of beam population deflection variability under loading Po is given in "discrete-direct" form in Table 2 as determined from calibration to four sample beams of similar material and applied loading but different dimensions as summarized in Section II. Note that the beam loading in the calibration experiments need not be the same as in the validation experiments. This is an artifact of the challenge problem [37] to allow some particular investigations and comparisons to be made. Note also that the temperature of the calibration beams is Tcal = 20C whereas the temperature of the validation beams is Tval = 60C and that the model does not have a mechanism to account for temperature effects on material strength.

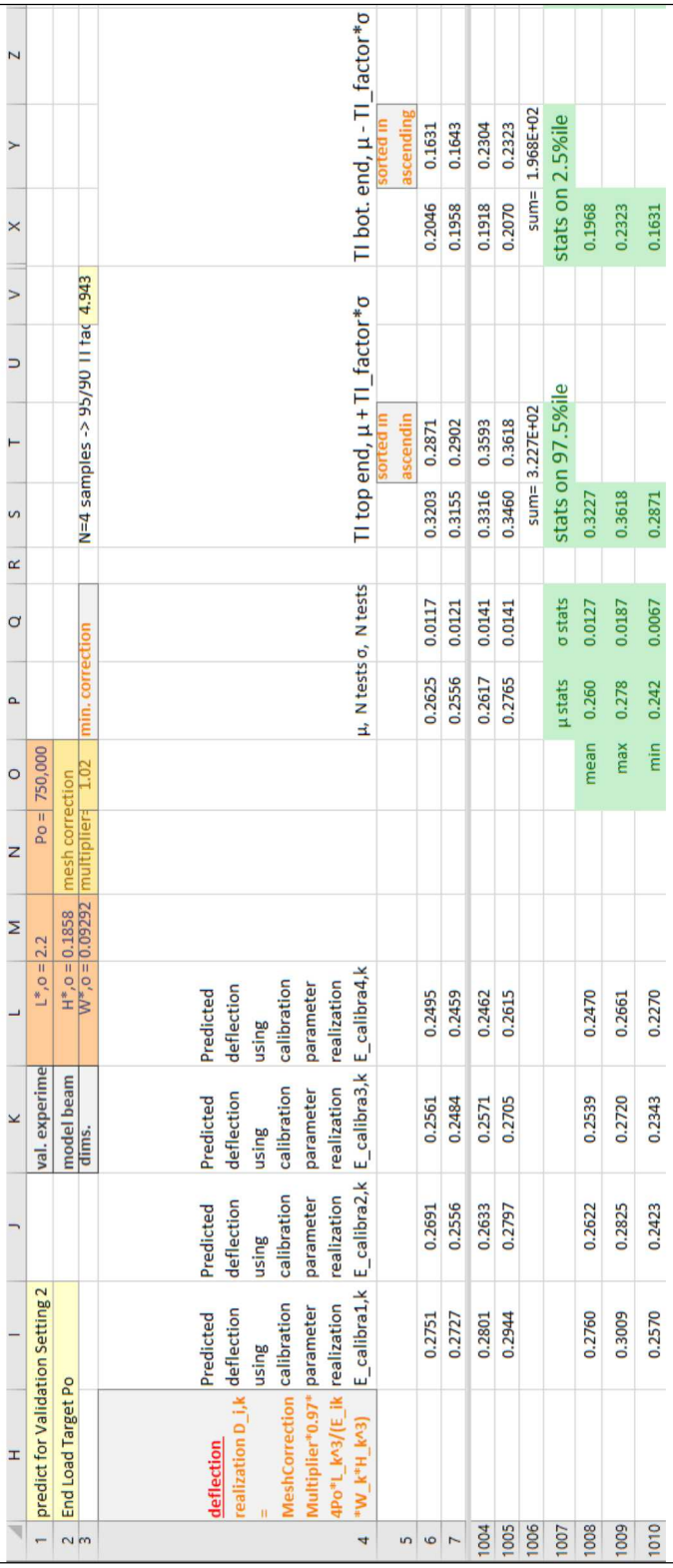
The 1K data rows in Table 2 each contain four plausible sample values for the stochastically varying material strength property E, with which to estimate or bound the material variability over the whole population of beams. For each data row in the table, the four plausible random variations E_i (i = 1 to 4) are used in the prediction model Eqn. 8 with a discretization related correction factor within the uncertainty range in Eqn. 9 to predict deflection under the reference conditions stated in the bullets at the start of section III.B. Four corresponding deflection predictions are

presented per data row in Figure 6, in columns I – L. The results in this figure are for the minimum value 1.02 of the discretization correction factor, per cell O3.

For each row's set of four predicted plausible beams deflections in Figure 6, a mean $\tilde{\mu}$ and standard deviation $\tilde{\sigma}$ are calculated. These are given in columns P and Q. For a 95/90 TI and four samples, f=4.943 is the value of the factor that multiplies the calculated standard deviation $\tilde{\sigma}$. This value of f comes from a look-up table in [41] reproduced from [42].

The top and bottom ends of 95/90 TIs created for each row of deflections are listed in columns S and X in Figure 6. For the 1000 TIs created, the uncertainty range of the top ends of the TIs (the range of 1000 possibilities below which the 97.5 percentile of deflection lies with ~90% confidence) is given by the max and min values in cells S1009 and S1010. Cells X1009 and X1010 give the uncertainty range for the bottom ends of the 1000 TIs (the range of possibilities above which the 2.5 percentile of deflection lies with ~90% confidence). These min-to-max ranges of the conservatively estimated 2.5 and 97.5 percentiles are plotted at right in Figure 7 with the designation 'Mesh-UQ Lower Extreme'. Analogous results obtained with Eqn. 9's maximum value 1.05 of the discretization correction factor are also plotted.

Figure 6. Predicted deflection possibilities from four calibrations to test beams of same material and reference load Po as in validation tests, but somewhat different dimensions and significantly different temperature. Associated 90% confident TI realizations are presented for predicted central 95% range of deflection in an asymptotically large population of validation beams under the reference load Po.



	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	X	Y	Z
1	predict for Validation Setting 2			val. experime	L*,o =	2.2	Po =	750,000										
2	End Load Target Po			model beam	H*,o =	0.1858	mesh correction											
3				dims.	W*,o =	0.09292	multiplier=	1.02	min. correction			N=4 samples -> 95/90 TI fac			4.943			
4	deflection realization D_j,k = MeshCorrection Multiplier*0.97* 4Po*L_k^3/(E_ik *W_k*H_k^3)	Predicted deflection using calibration parameter realization E_calibra1,k	Predicted deflection using calibration parameter realization E_calibra2,k	Predicted deflection using calibration parameter realization E_calibra3,k	Predicted deflection using calibration parameter realization E_calibra4,k				μ, N tests	σ, N tests		TI top end, μ + TI_factor*σ				TI bot. end, μ - TI_factor*σ		
5													sorted in ascendin				sorted in ascending	
6		0.2751	0.2691	0.2561	0.2495				0.2625	0.0117		0.3203	0.2871			0.2046	0.1631	
7		0.2727	0.2556	0.2484	0.2459				0.2556	0.0121		0.3155	0.2902			0.1958	0.1643	
1004		0.2801	0.2633	0.2571	0.2462				0.2617	0.0141		0.3316	0.3593			0.1918	0.2304	
1005		0.2944	0.2797	0.2705	0.2615				0.2765	0.0141		0.3460	0.3618			0.2070	0.2323	
1006													sum= 3.227E+02				sum= 1.968E+02	
1007									μ stats	σ stats		stats on 97.5%ile				stats on 2.5%ile		
1008		0.2760	0.2622	0.2539	0.2470			mean	0.260	0.0127		0.3227				0.1968		
1009		0.3009	0.2825	0.2720	0.2661			max	0.278	0.0187		0.3618				0.2323		
1010		0.2570	0.2423	0.2343	0.2270			min	0.242	0.0067		0.2871				0.1631		

III.D. Validation Comparisons and Interpretation of Experimental and Simulated Deflections

Figure 7 shows some example results comparing the 2.5 and 97.5 percentiles of experimental and simulated deflections. The predicted deflection percentiles are generally to the low side of the corresponding experimental percentiles. This is attributed to the experimental beams experiencing material softening and strength loss (thus higher deflections for a given load) at the higher validation temperatures (60C) than what the prediction model accounts for with the calibrated material properties from the beam tests at 20C. There is also a much larger potential range between the experimental 2.5 and 97.5 percentiles than between the predicted percentiles.

Within the present uncertainties, the experimental population of deflections could have significantly higher or lower variance than the predicted population. That is, predicted variability as signified by the central 95% of response between the 2.5 and 97.5 percentiles does not bound the potential extent of central 95% variability of experimental deflections conservatively estimated from the experimental sample data. In this way, the validation assessment does not establish that the model is a conservative predictor; it could significantly under-predict the true variability. Thus, it may not be a suitable tool for design and safety analysis purposes. Such purposes would usually not be well served by a model that significantly under-predicts system and response variability. This finding stands for either extreme of mesh-related prediction uncertainty treatment.

For the 97.5 percentile of response, the experimental uncertainty bar (0.383 to 0.701) constructed from the experimental sample data and its uncertainty comes with ~90% confidence that the true experimental 97.5 percentile of deflections does not lie above the top of the uncertainty bar. The model-predicted uncertainty bars in the figure are calibration projections of nominal 90% confidence that the experimental 97.5 percentile of response will not lie above the predicted uncertainty bars. But the experimental uncertainty bar lies above the predicted uncertainty bars, so the predictions are not demonstrably conservative. The experimental uncertainty bar establishes that a 90% confidence limit on the true 97.5 percentile of deflection can be as high as 0.701. The predicted uncertainty bars only project upper values about half as high: 0.362 and 0.372 for the lower and upper extremes of potential mesh related prediction error. Thus, the model is not established to be a conservative predictor for the larger potential deflections in the beam population that this upper percentile of response (97.5) coincides with. Since design and safety related performance targets for the beams would most likely be assessed against the larger potential deflections in the population, the validation-assessed model as it presently stands may not be a suitable prediction tool for design and safety analysis purposes. Indeed, the true 97.5 percentile is plotted in the figure for comparison and has a value of ~0.41 and is higher than the tops of the prediction uncertainty bars.

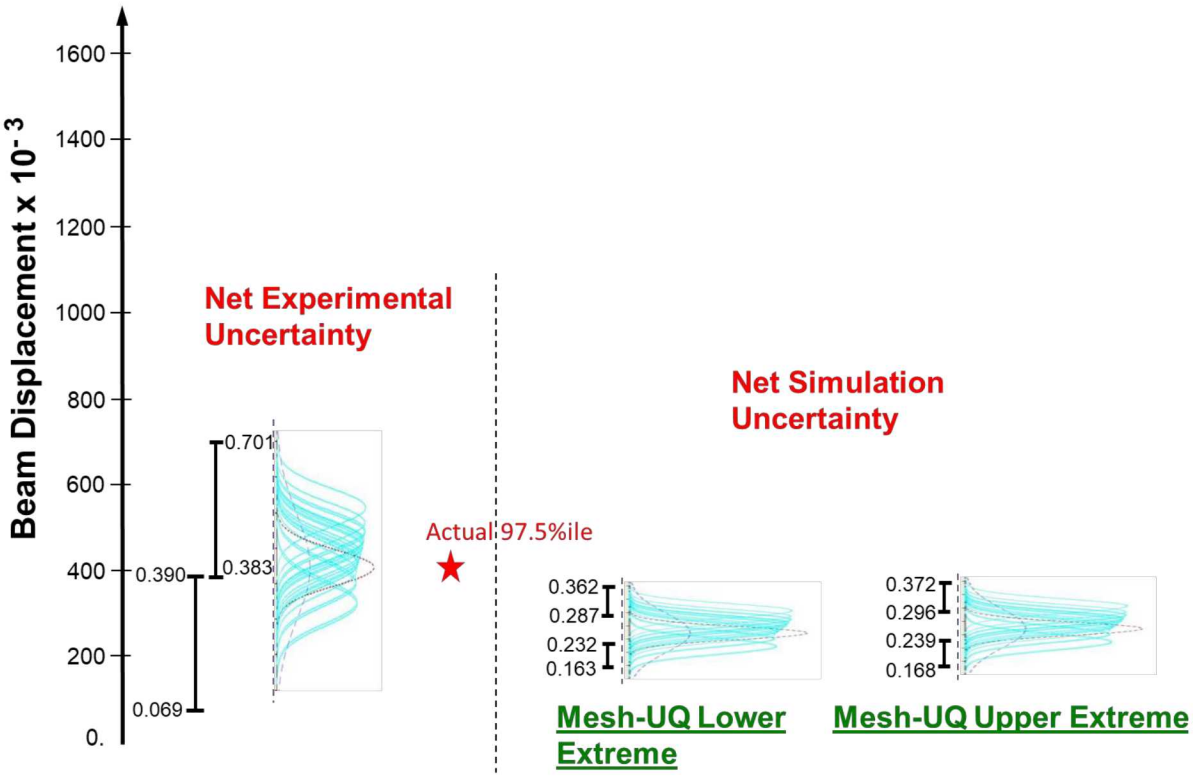


Figure 7. Model validation comparisons of similarly conservatively estimated 2.5 and 97.5 percentiles of experimental and simulated deflections. An uncertainty range exists for each percentile value because of the epistemic uncertainty sources in the validation assessment.

For the 2.5 percentile of response, the experimental uncertainty bar (0.069 to 0.390) indicates with ~90% confidence that the true experimental 2.5 percentile of deflections does not lie below the bottom of uncertainty bar. The bottom of the model-predicted uncertainty bars with either extreme of mesh-related prediction uncertainty treatment do not lie below the bottom of experimental 2.5 percentile uncertainty bar. Therefore, the model predictions do not necessarily bound from below. However, is may not be useful to establish a lower-bound for a deflection type quantity, so there may be little significance to this validation finding.

IV. Predictor-Corrector Extrapolative Prediction Methodology and Results

The Predictor-Corrector extrapolation approach uses a dual model prediction approach as mentioned in the Introduction and illustrated in Figure 8. The methodology is explained and demonstrated below.

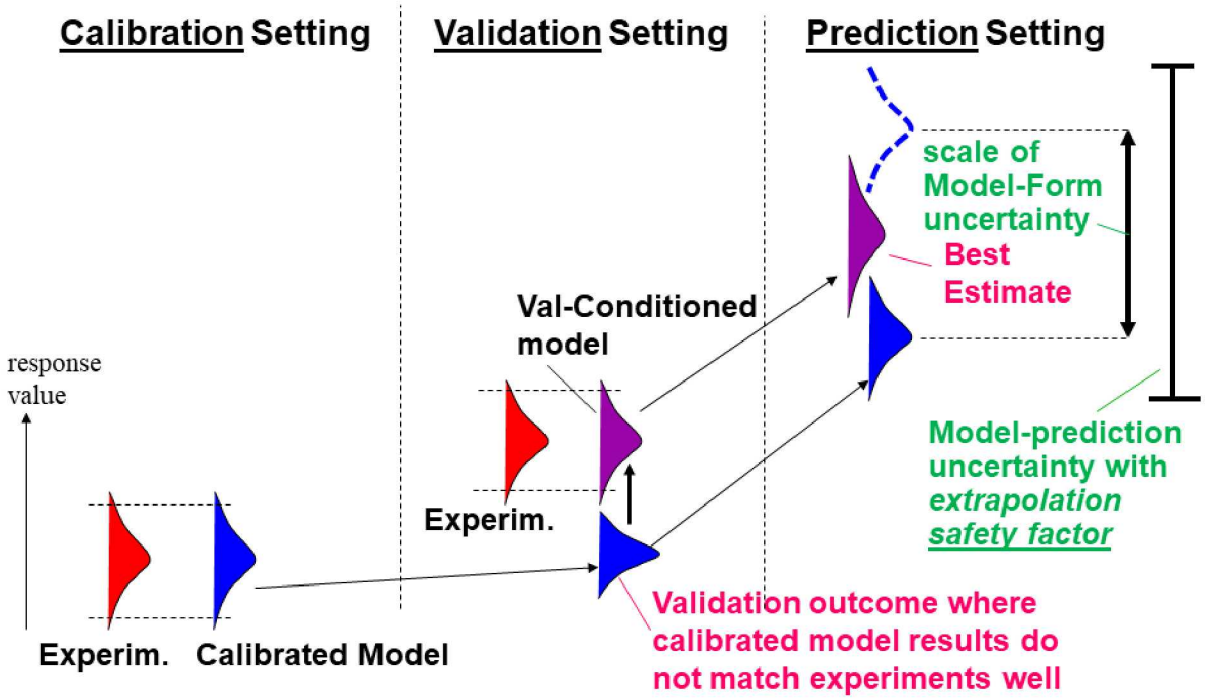


Figure 8. Predictor-Corrector extrapolation paradigm with extrapolation-scaled UQ and Safety Factor.

Figure 9 shows results of the approach applied to predict the 97.5 percentile of the deflection distribution for beams with the service dimensions validated at (Table 1) and the same end-load Po=750,000 but 20C higher temperature. The figure also shows the actual material-strength weakening function vs. temperature for the physical beams and plots the large temperature effect on the centering and spread of deflection behavior in an asymptotically large population of beams. The physical nonlinearities and magnitude of effects on the particular QOIs in this problem are substantial.

The Predictor-Corrector (PC) extrapolation method starts from a Real-Space validation comparison as shown in Figure 7. A validation finding of relevance was that the experimental uncertainty bar establishes a ~90% confidence limit on the true 97.5 percentile of response that can be as high as 0.701. The predicted uncertainty bars only project upper values about half as high: 0.362 and 0.372 for the lower and upper extremes of potential mesh related prediction error. In general it would be desirable to enact a bias correction that corresponds to a mesh-converged prediction so that the bias correction has no mesh dependency. Then mesh effects are ostensibly lessened when the bias correction is extrapolated to other conditions. However, solution convergence is rarely achievable in practice, so a prudent approach is to make the bias correction large enough to be conservative with respect to mesh effects. Here this means enacting a bias-correction for the mesh-UQ low-extreme case. Then the correction will be sufficiently large no matter what the converged solution result is.

An ideal correction would translate upwards the top end of the said prediction uncertainty bar at 0.362 in Figure 9 to be even with the top end of the experimental uncertainty bar at 0.701 in the figure. This would be a target translation of 0.339 deflection units. The top end of the said prediction uncertainty bar is set by the 95/90 TI from the 127th row of realizations in the spreadsheet in Figure 6. This TI is based on the variability in the four deflection predictions that incorporate the four realizations of material property strength parameter E_i (i = 1 to 4) in 127th row of realizations in the calibration data spreadsheet in Figure 2. The four values E_i in this row are all shifted by the same trial value ΔE, where the value ΔE is iterated to find the value that brings the top of the corresponding 95/90 TI[5] to the top end of the experimental uncertainty bar (0.701) or just above this. A shift ΔE= -7.8e-10 is found that brings the top of the said TI to 0.707 as indicated in the figure. This deflection shift is 0.345 as indicated in the figure.

Thus, a model bias correction is enacted that yields an upper-bound on a ~90% confidence limit for the 97.5 percentile of response that is just slightly more conservative than the experimentally inferred upper-bound ~90% confidence limit at the validation conditions. The model has been tuned for this QOI agreement and nothing else explicitly, so agreement with other statistical measures of response such as the central 95% of response may be not good even at the validation conditions—much less the extrapolative prediction conditions. This is particularly true for other statistical QOIs such as central 95% of response that were substantially mis-predicted with the original model in the validation assessment. More research must be conducted on the prospect of enacting one or more model adjustments to satisfactorily bias-correct multiple QOIs for extrapolative prediction. These could include multiple statistical QOIs such as 97.5 percentile of response *and* central 95% of response, potentially for multiple physical outputs such as displacement at multiple points on the beam and/or displacement *and* strain quantities (that are measurable for the validation comparisons).

Because the QOI-specific correction applied here not baing a general fix to the physics model, extrapolation of the bias correction is not expected to be perfect, either. Therefore, PC extrapolation employs both the bias-corrected model and a reasonable "model basis" lower-limit to the correction, in order to estimate upper and lower bounds to the QOI correction at the extrapolation conditions. As indicated in Figure 8, an optional factor-of-safety on the uncertainty estimate for the predicted QOI may be defined (by judgment from subject matter experts) that also scales with the extrapolation. However, this is not done in the current demonstration.

In the current problem the corrective shift ΔE= -7.8e-10 applied to the selected 20C calibration data E_i$_{127}$ (i = 1 to 4) achieved an appropriate predictive correction for the 60C validation conditions for the said QOI. The extrapolation prediction conditions are at 20C higher, or 80C. Thus, the corrective shift ΔE= -7.8e-10 is ratio-ed to the new prediction conditions by multiplying ΔE by 1.5. (The original shift ΔE for a 40C temperature change from 20C to 60C becomes 1.5ΔE for an additional temperature change of 20C to get to 80C. Then the total temperature change from calibration conditions is 60C, which is 50% larger than the original 40C change from the calibration conditions to the validation conditions.) Using 1.5ΔE in the model, the predicted QOI (upper-bound on a ~90% confidence limit for the 97.5 percentile of

[5] Here the model inputs are the reference values from the validation analysis (i.e., the two bullets just below the start of section III.B). Normalizing the experimental data to a single reference set of experimental inputs facilitates the prediction bias correction procedure.

response) takes a value of 1.278 as indicated in the figure. This is a deflection shift of 0.571 as noted in the figure. This shift for a 20C increment from validation to extrapolation conditions is much larger than the 0.345 deflection shift corresponding to the original 1.0ΔE correction for a 40C increment in temperature. This illustrates a nonlinear correction to the model prediction even though the corrective variable itself has a linear adjustment function based on the two input scenario points at calibration the validation conditions.

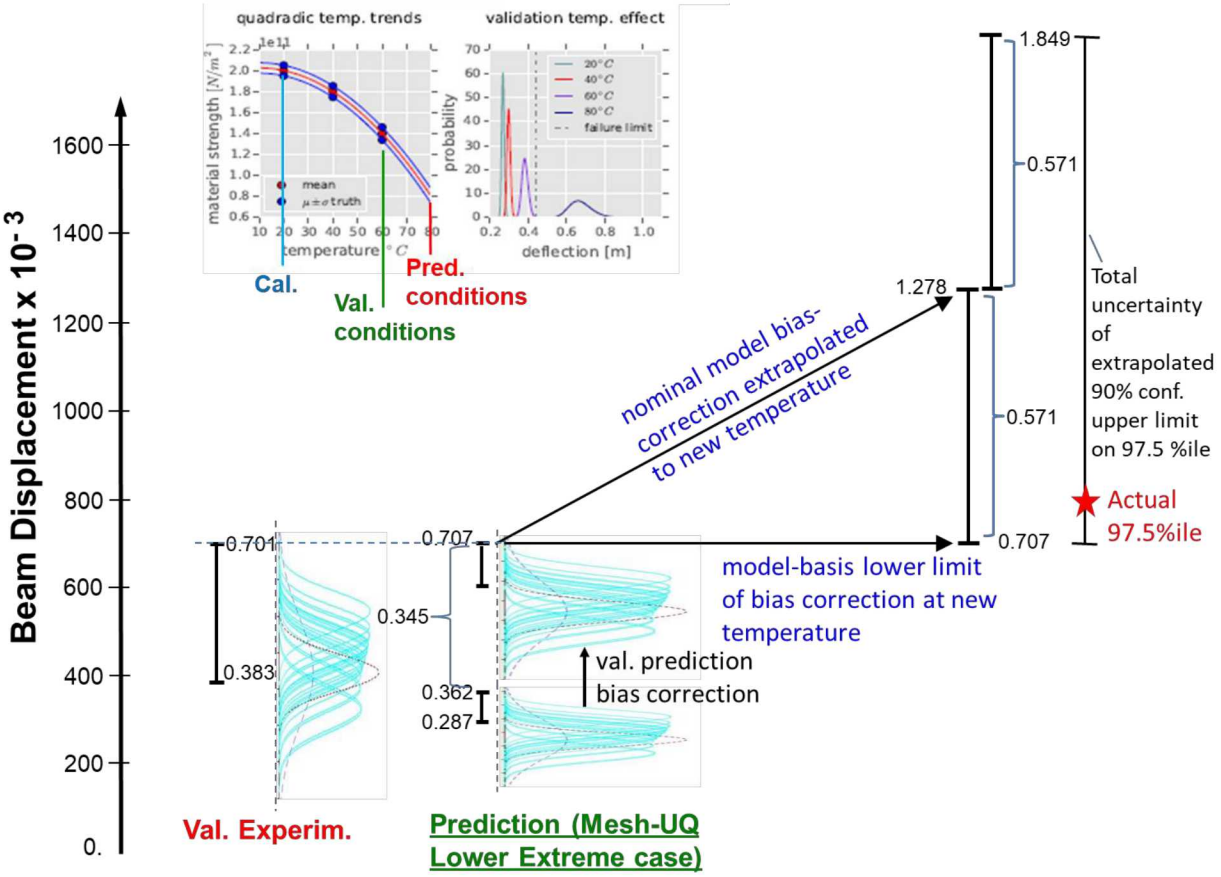


Figure 9. Results of Predictor-Corrector extrapolation in temperature (QOI is 97.5 percentile of deflection response), with no added factor of safety for this extrapolation.

If a submodel for temperature effect existed in the beam model, then the multiplier 1.5 would not be used because the submodel would be an explicit function of temperature so would pick up that the extrapolation involves a temperature change. The model would react to the temperature change, even if the submodel is not perfect. The current model has no such submodel that would be aware of or react to temperature change, so the need to actively set the multiplier (to 1.5 in this case) to account for the known temperature change at the prediction conditions.

The original mapped correction ΔE is surmised to be relatively robust if the extrapolation does not involve a temperature change, but instead involves different geometry or loading at the same temperature conditions. An initial indication of such robustness for a related beam problem is

found in [3] and [4]. Additive or multiplicative corrections to predicted response values under different geometry and loadings do not appear to be as robust in this way.

A reasonable lower-limit to the extrapolated correction is needed to size the extrapolation uncertainty in the PC method. Here the validation-sized 1.0ΔE corrective model is used. It does not have the 1.5ΔE correction for temperature extrapolation, so will estimate significantly smaller deflections at the extrapolation conditions and is a reasonable model-basis lower-bound limit to an appropriate correction at the extrapolation conditions. The predicted QOI remains unchanged (0.707) at the extrapolation conditions as shown in the figure.

A projected upper limit to the extrapolated correction is here simply taken as an upside-reflected version of the downside uncertainty just described (see the figure). This presents a somewhat arbitrary but seemingly reasonable scale of uncertainty for an upper limit to the extrapolated bias correction.

The farthest-right uncertainty bar in the figure gives the total uncertainty estimate for extrapolative prediction of an upper-bound on a 90% confidence limit on the 97.5 percentile of response at 80C. Even though everything but temperature is the same as in the validation setting, the extrapolation uncertainty bar is relatively large. If a somewhat inaccurate but reasonable submodel for temperature effect existed in the beam model, then the extrapolation uncertainty bar would be smaller by the construction approach used. It turns out that the current uncertainty bar's upper limit easily bounds the true (beam population's) 97.5 percentile of response (~0.8) indicated by the red star.

Planned future work is to extrapolate in both temperature (parametric extrapolation) and loading configuration (non-parametric extrapolation, e.g. switch to uniform distributed load in the 80C prediction conditions). A strength of the PC method is that it applies to extrapolation for both parametric and non-parametric changes to prediction conditions. It will also be illuminating to explore various imperfect submodels for temperature effects on beam material strength.

As discussed and illustrated here, model prediction bias-correction and extrapolation thereof is a difficult and challenging prospect. This is an area ripe for future research with the proposed PC method and others in the literature.

V. Closing Remarks

A novel set of coordinated methods comprising a systems approach to experimental data UQ and model calibration, validation, and extrapolation has been introduced and illustrated. The DD calibration and propagation, RS validation, PC extrapolation, and sparse-data UQ methods were developed on industrial-scale application problems, where practical relevance and feasibility have been demonstrated. Their quantitative performance in terms of robustness, reliability, and efficiency in being "conservative but not overly conservative" is being assessed with specialized performance metrics [43] over many random trials on various test problems.

The Sandia Cantilever Beam End-to-End UQ problem is a one such test problem with scalar inputs and outputs. It has many UQ features reflective of real experimental data and model calibration, validation, and extrapolative prediction problems. It provides a useful test bed for evaluating the applicability and performance of proposed end-to-end UQ methodologies and frameworks. A full characterization of the DD, RS, PC and sparse-data methods' performance on the beam problem will require testing and evaluation in many trials involving random draws of the slightly varying calibration and validation beams from their populations, and how these and the draws from experimental load control variability in the tests, and random and systematic measurement errors on test inputs and outputs, affect the performance success of the methods.

The DD, RS, PC, and sparse-data methods are relatively simple and straightforward as applied to the Cantilever Beam problem. All have been applied with standard functions in EXCEL spreadsheets. Initial examination and application of more elaborate methods tried from the literature have not shown better success or promise under the conditions of the Beam problem. In fact, they have exhibited debilitating difficulties on some aspects of the beam problem in testing so far. Nonetheless, a tremendous amount of further testing is needed with other methods and test problems so that even better hybrid techniques can be developed and appropriately characterized in terms of expected performance on real problems.

References

[1] [25] Romero, V.J., "A Paradigm of Model Validation and Validated Models for Best-Estimate-Plus-Uncertainty Predictions in Systems Engineering," paper 2007-01-1746 for Soc. Automotive Engineers 2007 World Congress April 16–20, Detroit, Michigan.

[2] Romero, V.J., "Validated Model? Not So Fast. The Need for Model 'Conditioning' as an Essential Addendum to Model Validation," paper AIAA-2007-1953, 9th Non-Deterministic Approaches Conference, Honolulu, HI, April 23-26, 2007.

[3] Romero, V.J., "Type X and Y Errors and Data & Model Conditioning for Systematic Uncertainty in Model Calibration, Validation, and Extrapolation," SAE paper 2008-01-1368 for Society of Automotive Engineers 2008 World Congress, April 14-17, 2008, Detroit, MI.

[4] Romero, V.J., "Comparison of Several Model Validation Conceptions against a "Real Space" End-to-End Approach," *Soc. Automotive Engrs. Intn'l. J. of Materials and Manufacturing*, article 2011-01-0238, vol.4 (no 1), 2011, pp. 396-420, doi:10.4271/2011-01-0238.

[5] Romero, V.J., "Elements of a Pragmatic Approach for dealing with Bias and Uncertainty in Experiments through Predictions: •Data and Model Conditioning; •"Real Space" Model Validation and Conditioning; •Hierarchical Modeling and Extrapolative Prediction," Sandia National Laboratories report SAND2011-7342, Nov. 2011.

[6] Romero, V.J., "Uncertainty Quantification and Sensitivity Analysis—Some Fundamental Concepts, Terminology, Definitions, and Relationships," Chapter 5 of Joint Army/Navy/NASA/Air Force (JANNAF) e-book: *Simulation Credibility—Advances in Verification, Validation, and Uncertainty Quantification*, U. Mehta (Ed.), D. Eklund, V. Romero, J. Pearce, N. Keim, document NASA/TP-2016-219422 and JANNAF/GL-2016-0001, Nov. 2016.

[7] Romero, V., "Discrete-Direct Model Calibration and Propagation Approach addressing Sparse Replicate Tests and Material, Geometric, and Measurement Uncertainties," Soc. Auto. Engrs. 2018 World Congress (WCX18) paper 2018-01-1101 (doi:10.4271/2018-01-1101), April 10-12, Detroit, MI.

[8]. Department of Defense Instruction 5000.61: *Modeling and Simulation Verification, Validation, & Accreditation (VV&A)*, Defense Modeling and Simulation Office, Office of the Director of Defense Research and Engineering, dated April 29, 1996.

[9] American Institute of Aeronautics and Astronautics (AIAA) *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, AIAA-G-077-1998, Reston, VA.

[10] Department of Energy – Defense Programs, Accelerated Strategic Computing Initiative Program Plan, 2000, DOE/DP-99-000010592American Society of Mechanical Engineers (ASME), *Guide for Verification and Validation in Computational Solid Mechanics*, ASME V&V 10-2006.

[11] NASA, *Technical Standard for Models and Simulations*, report NASA-STD-7009, July 11, 2008.

[12] American Society of Mechanical Engineers document V&V 20-2009, *Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer*.

[13] American Society of Mechanical Engineers document V&V 10.1-2012, *An Illustration of the Concepts of Verification and Validation in Computational Solid Mechanics.*

[14] Coleman, H. W., and Stern, F., "Uncertainties in CFD Code Validation," *Journal of Fluids Engineering*, Dec. 1997, vol. 119, pp. 795-803.

[15] Roache, P. J., *Verification and Validation in Computational Science and Engineering*, Hermosa Publishing, 1998.

[16] Anderson, M. G., and Bates, P. D., editors, *Model Validation—Perspectives in Hydrological Science*, Wiley, 2001.

[17] Trucano, T. G., M. Pilch, and W. L. Oberkampf, "General Concepts for Experimental Validation of ASCI Code Applications," Sandia National Laboratories Report SAND2002-0341, March 2002.

[18] Chen, W., Y. Xiong, K-L. Tsui, S. Wang, "Some Metrics and a Bayesian Procedure for Validating Predictive Models in Engineering Design," ASME Design Technical Conference, Design Automation Conf., Philadelphia, PA, Sept. 10-13, 2006.

[19] Bayarri, M., Berger, J., Rui, P., Sacks, J., Cafeo, J., Cavendish, J., Lin, C-H., and Tu, J. "A Frame-work for Validation of Computer Models," *Technometrics*, May 2007, Vol. 49, no. 2, pp. 138 – 154.

[20] Babuska, I., F. Nobile, R. Tempone, "Reliability of Computational Science," *Num. Methods in Partial Differential Equations*, (2007) Vol. 23, pp. 753-784.

[21] Rebba, R, and S. Mahadevan, "Computational methods for model reliability assessment," *Reliability Engineering and System Safety*, Vol. 93 (2008), pp. 1197-1207.

[22] Ferson, S., W. L. Oberkampf, L. Ginzburg, "Model Validation and Predictive Capability for the Thermal Challenge Problem," *Comput. Methods in Applied Mechanics and Engrng*., Vol. 197, 2009, pp. 2408–2430.

[23] Liu, Y., W. Chen, P. Arendt, H-Z Huang, "Towards a Better Understanding of Model Validation Metrics," paper AIAA-2010-9240, 13th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Sept. 12-15, 2010, Fort Worth, TX.

[24] Romero, V.J., A. Luketa, M. Sherman, "Application of a Versatile "Real Space" Validation Methodology to a Fire Model" *AIAA J. of Thermophysics and Heat Transfer*, Vol. 24, No. 4, Oct. – Dec. 2010, pp. 730-744.

[25] Oberkampf, W.L., and Roy, C.J., *Verification and Validation in Scientific Computing*, Cambridge University Press, 2010.

[26] Roy, C.J., and W.L. Oberkampf, "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing." *Computer Methods in Applied Mechanics and Engineering*, 200.25-28 (2011): 2131-2144.

[27] National Research Council, *Assessing the reliability of complex models: The mathematical and statistical foundations of verification, validation, and uncertainty quantification*, 2012.

[28] Romero, V., F. Dempsey, B. Antoun, G. Wellman, M. Sherman, "Handling Bias and Uncertainty in Model Verification and Validation associated with Heated Pipes Pressurized to Failure," paper AIAA2014-0811, 16th AIAA Non-Deterministic Approaches Conference, AIAA SciTech 2014, Jan 13-16, 2014, National Harbor, MD.

[29] Mullins, J., and S. Mahadevan, "Bayesian uncertainty integration for model calibration, validation, and prediction." *Journal of Verification, Validation and Uncertainty Quantification* 1.1 (2016): 011006.

[30] Romero, V., A. Black, N. Breivik, G. Orient, J. Suo-Anttila, B. Antoun, A. Dodd, "Advanced UQ and V&V Procedures applied to Thermal-Mechanical Response and Weld Failure in Heated Pressurizing Canisters," Sandia National Laboratories document SAND2015-3005C presented at Soc. Auto. Engrs. 2015 World Congress, April 21-23, 2015, Detroit, MI.

[31] Romero, V.J., Shelton, J.W., and Sherman, M.P., "Modeling Boundary Conditions and Thermocouple Response in a Thermal Experiment," 2006 ASME Int'l. Mechanical Engineering Congress and Exposition, Nov. 5-10, 2006, Chicago, IL.

[32] McFarland, J., S. Mahadevan, V. Romero, L. Swiler, "Calibration and Uncertainty Analysis for Expensive Computer Simulations with Multivariate Output," *AIAA Journal*, vol. 46, no. 5, May 2008, pp. 1253-65.

[33] Jamison, R., V. Romero, M. Stavig, T. Buchheit, C. Newton, "Experimental Data Uncertainty, Calibration, and Validation of a Viscoelastic Potential Energy Clock Model for Inorganic Sealing Glasses," Sandia National Laboratories document SAND2016-4635C presented at ASME Verification & Validation Symposium, May 18-20, 2016, Las Vegas, NV.

[34] Romero, V.J., M.P. Sherman, J.F. Dempsey, J.D. Johnson, L.R. Edwards, K.C. Chen, R.V. Baron, C.F. King, "Development and Validation of a Component Failure Model," paper AIAA-2005-2141, 45th AIAA/ASME/ ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, April 18-21, 2005, Austin, TX.

[35] Romero, V., B. Schroeder, J.F. Dempsey, N. Breivik, G. Orient, B. Antoun, J.R. Lewis, J. Winokur, "Simple Effective Conservative Treatment of Uncertainty from Sparse Samples of Random Variables and Functions," ASCE-ASME Journal of Uncertainty and Risk in Engineering Systems: Part B. Mechanical Engineering, DOI 10.1115/1.4039558, April 30, 2018 (online), Dec. 2018, vol. 4 (print) pp. 041006-1 – 041006-17.

[36] Romero, V., M. Bonney, B. Schroeder, V.G. Weirs, "Evaluation of a Class of Simple and Effective Uncertainty Methods for Sparse Samples of Random Variables and Functions," Sandia National Laboratories report SAND2017-12349, Nov. 2017.

[37] Romero, V., B. Schroeder, M. Glickman, "Cantilever Beam End-to-End UQ Test Problem: Handling Experimental and Simulation Uncertainties in Model Calibration, Model Validation, Extrapolative Prediction, and Risk Assessment," Sandia National Laboratories document SAND2017-4689 O, version BeamTestProblem-33.docx, 2017.

[38] Byars, E.F., R.D. Snyder, H.L. Plants, "*Engineering Mechanics of Deformable Bodies*," 4th ed. (1983), Harper & Row, New York.

[39] Romero, V.J., "Model-Discretization Sizing and Calculation Verification for Multipoint Simulations over Large Parameter Spaces," paper AIAA2007-1953, 9th AIAA Non-Deterministic Methods Conference, April 23 - 26, 2007, Honolulu, HI.

[40] Kline, S., and McClintock, F., "Describing Uncertainties in Single-Sample Experiments," *Mechanical Engineering*, Vol. 75, 1953. pp. 3-8.

[41] Coleman, H.W., and Steele, Jr., W.G., *Experimentation and Uncertainty Analysis for Engineers*, 2nd Edition, John Wiley & Sons, New York, NY, 1999.

[42] Montgomery, D.C., and Runger, G.C., *Applied Statistics and Probability for Engineers*, Wiley & Sons, 1994.

[43] Romero, V., B. Schroeder, M. Glickman, J. Winokur, "Cantilever Beam End-to-End UQ Test Problem and Evaluation Criteria for UQ Methods Performance Assessment," Sandia National Laboratories document SAND2017-4592 C, presented at ASME Verification & Validation Symposium, May 3-5, 2017, Las Vegas, NV.