2152
LA-UR- 03-XXXX

Approved for public release; distribution is unlimited.

Title: ESTIMATING THE ERROR IN SIMULATION PREDICTION OVER THE DESIGN SPACE (U)

Author(s): Rachel Shinn, Embry-Riddle Aeronautical University

François M. Hemez, LANL, ESA-WR

Scott W. Doebling, LANL, ESA-WR

Submitted to: 44th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics and Materials Conference, Norfolk, Virginia, April 7-10, 2003.

42

Los Alamos
NATIONAL LABORATORY

Proceedings of the 44th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conference, Norfolk, Virginia, April 7-10, 2003.

ESTIMATING THE ERROR IN SIMULATION PREDICTION OVER THE DESIGN SPACE

Rachel Shinn[1]
Embry-Riddle Aeronautical University
Aerospace Engineering, College of Engineering
3200 Willow Creek Road, Prescott, Arizona 86301

François M. Hemez[2], **Scott W. Doebling**[3]
Los Alamos National Laboratory
Engineering Sciences and Applications, ESA-WR
Mail Stop P946, Los Alamos, New Mexico 87545

ABSTRACT

This study addresses the assessment of accuracy of simulation predictions. A procedure is developed to validate a simple non-linear model defined to capture the hardening behavior of a foam material subjected to a short-duration transient impact. Validation means that the predictive accuracy of the model must be established, not just in the vicinity of a single testing condition, but for all settings or configurations of the system. The notion of validation domain is introduced to designate the design region where the model's predictive accuracy is appropriate for the application of interest. Techniques brought to bear to assess the model's predictive accuracy include test-analysis correlation, calibration, bootstrapping and sampling for uncertainty propagation and metamodeling. The model's predictive accuracy is established by training a metamodel of prediction error. The prediction error is not assumed to be systematic. Instead, it depends on which configuration of the system is analyzed. Finally, the prediction error's confidence bounds are estimated by propagating the uncertainty associated with specific modeling assumptions. This publication has been approved for unlimited, public release on March ??, 2003. LA-UR-03-????. **Unclassified.**

I. INTRODUCTION

The objective of this paper is to illustrate the assessment of accuracy of simulation predictions based on minimal testing. The paper demonstrates the validation of a simple non-linear model defined to capture the hardening behavior of a foam material as it is subjected to a short-duration transient impact.

It is assumed that a numerical model is developed to simulate the response of a structural system and predict specific features of the response. When measurements obtained from physical testing are available, the simulation results can be compared to the test results to attempt to quantify the level of fidelity provided by the model. If the agreement between measurements and predictions is not deemed appropriate, it is common practice in engineering sciences to tune the model to achieve small errors on some given response features.[1] This is generally referred to as tuning, model updating or calibration.

However, these same simulations need to predict the system behavior at points in the design domain or operational space other than those for which the simulations have been tuned. For example, a model of wing flutter might be calibrated to reproduce test data available for several combinations of speeds and angles of attack but calibration does not necessarily provide confidence that the predictions will be accurate away from those combinations. There will certainly be some error when these models are extrapolated to different parameter values within the design or operational space. This paper aims to give a method for assessing the magnitude of these errors within the design space.

The procedure is summarized as follows. First, test-analysis correlation is performed to assess the fidelity of predictions at those settings where impact tests have been performed. Second, the calibration variables of the non-linear model are optimized to improve the model's fidelity to test results. Third, statistical metamodels of prediction accuracy are trained to estimate the model's expected errors throughout the design domain.

[1]Professor of Aerospace Engineering, 928-777-6979 (Phone), shinnr@erau.edu.
[2]Technical staff member, ESA-WR, Sensor Technology Team, 505-663-5204 (Phone), hemez@lanl.gov.
[3]Technical staff member, ESA-WR, Validation Methods Team Leader, 505-667-6950 (Phone), doebling@lanl.gov.

The prediction error is not assumed to be systematic. It varies within the design space to reflect the fact that the model may be more appropriate in some regions than others. In this work variability and uncertainty are propagated using bootstrapping and Monte Carlo sampling.

Finally the assumptions made during modeling are assigned prior probabilities based on the analyst's confidence that they are correct. The prior knowledge is updated using the Bayes theorem of aggregation that accounts for fidelity to test data. Posterior probabilities result from the analysis. They are sampled to estimate the confidence bounds of predictive accuracy. The end result is an estimation of prediction errors anywhere in the design space, together with intervals that bound these errors at a given confidence level. The confidence bounds capture the prediction's uncertainty introduced by the modeling assumptions.

II. TESTING OF A FOAM MATERIAL

The example used to illustrate the assessment of predictive accuracy within a design space is an impact test modeled by a Single Degree of Freedom (SDOF) system. The test hardware consists of a steel cylinder, sandwiching a foam pad with a steel plate, which is the carriage or impact table.[2] Figure 1 illustrates the assembly of steel cylinder and foam pad mounted on the carriage. The steel cylinder and foam pad are held to the carriage through a bolted attachment.

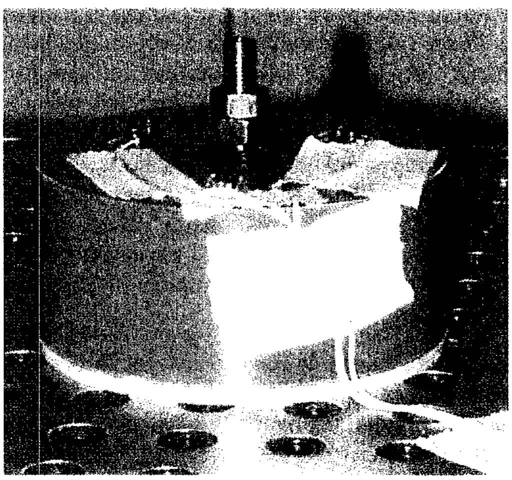

Figure 1. Drop test assembly.

The system is assembled using quarter-inch or half-inch thick foam pads and dropped from a 13-inch or 155-inch height. Typical acceleration measurements measured on top of the carriage and steel cylinder are pictured in Figure 2 for a configuration of the system featuring the quarter-inch thick pad and 13-inch drop height. Each configuration is tested several times to estimate the effect of environmental variability and uncontrolled testing conditions on the response, as can be observed in Figure 2.

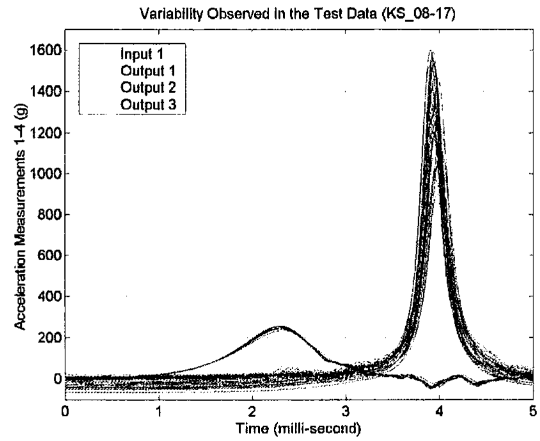

Figure 2. Measured acceleration signals.

Physical testing provides data sets of input and output acceleration signals. Input refers to the acceleration applied to the carriage when it hits the ground. Output refers to the acceleration transmitted through the foam pad and measured on top of the steel cylinder. However physical testing does not provide an explanation of how the foam material behaves, which is the reason why a material model is developed. The model development is overviewed in Section III.

III. SDOF IMPACT MODEL

Even though finite element models have been developed to study the system,[2,3] the analysis reported here involves a simple SDOF model. The reason is because we are primarily interested in demonstrating the concepts of design domain, uncertainty assessment and predictive quality evaluation. This is best achieved with a model for which the modeling complexity does not confound the uncertainty assessment.

The assembly of steel cylinder and foam pad is modeled as a spring-mass system with a moving base, where the mass corresponds to the steel cylinder, the spring/damper corresponds to the combined stiffness of the foam and any additional stiffness of the system, and the moving base is the steel plate on which the foam pad and steel cylinder are mounted. The equation of motion is simply written as:

$$m\ddot{x}(t) + kx(t) + F(t) = m\ddot{x}_{base}(t) \qquad (1)$$

where m denotes the steel cylinder's mass; k denotes the foam pad's linear stiffness; and F represents a non-linear forcing function. The base acceleration applied to the right-hand side of equation (1) represents the input acceleration measured during one of the tests.

Because of the high accelerations applied to the system, it is determined that the spring stiffness must be highly non-linear to show the same amplification effect observed from impact measurements. The linear stiffness contribution of equation (1) is augmented with a non-linear forcing function F defined as:

$$F(t) = k_{nl}(x(t))^p \tag{2}$$

where x denotes the spring displacement. Note that the non-linear stiffness coefficient k_{nl} and exponent p are introduced by our modeling assumption. Other models and solution procedures—such as finite elements—introduce other parameterizations. For obvious reasons the parameters defined by our modeling choices are collectively referred to as calibration variables. Their initial values are left to the analyst's best judgment.

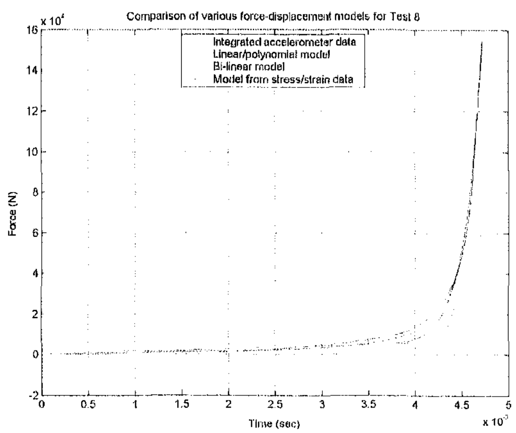

Figure 3. Examples of non-linear stiffness models.

The non-linear model defined in equation (2) is a higher-order polynomial. Some of the other models investigated involve a linear part followed by an exponential or a linear part followed by a second linear part of a different slope. Figure 3 illustrates several non-linear stiffness models.

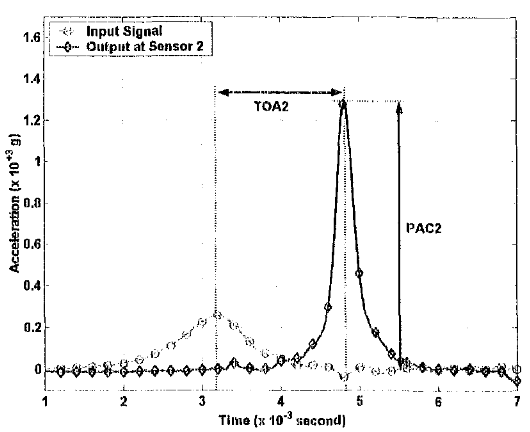

Figure 4. Features of the acceleration response.

Finally it is mentioned that, although the entire time response can be simulated with the SDOF model, only a few features of the response are considered in the subsequent analysis. They are illustrated in Figure 4 and defined as the peak acceleration (denoted by PAC2) and the time it takes the shock wave to travel from the input sensor location to the output sensor location (or time of arrival, denoted by TOA2).

It is determined that the factors that most affect the simulation results are the input acceleration profile (that is, the acceleration signal inputted to the base of the SDOF model) and the stiffness model. The input acceleration profile is modeled by a hyperbolic secant function and it is demonstrated in Section IV that this curve fits the data very well. The unknown parameters $(m;k;k_{nl};p)$ of the equation of motion are calibrated to improve the fidelity of the model, as discussed below.

IV. TEST-ANALYSIS CORRELATION

The study starts by investigating the merits of various models in terms of matching the features PAC2 (peak amplification) and TAC2 (time delay) extracted from test measurements. The linear model obtained by setting $F=0$ in equation (1) does not reproduce the measured features well. The non-linear models capture these characteristics with various degrees of accuracy.

The study discussed in the remainder is restricted to the SDOF model with polynomial non-linearity (2). This choice introduces four calibration variables p:

$$p = \{m; k; k_{nl}; p\}^T \tag{3}$$

To improve the model's predictive accuracy, the previous parameters are calibrated with the objective of minimizing the difference between measurements and predictions. This difference is measured as a weighted L^2 norm of the prediction error with a penalty term that avoids drifting too far away from the nominal model. In the context of Bayesian inference where variability of inputs p and outputs y is represented with Gaussian probability laws,[4] the same cost function is defined as:

$$J(p) = \left(y^{Test} - y(p)\right)^T S_{yy}^{-1} \left(y^{Test} - y(p)\right) + \left(p - p_o\right)^T S_{pp}^{-1} \left(p - p_o\right) \tag{4}$$

where vector y^{Test} collects the mean of output features measured during a series of replicate impact tests and S_{yy} denotes the corresponding covariance matrix. The vector y collects the output features PAC2 and TOA2:

$$y = \{PAC2; TOA2\}^T \tag{5}$$

For a given quadruplet (3) the output acceleration signal is simulated with the equation of motion (1-2), the features (5) are extracted from the response, and the objective (4) is computed. The features y^{Test} used as reference are extracted from the impact tests conducted with a quarter-inch thick foam pad and 13-inch drop height. A constrained, gradient-based optimization solver is wrapped around these steps to calibrate the inputs p that minimize the objective $J(p)$.

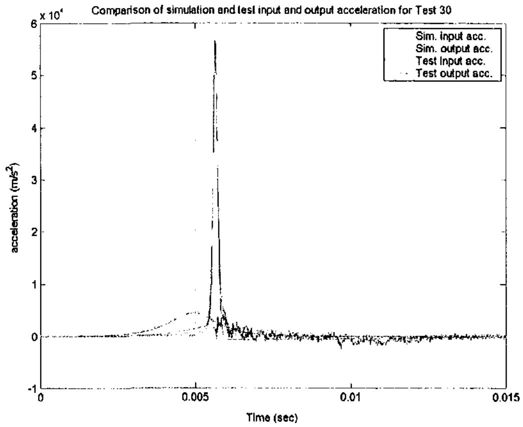

Figure 5. Test-analysis correlation of a calibrated non-linear model $(m;k;k_{nl};p)$.

Figure 5 illustrates the prediction of a non-linear model $(m;k;k_{nl};p)$. The model not only reproduces with good accuracy the two features PAC2 and TOA2, but it also captures other characteristics of the response of the system. Although these models can be made to fit the test data almost perfectly, the question arises of how to extrapolate the model to other configurations of foam thickness and drop height values.

V. THE DOMAIN OF VALIDATION

The test-analysis correlation and model updating discussed previously implicitly assume that a particular configuration of the system is investigated. The results presented in Section IV are obtained when the drop height d and foam thickness h are kept constant ($d=13''$ and $h=\frac{1}{4}''$). The non-linear model that results from calibration may be appropriate to predict the material behavior at these settings, but the question of its adequacy for other regimes remains open.

The parameters k_{nl} and p that represent the foam material can be calibrated for each configuration tested. This typically results in different values of k_{nl} and p for each drop height, which some may argue is not consistent because the foam material does not change between tests. On the other hand refining a numerical model to provide acceptable accuracy for a single configuration or regime is not an exercise of great value. The activity of modeling is justified whenever models can be applied to a wide variety of situations, especially those that cannot be tested for economic, safety or practical reasons. Instead of attempting to calibrate the model to "perfection," it appears more useful to estimate the model's predictive accuracy for all configurations of the system. Introducing the notion of validation domain helps with this assessment.

The design domain (or validation domain) refers to the space of all potential settings, regimes or environments at which the system may be required to operate. Our opinion is that the predictive accuracy of the model must be assessed within the entire design domain, not just in the vicinity of a single physical experiment. Conversely the region of the design space where the accuracy is deemed appropriate for the application of interest defines the domain of validation.

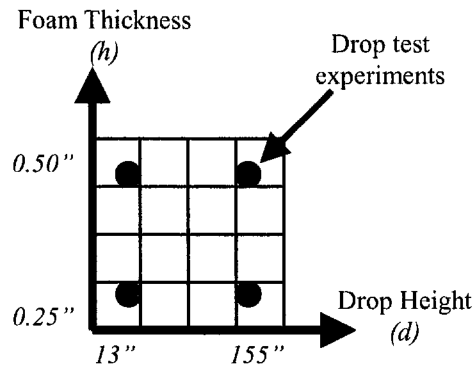

Figure 6. Illustration of the domain of validation. (Red dots represent the configurations (h;d) tested.)

In the example the domain of validation is two-dimensional and consists of the drop height d and foam thickness h, as shown in Figure 6. The design domain is bounded by the tested configurations: $13'' \leq d \leq 155''$ and $\frac{1}{4}'' \leq h \leq \frac{1}{2}''$. It is re-emphasized that the goal of the modeling effort is to provide a model capable of predicting the response of the system throughout the design space, not just for a single configuration $(h;d)$.

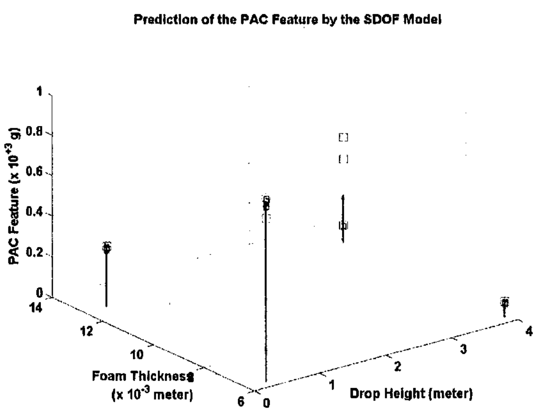

Figure 7. Prediction of PAC2 features.

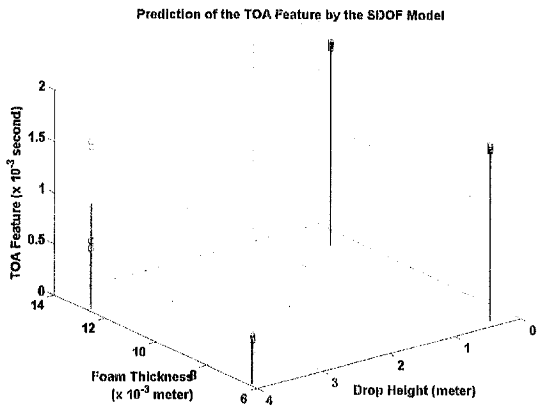

Figure 8. Prediction of TOA2 features.

The predictions of PAC2 and TOA2 features are shown in Figures 7-8 when the non-linear SDOF model is exercised at the four settings of drop height and foam thickness. The horizontal plane represents the domain of validation and the vertical axis shows the feature.

The PAC2 and TOA2 features are computed for each configuration and replicate experiment. The thin foam pad (*h=¼"*) is impacted ten times and the thick foam pad (*h=½"*) is impacted five times. It can be observed from Figures 7-8 that the PAC2 and TOA2 features are mostly consistent with each other, except for the configuration (*h=½"; d=155"*). Note that these predictions are obtained using a single model: the one previously calibrated with respect to the configuration (*h=¼"; d=13"*) of the system. Next the accuracy of the non-linear model is estimated throughout the design domain based on the predictions shown in Figures 7-8.

VI. DETERNIMISTIC METAMODELS OF PREDICTIVE ACCURACY

In this Section metamodels of predictive accuracy are discussed. The goal is to estimate the model's prediction error within the previously defined domain of validation.

Clearly the predictive accuracy of the model can only be estimated at those settings *(h;d)* where physical experiments have been conducted. In the foam impact example only the four settings shown in Figure 6 are tested. Elsewhere an extrapolation of prediction error must be performed. Polynomial metamodels are chosen to perform such extrapolation, mostly for simplicity. Other metamodels could be deployed, such as neural networks[5] or statistical Kriging-like models.[6]

It may be argued that there is no reason why the prediction error should vary according to a polynomial model. The assumption is reasonable because, for this application, the prediction error is expected to vary continuously and "smoothly" as a function of the configuration *(h;d)*. It is nevertheless recognized that assumptions are generally made to deal with—and, often, reduce—the uncertainty. For example assuming a polynomial error model lets us eliminate from the study the fact that the "true" behavior of the prediction error is unknown. To be thorough, such assumptions should be questioned. A probabilistic framework is discussed in Sections VIII and IX to handle the uncertainty reflected in modeling assumptions.

Table 1. Types of prediction error metamodels.

Type	Definition	Unknowns
1	Linear polynomial	3
11	Bi-linear polynomial	4
2	Full quadratic polynomial	6
3	Full cubic polynomial	10

Table 1 defines the prediction error metamodels studied in this Section. It is a family of polynomials:

$$\begin{cases} e(h;d) = c_1 + c_2 h + c_3 d \\ e(h;d) = c_1 + c_2 h + c_3 d + c_4 hd \\ e(h;d) = c_1 + c_2 h + c_3 d + c_4 hd + c_5 h^2 + c_6 d^2 \\ e(h;d) = c_1 + c_2 h + c_3 d + c_4 hd + c_5 h^2 + c_6 d^2 \\ \qquad + c_7 dh^2 + c_8 hd^2 + c_9 h^3 + c_{10} d^3 \end{cases} \tag{6}$$

The unknown coefficients c_1 to c_{10} are best fitted to the errors *e(h;d)*. One prediction error is defined for each configuration tested, that is, only four data points are available within the validation domain to train the polynomials. To obtain a single prediction error *e(h;d)* that accounts for both features PAC2 and TOA2, the Mahalanobis metric is computed:

$$e(h;p) = \left(y^{Test} - y(h;p)\right)^T S_{yy}^{-1} \left(y^{Test} - y(h;p)\right) \tag{7}$$

where y is the feature vector defined in equation (5), y^{Test} denotes the mean features extracted from replicate impacts and S_{yy} is the corresponding covariance matrix.

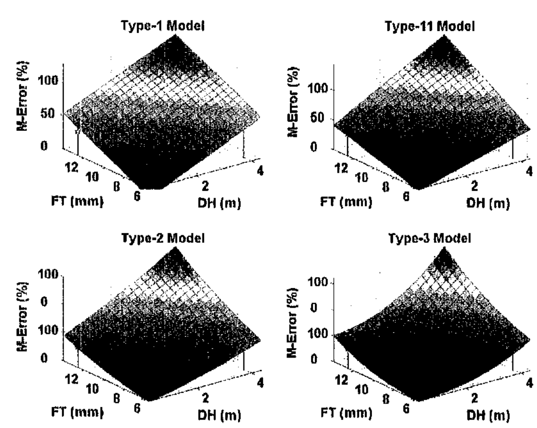

Figure 9. Response surfaces of prediction errors.

The unknown coefficients are computed for each error metamodel defined in Table 1 as the least-squares solution. A numerical solver based on the Singular Value Decomposition (SVD) is implemented to mitigate the adverse effects of ill conditioning. Non-linear optimization is also performed, although the results show that it does not significantly improve the goodness-of-fit. The four error metamodels hence obtained are pictured in Figure 9 as response surfaces over the entire domain of validation.

Tables 2 and 3 provide the coefficients identified for the linear (type-1) and fully quadratic (type-2) error models. One observation is that the linear model does not fit the four data points well, as shown by the large RMS fitting error. The other three metamodels provide excellent fits. The cubic polynomial gives indications of over-fitting the data points.

Table 2. Type-1 prediction error metamodel.

Coefficient	Effect	Value
1	*1*	-661.7
2	*h*	190.5
3	*d*	81.5
RMS Fitting Error		171.6

Table 3. Type-2 prediction error metamodel.

Coefficient	Effect	Value
1	*1*	-7.1
2	*h*	0.4
3	*d*	-29.7
4	$h*d$	15.0
5	h^2	11.1
6	d^2	0.4
RMS Fitting Error		7.10×10^{-13}

The significance of the error metamodels pictured in Figure 9 is emphasized. They provide an estimation of the error committed when predicting the peak acceleration and time of arrival of the impact wave, without having to perform the calculation itself. Two issues that have not been addressed so far are 1) how to account for the experimental variability; and 2) how to select the appropriate functional form of a prediction error metamodel. These issues are addressed next.

VII. STATISTICAL METAMODELS OF PREDICTIVE ACCURACY

Deterministic error metamodels were previously sought for simplicity. Deterministic models, however, cannot account for the variability encountered during physical experimentation, nor can they include other sources of uncertainty.

Another limitation of deterministic polynomials is that they only provide deterministic error estimations. The Mahalanobis error metric employed here can be related to a statistical test providing that assumptions are made about the probability distribution laws of the underlying random processes. The chi-square statistical test, for example, assumes Gaussian laws. Assuming that the joint probability distribution of the PAC2 and TOA2 features is Gaussian based on only five or ten replicates may be more than what is actually known.

To avoid introducing unjustifiable assumptions, no assumption must be made regarding the probability laws of the measured or predicted features. In doing so we comply with the principle of "*accounting for the uncertainty, all of it, but no more.*" Error metamodels are developed using the same procedure as outlined in Section VI, with the exception of two differences. First the errors are defined in physical units as:

$$e(h;p) = y^{\text{Test}} - y(h;p) \tag{8}$$

where y^{Test} and y represent one of the PAC2 or TOA2 features. One error model is trained for each feature.

The second difference is that the features used as the reference in y^{Test} are not averaged from the five or ten replicates as before. Instead one of the replicate tests is selected randomly for each setting *(h;d)*. The four selected features—each drawn randomly from its own "pool" of replicate tests—are used to calculate the prediction errors (8). The prediction error polynomials defined in Table 1 are best fitted to the data. The whole procedure is repeated a thousand times, hence, defining a bootstrapping algorithm. It is emphasized that the random drawings do not constitute Monte Carlo runs because no probability law is sampled. Figure 10 shows the training errors of the PAC2 (red squares) and TOA2 (blue circles) bilinear polynomials.

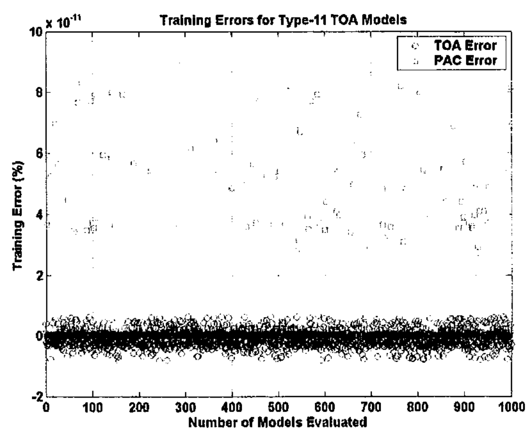

Figure 10. PAC2 and TOA2 training errors.

The result of bootstrapping is a thousand error models for each feature and each polynomial type. The distributions of bilinear coefficients for the PAC2 and TOA2 prediction errors are shown in Figures 11 and 12. As previously noted the linear model (type-1) does not fit the data points well, while the other models do. Figure 11 shows that the coefficients of PAC2 errors are distributed according to a Gaussian-like law. On the other hand the coefficients of TOA2 errors seem to be distributed according to bi-modal probability laws.

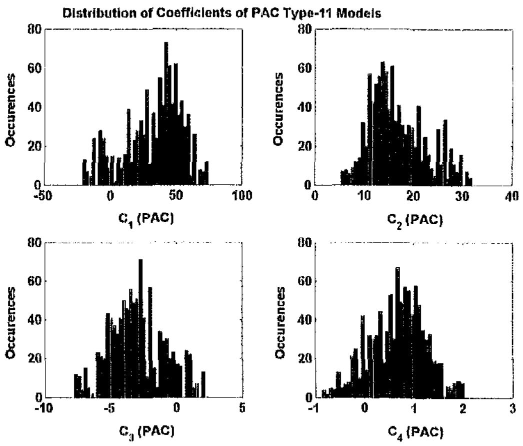

Figure 11. Distribution of coefficients for the bilinear polynomials of PAC2 prediction errors.

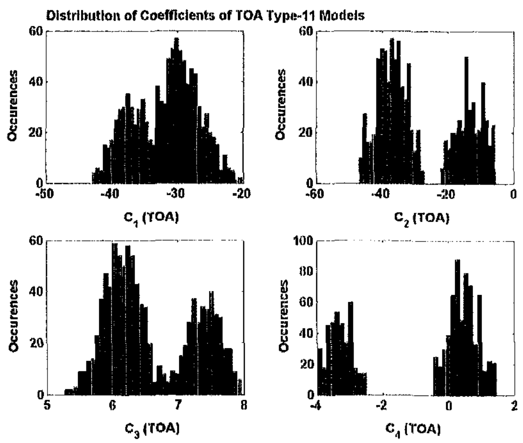

Figure 12. Distribution of coefficients for the bilinear polynomials of TOA2 prediction errors.

Tables 4 and 5 list the statistics (mean, standard deviation) of coefficients for the bilinear polynomials of PAC2 and TOA2 prediction errors. In contract to the metamodels developed in Section VI—that provide a deterministic prediction error at each setting *(h;d)*—the coefficients of the statistical metamodels are sampled according to their probability laws in Figures 11, 12 or Tables 4, 5 to generate a population of error response surfaces. Next statistics of the errors are computed.

An illustration is provided in Figure 13. It shows the expected PAC2 error (blue solid line) and the two-standard deviation confidence intervals (red dashed lines). The top sub-figure illustrates the error as a function of foam thickness when the drop height is set to 75 inches. The bottom sub-figure illustrates the error as a function of drop height when a 0.39-inch thick pad is simulated. The confidence intervals capture the spread of PAC2 errors caused by environmental variability. Note how the prediction error's uncertainty amplifies with drop height, while it remains constant with as a function of the foam pad's thickness. This observation suggests that controlling the environmental variability inputted to the model is more critical if varying drop heights are simulated with the model.

Table 4. Type-11 PAC2 prediction error metamodel.

Effect	Mean PAC2 Coefficient	Standard Deviation	Percent Variation
1	32.9	22.1	67.2%
h	16.9	5.7	34.1%
d	-3.0	2.1	70.6%
*h*d*	0.7	0.6	83.9%

(Percent variation = 100 x standard deviation / mean.)

Table 5. Type-11 TOA2 prediction error metamodel.

Effect	Mean TOA2 Coefficient	Standard Deviation	Percent Variation
1	-31.7	4.9	15.3%
h	-27.9	12.4	44.4%
d	6.6	0.7	10.1%
*h*d*	-0.9	1.9	196.0%

(Percent variation = 100 x standard deviation / mean.)

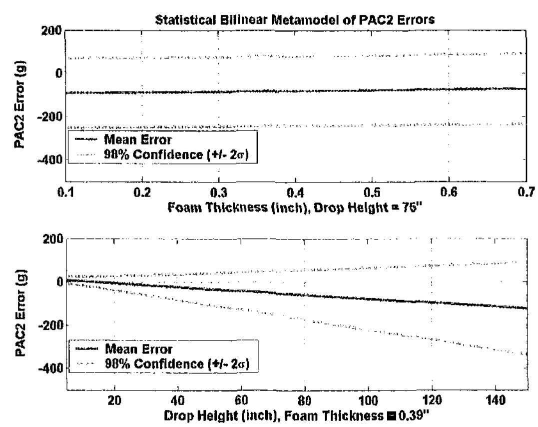

Figure 13. Statistics of PAC2 prediction errors.

Error metamodels such as illustrated in Figure 13 allow the analyst to estimate the error associated with a prediction at points *(h;d)* in the design space where testing has not been conducted. Conversely regions of the design space can be identified that yield less than a predefined error level. The bootstrapping procedure demonstrates how to propagate experimental variability from test measurements to metamodels of predictive accuracy. The same approach can propagate other sources of variability and uncertainty. Monte Carlo simulations can just as well be implemented, if it is believed that sufficient evidence is available to define probability laws that can be sampled.

VIII. PRIORS AND POSTERIORS OF PREDICTION ERROR MODELS

The work has so far focused on the development of metamodels of predictive accuracy. In Section VII a general procedure is suggested to develop error models that account for sources of variability and uncertainty. Sections VIII and IX illustrate how to combine several metamodels, which addresses the issue of model form uncertainty and selection.

As discussed previously the correct form of the error metamodel is unknown. We have assumed a family of polynomials but other choices, such as neural networks, are equally likely. So far this uncertainty has not been accounted for in the analysis. For example it is assumed in Section VII that the correct form of the error model is bilinear. The advantage of formulating such assumption is that it removes uncertainty about the model form from the analysis. Accounting for the uncertainty is nevertheless critical when the objective is to assess the total prediction error of a simulation. The same remark applies to the functional form (2) of the non-linear forcing function and, in general, any modeling assumption. For simplicity the only source of modeling uncertainty considered in the remainder is the polynomial form of the error metamodel. The approach discussed below generalizes to other uncertainties.

Instead of postulating an assumption to eliminate this modeling uncertainty, we seek to estimate its effect on the predictive accuracy. Because assumptions such as "*the metamodel is linear*" or "*it is quadratic*" lead to different error metamodels, our approach is to assign probabilities to the various assumptions and propagate the uncertainty by means of a Monte Carlo simulation.

The procedure starts by seeking a mechanism to assign probabilities to the modeling assumptions or, equivalently, combine the prediction error metamodels. Here metamodels are combined using the framework provided by the Bayes Theorem of conditional probabilities.[7] In Bayesian analysis the probability of a model $y=M(p)$ is updated and conditioned on the test data y^{Test} by accounting for the ability of this model to reproduce the available data:

$$\Pr(p\,|\,y^{Test}) = \frac{\Pr(y^{Test}\,|\,p)\Pr(p)}{\int_{p\in\Omega}\Pr(y^{Test}\,|\,p)\Pr(p)} \qquad (9)$$

where $Pr(p)$ represents the prior knowledge, that is, what the analyst knows before the model is put to the test of comparing its predictions to test data. Quantity $Pr(p|y^{Test})$ denotes the posterior probability sought to assign a probability to each modeling assumption. The likelihood function $Pr(y^{Test}|p)$ can be defined a number of ways. For all practical purposes it designates the agreement between predictions y of the model and test data y^{Test}, which is why it is here calculated as:

$$\Pr(y^{Test}\,|\,p) = k e^{-S(p)} \qquad (10)$$

where S is a "misfit" function[8] simply defined in the example as the Mahalanobis distance (7). Note that the likelihood function is not a probability. It does not need to satisfy the Kolmogorov axioms of probability, which is why the scaling factor k included in equation (10) is later omitted ($k=1$).

Table 6 lists the prior probabilities, likelihood values, and posterior probabilities of the four error polynomials and Figure 14 presents the results of the Bayesian updating. Originally the linear, bilinear, and quadratic models are assigned equal probabilities of 30%. The cubic model is assigned a 10% probability because the investigation shows that it tends to over-fit the data. The likelihood values are computed by combining equations (7) and (10) where $S=e(h;d)$. Priors and likelihood values are multiplied together and normalized according to equation (9) to calculate the posteriors listed in Table 6.

Table 6. Prior and posterior probabilities.

Type	Priors	Likelihood	Posteriors
1	30.0%	4.59×10^{-5}	1.51%
11	30.0%	1.28×10^{-3}	42.21%
2	30.0%	1.28×10^{-3}	42.21%
3	10.0%	1.28×10^{-3}	14.07%

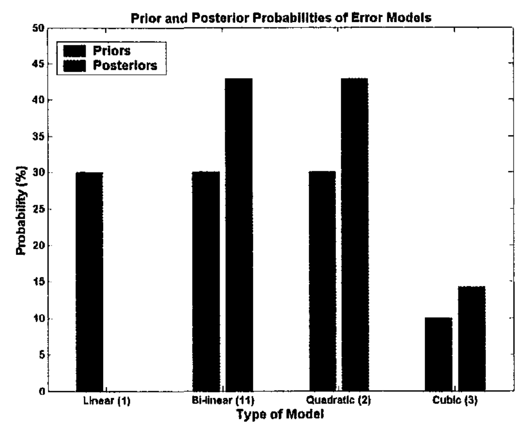

Figure 14. Prior and posterior probabilities.

The results confirm that the bilinear and quadratic polynomial forms are most appropriate because their posterior probabilities increase. The linear polynomial becomes unlikely due to its lack of fit to the data, while the cubic polynomial's posterior increases slightly. In conclusion a probability law is assigned to represent the uncertainty associated with the functional form of the prediction accuracy metamodels. This example illustrates how an epistemic lack of knowledge can be replaced by a probability distribution instead of being somewhat arbitrarily eliminated from the analysis by formulating an assumption.

IX. BOUNDS OF MODELING ERRORS

The concepts of deterministic and statistical error models have previously been illustrated. The statistical metamodels developed in Section VII accounted for the experimental variability but their functional form was not questioned. Uncertainty regarding their form is now taken into account.

The end result is an assessment of predictive accuracy throughout the domain of validation, together with confidence intervals—or the "error of the error"—that reflect the experimental variability and modeling uncertainty. This is here referred to as bounding the modeling error. We seek to obtain an estimation of prediction error $e(h;d)$ at any point $(h;d)$ of the design domain, and also intervals $[e_{Min};e_{Max}]$ that bound these errors at a given confidence level C_E. The statistical interpretation is that, if a hundred simulations are performed with all sources of uncertainty represented, then the prediction errors would fall C_E times out of a hundred within the interval $[e_{Min};e_{Max}]$.

Because the unknown polynomial model form is represented as a probability law in Section VIII, studying the effect of this uncertainty on the prediction accuracy simply becomes a sampling issue. Basically the posterior probabilities listed in Table 6 are sampled, that is, a polynomial type (linear, bilinear, quadratic or cubic) is randomly selected according to the posteriors. Once the error metamodel selected its coefficients are sampled to account for the effect of environmental variability. Figures 15 and 16 illustrate forty randomly selected error models for predicting PAC2 and TOA2.

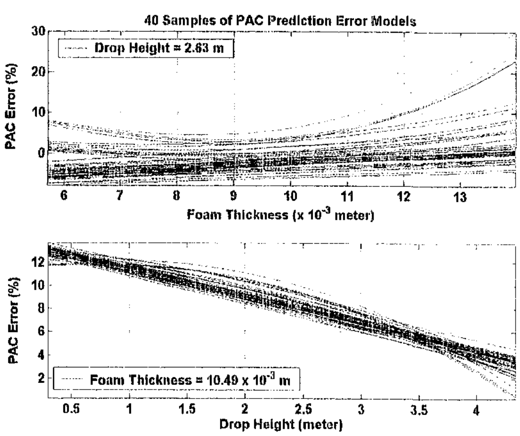

Figure 15. 40 realizations of PAC2 error metamodels.

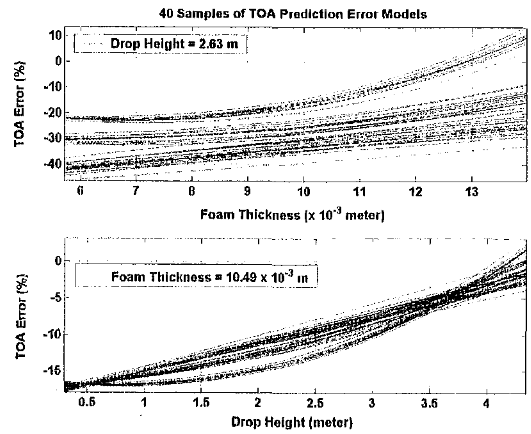

Figure 16. 40 realizations of TOA2 error metamodels.

In Figures 15 and 16 the error models are shown as one-dimensional curves for clarity, even though the information is available in the plane $(h;d)$. Some of the metamodels are bilinear polynomials, while others are quadratic or cubic. The polynomials of same type do not necessarily possess the same coefficients because these are sampled to reflect the test variability.

In essence this approach defines a hierarchical Monte Carlo simulation. The outer loop samples the model form and the inner loop samples the polynomial coefficients. A hundred outer runs and a thousand inner runs are performed, for a total of 100,000 simulations. The algorithm is not demanding because evaluating the error metamodels can be performed in a fraction of a second on a regular desktop computer.

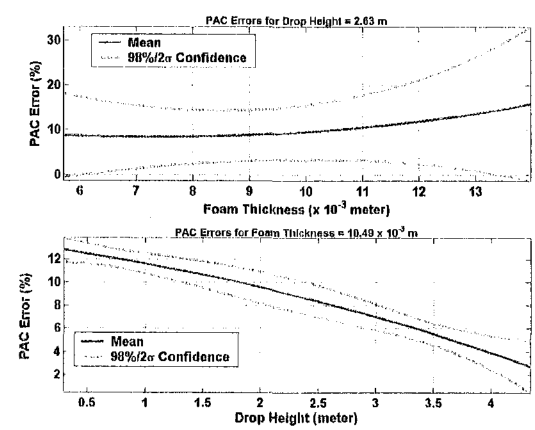

Figure 17. Final PAC2 prediction errors and uncertainty bounds of uncertainty.

Figure 17 pictures the final metamodel of PAC2 prediction error as well as its uncertainty bounds. The mean error (blue solid line) and two-standard deviation intervals (red dashed lines) are shown. Figure 18 shows the same information for the TOA2 feature. Slices of constant foam thickness or drop height are shown for clarity. A "necking" pattern can be observed in the

Figures. It expresses that the assessment of prediction error is more confident in the vicinity of settings *(h;d)* where physical tests have been performed. Away from the tested configurations the prediction error does not necessarily grow but the uncertainty of the assessment does. Should such uncertainty become too important to prevent any meaningful decision-making, action such as additional testing or further model revision would be recommended to attempt to keep it under control.

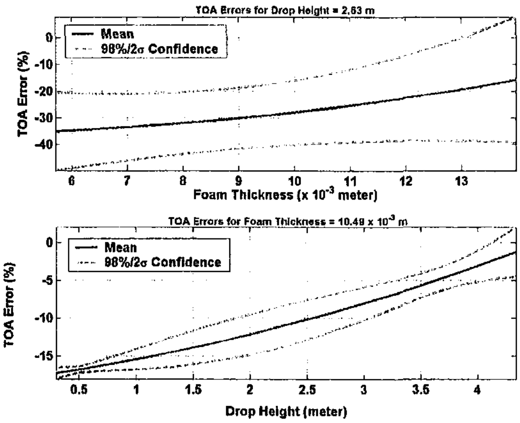

Figure 18. Final PAC2 prediction error metamodels and their bounds of uncertainty.

In the example the variability of impact tests and the uncertainty about the functional form of the error metamodel are considered. Nevertheless other sources of uncertainty—such as the functional form of the non-linear force, its parameters, the boundary and initial conditions—should be propagated in a similar manner if they are included in the analysis. Although an exhaustive quantification of uncertainty is not provided in this study, the main elements are presented. One can easily understand how to apply the method to a more complete assessment of predictive accuracy.

X. CONCLUSION

In the context of model validation the ultimate goal of uncertainty quantification is to construct an uncertainty model for every component of the simulation, which taken all together summarize how well the predictions agree with all available experimental results. It is emphasized that uncertainty models must not necessarily be statistical in nature. Other frameworks, such as the non-probabilistic theory of information gap, might be more appropriate in cases of extreme uncertainty or scarce data.[9]

Being able to assess the predictive error within the design domain is a necessary step to answer other important questions in model validation. The first one of them is to decide how many physical experiments are required to provide a reliable estimate of the prediction error. Procedures must be developed to decide how many experiments are required and where in the design domain to perform the physical testing.

Another important question is to assess how "good" across the design domain must the model be for a particular application. The estimation of predictive accuracy also makes it possible to quantitatively rank competitive models against each other. These questions will be the thrust of future research.

XI. REFERENCES

[1]Mottershead, J.E., Friswell, M.I., "Model Updating in Structural Dynamics: A Survey," *Journal of Sound and Vibration*, Vol. 162, No. 2, 1993, pp. 347-375.

[2]Hemez, F.M., Wilson, A.C., Doebling, S.W., "Design of Computer Experiments for Improving an Impact Test Simulation," *19th SEM International Modal Analysis Conference*, Kissimmee, FL, Feb. 5-8, 2001, pp. 977-985.

[3]Schultze, J.F., Hemez, F.M., Doebling, S.W., Sohn, H., "Statistical-based Non-linear Model Updating Using Feature Extraction," *19th SEM International Modal Analysis Conference*, Kissimmee, FL, Feb. 5-8, 2001, pp. 18-26.

[4]Hemez, F.M., Doebling, S.W., "A Validation of Bayesian Finite Element Model Updating for Linear Dynamics," *17th SEM International Modal Analysis Conference*, Kissimmee, FL, Feb. 8-11, 1999, pp. 1545-1555.

[5]Bishop, C.M., **Neural Networks for Pattern Recognition**, Oxford University Press, New York, NY, 1998.

[6]Rutherford, B., "Response Modeling for the Design of Computer Experiments," *20th SEM International Modal Analysis Conference*, Los Angeles, CA , Feb. 4-7, 2002, pp. 683-688.

[7]Hanson, K.M., "A Framework for Assessing Uncertainties in Simulation Predictions," *Physica D*, Vol. 133, 1999, pp. 179-188.

[8]Mosegaard, K., Tarantola, A., "Monte Carlo Sampling of Solutions to Inverse Problems," *Journal of Geophysical Research*, Vol. 100, No. B7, July 1995, pp. 12431-12447.

[9]Ben-Haim, Y., **Information-Gap Decision Theory: Decisions Under Severe Uncertainty**, *Series on Decision and Risk*, Academic Press, London, U.K., 2001.

LA-UR- 03-XXXX

Approved for public release; distribution is unlimited.

Title: ESTIMATING THE ERROR IN SIMULATION PREDICTION OVER THE DESIGN SPACE (U)

Author(s): Rachel Shinn, Embry-Riddle Aeronautical University

François M. Hemez, LANL, ESA-WR

Scott W. Doebling, LANL, ESA-WR

Submitted to: 44th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics and Materials Conference, Norfolk, Virginia, April 7-10, 2003.

Los Alamos
NATIONAL LABORATORY