LA-UR- 01-6574

Approved for public release; distribution is unlimited.

Title:

DISCUSSION OF MODEL CALIBRATION AND VALIDATION FOR TRANSIENT DYNAMICS SIMULATION (U)

Author(s):

François M. Hemez, Scott W. Doebling, Amanda C. Wilson; Los Alamos National Laboratory, ESA-WR

Submitted to:

20th International Modal Analysis Conference (IMAC-XX), Los Angeles, California, February 4-7, 2002

DISCUSSION OF MODEL CALIBRATION AND VALIDATION FOR TRANSIENT DYNAMICS SIMULATION

François M. Hemez,[1] Scott W. Doebling[2] and Amanda C. Wilson[3]
Los Alamos National Laboratory
Engineering Sciences & Applications Division
Weapon Response Group (ESA-WR)
P.O. Box 1663, M/S P946, Los Alamos, New Mexico 87545

ABSTRACT

Model calibration refers to a family of inverse problem-solving numerical techniques used to infer the value of parameters from test data sets. The purpose of model calibration is to optimize parametric or non-parametric models in such a way that their predictions match reality. In structural dynamics an example of calibration is the finite element model updating technology. Our purpose is essentially to discuss calibration in the broader context of model validation. Formal definitions are proposed and the notions of calibration and validation are illustrated using an example of transient structural dynamics that deals with the propagation of a shock wave through a hyper-foam pad. An important distinction that has not been made in finite element model updating and that is introduced here is that parameters of the numerical models or physical tests are categorized into input parameters, calibration variables, controllable and uncontrollable variables. Such classification helps to define model validation goals. Finally a path forward for validating numerical model is discussed and the relationship with uncertainty assessment is stressed.

1. INTRODUCTION

Today's computational resources make it more than ever possible to model and analyze phenomena characterized by complex geometries and boundary conditions, multi-physics, nonlinear effects and variability. An example of such resource is the U.S. Department of Energy's Accelerated Strategic Computing Initiative (ASCI) that has developed several platforms able to sustain over 3 Tera-OPS, that is, $3 \times 10^{+12}$ floating point operations per second, by distributing computations over arrays of more than 6,000 processors. Reference [1] discusses the overall ASCI program and its objectives. Examples of problems requiring access to these multi-physics codes and massively parallel architectures include global climate prediction, epidemics modeling, computational molecular dynamics, thermo-nuclear physics and complex engineering simulations.

Obviously the hypothesis sustaining the development of ASCI-class computing resources is that predictive accuracy can be achieved if enough details and physics can be included in the simulation. For example constitutive models at the microscopic and nano-scale levels based on "first principle physics" such as statistical quantum mechanics are increasingly investigated. The intent is to capture the physics of interest at its source rather than relying on global and somewhat arbitrary approximations such as, for example, modal damping ratios in solid mechanics.

In the field of structural dynamics computational models are developed for predicting the response of a system when the phenomenon is not accessible by direct measurement or numerical simulations are cheaper than testing. To develop high-fidelity models analysts increasingly account for nonlinear behaviors and variability. However implementing sophisticated models does not guarantee the accuracy of their predictions. It must be verified that the discretization, mathematical idealization, computational errors and other assumptions involved yield a satisfactory solution. This is usually referred to as "model validation" and carried out by comparing the model's prediction to test data. If the agreement between measurements and predictions is not satisfactory, input parameters are optimized to improve the model's predictive quality.

[1] Technical staff member, ESA-WR, **hemez@lanl.gov**.
[2] Technical staff member, ESA-WR, **doebling@lanl.gov**.
[3] Graduate student, Georgia Tech University, **awilson42@yahoo.com**.

In this publication such definition of "model validation" is challenged. The reason is because predicting a measured response does not necessarily provide accuracy throughout the design or operational space. The notion of ***design space*** is illustrated in Figure 1 where, for simplicity, the structural dynamics model is denoted by the input-output relationship

$$y = M(p_1;p_2) \qquad (1)$$

where p_1 and p_2 denote two of the model's input parameters and y denotes a scalar prediction, also referred to as the output feature. In modal analysis, for example, the input parameters p_1 and p_2 might represent a beam's moments of flexure EI and the output feature might represent the first bending frequency. Models considered in equation (1) can range from general-purpose finite element analyses to simple polynomial models. For computational efficiency and visualization simplicity, it is often advantageous to replace physics-based models with surrogates as discussed in Reference [2].

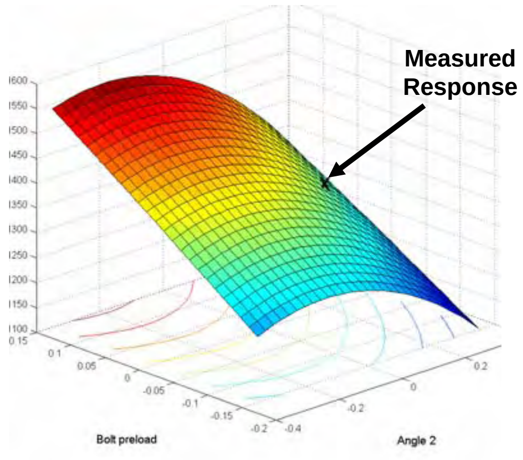

Figure 1. Conceptual illustration of a model and comparison between measurement and prediction.

Figure 1 illustrates a good agreement between model prediction and test data at one location $(p_1;p_2)$ in the design space. The conventional paradigm of "model validation" is that, to obtain an accurate prediction, the parameters $(p_1;p_2)$ or the form of the computational model can be optimized in such a way that the distance between measured and predicted features is minimized. Here such optimization procedure is called ***model calibration***.

Calibration is the main concept behind the development of finite element model updating methods. Its main drawback is that, in general, no conclusion about the model's predictive quality can be made away from the calibration point. This issue has been addressed by repeating calibration experiments at various points in the design space. One example in structural dynamics is pseudo-testing where receptance functions are modified to generate additional data sets without requiring further testing [3, 4]. However this tends to reduce the concept of model validation to a series of calibration experiments, which we claim it is not. In addition calibration techniques, with a few exceptions, do not provide any statistical assessment of the prediction's accuracy.

The first objective of this publication is to discuss parametric calibration in the broader context of model validation. To stress the difference between calibration and validation, we find it necessary to separate the model's variables into input parameters and calibration variables. Calibration variables include controllable, uncontrollable or measured variables. Establishing a clear distinction between input parameters and calibration variables is not generally addressed in the field of finite element model updating although it is critical to the success of a model validation experiment. The distinction is illustrated in section 2 that briefly introduces a numerical simulation of transient structural dynamics. A similar discussion can be obtained from Reference [5] that deals with applications in hydrology and radiation management. Calibration and validation are formally defined in sections 3 and 4. Finally a path forward for model validation is discussed and the relationship with uncertainty assessment is stressed.

2. HYPER-FOAM IMPACT TESTING

One example of numerical simulation for transient dynamics is the Los Alamos impact experiment, or drop test, discussed in References [6] and [7]. This application involves the transmission of a shock wave through an assembly that consists of a steel cylinder and a layer of elastomeric, or hyper-foam, material.

Figure 2 pictures the hardware involved during impact testing. An assembly of elastomeric layer and steel cylinder is mounted on an impact table and dropped to generate the shock wave—the elastomeric layer sits underneath the steel cylinder and is barely visible in Figure 2. The input acceleration and three output accelerations are measured. The input acceleration is collected on the drop table and represents the acceleration inputted to the elastomeric layer-cylinder assembly. Three output accelerations are collected on top of the steel cylinder (see Figure 2).

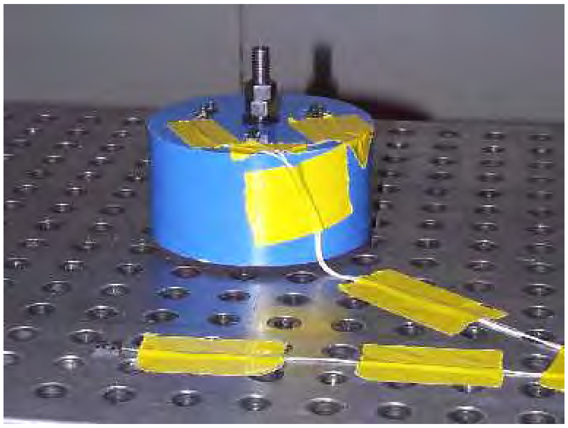

Figure 2. Impact test setup.

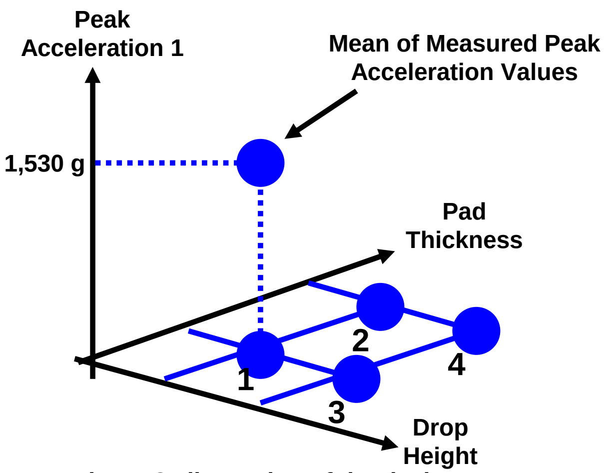

Figure 3. Illustration of the design space.

Table 1. Drop test configurations.

	Low drop (0.3 m)	High drop (4.0 m)
Thin pad (6.3 mm)	Configuration 1 (10 replicates)	Configuration 3 (5 replicates)
Thick pad (12.6 mm)	Configuration 2 (10 replicates)	Configuration 4 (5 replicates)

Foam layers of different thickness and several drop heights are considered during impact testing. The first configuration tested is defined with a 0.25-inch (6.3 mm) thick elastomeric pad and a 13-inch (0.3 m) drop height. Another pad of the same material but different thickness—0.50 inches or 12.6 mm—is used to perform additional physical experiments. Similarly testing is performed at a second drop height of 155 inches (4.0 m). We refer to the pad thickness and drop height as the two input parameters p_1 and p_2, respectively. Combinations $(p_1;p_2)$ of these parameters define a two-dimensional space illustrated in Figure 3. The two output features of interest are the peak acceleration value and the corresponding delay time between peak input and peak output. In Figure 3 the peak acceleration recorded at channel 1 is used as an example. Physical testing provides measurements for the four configurations identified with the numbers 1-4 in Figure 3 and defined in Table 1.

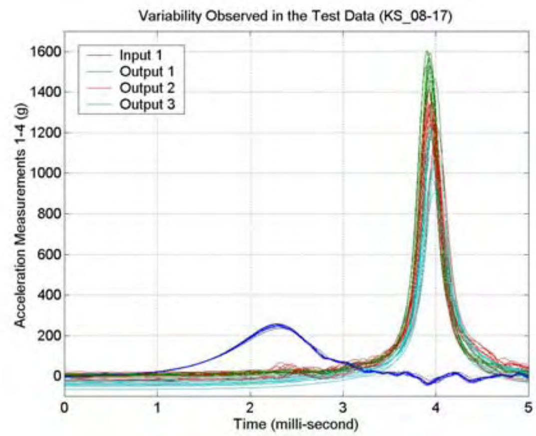

Figure 4. Variability obtained during testing.

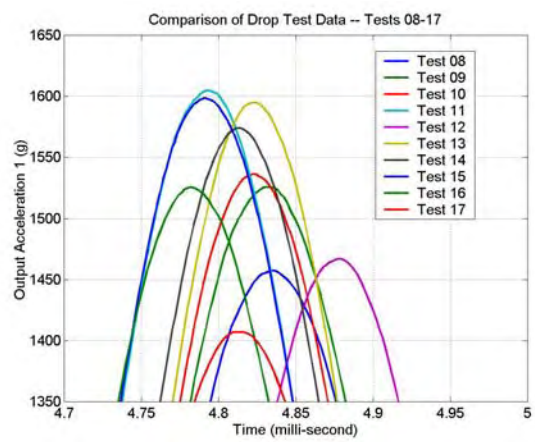

Figure 5. Variability of peak output at channel 1.

In addition to testing various configurations each experiment is replicated several times to estimate the environmental variability. Table 1 provides the number of replicate experiments performed for each configuration. For example the first configuration where p_1 = 6.3 mm and p_2 = 0.3 m is tested ten times. Figures 4 and 5 illustrate the response variability obtained when “identical” tests are repeated for the first configuration. Figure 4 shows the input acceleration signal and three output acceleration signals collected during the ten replicates. Figure 5 enlarges the view in the vicinity of output channel 1’s peak acceleration.

The fact that significant variability is obtained when the “same” experiment is repeated ten times means that the recorded output varies randomly; that other, uncontrolled input parameters influence the output result; or a combination of both. None of these sources of variability can be ruled out ***a priori***. An example of the first source—random output variability—is measurement error. Examples of the second source—input parameter variability—often encountered in experimental sciences are temperature and humidity. Indeed several input parameters were not controlled nor measured during the drop tests. They include the preload applied by the two tightening bolts visible on Figure 2 and the angles at which the carriage impacts the ground. It was later demonstrated that these parameters significantly influence the acceleration signals and are responsible for partially explaining the observed variability. Details of this analysis are provided in References [6] and [7]. The impact angles are referred to as uncontrolled parameters because they vary randomly from test-to-test and their variation is not controlled nor measured.

Table 2. Parameters of the numerical model.

Factor	Description	Unit
1	Pad thickness	Millimeter
2	Drop height	Meter
3	First angle of impact	Degree
4	Second angle of impact	Degree
5	Bolt preload	N/m^2
6	Hyper-foam coefficient 1	Unit-less
7	Hyper-foam coefficient 2	Unit-less
8	Input amplitude	Unit-less
9	Static friction coefficient	Unit-less
10	Bulk viscosity coefficient	Unit-less

In the drop test example the main objective of numerical modeling is to develop a finite element representation of the system capable of predicting the system's response with "acceptable" accuracy, not just at the four points in Figure 3 where physical tests were performed, but throughout the input space. Prediction accuracy is deemed sufficient if the numerical model can reproduce the observed data within the level of uncertainty visible in Figures 4-5. At the very least the numerical model should reproduce the mean behavior observed during testing as well as some of the statistics such as the response's total variance and covariance structure. To achieve this result it is necessary that the most important sources of variability encountered during physical experimentation be taken into account in the model. Our simulation of the drop test is therefore parameterized with ten parameters listed in Table 2.

It may seem like, starting from a two-dimensional space, we now have to account for ten—and maybe more—input parameters. However it is important to emphasize that the problem remains essentially two-dimensional. The reason is because the main purpose of the simulation is to predict the system's response as a function of pad thickness and drop height. Input parameters 3-10 in Table 2 represent additional variables introduced by the modeling effort. In the remainder they are referred to as ***calibration variables*** to stress the distinction with the two input parameters p_1 and p_2 we are genuinely interested in. Table 3 classifies the ten parameters in four sets according to whether they are genuine input parameters of the problem, controllable calibration variables, uncontrollable calibration variables or measured variables.

Table 3. Current classification of input parameters.

Factor & Description	I	C	U	M
1, Pad thickness	X			
2, Drop height	X			
3, First angle of impact			X	
4, Second angle of impact			X	
5, Bolt preload			X	
6, Hyper-foam coefficient 1		X		
7, Hyper-foam coefficient 2		X		
8, Input amplitude				X
9, Static friction coefficient		X		
10, Bulk viscosity coefficient		X		

I: input parameter; **C**: controlled; **U**: uncontrolled; **M**: measured.

Table 4. Ideal classification of input parameters.

Factor & Description	S	I	C	U	M
1, Pad thickness	++	X			
2, Drop height	++	X			
3, First angle of impact	++				X
4, Second angle of impact	+				X
5, Bolt preload	++				X
6, Hyper-foam coefficient 1	-		X		
7, Hyper-foam coefficient 2	-		X		
8, Input amplitude	++				X
9, Static friction coefficient	-		X		
10, Bulk viscosity coefficient	-		X		

S: indicator of global sensitivity ("++" is high; "+" is medium; "-" is low); **I**: input parameter; **C**: controlled; **U**: uncontrolled; **M**: measured.

Ideally all calibration variables should be eliminated from the analysis to reduce its dimensionality and leave only the relevant input parameters. Variables can be eliminated in three main ways. First parameter calibration techniques can be implemented to infer the value of a variable from experimental data sets. Secondly the physical experiment can, in some cases, be modified to provide more control over a previously uncontrolled variable. The best way to eliminate a variable, however, remains through direct measurement. This is reflected in Table 4 that illustrates an "ideal" situation where all significant calibration parameters are measured. By "significant" it is meant a parameter that largely contributes to the total output variability. Statistical techniques such as the analysis of variance (ANOVA) can assess the global influence of an input parameter throughout the design space [8]. In the case of the drop test, ANOVAs have demonstrated that calibration variables 3, 4, 5 and 8 contribute to the output variability more than any other. Ideally they should be measured during the experiment. Other variables, such as the material coefficients 6-7 and numerical coefficients 9-10, cannot be measured directly. However one advantage is that they characterize intrinsic properties of the hyper-foam material and are not expected to vary within the design space. The second best option would be to infer, or calibrate, their values from experimental data sets.

Additional drop tests are being planned to reflect the ideal testing setup illustrated in Table 4 as opposed to the definition of past tests (Table 3). In particular the preload will be measured directly by instrumenting the bolt or inserting a piezoelectric washer. Values of the potentially non-zero angles of impact will be inferred through the procedure discussed in section 3.

3. MODEL CALIBRATION

Model calibration is defined as the optimization of input parameters and/or calibration variables such that the agreement between the measured and predicted responses is improved. In the field of structural dynamics such inverse problems are generally formulated as parametric optimization problems although other approaches are available for non-parametric optimization or two-point boundary value problems [9]. The discrepancy between measured and predicted responses is expressed as a distance vector **{e}** such as

$$\{\mathbf{e}\} = \{\mathbf{y}_{\mathbf{Measured}}\} - \{\mathbf{y}_{\mathbf{Predicted}}\} \qquad (2)$$

For example **{e}** might collect the differences between measured and computed modal frequencies in linear structural dynamics. Then a cost function *J* is defined for minimization

$$J(p+dp) = \{\mathbf{e}\}^{\mathbf{T}}[\mathbf{S}_{\mathbf{EE}}]^{-1}\{\mathbf{e}\} + \{\mathbf{dp}\}^{\mathbf{T}}[\mathbf{S}_{\mathbf{PP}}]^{-1}\{\mathbf{dp}\} \qquad (3)$$

The definition of the cost function can be purely deterministic or include a representation of the uncertainty associated to the test data and variability of the calibration parameters. One such example is the Bayesian parametric inference documented, among others, in References [10] and [11]. The cost function can also represent a statistical test, such as the Kullback-Leibler entropy used in Reference [12] for discrimination and clustering analysis. Reviews of finite element model updating in structural dynamics are available from References [9] and [13].

In any case values of the model's parameters are inferred from test data at one point in the design space. Inference does not provide information about the parameters away from the design point where the calibration experiment is performed. This is illustrated in Figure 6. At best calibration improves the predictive accuracy of the numerical model in the vicinity of one combination *(p_1;p_2)*. In addition it is emphasized that the overall predictive accuracy of the model cannot be assessed through calibration only.

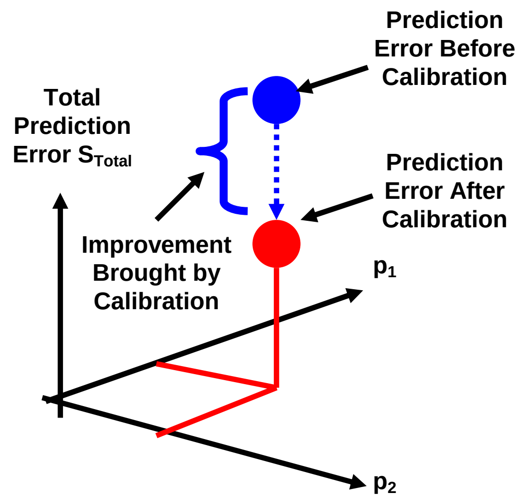

Figure 6. Prediction improvement via calibration.

An example of model calibration is provided below for the hyper-foam application introduced previously. The purpose of this example is to emphasize that calibration does not validate model predictions in any way and that calibration results should be confirmed with independent investigations as much as possible.

In the case of the hyper-foam impact application, surrogate models are developed to establish a simple, polynomial relationship between the eight calibration variables of Table 2 and six output features defined as the peak accelerations and times-of-arrival at the three output sensors [6-7]. Figure 7 illustrates one of these surrogate models for configuration 1 that, after careful design of experiments and ANOVA variable screening, takes the form of a quadratic polynomial.

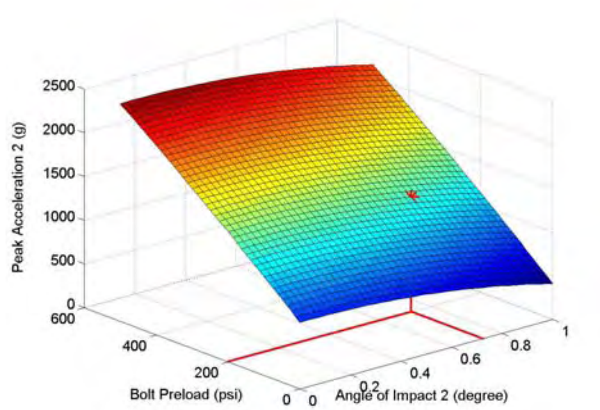

Figure 7. Surrogate polynomial model.

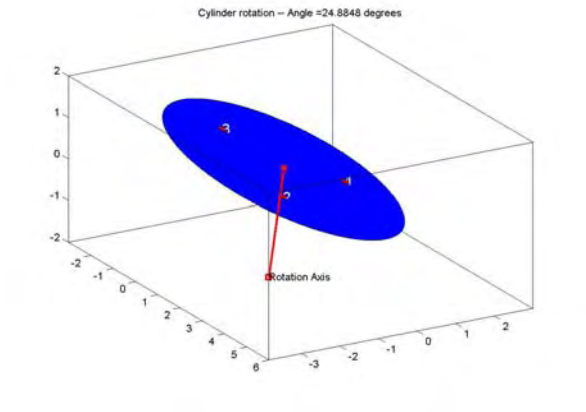

Figure 8. Inference of the angles of impact.

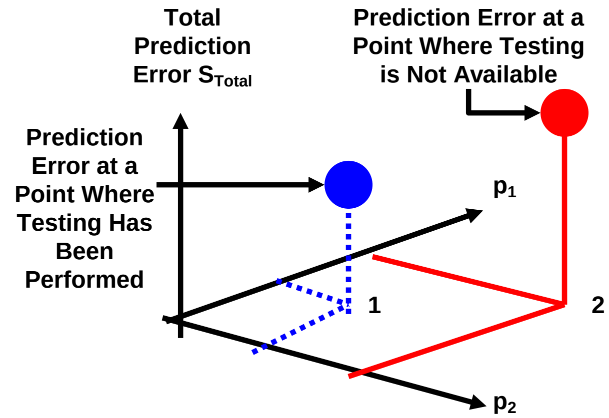

Figure 9. Concept of uncertainty quantification.

Variables of the finite element model are then calibrated using the six mean features obtained from the drop test measurements as the reference. The optimization provides values for the four most sensitive calibration variables—for example Figure 7 shows that the inferred value of the second impact angle is equal to 0.7 degree—but such information is of little value to the analyst unless it can be verified by independent means. To verify that the calibration exercise provides reasonable results, the measured output acceleration signals are integrated numerically to obtain the positions versus time. Then a plane is fit through the displacement history of the three accelerometers as illustrated in Figure 8. The plane's inclination provides an independent verification of the angle of impact. Remarkably a value equal to 0.65 degree is obtained at the time of impact. This result indicates that our methodology of fitting surrogate models through an appropriate design of experiments [8, 14] and calibrating the unknown parameters seems to provide accurate results.

4. MODEL VALIDATION

Validation may be defined as the process of determining the degree to which the output of the simulation code agrees with the actual behavior of the physical system in a specified application. A formal definition is given in Reference [15] as

> *"The substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended applications of the model."*

This definition clearly identifies the three key issues of model validation: 1) A model is defined throughout a domain of applicability, or design space, and not just at a single operating point. 2) The application intended must be consistent with the model's original purpose. 3) Validation must be established through the assessment of confidence that the predictions are accurate.

The goal of model validation is therefore to specify the uncertainty in a prediction made by a simulation code for a hypothetical new experimental situation. Such definition confers a central role to the assessment, or quantification, of uncertainty. The concept is illustrated in Figure 9. At the design point 1, a validation experiment is performed and the total error between measured and predicted responses can be estimated. The central question is to estimate the total prediction error at the second design point where no experimental data are available.

Uncertainty quantification (UQ) of a code's output provides the metric needed to specify the degree of agreement between the simulation prediction and reality. The ultimate goal of the UQ process is to construct an uncertainty model for every component of the simulation code, which taken all together summarize how well the code's predictions agree with all available experimental results. This same set of uncertainty models is used to estimate the uncertainties in code prediction of a new application (see Figure 9). It is emphasized that uncertainty models do not necessarily have to 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 experimental data [16]. In the remainder, however, statistical models of uncertainty are assumed for simplicity.

We have seen that a pre-requisite to model validation is that the total error between physical observation and model prediction be characterized. A possible path forward is now discussed. One approach is to break down the total error into individual components and estimate their probability information

$$\begin{aligned} y_{Measured} &= y_{Predicted} + e_{Total} \\ y_{Predicted} &= M(p_1;p_2) \\ e_{Total} &= N(0;S_{Total}) \end{aligned} \tag{4}$$

For simplicity the error model is assumed Gaussian in equation (4). The main difficulty is that such error model

$N(0;S_{Total})$ must be derived over the entire design space—which means that the total variance S^2_{Total} is a function of the model's input parameters $(p_1;p_2)$—with limited validation experiments. The analysis of a single test, which is typical of a calibration experiment, will not permit to derive an error model valid over the entire design space. The total variance S^2_{Total} between measurements and predictions can be decomposed if independent Gaussian processes are assumed

$$S^2_{Total} = S^2_T + S^2_D + \sum_{i=1,2} S^2_i + S^2_M \qquad (5)$$

Equation (5) states that there are several independent components that contribute to the total error. For example the total error might include a measurement error of variance S^2_T, discretization error of variance S^2_D and parametric variability of variance S^2_j—S_j denotes the output feature's standard deviation due to variability of the j^{th} parameter p_j. The total variance S^2_{Total} can be obtained from a comparison of measured and predicted responses for a design of experiments that attempts to explore the input space as much as possible. Components such as S^2_T and S^2_D are estimated by investigating the measurement system and mesh convergence properties, respectively. The variability S^2_j of the output due to input parameter uncertainty is typically identified through an input-output effect analysis [8, 14, 17]. In equation (5) the only term that remains unknown, S_M, represents the residual sources of uncertainty that include, for example, model form error. Obtaining an estimation of model form error is critical to assess the validity of the numerical model over its domain of applicability. Once available the probability information $N(0;S_M)$ can be combined with the code's output to assess confidence bounds associated with a new prediction of the model.

5. CONCLUSION

Model calibration—also known as finite element model updating in structural dynamics—is discussed in the broader context of model validation. Formal definitions are proposed and the notions of calibration and validation are illustrated using one example of transient structural dynamics. A distinction is introduced between input parameters and calibration variables. The former define the input space in which the model must be exercised and validated. The latter are generally introduced by the modeling process and should be eliminated through calibration and direct measurement as much as possible.

A possible formulation of model validation is introduced. It is based on the assessment of total discrepancy between test data and numerical predictions. The total error is broken down into independent components, each evaluated over the entire design space if enough experiments are available. This methodology is currently being pursued with the hyper-foam impact experiment and preliminary results will be reported in future publications.

REFERENCES

[1] Garcia, J., "The Need For Computational Model Validation," *Experimental Techniques*, Vol. 25, No. 2, Mar.-Apr. 2001, pp. 31-33.

[2] Doebling, S.W., Hemez, F.M., Schultze, J.F., Cundy, A.L., "A Metamodel-based Approach to Model Validation for Nonlinear Finite Element Simulations," *20th International Modal Analysis Conference*, Feb. 2002.

[3] Lallement, G., Cogan, S., "Reconciliation Between Measured and Calculated Dynamic Behaviors: Enlargement of the Knowledge Space," *10th International Modal Analysis Conference*, Feb. 1992, pp. 487-493.

[4] Mottershead, J.E., Lallement, G., "Vibration Nodes and the Cancellation of Poles and Zeros by Unit-rank Modifications to Structures," *Journal of Sound and Vibration*, Vol. 222, 1999, pp. 833-851.

[5] Kennedy, M.C., O'Hagan, A., "Bayesian Calibration of Computer Models," *Journal of the Royal Statistical Society (Series B*), Jan. 2001. Download at **http://www.shef.ac.uk/~st1ao/**.

[6] Schultze, J.F., Hemez, F.M., Doebling, S.W., Sohn, H., "Statistical-based Nonlinear Model Updating Using Feature Extraction," *19th International Modal Analysis Conference*, Feb. 2001, pp. 18-26.

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

[8] Myers, R.H., Montgomery, D.C., **Response Surface Methodology: Process and Product Optimization Using Designed Experiments**, Wiley Inter-science, New York, 1995.

[9] Hemez, F.M., Doebling, S.W., "Review and Assessment of Model Updating for Nonlinear, Transient Dynamics," *Journal of Mechanical Systems and Signal Processing*, Vol. 15, No. 1, Jan. 2001, pp. 45-74.

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

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

[12] Kakizawa, Y., Shumway, R.H., Taniguchi, M., "Discrimination and Clustering for Multivariate Time Series," *Journal of the American Statistical Association*, Vol. 93, No. 441, Mar. 1998, pp. 328-340.

[13] 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.

[14] Hedayat, A.S., Sloane, N.J.A., Stufken, J., **Orthogonal Arrays: Theory and Applications**, Springer-Verlag, New York, 1999.

[15] Schlesinger, S., "Terminology for Model Credibility," *Simulation*, Vol. 32, No. 3, 1979, pp. 103-104.

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

[17] Chipman, H., Hamada, M., Wu, C.F.J., "A Bayesian Variable Selection Approach for Analyzing Designed Experiments With Complex Aliasing," *Technometrics*, Vol. 39, No. 4, Nov. 1997, pp. 372-381.