LA-UR-01-2492
Approved for public release; distribution is unlimited.

Title: **UNCERTAINTY, VALIDATION OF COMPUTER MODELS AND THE MYTH OF NUMERICAL PREDICTABILITY**

Author(s): François M. Hemez and Scott W. Doebling

Submitted to:

http://lib-www.lanl.gov/la-pubs/00818314.pdf

UNCERTAINTY, VALIDATION OF COMPUTER MODELS AND THE MYTH OF NUMERICAL PREDICTABILITY[1]

François M. Hemez[2] and **Scott W. Doebling**[3]

Engineering Analysis Group (ESA-EA)

Los Alamos National Laboratory

P.O. Box 1663, Mail Stop P946, Los Alamos, New Mexico 87545

ABSTRACT

This publication addresses the issues of modeling, uncertainty quantification, model validation and numerical predictability. With the increasing role of numerical simulation in science, technology as well as every day decision-making, assessing the predictive accuracy of computer models becomes essential. Conventional approaches such as finite element model updating or Bayesian inference are undeniably useful tools but they do not fully answer the question: ***How accurately does the model represent reality?*** First, the evolution of scientific computing and consequences in terms of modeling and analysis practices are discussed. The intimate relationship between modeling and uncertainty is explored by defining uncertainty as an integrate part of the model, not just parametric variability or the lack of knowledge about the physical system being investigated. Examples from nuclear physics, climate prediction and structural dynamics are provided to illustrate issues related to uncertainty, validation and predictability. Feature extraction or the characterization of the dynamics of interest from time series is also discussed. Finally, a general framework based on response surface methodology is proposed for the fusion of model predictions, validation data sets and uncertainty analysis.

---

[1] Approved for unlimited, public release on May 14th, 2001. LA-UR-01-2492. This publication is unclassified.

[2] Technical Staff Member, hemez@lanl.gov (E-mail), 505-665-7955 (Phone), 505-665-7836 (Fax).

[3] Technical Staff Member, doebling@lanl.gov (E-mail), 505-667-6950 (Phone), 505-665-2137 (Fax).

I. INTRODUCTION

Today's computational resources make it more than ever possible to model and analyze complex 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 x 10^{+12} operations per second (or 3 TeraOps) by distributing computations over arrays of more than 6,000 processors. The next generation of ASCI computers is expected to reach 30 TeraOps by the year 2004 with the goal of approaching 100 TeraOps a few years later. 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. Reference [1] discusses the overall ASCI program and its objectives.

In addition to improving computational resources, as we know them today, groundbreaking discoveries are being made in the area of quantum computing, a field thought to be an elegant but impractical theory only a few years ago. This technology enables scientists to store information (bits of zeros and ones) as positive or negative spins of elementary particles that form the building blocks of molecules. Immediate and obvious advantages are infinitely large memory sizes and rapidity of access to the information bounded only by the speed of light. Moreover, the theory of quantum mechanics states that an elementary particle may feature positive and negative spin values simultaneously. Thus, a single particle may potentially store two bits of information at once. Just like a "conventional" computer combines analog bits to perform an operation, a quantum computer would combine the spin values of its elementary particles to add and multiply numbers or search a database. Since one particle can store two pieces of information, two particles can access 2^2 bits. If a very small number of particles can be stabilized, say, no more than 1,000, then this quantum computer could potentially access 2-to-the-power-10^{+3} bits simultaneously. Assuming that the multiplication of two 256-digit numbers involves 10^{+5} bits of information, this translates into 10^{+48} TeraOps of computing power for a single molecule! Practical difficulties such as verification algorithms and the instability of this information storage

technology (typically, a particle may randomly change its spin value as often as every 10^{-15} seconds) are currently being addressed at Los Alamos and other national laboratories and university research centers.

Quantum computing will probably not offer any practical outcome for several decades but it is undeniable that unprecedented computational resources are becoming available. What will be the impact on our modeling capabilities and analysis practices?

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 numerical simulation. For example, physicists and mecanicians are increasingly involved in the development and implementation of constitutive models at the microscopic and nano-scale levels based on basic physics (or "first principles") such as statistical quantum mechanics. The intent is to capture the physics of interest at the source rather than relying on global and somewhat arbitrary quantities generally defined in solid mechanics such as modal damping ratios.

Much of the effort within ASCI-like programs is currently devoted to proving that complex phenomena can be modeled mathematically, adequate solution procedures can be implemented and computer systems are robust and stable enough to sustain the amount of computations requested. In addition, scientists are increasingly becoming concerned with the predictive accuracy of their numerical models. This emerging field is often referred to as ***model validation***. Here, the central question is: ***How accurately does the model represent reality?*** It consists of determining the predictive quality of numerical simulations and assessing the degree of confidence with which models can be analyzed outside of their nominal operating conditions. Caution must however be exercised to avoid the common confusion between model ***verification*** and model ***validation***. For example, consider ancient Greek astronomy. The models developed by Pythagoras, Aristotle and Ptolemy between 500 BC and 300 BC dominated Western astronomy for nearly 2,000 years. They exhibited various levels of complexity but all shared the characteristic that our Sun and other planets of our solar system did gravitate around the Earth. These models were somewhat consistent with each other, they fulfilled their purpose of predicting

with remarkable accuracy the cycles of seasons and they even matched physical observation. Yet, they were not accurate representation of reality.

In this work, we conform to the U.S. Department of Energy's definition of verification and validation where, basically, ***verification*** consists of verifying that equations are solved correctly while ***validation*** consists of verifying that the equations implemented provide an acceptable representation of reality. This publication discusses the concepts of modeling, uncertainty, model validation and predictability in the context of large-scale numerical simulations. The discussion is illustrated using an engineering application currently dealt with at Los Alamos National Laboratory. References [2-4] offer additional details regarding the particular analysis techniques and results to which the discussion refers.

II. CONCEPTUAL VIEW OF MODELING AND UNCERTAINTY

Uncertainty is defined as the omitted or unknown part of a mathematical model. This definition is somewhat different from the parametric variability or lack-of-knowledge views generally agreed upon in the scientific community. It is also implied that numerical simulations should always include a representation of the uncertainty associated with a particular model. This is consistent with our approach to model validation that states that there is no such thing as model "validation" because all that statistical testing can assess is the degree to which a model breaks down, not the degree to which a model works.

2.1 Where Does Uncertainty Come From?

To illustrate how model order truncation and uncertainty are related to the process of modeling a given phenomenon, we consider the example of two elementary particles interacting with each other. This is a common problem in quantum mechanics further complicated by extreme uncertainty (the well-known Heisenberg principle states that position accuracy is bounded by momentum uncertainty, and vice-versa) and large scaling differences (heavy particles interact with much lighter particles). Credit must be given to Reference [5] for originally discussing this example in the context of uncertainty

analysis. Our system is formed of two particles, $\mathbf{X_1}$ and $\mathbf{x_2}$, interacting with each other. The first particle denoted by $\mathbf{X_1}$ is the primary degree-of-freedom of interest. Its dynamic is influenced by a secondary particle denoted by $\mathbf{x_2}$. We have no real interest in predicting the behavior of $\mathbf{x_2}$ but it is included in the equations because of the interaction with the main degree-of-freedom. The equations considered are:

$$\frac{\partial^2 \mathbf{X}_1}{\partial \mathbf{t}^2} + \mathbf{?}_1^2(\mathbf{1} + \mathbf{x}_2^2)\mathbf{X}_1 = \mathbf{0}, \qquad \mathbf{e}^2 \frac{\partial^2 \mathbf{x}_2}{\partial \mathbf{t}^2} + \mathbf{?}_2^2(\mathbf{1} + \mathbf{X}_1^2)\mathbf{x}_2 = \mathbf{0} \tag{1}$$

In the numerical application, values of $\mathbf{?}_1$ and $\mathbf{?}_2$ are kept constant and equal to one. The initial conditions of particle $\mathbf{X_1}$ are set to one for displacement and zero for velocity. Initial conditions in both displacement and velocity for particle $\mathbf{x_2}$ are uncertain and vary uniformly among 13 discrete values $\{10^{-1}; 3\text{x}10^{-1}; 7\text{x}10^{-1}; 1; 3; 7; 10^{+1}; 3\text{x}10^{+1}; 7\text{x}10^{+1}; 10^{+2}; 3\text{x}10^{+2}; 7\text{x}10^{+2}; 10^{+3}\}$. Another important characteristic of the system of equations (1) is that the secondary particle is very light compared to the primary one. The mass of the $\mathbf{x_2}$ is chosen $\mathbf{e^2 = 10^{-6}}$ times smaller than the mass of $\mathbf{X_1}$. This introduces ill-conditioning and convergence difficulties when, for example, Runge-Kutta finite differences are implemented to solve the system of partial differential equations (1).

Figure 1 illustrates the output when equations (1) are integrated numerically in time. The top figure shows the position $\mathbf{X_1}$ and the bottom figure shows the position $\mathbf{x_2}$. Note the large amplitude difference. Displacements of the lighter particle $\mathbf{x_2}$ are approximately three orders of magnitude smaller. As the **e**-parameter tends to zero (**e ? 0**), the second mass becomes smaller relative to the first one and the response $\mathbf{x_2}$ becomes rapidly varying. This makes it suitable for statistical treatment, as illustrated in the following.

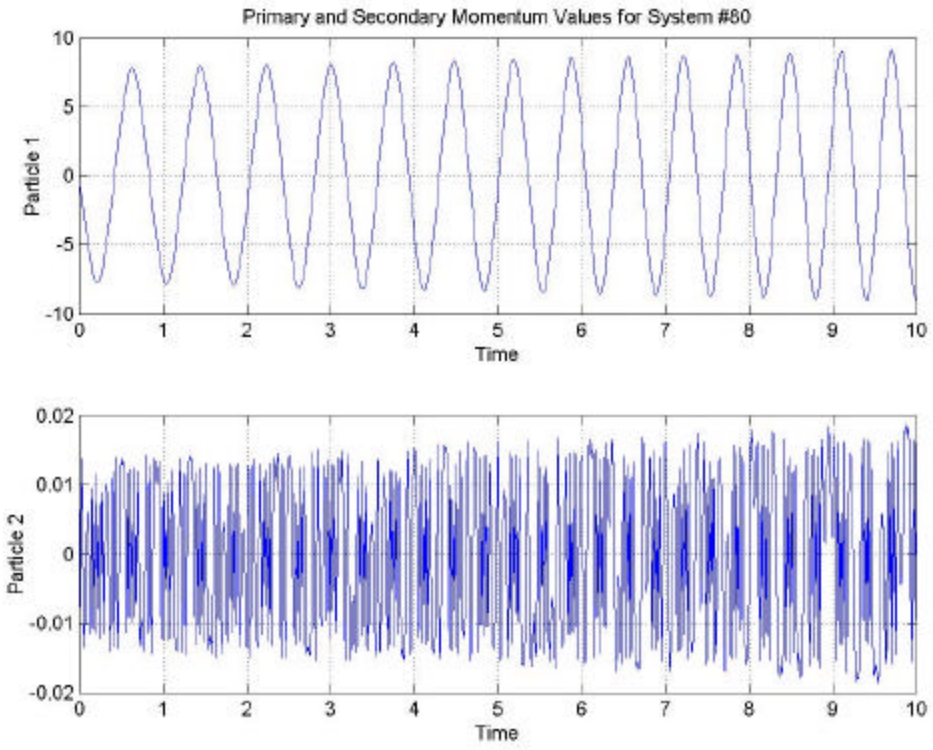


Figure 1. Responses of the Two Degrees-of-freedom System.

(Top: Displacement of the heavy particle X_1. Bottom: Displacement of the light particle x_2. Initial conditions are set to {1; 7} for displacements and {0; 0.3} for velocities.)

Clearly, equation (1) for the heavy particle is analogous to a simple mechanical oscillator to which an internal force is added:

$$\frac{\partial^2 \mathbf{X}_1(t)}{\partial t^2} + ? \,{}_1^2 \mathbf{X}_1(t) + \mathbf{F}_1(t) = \mathbf{0} \qquad (2)$$

The internal force $\mathbf{F_1(t)}$ might be nonlinear and possibly stochastic. In general, the relationship between the degrees-of-freedom of interest (in this case, the heavy particle $\mathbf{X_1}$) and their environment is not exactly known. Similarly, it may be concluded from empirical observation that other variables have little to no significant influence. In any case, a widely accepted modeling practice is to approximate or simply

omit the coupling term $\mathbf{F_1(t)}$. Thus, uncertainty can originate from several actions commonly taken during modeling:

(a) Selecting an inadequate model form:

If the nature of the internal force is not known precisely, an inadequate model form is likely to be implemented. For example, a linear model $\mathbf{F_1(t) = k\ X_1}$ or a cubic model $\mathbf{F_1(t) = k\ X_1^3}$ might be implemented that introduces some discrepancy between the "true" system (1) and its mathematical representation (2).

(b) Truncating the model order:

Secondary dynamics are generally truncated when they are not believed to influence significantly the primary dynamics. Order truncation is a common practice in numerical modeling. Truncation in our example would consist of restricting the degrees-of-freedom to $\mathbf{X_1}$ only and condensing the information represented by the second of equations (1) into $\mathbf{F_1(t)}$.

(c) Approximating equations:

Equations are often approximated for computational efficiency, model order truncation or because the exact functional relationship is somewhat unknown. For example, the internal force $\mathbf{F_1(t)}$ might be defined as the expected value of the coupling term $\mathbf{(x_2)^2 X_1}$ between $\mathbf{X_1}$ and $\mathbf{x_2}$ given the knowledge of the system's position-momentum states (and $\mathbf{? _1 = ? _2 = 1}$):

$$\mathbf{F_1(t)} = \mathbf{E}\left[\mathbf{x_2^2 X_1} \mid \mathbf{X_1; \dot{X}_1}\right] = \frac{\mathbf{2} - \left(\mathbf{X_1^2} + \mathbf{\dot{X}_1^2}\right)}{\mathbf{2}\left(\mathbf{1} + \mathbf{X_1^2}\right)}\mathbf{X_1} \tag{3}$$

This approximation is justified because the rapidly varying degree-of-freedom $\mathbf{x_2}$ can be estimated statistically instead of carrying out the fully coupled resolution. The impact on computational requirements is significant even for this simple system, as noted in Reference [5]. Typically, integrating in time the coupled equations (1) requires a time step proportional to the mass of the light particle or $\mathbf{?t \tilde{} O(e^2)}$. In comparison, integrating equations (2-3) with the same absolute accuracy requires $\mathbf{?t \tilde{} O(e)}$ at the most. Combining equations (2) and (3) effectively defines an approximated system where the original differential equations (1) are solved "in average" instead of exactly.

(d) Introducing parametric uncertainty:

In equation (2), the (unknown) term $\mathbf{F_1(t)}$ might be replaced with a random process. This would be a likely modeling option when the functional form of the interaction between degrees-of-freedom $\mathbf{X_1}$ and $\mathbf{x_2}$ is not known precisely. For example, the internal force can be estimated as a Gaussian process, $\mathbf{F_1(t) \tilde{} N(\mu; s)}$, whose mean $\boldsymbol{\mu}$ and standard deviation $\mathbf{s}$ parameters are defined as the first and second statistical moments, respectively, of the coupling term $\mathbf{(x_2)^2 X_1}$. Information that defines such random processes might come from "first principles" physics, scaling arguments and other ad-hoc reasoning, physical experimentation or hypotheses and modeling assumptions. In any case, random processes generally depend on uncertain parameters. These hyper-parameters might include the mean, variance and structure of the covariance matrix. This is a typical example of parametric uncertainty. A specific model is agreed upon (probability, possibility, information-gap) and parameters of the model become the means by which uncertainty is introduced.

2.2 Complete Characterization of an Uncertain System

The coupled equations (1) are first solved for all possible combinations of initial conditions of particle $\mathbf{x_2}$. This analysis illustrates a situation where enough computational or experimental resources are available to conduct a full characterization of the uncertain system. The position and momentum of particle $\mathbf{x_2}$ can assume 13 discrete values each, which leads to a full factorial analysis of $13^2 = 169$

systems. Figure 2 illustrates the position of particle $\mathbf{X_1}$ versus time obtained for the 10th, 60th, 100th and 150th systems. It can be observed that the system of equations (1) spans a wide range of dynamics. Responses range from linear, single degree-of-freedom (top left) to linear with increasing damping (top right) and time varying, higher-frequency harmonics (bottom left). The forth response shown (bottom right) exhibits chaotic behavior and a component that eventually grows unstable. Being able these dynamics is an important step of model validation, as discussed in section V.

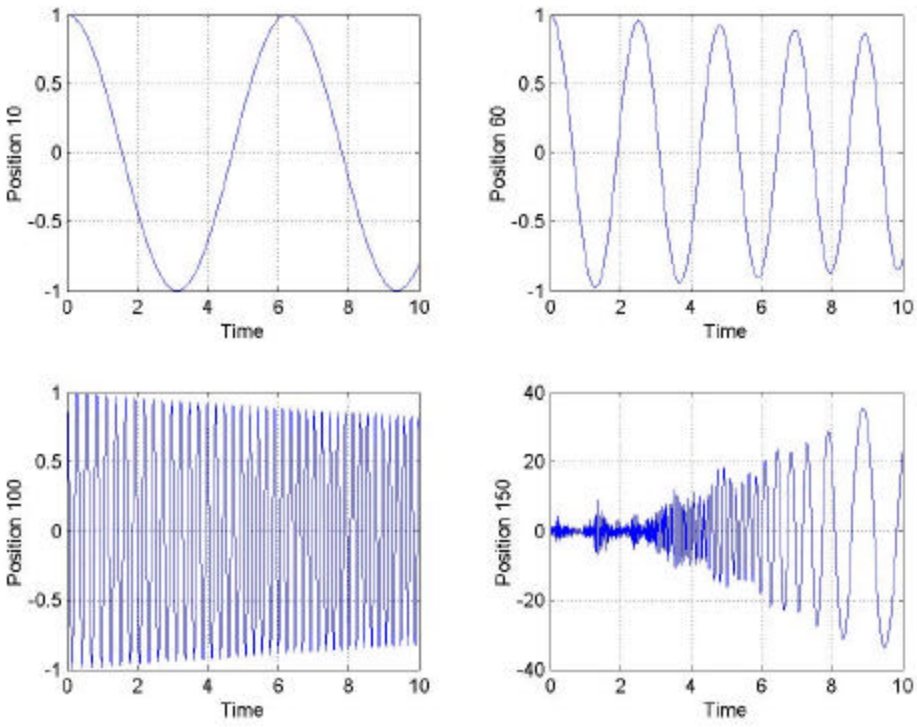


Figure 2. Four Typical Responses of the Two Degrees-of-freedom System.
(Initial conditions for X_1 and its momentum are kept constant and equal to (1; 0). Initial conditions for x_2 and its momentum are as follows. Top Left: IC = (10^{-1}; 10^{-3}). Top Right: IC = (1; $3x10^{-5}$). Bottom Left: IC = (10; $7x10^{-5}$). Bottom Right: IC = (10^{+2}; 10^{-5}).)

The information generated by solving the system of equations (1) for all possible combinations of initial conditions is summarized in Figure 3. It shows the most probable state of each particle in the position-momentum plane. The most probable states of primary particles $\mathbf{X_1}$ are shown on the top half

while those of secondary particles $\mathbf{x}_2$ are shown on the bottom half. Hence, Figure 3 illustrates the output variability obtained by propagating uncertain initial conditions through the forward calculation. Properties of the joint position-momentum distributions can be analyzed from the output generated by this computer experiment. Other design-of-experiment techniques are available to propagate variability such as fractional factorial, orthogonal array or Latin Hypercube sampling. It is emphasized however that, in general, a complete characterization of the output space such as illustrated with this example is not possible due to limited knowledge of the system and limited computational and experimental resources.

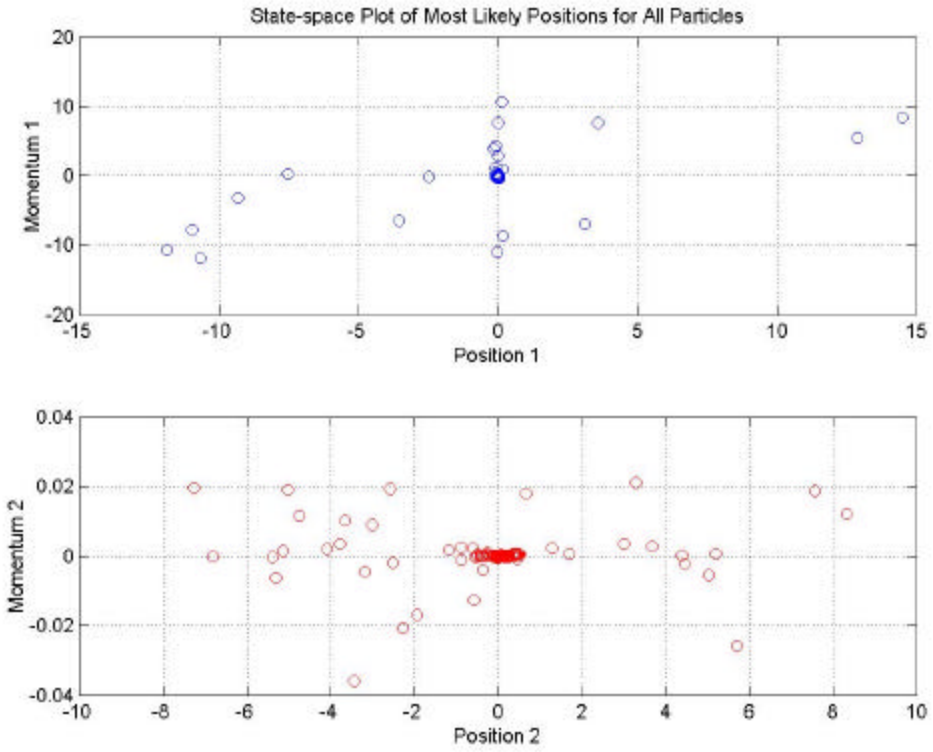


Figure 3. Most Probable Position-Momentum States of Particles X_1 and x_2.

(Top: Mean state values of particle $\boldsymbol{X_1}$. Bottom: Mean state values of particle $\boldsymbol{x_2}$. Note that numerical values are plotted on different horizontal and vertical scales. Also noticeable are the different correlation structures of the joint probability distributions of particles $\boldsymbol{X_1}$ and $\boldsymbol{x_2}$.)

2.3 Approximation of an Uncertain System

The dynamics of particle $\mathbf{X_1}$ is now represented as a single degree-of-freedom linear oscillator. The coupling with particle $\mathbf{x_2}$ is completely ignored and replaced by a random process. This illustrates situations where the model order is unknown or the interaction between some of the degrees-of-freedom is not modeled correctly. Obviously, integrating the time response of particle $\mathbf{X_1}$ based on a linear oscillator representation would result into large prediction errors. However, unavailable modeling information can advantageously be replaced by probabilistic information. Figure 4 pictures the most probable position-momentum states of the 169 systems for the full factorial analysis. Solutions obtained with uniform distributions are shown on the top half and solutions obtained with normal distributions are shown on the bottom half. Both results can be compared to the "true" solution shown in Figure 3 (top half). A comparison of Figures 3 and 4 shows that the correlation structure between position and momentum is lost. This is expected because the physics-based coupling is replaced with an arbitrary random process. Nevertheless, the approximated solutions are consistent with the true solutions. Another advantage is that the linear oscillator equation can be solved in a fraction of the time required to integrate the coupled equations (1) because it is a decoupled as well as well-posed problem.

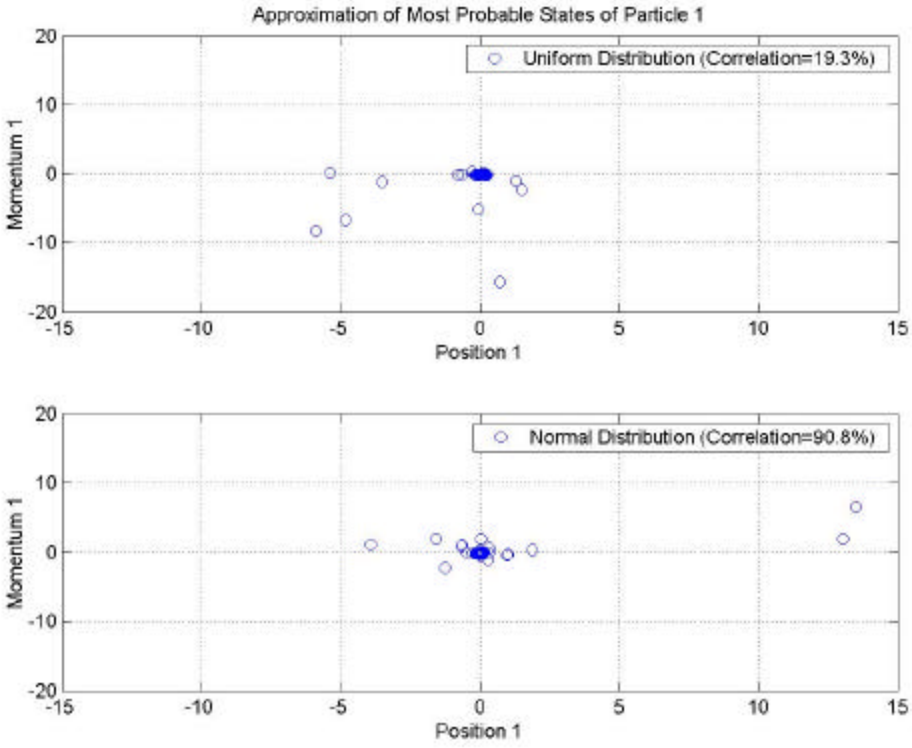


Figure 4. Most Probable States of Particle X_1 With Coupling Approximated.

(Mean position and momentum values of X_1 when a random process replaces the coupling with x_2. Top: Uniform, $F_1(t) \sim U[\mu-2s;\ \mu+2s]$. Bottom: Normal, $F_1(t) \sim N(\mu;s)$. For each one of the 169 systems analyzed, the statistics are computed as $\mu=E[(x_2)^2X_1]$ and $s=s[(x_2)^2X_1]$.)

The two-particle example illustrates that potentially missing information can be replaced by adequate statistical treatment. Since a random process can generally be parameterized, hyper-parameters such as the mean, variance, covariance structure and higher-order statistics can be calibrated to improve the predictability of the computer simulation. An illustration of model calibration via Bayesian inference is provided in section III.

This simple example also illustrates our opinion that uncertainty can never be dissociated from modeling. Modeling consists of formulating hypotheses and the likelihood of assumptions and modeling rules should be included in the analysis of the numerical simulation's output. The process of explaining a complex phenomenon by mathematical models generates uncertainty. This implicitly defines model

validation as an exercise where the consistency between model output and reality is assessed away from the model's nominal operating ranges.

We have mentioned that missing information can be replaced by probabilistic information. Other frameworks are available for quantifying and propagating uncertainty that may offer attractive alternatives to the theory of probability especially in the event of extreme uncertainty. Among them, we cite the Dempster-Shafer theory of possibility and belief [6], the theory of fuzzy sets [7], information gap theory and convex models of uncertainty [8]. In the remainder, uncertainty is represented by probability density functions. This is a reasonable assumption when dealing with physics or engineering applications where reasonably large amounts of test data are available and the systems investigated are governed by well-established formal theories.

III. PREDICTABILITY IN COMPUTATIONAL SCIENCES

Currently, all computational sciences are, to various degrees, struggling with the notion of numerical predictability, uncertainty quantification and model validation. The reason is because scientists are increasingly relying on numerical models to make predictions and replace physical measurements. As the computer models grow in size and complexity, so does the need to assess their validity especially when full-scale testing is not available. To illustrate these trends, two examples taken from communities other than structural dynamics are discussed. The first example addresses global climate prediction. The second one illustrates the inference from uncertain measurements of a time-varying parameter in nuclear physics.

3.1 Climate Modeling Via Large-scale Computing

The Parallel Ocean Program (POP) was developed at Los Alamos National Laboratory under the sponsorship of the U.S. Department of Energy in an attempt to bring massively parallel computers to the realm of climate modeling. The main purpose of these simulations is to predict global climate trends over

time periods of 10-to-1,000 years [9]. Clearly, this is not something that can be achieved through laboratory experiments. Such endeavor shares several characteristics of the ASCI-class simulations, namely, high-fidelity geometry, high-fidelity physics and lack of physical experiments.

The equations solved for ocean and atmospheric simulations are typically Navier-Stokes equations with coupling with other phenomena such as heat flux, chemical reactions and season varying climatology. Multi-physics modeling that aims at interfacing ocean, ice, continent and atmospheric models currently generates great interest in the scientific community. Because the computational domain can be as large as the entire planet and local turbulent phenomena must be captured to predict global trends, the discretized problems are generally very large. For example, the models currently analyzed feature discretization sizes typically equal to 0.28 degrees in longitude, 0.17 degrees in latitude, up to 20 non-uniformly spaced depth levels and realistic bottom topography, which yields more than 200 Million degrees-of-freedom. Another characteristic of climatology and oceanography is that useful physical measurements are very difficult to come by. Because of cost and difficulty of access, measurement stations are available with adequate density only over the land and in populated areas. This is however marginally useful because oceans are mostly responsible for governing global climate trends. Even when measurements can be made over the ocean, for example, through satellite observation, very rarely are they available through the ocean's thickness. When physical measurements are available, model validation generally takes the form illustrated in Figure 5.

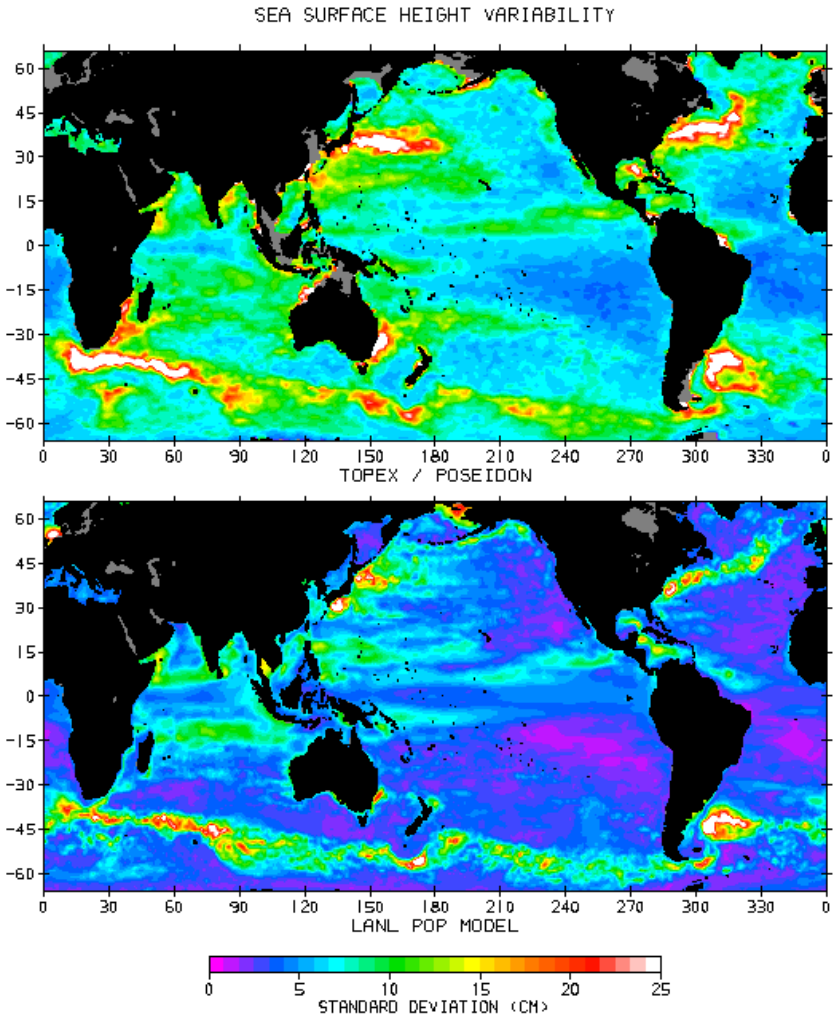


Figure 5. Comparison of TOPEX/Poseidon Measurement and LANL/POP Prediction.

(The figure shows the standard deviation of variability of sea surface heights. Top: Observation of the TOPEX/Poseidon satellite. Bottom: Prediction of the LANL/POP model. Discrepancies are located in regions where sea surface height variability is important, therefore, reflecting the fact that the numerical model captures mean behaviors better than extreme events.)

Figure 5 compares the RMS variability of the sea surface height measured by the TOPEX/Poseidon satellite (top half) and predicted by the LANL/POP model (bottom half), as reported in Reference [10]. Observed peak variability occurs in regions of strong currents such as the Gulf Stream, the Kuroshio Current and the Antarctic Circumpolar Current. The POP model has its maximum variability in the same general areas but amplitudes are too small by up to a factor of two. Nevertheless, the so-called "view-graph norm" where images or curves are compared visually provides little quantitative insight. For example in Figure 5, analysts could equally argue that the numerical model captures the main features of interest or, to the contrary, that some important characteristics are not predicted well enough. Low-dimensionality is necessary because analysts cannot visualize nor handle the vast amounts of data generated by large-scale simulations. Another reason is that the efficiency of multivariate statistical analysis rapidly decreases with the number of variables analyzed. This leads to the notion of ***feature extraction*** discussed in section V.

3.2 Bayesian Inference in Nuclear Physics

The second example illustrates a calibration experiment whose purpose is to infer the value of a time-varying parameter by maximizing the statistical consistency between physical observation and model output. This presentation summarizes the work originally presented in Reference [11].

When fissionable materials are assembled, the system can become critical, that is, neutron fluxes can grow exponentially. The measure of this criticality is a parameter known as the Rossi alpha. It is defined as the rate at which the neutron flux grows:

$$\mathbf{a(t)} = \frac{\mathbf{1}}{\mathbf{y(t)}} \frac{\partial \mathbf{y(t)}}{\partial \mathbf{t}} = \frac{\partial \mathbf{ln(y(t))}}{\partial \mathbf{t}} \tag{4}$$

In equation (4), the symbols **a(t)** and **y(t)** denote the Rossi alpha parameter and the neutron flux, respectively. To develop a numerical model of criticality, the value of **a(t)** must be inferred from the measurement of Rossi traces $\mathbf{y(\cos(2\pi f_R t))}$. Because the neutron flux increases rapidly during a

criticality experiment, the time variable is replaced by a pseudo-time $\mathbf{cos(2pf_R t)}$. Figure 6 illustrates the typical amplitude growth of a neutron flux.

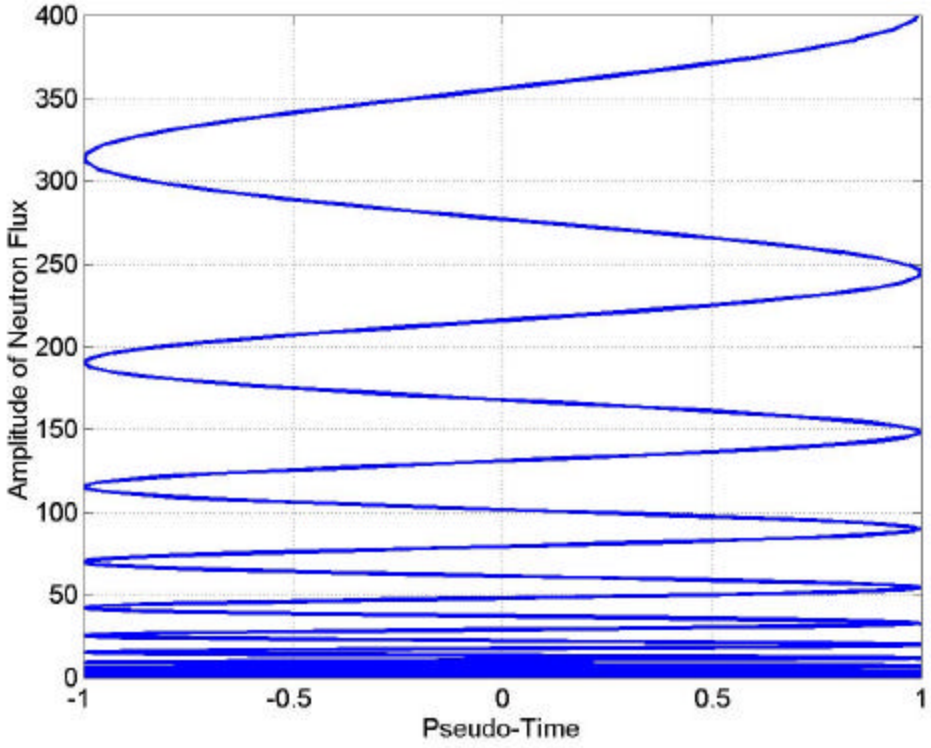


Figure 6. Amplitude Growth of a Neutron Flux During Criticality.

(The amplitude of a neutron flux is illustrated on a log-scale as a function of $\boldsymbol{cos(2pf_R t)}$ *where* $\boldsymbol{f_R}$ *denotes the Rossi frequency. The shape of the curve is characteristic of physical measurements, however, the values shown here are numerically simulated for this illustration.)*

The inverse problem thus consists of obtaining the value of $\mathbf{a(t)}$ that best reproduces the test data $\{\mathbf{x_k}; \mathbf{y_k}\}$. A formulation among others is Bayesian inference. It has shown great success for this particular application and details about the procedure can be obtained from References [11-12]. First, a parametric model is chosen for representing the unknown function $\mathbf{a(t)}$. This model translates prior knowledge about the Rossi alpha and depends on unknown parameters that are collectively denoted by the symbol $\mathbf{?}$. Next, the sources of uncertainty must be assessed and propagated. The main uncertainty

for this application resides in the placement of data points from measurement readings. Each point shown in Figure 6 is typically associated with uncertainty in the x-direction and y-direction. Standard deviation values corresponding to the placement of data $\mathbf{x_k}$ and $\mathbf{y_k}$ are denoted by $\mathbf{s_x}$ and $\mathbf{s_y}$, respectively. For simplicity, it is assumed that uncertainty in the placement of $\mathbf{x_k}$ and $\mathbf{y_k}$ points is uncorrelated. A natural metric that expresses the "distance" between test data and model output is defined by the minus-log likelihood of the observed data given the current model:

$$\mathbf{\log P(x_k; y_k \mid a)} = -\sum_{\mathbf{k=1\cdots N_{data}}} \left(\frac{\mathbf{x_k^{Test} - x_k(a)}}{\sqrt{2}\mathbf{s_x}} \right)^2 + \left(\frac{\mathbf{y_k^{Test} - y_k(a)}}{\sqrt{2}\mathbf{s_y}} \right)^2 \tag{5}$$

Equation (5) assumes that the statistical distributions are Gaussian, which is not a general requirement. Anther potential source of uncertainty, not accounted for in equation (5), is the nature of the parametric model used to represent the Rossi alpha **a(t)**. Smoothness parameters are typically included in the set of hyper-parameters **?** to control the prior knowledge. Prior knowledge plays the same role as "regularization" in the Tikonov theory of ill-posed problems. Inference of the model **a(t)** and, potentially, inference of the unknown hyper-parameters **?**, is achieved according to the Bayes law that states that the posterior distribution of the model given the available data or $\mathbf{P(a \mid x_k; y_k)}$ is equal to the likelihood function $\mathbf{P(x_k; y_k \mid a)}$ multiplied by the prior distribution **P(a)**:

$$\mathbf{P(a \mid x_k; y_k) = P(x_k; y_k \mid a)\, P(a)} \tag{6}$$

The objective naturally becomes to maximize the posterior distribution, which translates that the model sought is the one that is most consistent with test observation. Similarly, the posterior distribution of hyper-parameters given the data or $\mathbf{P(? \mid x_k; y_k)}$ can be maximized to infer the value of hyper-parameters **?** that are most consistent with test data. Any optimization solver can be implemented to maximize the posterior distributions $\mathbf{P(a \mid x_k; y_k)}$ and $\mathbf{P(? \mid x_k; y_k)}$. One particularly attractive choice is the Markov Chain Monte Carlo (MCMC) algorithm. The MCMC method generates a random sequence of parameters **a(t)** that samples the posterior distribution $\mathbf{P(a \mid x_k; y_k)}$. The main advantage of

the MCMC optimization is that sampling can be carried out independently of the distribution being sampled. In particular the assumption of normal probability distribution, which is encountered in many formulations for the only purpose of allowing tractable analytical derivations, is not required. For more details, the reader is referred to Reference [13] where a tutorial of MCMC methods is provided. Figure 7 illustrates three realizations of **a(t)** obtained through MCMC sampling of the Bayesian posterior distribution (6).

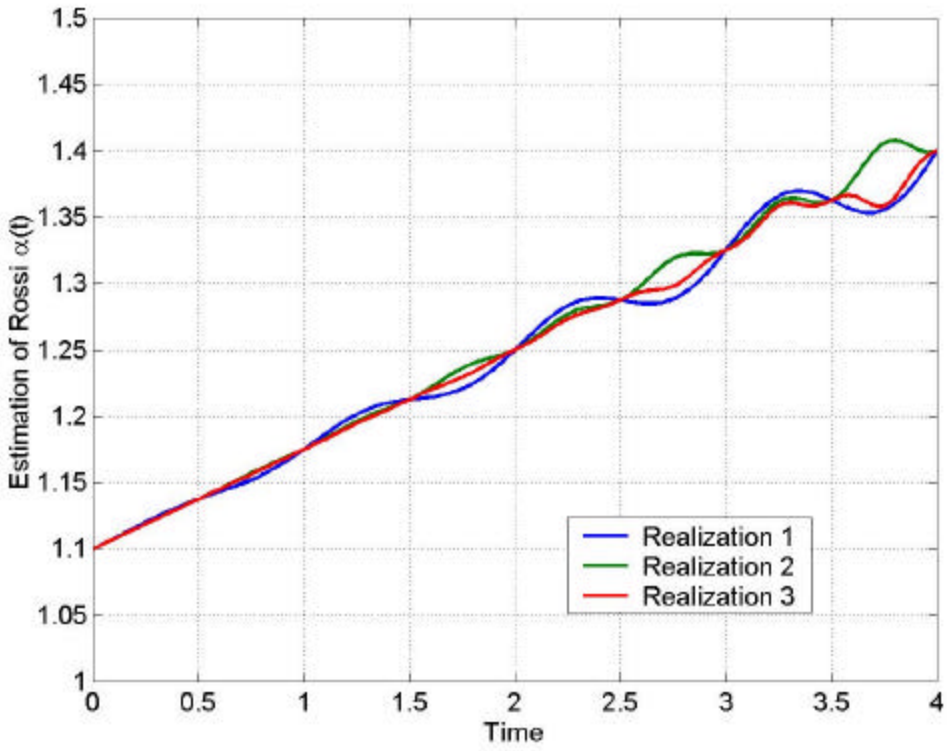


Figure 7. MCMC Inference of Rossi Alpha a(t) From the Bayesian Posterior.

*(The figure illustrates that several optimal solutions can be obtained that are statistically consistent with physical measurements. If enough independent samples can be drawn from the posterior distribution, basic statistics about the inferred parameter **a(t)** can also be estimated.)*

The purpose of this example is to illustrate inverse problem solving. In the presence of uncertainty, several optimal solutions can be obtained that remain consistent with the physical observation. Exploring

the posterior distribution therefore becomes critical. If enough independent samples can be drawn from the posterior probability distribution, basic statistics about the inferred parameters—mean, variance, covariance structure, etc.—can be estimated. Hence, the uncertainty observed through physical experimentation can be related to parametric variability of the model, which is important information for design and decision-making.

Bayesian inference represents one of many possible formulations among which we cite maximum likelihood, Mahanalobis hypothesis testing, Kullback-Leibner entropy and Chernov entropy. Statistical techniques developed for hypothesis testing or group classification can generally be applied to formulate inference problems. Methods such as finite element model updating, parameter identification and parameter calibration (also referred to as parameter "tuning") all fall under this general description. However, we stress that a ***calibrated*** model by no means constitutes a ***validated*** model, as commonly accepted in the finite element updating community. Validation is fundamentally a statement about ***predictability*** whereas calibrated models are, at best, capable only to match physical measurements at one or several design points.

IV. ASCI ENGINEERING APPLICATIONS

Quantifying shock transmission through complex, jointed structures has traditionally been possible only with experimental methods. These experiments are expensive and time-consuming and thus only a few cases can be studied. With the advent of large scale computing capabilities, estimation of the shock transmission with numerical models is becoming a tractable problem. The ASCI computing environment is being used at Los Alamos to study, among other things, the transmission of these shocks through complex, jointed structures. This on-going experiment is summarized to illustrate the validation of engineering applications in structural dynamics.

4.1 The Forward Mount Impulse Test

The test article consists of several components fabricated from a variety of materials. A titanium component designated the “mount” is shown in Figure 8 (top left). All other components are connected to the titanium mount. The upper payload mass simulator, which is fabricated from 6061-T6 aluminum, is bolted to the three feet on the upper end of the mount. The lower payload mass simulator, which is fabricated from carbon steel, is held inside the mount using a tapered tape joint. The tapered tape is fabricated from SS-304 stainless steel and is inserted through the thin, horizontal slot near the base of the mount. Separate pieces of the tapered tape are driven in, wedging the mass simulator against an inner retaining surface. All these components are pictured in Figure 8 (bottom left). The lower shell, fabricated from 7075-T4 aluminum, and then anodized, is placed over the titanium mount and its rim sits on a ledge just below the threaded portion of the mount. Next, a titanium retaining nut threads onto the titanium mount bearing against the upper surface of the lower shell rim. A specified torque value is applied. Finally, the upper shell, also fabricated from aluminum, is threaded onto the mount. As this second specified torque is applied, the load between the retaining nut and lower shell is somewhat reduced.

Figure 8. Forward Mount and Other Components of the Assembly.

(Top Left: Titanium forward mount. Bottom Left: Other components of the assembly. From left to right, lower shell; titanium mount and bolted mass simulator; retaining nut and upper cylindrical shell. Right: Instrumented system, explosive grid and testing fixture.)

Figure 8 (right) also shows that the test article is suspended using wire rope. This creates a pendulum with a length of about one meter. Pendulum motion is monitored using high-speed photography and fiber optic-based displacement sensing. An explosive source is developed to apply an impulsive load to a portion of the outside surface of the test article. The source is fabricated from strips of thin explosive sheet material. The explosive strips are simultaneously initiated using an explosive lens. The pressure at the surface of the test article is moderated with a buffer material made from solid neoprene. Prior to testing, the explosive load underwent a careful characterization to make sure that the correct impulse was measured. Two independent investigations indicated that the measurement system provided no more than a few percent errors.

The test article is instrumented with 33 strain gages and 6 accelerometers. The strain gages are attached to the inside surface of the titanium mount and have an active length of 0.8 mm to obtain localized effects. The six accelerometers are Endevco model 7270A-200k and they are located on either end of both payload mass simulators. Four are oriented laterally in the direction of the delivered impulse and two are oriented along the axis of the structure. Measured strains range up to 1.0% and have a frequency response of 100 KHertz. Peak accelerations after low-pass filtering at 50 KHertz range up to 10,000 g.

The comparison of strain and acceleration responses in Figure 9 (top) indicates that the shock wave rapidly propagates through the main joint. The ringing of the mass simulators is attenuated after 5-to-6 milliseconds. The sixth acceleration response is used to define the input excitation for shock response spectral (SRS) analysis. The SRS in Figure 9 (bottom) shows the peak acceleration response that would be witnessed by a single degree-of-freedom system whose fundamental frequency is set by the horizontal axis. Clearly, most of the resonant dynamics occur between 10 and 30 KHertz, making it impractical to analyze the system's response through modal superposition. The SRS estimates the acceleration levels that would be witnessed, for example, by an electronic component integrated at the location of the sixth accelerometer.

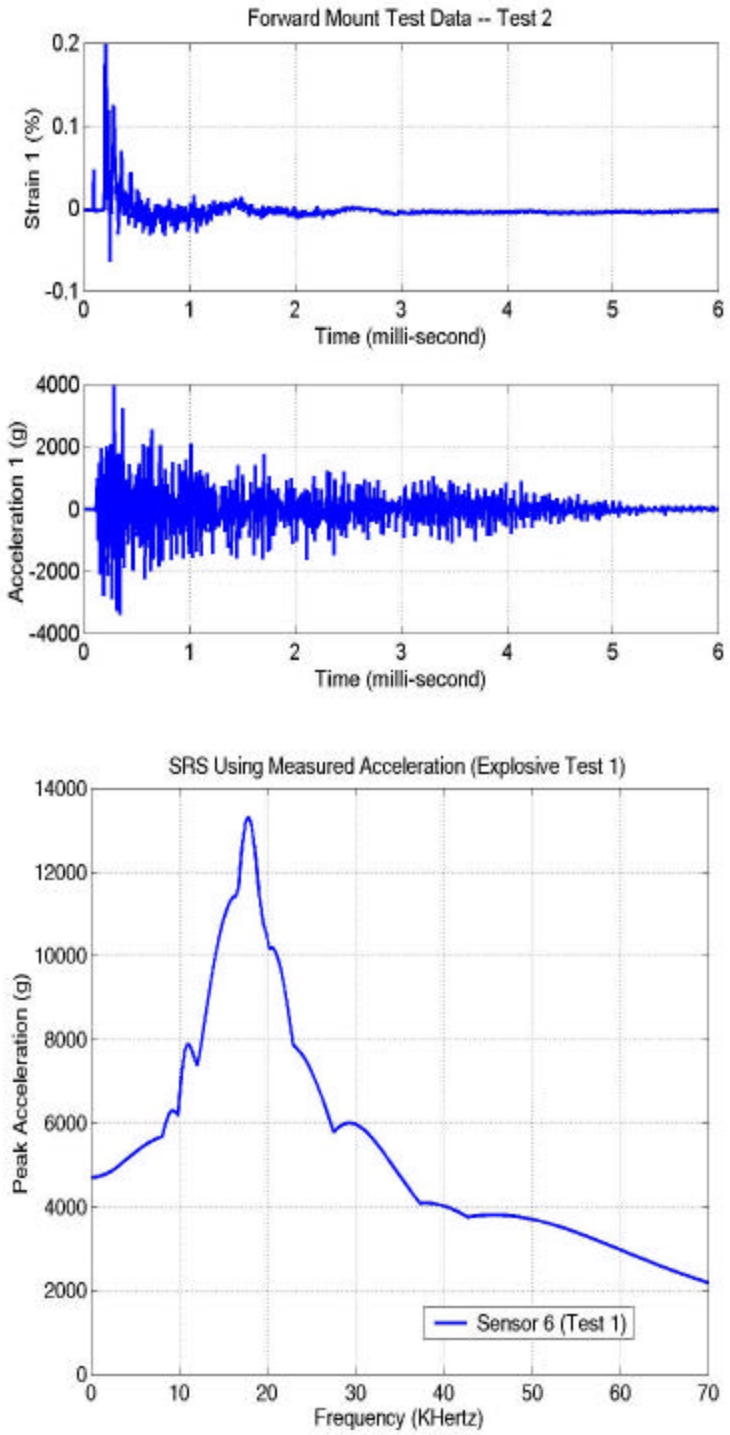


Figure 9. Impulse Response of the Forward Mount Test Article.

(Top: Strain gage 1 located behind the explosive grid and accelerometer 1 located on the lower mass simulator. Bottom: SRS of the 6th acceleration response with 2% modal damping.)

Figure 10 compares the strain responses collected at location 1 (inside the mount, behind the explosive grid) during two replicate tests. Only the first 400 microseconds of response after detonation are shown. Even though such "view-graph norm" assessment should be avoided as much as possible, it can be observed that the peak strain and low-frequency content (resonant mode at 12,000 Hertz, approximately) are very similar. The discrepancy between the two curves is attributed to the experiment's variability. The main source of variability is a tolerance and assembly-positioning threshold that is controlled during these two experiments. Other potential sources of variability are the temperature and humidity, total impulse applied, strain gage de-bounding and measurement errors. In addition to predicting the dynamics of interest, the numerical model should also reproduce with fidelity the variability inferred from physical observation when enough replicate data sets are available.

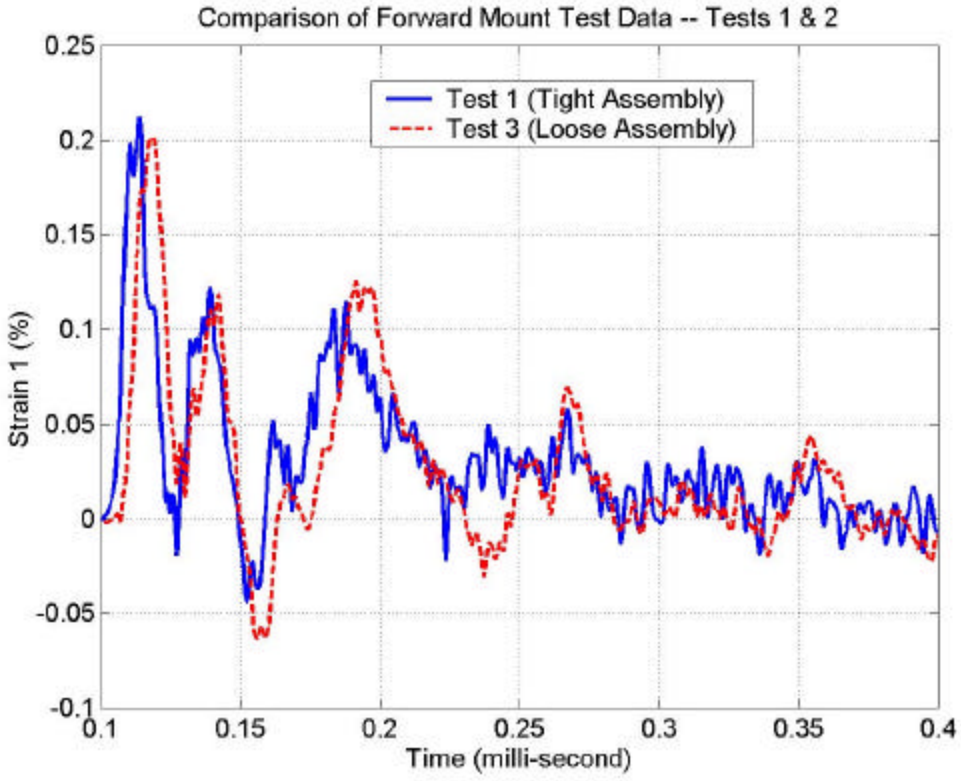


Figure 10. Comparison of Strain Responses Collected During Two Similar Tests.

(Responses of strain sensor 1 located behind the explosive grid. Solid, blue line: Response when the assembly is closely controlled. Dashed, red line: Response with a "loose" assembly.)

4.2 Finite Element Modeling and Analysis

An explicit finite element model of the test article is developed using the ParaDyn finite element code [14]. In general, explicit formulations are preferred for such problems because of their nonlinear nature and the fact that numerous contact conditions must be handled. The model has approximately 1.4 million 8-node hexahedral elements, 56,000 4-node shell elements, 480 contact surfaces and 1.8 million node points. It results into 6 million degrees of freedom that include structural displacements and Lagrange multipliers defined at the contact interfaces. The large number of contact pairs is required because each individual surface, usually circumferential in nature, has to be broken into several individual surfaces to accommodate efficient partitioning for the parallel code. Automatic contact capabilities that are currently under development in ParaDyn will obviate the need to break the contact into so many surfaces.

The computational mesh is illustrated in Figure 11. The finite element model is currently run on 504 processors on the Los Alamos Blue Mountain ASCI computer. Using this number of processors results in 1.3 CPU hours for each millisecond of simulation.

Preloading due to assembly of the threaded joints and the tape joint is accomplished in the model by implementing an orthotropic thermal coefficient of expansion in specific layers of elements. At the start of each analysis, the temperature is increased using a half-cosine time history over 0.2 milliseconds. The structure is then allowed to freely respond with no additional input for 0.1 milliseconds before applying the explosive impulse. The impulse is applied over the appropriate region of the test article as a pressure time history.

Because the contact involves interfaces among stainless steel, carbon steel, anodized 7075-T4 aluminum, 6061-T6 aluminum, and titanium, precise selection of static and kinetic coefficients of friction is not possible. Some of the variables that contribute to the coefficients of friction include surface finish, surface hardness and the presence of lubricants. Since these are not known, these coefficients of friction

are estimated by bounds and allowed to vary between specified limits. Lacking a precise definition of the coefficients of friction also leads to unknown levels of preloads. Therefore, the preloads are also allowed to vary between specified limits among the different runs. Based on engineering judgment, a total of twelve parameters are defined as having a relatively high uncertainty associated with their value. These parameters consist of three component preloads (labeled $\mathbf{p}_1$ to $\mathbf{p}_3$), four static coefficients of friction (labeled $\mathbf{p}_4$ to $\mathbf{p}_7$), four kinetic coefficients of friction (labeled $\mathbf{p}_8$ to $\mathbf{p}_{11}$) and the magnitude of the explosive impulse (labeled $\mathbf{p}_{12}$).

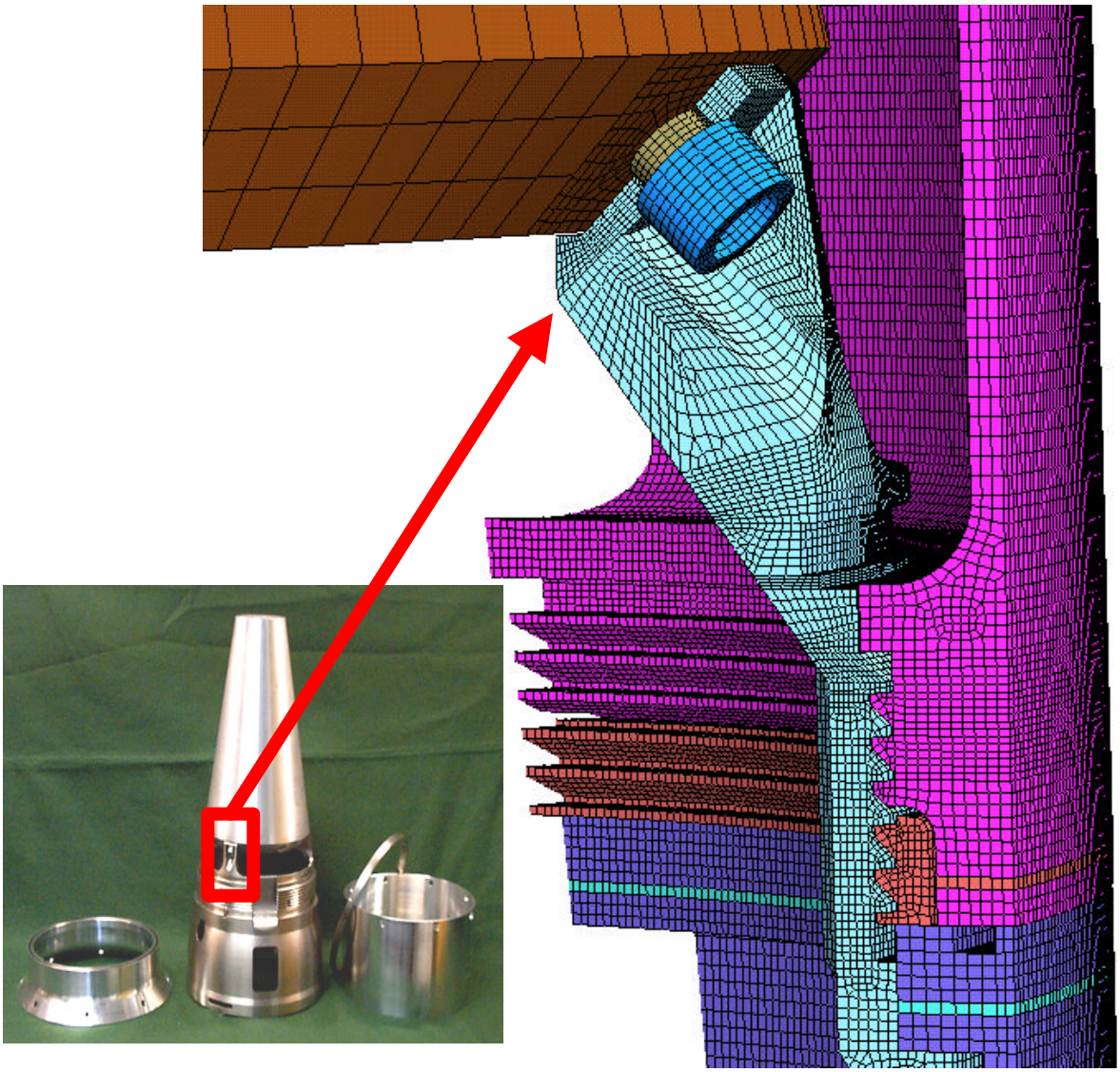


Figure 11. Computational Mesh of the Threaded Joint and Bolts.

(Viewed from the inside of the titanium mount, one of the upper mass simulator's bolts and a section of the retaining nut are visible between the assembly of upper and lower shells.)

4.3 Uncertainty Analysis

In this section, we illustrate one of the steps that would be involved in a typical validation experiment. A parameter effect analysis is performed to determine which of the twelve input parameters $\mathbf{p_1}$ to $\mathbf{p_{12}}$ are most responsible for explaining the total variability of the output. Other steps, not discussed here, would include parametric calibration, characterization of the model output's probability information and assessment of the model's predictive quality.

Based on engineering judgment and empirical observation, it is decided that four of the twelve input parameters can take two possible values and the remaining eight parameters can take three possible values. The total number of runs that would be required to build a full factorial analysis is therefore equal to $2^4 \times 3^8 = 104{,}976$. Clearly, this analysis is impossible because it would require 8 years of computational time with a 3 TeraOps ASCI platform. To limit the required simulation time, a subset of 48 runs is completed from parameter samples selected using the Taguchi orthogonal array technique [15]. After these 48 runs have been completed and the appropriate features have been extracted, statistical ***response surface models*** (RSM) of the form

$$\mathbf{y} = \mathbf{a}_o + \sum_{i=1\cdots12} \mathbf{a}_i \mathbf{p}_i + \sum_{i=1\cdots12} \sum_{j=1\cdots12, j\geq i} \mathbf{ß}_{ij} \mathbf{p}_i \mathbf{p}_j \qquad (7)$$

are fitted to the computer data for each feature of interest. RSM are developed as a substitute to the large-scale, high fidelity numerical simulations [16]. Because the transmission of shock across the mount to the payload components is the primary event of interest, errors between the predicted and measured statistical moments of the time history, shock response spectrum and power spectral density (PSD) at each accelerometer location are used as features. Time histories are restricted to the first four milliseconds following detonation and the SRS and PSD analyses are performed from 0-to-50 KHertz

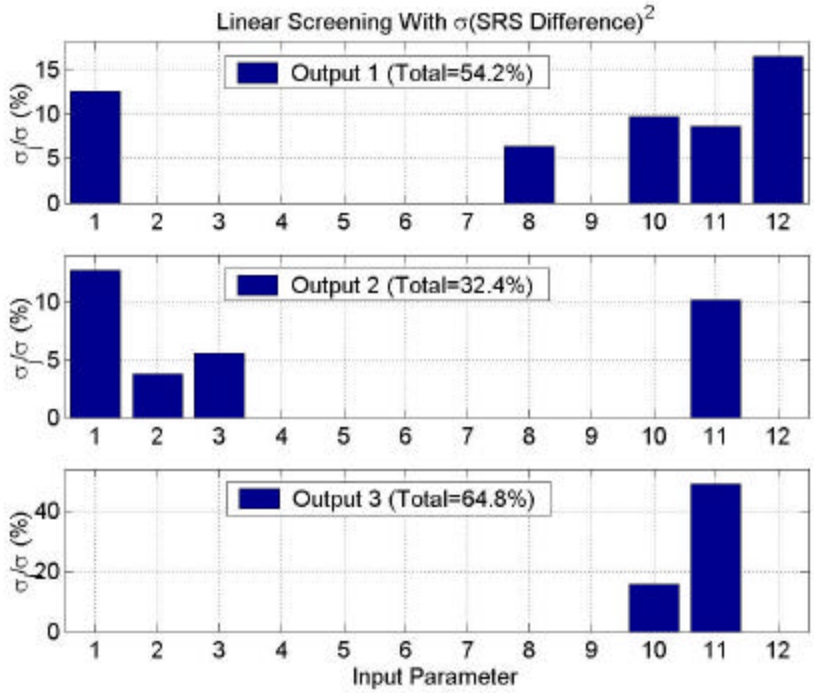
with a uniform 2% critical damping for the SRS. The first and second statistical moments (mean and standard deviation) are investigated. Hence, a total of 36 output features are defined (6 accelerometers by 3 output criteria by 2 moments).

In this application, statistical RSM are adopted to investigate if the number of input parameters can be down-selected from twelve to 5 or 6. First-order RSM that account for linear effects only are generally considered sufficient for the purpose of down-selection. For each one of the 36 features, a linear statistical RSM is constructed by ignoring higher-order effects β_{ij} in equation (7). Statistical tests are implemented to assess the global contribution of each parameter to the total variability observed from the computer simulations. A popular example is the R-square (R^2) statistics that estimates Pierson's correlation ratio. It is defined as the ratio between the variance that can be attributed to a given effect and the total variance of the data [17]:

$$R^2 = 1 - \frac{\sum_{l=1\cdots N_{level}} \sum_{j=1\cdots N_{data}^{(l)}} \left(y_j^{(l)} - \bar{y}^{(l)}\right)^2}{\sum_{j=1\cdots N_{data}} \left(y_j - \bar{y}\right)^2} \tag{8}$$

The features are evaluated based on whether a) the total contribution of the individual parameters is significant, e.g. the feature is significantly sensitive to at least one of the parameters; and b) whether the feature is amenable to a linear RSM fit, i.e. the linear fit has a high R^2 value. A few of the features either are not amenable to linear screening or do not demonstrate significant sensitivity to any of the parameters. Generally, however, a trend is observed for features from all sensors indicating significant effects due to the following six parameters: one preload (p_1), one static coefficient of friction (p_5), three kinetic coefficients of friction (p_8, p_{10}, p_{11}) and the impulse magnitude (p_{12}). The R^2 values obtained by analyzing the standard deviation of SRS errors are pictured in Figure 12. The six acceleration locations exhibit global sensitivity to at least one of the five parameters p_1, p_8, p_{10}, p_{11} and p_{12}. For location 2, the other two preloads p_2 and p_3 are significant but this linear RSM only explains 32.4% of the output feature's total variance. Linear models that do not explain at least 50% of the total variance are

considered suspect and higher-order parameter interactions should be included. This analysis demonstrates that the parameter space can be reduced from a dimension twelve to a dimension six, therefore, allowing realistic generation of a higher-order RSM.



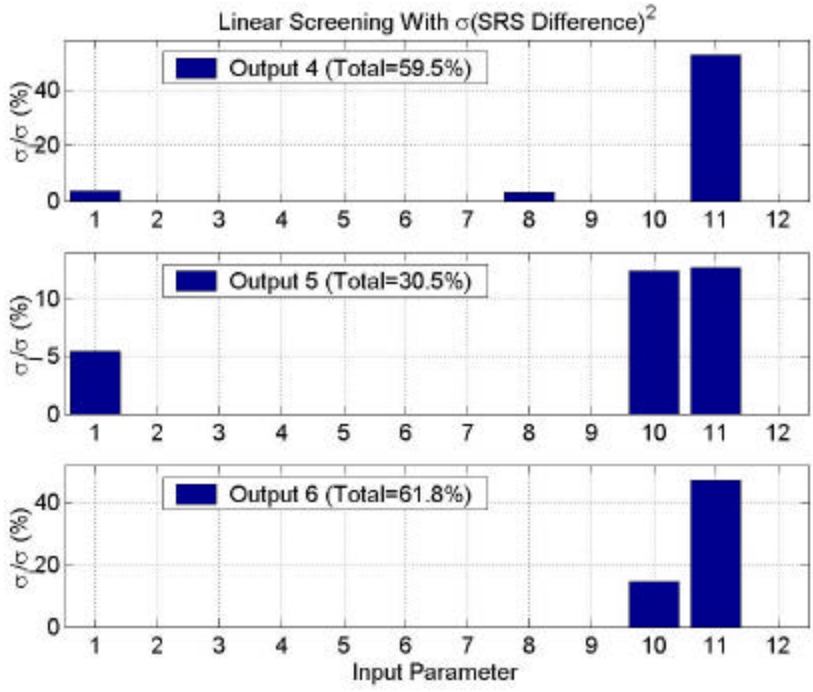


Figure 12. Parameter Contributions Calculated From the Analysis of Variance (ANOVA). *(Left: ANOVA of accelerations 1-3. Right: ANOVA of accelerations 4-6. The feature analyzed is the variance s^2 of the difference between predicted and measured SRS. Contributions of each input parameter to the variance of the output feature are shown in percentage of total variance.)*

When no other option is available but an extremely sparse sampling of the input space, special care must be brought to selecting a design matrix that avoids ***aliasing***. Aliasing in statistical modeling refers to contamination of main (linear) effects by secondary (higher-order) effects and is caused by a too sparse sampling matrix. Obviously, any sampling matrix must provide the ability to distinguish the variance associated with the linear effect of a variable from the other variables. But it can happen that linear effects (such as α_i) are confounded or aliased with second order or higher effects (such as β_{ij}). With a Taguchi orthogonal array, columns of the design matrix are not correlated with other columns and, in addition, are free of interaction with second-order effects.

One difficulty of the forward mount impulse test is that shock transmission is highly sensitive to contact modeling. In turn, contact forces across an interface depend on the product between coefficient of friction and preload. Therefore, second-order effects $(p_i p_j)$ where p_i denotes a coefficient of friction and p_j denotes a preload, respectively, play a significant role. This explains why the linear RSM rarely account for more than 60% of the total output feature variability. Currently, the predictive quality of the high fidelity model is being assessed using Taguchi orthogonal array designs capable of capturing, without aliasing, all quadratic interactions of the six input parameters based on 64 or 128 runs.

V. FEATURE EXTRACTION

In this section, the notion of ***feature extraction*** is discussed. Features are defined as small-dimensional quantities that extract information from physical observation or computer output. Obviously, their definition is application-dependent and they should satisfy two other criteria. First, a feature must

provide insight regarding the physics investigated. Second, it must be sensitive to changes in the input parameters.

In linear structural dynamics, conventional features are resonant frequencies, modal damping ratios and mode shape vectors. Nevertheless, their application is restricted to stationary and periodic signals generated from the response of linear systems. Because non-linearity is increasingly investigated, attempts are made to generalize the notion of modal superposition to nonlinear dynamics, for example, using the wavelet transform and the proper orthogonal decomposition. Success is documented in Reference [18] (wavelet transform) and Reference [19] (Karhunen-Loève transform), among others. However, these analysis techniques remain based on the assumption of linearity and their application to non-stationary data sets or nonlinear systems is necessarily limited. Our opinion is that tools commonly used in other scientific communities (e.g. physics, statistical sciences and pattern recognition) are not exploited to their full potential.

The applicability of non-conventional features is first illustrated using an Earthquake example that has recently been publicized and widely distributed over the Internet [20]. On February 28, 2001, a magnitude 6.8 Earthquake located thirty miles below the surface and a few miles away from Olympia, Washington, moved the ground for 30-to-40 seconds. The recorded Earthquake waveform is shown in Figure 13. A sand-tracing pendulum located in the vicinity produced the patterns depicted in Figure 14. The smooth curves seen to the outside of the Earthquake “rose” are what is normally observed when someone sets the pendulum in motion to make a tracing. The pattern produced when the pendulum was started prior to the Earthquake is still visible. It was then overwritten by another pattern resulting from ground motion. Clearly, the difference between these two patterns indicates different dynamics more so than a direct comparison between, for example, time series or shock response spectra.

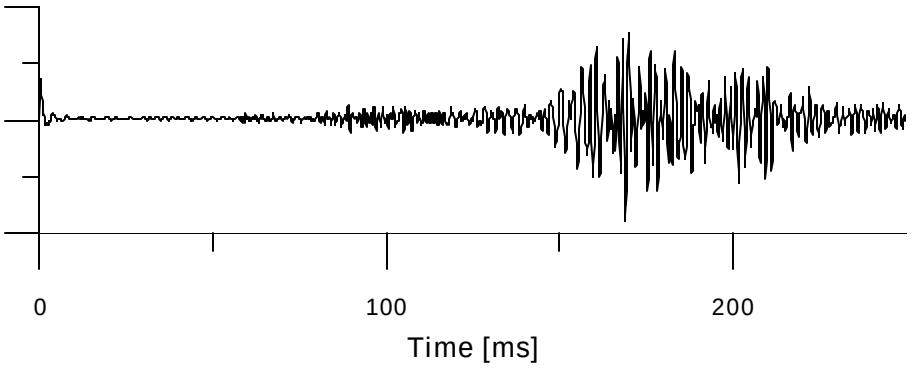


Figure 13. Typical Earthquake Record.

(Record collected during a magnitude 6.8 Earthquake in Washington State that occurred in February 2001 near Olympia, Washington.)

Figure 14. Patterns Recorded by a Sand-tracing Pendulum During the Earthquake.

(Left: Sand-tracing pendulum. Right: Patterns produced by the pendulum. A steady-state periodic signal produces the circle-like pattern seen on the outside. The transient Earthquake signal produces the twisted pattern at the center.)

The mathematical transformation implemented by the pendulum is very similar to the state-space representation of a symmetric dot pattern transform. Its efficiency for characterizing complex dynamics is further illustrated with the two-particle interaction model. Consider the four signals **y(t)** shown in

Figure 2. They can be transformed into $\mathbf{z(t) = a(t)e^{jf(t)}}$ where $\mathbf{j^2 = -1}$ and the amplitude **a(t)** and phase **f (t)** components are defined as:

$$\mathbf{a(t)} = \frac{\left(\mathbf{y(t)} - \mathbf{y_{min}}\right)}{\left(\mathbf{y_{max}} - \mathbf{y_{min}}\right)}, \quad \mathbf{f\ (t)} = \mathbf{?} + \mathbf{?}\frac{\left(\mathbf{y(t+t)} - \mathbf{y_{min}}\right)}{\left(\mathbf{y_{max}} - \mathbf{y_{min}}\right)} \tag{9}$$

The symmetric dot pattern method transforms the correlation between values **y(t)** and **y(t+t)** distant of a time shift **t** into polar coordinates. This transformation is used in the field of speech recognition to express visually, in an easy-to-understand figure, the changes in amplitude and frequency content of sound signals. An application to fault diagnosis is presented in Reference [21] where the transformation is implemented to diagnose changes in sound signals between health and faulty bearings. In our example, the angle shift **?** is set to 60 degrees, the amplification factor **?** is set to 10.0 and the time shift **t** is equal to $1/500^{th}$ of the time record's length. These parameters are determined somewhat arbitrarily. When transformed, the previous four time series produce the patterns shown in Figure 15. An immediate advantage is normalization. While the horizontal and vertical scales are all different in Figure 2, the four subplots of Figure 15 are contained between –1 and +1, which makes for a convenient comparison. Significant differences can be observed between the patterns produced by the linear signals (top) and the nonlinear signals (bottom). The chaotic response is also different from the other three as indicated by the accumulation of data points in the vicinity of coordinates **(½; 0)**.

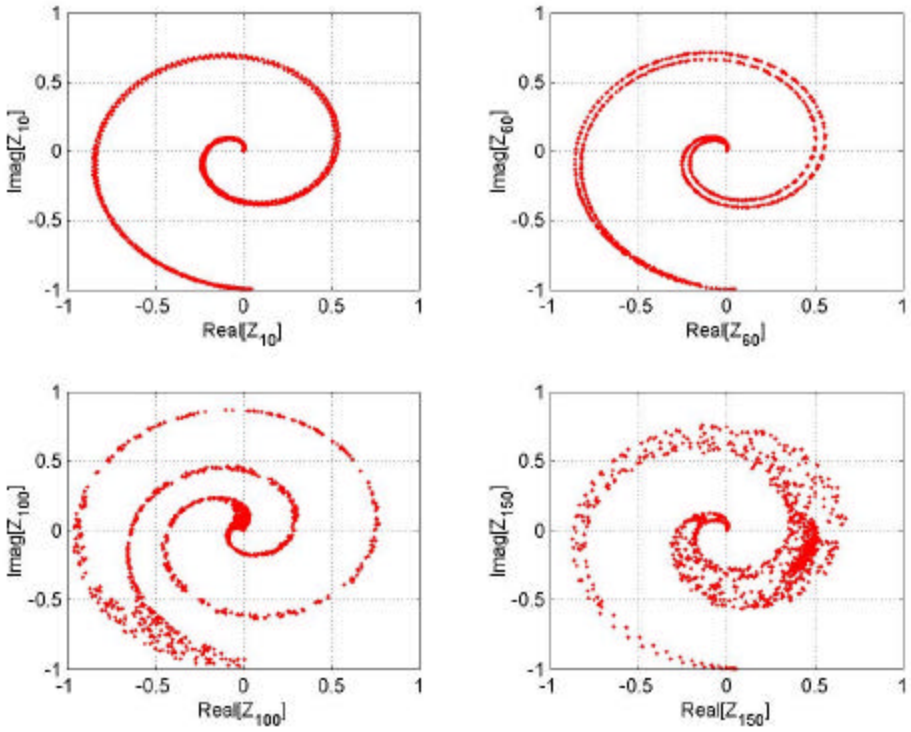


Figure 15. Symmetric Dot Patterns From the Two-particle Systems.

(Upper Left: System 10; single mode, low-frequency response. Upper Right: System 60; single mode, lightly damped response. Lower Left: System 100; high-frequency response with time-varying frequency. Lower Right: System 150; chaotic and unstable response.)

If less amplification is used (**?~1**) and several shift angles **?** are defined (for example, 0, 60, 120, 180, 240 and 300 degrees), then the patterns shown in Figure 15 are replicated six times and the resulting figure is very similar to that produced by the sand-tracing pendulum in Figure 14.

Comparisons such as those illustrated in Figures 2 and 15 are visually appealing but they do not provide a quantitative assessment of the difference between data sets. Quantitative measures are needed for inference and parametric calibration. One solution is to train surrogate models to recognize the difference between images and to relate the changes observed to characteristics of the original models or experiments. This is typically how pattern and image recognition techniques proceed [22]. The alternative is to further condense the information into low-dimensional features. Although many signal

processing techniques cannot accommodate nonlinear and non-stationary behaviors, tools such as fractal analysis are available that can.

The fractal of a signal is defined as one of its topological dimension. It represents the number of degrees-of-freedom of the corresponding data set in a multi-dimensional space. Fractal models are appropriate for analyzing signals that exhibit some form of self-similarity (for example, statistical), strong irregularity and structure on a global scale as well as arbitrarily fine scales. An application is detailed in Reference [23] where an input-output model of cavitation diagnosis is constructed between the fractal dimension of acoustic pressure measurement and the degree of cavitation in a pump. One estimate of the fractal dimension is provided by the Higuchi method that models the average evolution of the signal's increment coefficients $?_{k,p}$. Increments for samples distant of **k** intervals in time are computed as:

$$?_{k,p} = \frac{N-1}{k^2 ?(N;k;p)} \sum_{i=1\cdots ?(N;k;p)} \left| y(p+ik) - y(p+(i-1)k) \right| \tag{10}$$

where **?(N;k;p)** is defined as the lower integer part of **(N-p/k)**. The index denoted by **p** in equation (10) allows multiple estimates on a single time series and **k** denotes the time shift considered. The increment coefficients $?_{k,p}$ are averaged over the values of indices **p** and the Higuchi model assumes a linear relationship (on a log-log scale) between $\langle ?_k \rangle$, the averaged increment coefficient at index **k**, and the time shift index **k**:

$$\langle ?_k \rangle = a k^{-D} \tag{11}$$

A simple least squares fit can be performed to estimate the fractal dimension **D**. Note that the fractal dimension must, by definition, be an integer even if its numerical estimation is not. Figure 16 (top) illustrates the data points $\langle ?_k \rangle$ as a function of the time shift **k**. The other half of Figure 16 (bottom) shows the fractal dimension **D** estimated for each system. The curves shown on the top half of Figure 16 correspond to the four responses of Figures 2 and 15. The model assumed in equation (11) states

that straight lines of slopes equal to **–D** should be observed. For each system, the mean $\langle ?_k \rangle$ is plotted in solid line and the 2-s confidence intervals are shown in dashed lines. Very little variability can be observed, meaning that the fractals **D** are consistent no matter which time window is analyzed. One exception is the 150th system. When the end of the time record is analyzed, large variations of the fractal dimension are obtained because the system transitions very rapidly from nonlinear to chaotic and unstable behaviors.

On the bottom half of Figure 16, three main categories of dynamics can be observed. When the initial momentum ranges from 10^{-1} to 1, little energy is provided by the secondary particle $\mathbf{x_2}$ and the two-particle model behaves as a single degree-of-freedom. The corresponding fractal dimension is logically found close to one. The second category of dynamic systems corresponds to cases where the energy inputted by the light particle $\mathbf{x_2}$ is significant enough to influence the heavy particle $\mathbf{X_1}$. Then, the system is truly a two-particle system. Figure 16 shows that the fractal dimension is close to two when the initial momentum ranges from $3x10^{+1}$ to $3x10^{+2}$. The last category of dynamic behaviors exhibits a fractal dimension that oscillates between one and two. This is because the dynamics of these systems transition from single degree-of-freedom linear (**D=1**) to multiple degrees-of-freedom nonlinear (**D=2**). This third category corresponds to initial momentum values equal to $7x10^{+2}$ or greater. Except for the last case, the estimated fractal is insensitive to the initial position and only the level of energy inputted by particle $\mathbf{x_2}$ matters.

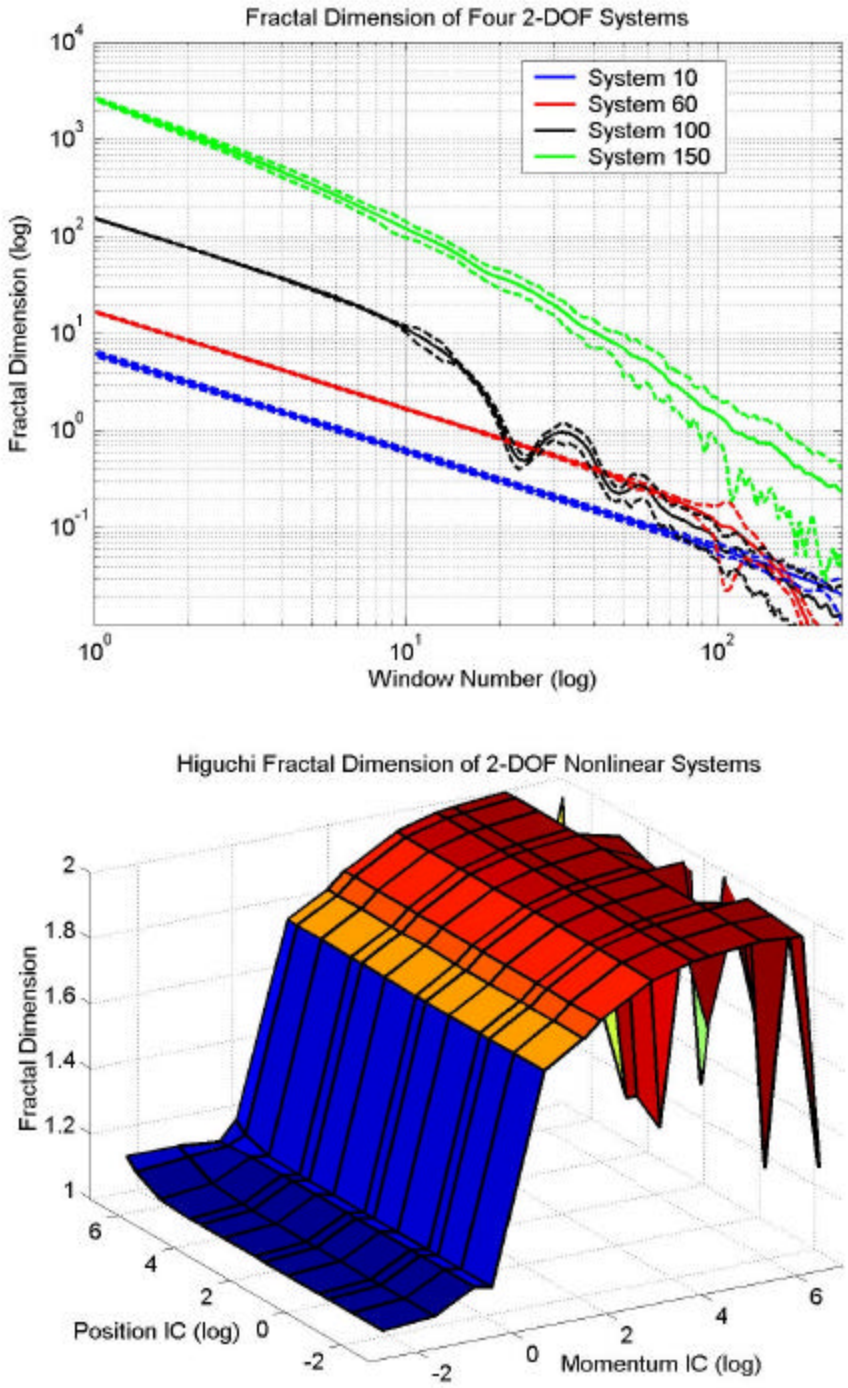


Figure 16. Increments and Fractal Dimensions of the Two-particle Systems.

*(Left: Signal increments <?$_k$> versus time shift **k**. Right: Fractal dimensions **D** for each system. Shown on the left and from bottom to top are the average increment curves for the 10th, 60th, 100th and 150th systems. Shown on the right are the fractal dimensions for all 169 systems.)*

This example illustrates that powerful features can be implemented that can effectively characterize a wide variety of dynamic behaviors. Of course, we cannot stress enough the importance of selecting features that are relevant to the application investigated and useful to the analyst. Time-frequency analysis techniques encountered in structural dynamics include the power spectral density, power cepstrum, cyclo-stationarity analysis, Willer-ville transform, wavelet transform, spectrogram and Karhunen-Loève decomposition. General-purpose features can alternatively be extracted from probability density functions, shock response spectra, temporal and statistical moments (mean, variance, energy, kurtosis, etc.) and fractal analysis. In addition, features can be extracted from parametric models best fitted to data sets. The model fitting techniques that have been applied to engineering mechanics problems with success include linear regression (such as AR, ARX), nonlinear regression (such as ARMA), neural networks, statistical models and kriging models.

VI. RESPONSE SURFACE MODELS AND PREDICTABILITY

Visualizing the output of large-scale numerical simulations can be a cumbersome process. Efficient and user-friendly tools must be developed to make model validation useful to the analysts. This means that output visualization, parameter sensitivity, test-analysis correlation, model updating, uncertainty analysis and predictability assessment must be integrated within the same framework and possibly so in a graphical manner. To fulfill these requirements, the methodology of response surface models is adopted, as introduced in Reference [16].

Response surface is a general methodology for representing the dependent variables of a process as a graphical function of the independent inputs. Visualization is accomplished by selecting a single dependent variable and by plotting its response against one or two independent variables. If the functional relationship between variables is continuous, the resulting object is a 1D plot or a 2D surface as illustrated by Figure 17. More commonly, the dependent variable is known at some discrete points representing combinations of particular values of independent variables. In this case, it is common

practice to best fit a multi-dimensional response surface (e.g. polynomial, exponential, neural network, statistical) through these points to approximate the unknown model.

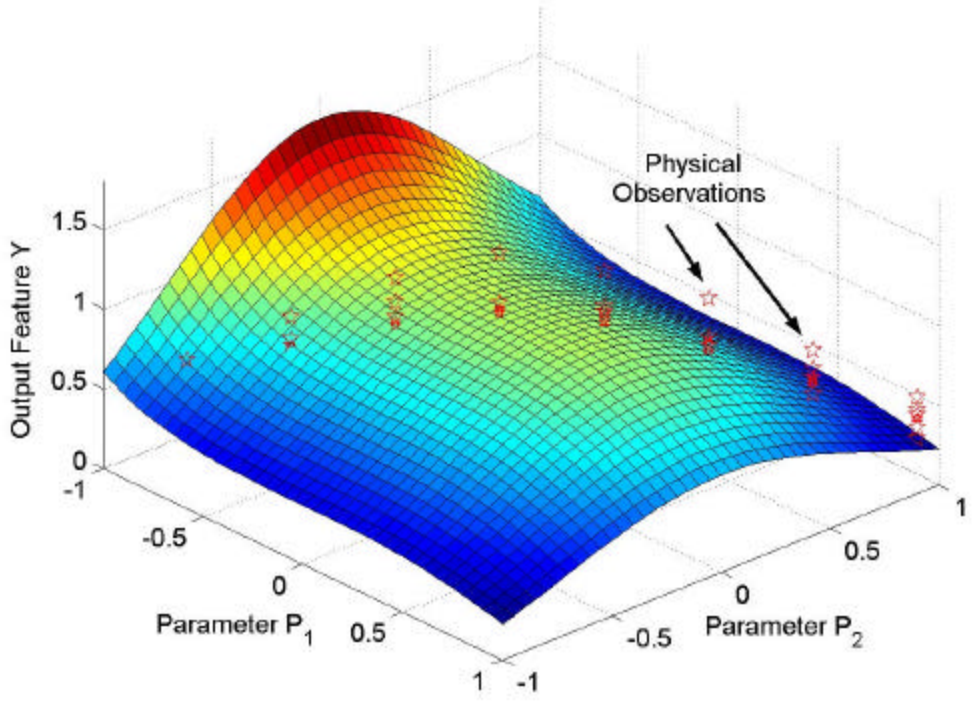


Figure 17. Illustration of a Response Surface for Surrogate Modeling.

(The figure illustrates a typical response surface generated from a model $\mathbf{y = M(p_1; p_2)}$ *when the input parameters* $\mathbf{p_1}$ *and* $\mathbf{p_2}$ *are uniformly varied from –1 to 1. Stars represent measurements from physical experiments performed at eight different values of* $\mathbf{(p_1; p_2)}$*. The experiments are replicated so that measurement variability can be inferred from the distribution of data points.)*

In the problems we are more particularly interested in, the "process" is a large, stochastic finite element simulation. The dependent response variables are features **y** of the simulation outputs (e.g. peak values, modal frequencies) and the independent variables are input parameters **p** that define the finite element simulations (e.g. material definition, initial conditions, preloads). The information that needs to be combined during model validation consists of output from the high-fidelity numerical simulation, output from the surrogate models and physical observation:

	High-fidelity Model	*Surrogate Models*	*Physical Observation*
Model	$\mathbf{y} = \mathbf{M}(\mathbf{p})$	$\hat{\mathbf{y}} = \hat{\mathbf{M}}(\mathbf{p})$	$\mathbf{y}^{\mathbf{Test}}(\mathbf{p}_1), \ldots, \mathbf{y}^{\mathbf{Test}}(\mathbf{p}_N)$
Input PDF	$\mathbf{f}_p(\mathbf{x}) = \frac{\mathbf{dP}}{\mathbf{dx}}[\mathbf{p} \le \mathbf{x}]$	$\hat{\mathbf{f}}_p(\mathbf{x}) = \frac{\mathbf{dP}}{\mathbf{dx}}[\hat{\mathbf{p}} \le \mathbf{x}]$	$\mathbf{f}_p^{\mathbf{Test}}(\mathbf{x}) = \frac{\mathbf{dP}}{\mathbf{dx}}[\mathbf{p}^{\mathbf{Test}} \le \mathbf{x}]$
Output PDF	$\mathbf{h}_y(\mathbf{x}) = \frac{\mathbf{dP}}{\mathbf{dx}}[\mathbf{y} \le \mathbf{x}]$	$\hat{\mathbf{h}}_y(\mathbf{x}) = \frac{\mathbf{dP}}{\mathbf{dx}}[\hat{\mathbf{y}} \le \mathbf{x}]$	$\mathbf{h}_y^{\mathbf{Test}}(\mathbf{x}) = \frac{\mathbf{dP}}{\mathbf{dx}}[\mathbf{y}^{\mathbf{Test}} \le \mathbf{x}]$

(12)

In equation (12), the symbols $\mathbf{f_p}$ and $\mathbf{h_y}$ denote the input parameter's multi-dimensional probability density function (PDF) and the output feature's PDF, respectively. When large-scale numerical simulations are involved, only an estimate of the output probability information $\mathbf{h_y}$ can be obtained, based either on surrogate models or approximate analysis such as fast probability integration. One advantage of the RSM approach is that probability information can be combined to the simulation output and visualized through the definition of confidence regions. As mentioned previously, integrating in a user-friendly environment the model outputs **y**, predictions of the surrogate models **y** and physical observations $\mathbf{y^{Test}(p_k)}$ can be a complicated process. This becomes even more challenging when the analysis must account for input and output uncertainty. The response surface methodology can offer several beneficial effects:

(a) Analyzing the finite element output:

When outputs of the simulation code are available, surrogate models can be trained to improve the fit to the simulation data. Figure 17 illustrates a 2D response surface that could typically serve as basis for all subsequent analyses. Beyond the graphical representation, this exercise can provide valuable information regarding the simulation. For example, investigating the most appropriate model form might reveal correlation between the parameters and features.

(b) Sensitivity and effect analysis:

The analyst can examine the location of a design point on the RSM and make some useful observations. For example, the design point might be in a region of relatively high slope. Thus, a small perturbation in either $\mathbf{p}_1$ or $\mathbf{p}_2$ can result in a significant change in the feature $\mathbf{y}$. If the application of interest is a stress analysis with $\mathbf{y}$ defined as the peak stress and $\mathbf{p}_1$ is a material constant with relatively high uncertainty, then the RSM indicates to the analyst that there may be serious concerns about the ability of a population of manufactured units to stay under the maximum stress limit. Beyond local sensitivity, statistical RSM can be developed for global parameter effect study and analysis of variance. Such techniques estimate the global influence of input parameters on output features over the entire design space.

(c) Test-analysis correlation:

Model validation and calibration experiments always require comparison between model output and physical observation. Test data sets can be graphically compared to the response surface, as illustrated in Figure 17. In this illustration, the results from eight physical experiments are plotted. It can be instantly observed that some observations are consistent with model predictions and others are not. Quantitative subspace agreement and test-analysis correlation metrics are easy to compute once RSM are fitted to the simulation data.

(d) Parametric calibration:

Parametric calibration experiments are generally formulated as optimization problems. An objective function parameterized by the model's input parameters is defined and minimized. Of course, evaluation of the objective function requires model output. It is therefore unrealistic to believe that large-scale, multi-physics computational codes can be embedded within an optimization procedure. Instead,

parametric calibration based on RSM is immediate because function, gradient and Hessian evaluations are straightforward.

(e) Uncertainty analysis:

Figure 17 illustrates the case where several replicate observations are collected at each design point for which an experimental measurement is carried out. Results are located at the same location in the $(p_1; p_2)$ plane and the vertical spread translates experimental variability. When enough replicates are available, statistical parameters (e.g. mean value, 2-s interval confidence) can be inferred and represented graphically as confidence intervals. If the controllability of parameters p_1 and p_2 is uncertain during testing, then the placement of replicate measurements in the $(p_1; p_2)$ plane becomes uncertain as well. Then, the confidence intervals become confidence regions in the 3D space (e.g. ellipsoids). Both geometrical entities can easily be plotted and compared to the response surface. Uncertainty associated with the numerical simulations and RSM can be represented in a similar manner.

(f) Design of validation experiments:

RSM and global effect analysis help design the test matrix, that is, the combinations of input parameters $(p_1; p_2)$ for which it is recommended to perform a measurement or series of replicate measurements. Physical experiments are recommended, for example, in areas of the design space where variability seems important, sensitivities are high or system reliability might be an issue. Analysis of variance techniques based on RSM can down-select input parameters as illustrated in section IV. It is emphasized that numerical simulations and RSM should guide, not replace, physical experiments. Experimental practices are likely to change from qualification testing (that is, ensuring that the integrated system satisfies a specified performance criterion) to validation testing (that is, ensuring that the numerical model predicts with acceptable accuracy the response of the system in a large region of the design and operating condition spaces).

(g) Predictability:

The ultimate goal of model validation is to guarantee that numerical models accurately represent reality, especially when testing is not an option. This, for example, corresponds to analyzing the RSM on Figure 17 away from any combination of input parameters $(p_1; p_2)$ where physical observations are available. Confidence in the prediction must therefore be assessed. Assessing predictability based on validation experiments is an area of open research to a great extent. Reference [23] develops a Bayesian melding framework for statistical inference of simulation models that integrates diagnostic checking, model validation, hypothesis testing and model selection methods. The approach proposed builds on conventional Bayesian inference, goes beyond parametric calibration but stops short of addressing the predictability issue. Another interesting attempt is made in Reference [24]. The authors present a calibration technique that integrates the notion of predictive confidence regions by quantifying and propagating residual errors between calibrated models and experimental data sets. By systematically quantifying all sources of uncertainty, their procedure can assess the prediction's confidence regions and monitor model inadequacy errors.

To the question ***"Can numerical predictability be assessed?"*** our opinion is therefore a cautious yes given that adequate validation experiments are performed. However, this is a difficult problem, far from being resolved. The aforementioned techniques are currently being investigated at Los Alamos for assessing the predictive accuracy of computer codes for hydrodynamic and structural mechanics applications. Given our current computational resources, it is unrealistic to envision that predictability assessment techniques can be efficiently integrated to large-scale computational codes. To the best of the authors' knowledge, no such work has ever been attempted without, first, developing an appropriate RSM.

VII. CONCLUSION

This publication discusses the concepts of modeling, uncertainty quantification, model validation and numerical predictability. Not only does uncertainty refer to parametric variability and lack of knowledge, but it can also be seen as an integral component of the numerical model. In light of this conception of uncertainty, model validation is defined as an attempt to identify regions of the design space where the model "breaks down" or uncertainty is too important. Predictability refers to the ability to quantify the accuracy of the model in regions of the design space where physical observations are not available. Model validation must rely on carefully planed experiments that provide an assessment of reality for the largest possible array of configurations and operational conditions. Parametric calibration is a pre-requisite that permits to reduce some of the discrepancy between physical observation and model output but under no circumstance should a calibrated model be considered validated. The issue of feature extraction, or how to characterize the dynamics of time series, is also discussed and the efficiency of a few techniques is illustrated with a simple model of elementary particle interaction.

A structural dynamics application is also presented that features the transmission of a shock through a complex jointed assembly. This highly transient event is analyzed on a 3 TeraOps computer using a multi-million degree-of-freedom explicit finite element model. Work to date has indicated that it is possible to reduce a high-dimension parameter space to a reasonable dimension using a moderate number of runs. Future work includes creating higher-order response surface models using this lower dimension parameter space. The higher-order models will then be used for parameter effect analysis and calibration. A second round of validation experiments will then be designed to further explore the parameter space and define the regime of validity of the model. The ultimate objective is to demonstrate that the numerical model can be used with some confidence (also to be quantified) to predict events outside the regime of practical testing.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contribution of Thomas A. Butler and John F. Schultze (ESA-EA, Engineering Analysis). They are responsible for most of the ASCI engineering application

work discussed in this publication. The authors are grateful to Mark C. Anderson (DX-5, Experimentation and Diagnostic Design), Kenneth M. Hanson (CCS-2, Scientific Computing) and Michael D. McKay (D-1, Statistical Sciences) from Los Alamos for sharing their thoughts and expertise on model validation. We also acknowledge the support of the weapons program and the ASCI project at Los Alamos for supplying the resources needed to support this work. Los Alamos National Laboratory is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36.

REFERENCES

1. Garcia, J., "The Need For Computational Model Validation," *SEM Experimental Techniques*, Vol. 25, No. 2, March-April 2001, pp. 31-33.

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, Florida, February 5-8, 2001, pp. 977-985.

3. Butler, T.A., Hemez, F.M., Schultze, J.F., Sohn, H., Doebling, S.W., "Model Validation For a Complex Jointed Structure," *19th SEM International Modal Analysis Conference*, Kissimmee, Florida, February 5-8, 2001, pp. 1318-1324.

4. Doebling, S.W., Butler, T.A., Schultze, J.F., Hemez, F.M., Moore, L.M., McKay, M.D., "Validation of Transient Structural Dynamics Simulations: An Example," *3rd International Conference on Sensitivity Analysis of Model Output*, Madrid, Spain, June 18-20, 2001, to appear.

5. Kast, A., "Dynamic Uncertainty," *19th SEM International Modal Analysis Conference*, Kissimmee, Florida, February 5-8, 2001. Special invited presentation at the SD-2000 session on uncertainty quantification and model validation developed by Hemez, F.M., Doebling S.W., pp. 1153-1158.

6. Yager, R.R., Kacprzyk, J., Fedrizzi, M., **Advances in the Dempster-Shafer Theory of Evidence**, John Wiley & Sons, New York, 1996.

7. Dumitrescu, D., Jain, L.C., Lazzerini, B., **Fuzzy Sets and Their Application to Clustering and Training**, International Series on Computational Intelligence, CRC Press, New York, 2000.

8. Ben-Haim, Y., Cogan, S., Sanseigne, L., "Usability of Mathematical Models in Mechanical Decision Processes," Mechanical Systems and Signal Processing, Vol. 12, No. 1, January 1998, pp. 121-134.

9. Maltrud, M.E., Smith, R.D., Semtner, A.J., Malone, R.C., "Global Eddy Resolving Ocean Simulations Driven by 1985-1995 Atmospheric Winds," *Journal of Geophysics Res.*, Vol. 103, No. C13, 1998, pp. 30825-30853.

10. Bryan, F.O., Smith, R.D., Maltrud, M.E., Hecht, M.W., "Modeling the North Atlantic Circulation: From Eddy Permitting to Eddy Resolving," *LANL Web Report*, Los Alamos National Laboratory, Los Alamos, New Mexico, March 2000. http://www.cgd.ucar.edu/oce/bryan/woce-poster.html.

11. Hanson, K.M., Booker, J., "Inference From Rossi Traces," *Oral Presentation at the Los Alamos Uncertainty Quantification Working Group*, Los Alamos National Laboratory, Los Alamos, New Mexico, September 2000. http://www.lanl.gov/home/kmh/.

12. Hanson, K.M., Cunningham, G.S., "Posterior Sampling With Improved Efficiency," *Proceedings of the SPIE Conference*, Vol. 3338, 1998, pp. 371-382.

13. Gilks, W.R**., Markov Chain Monte Carlo in Practice**, Chapman and Hall, New York, 1996.

14. Hoover, C.G., De Groot, A.J., Sherwood, R.J., "ParaDyn: A Parallel Nonlinear, Explicit, Three-Dimensional Finite-Element Code for Solid and Structural Mechanics—User Manual," *LLNL UCRL-MA Report*, Lawrence Livermore National Laboratory, Livermore, California, 2000.

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

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

17. McKay, M.D., Beckman, R.J., Conover, W.J., "A Comparison of Three Methods For Selecting Values of Input Variables in the Analysis of Output From a Computer Code," *Technometrics*, Vol. 21, No. 2, 1979, pp. 239-245.

18. Dalpiaz, G., Rivola, A., "Condition Monitoring and Diagnostics in Automatic Machines: Comparison of Vibration Analysis Techniques," *Mechanical Systems and Signal Processing*, Vol. 11, No. 1, January-February 1997, pp. 53-73.

19. Lenaerts, V., Kerschen, G., Golinval, J.C., "Parameter Identification of Nonlinear Mechanical Systems Using Proper Orthogonal Decomposition," *18th SEM International Modal Analysis Conference*, San Antonio, Texas, February 7-10, 2000, pp. 133-139.

20. MacLeod, N., "The Earthquake Rose," *Web Report*, March 2001. http://www.gaelwolf.com/pendulum.html.

21. Shibata, K., Takahashi, A., Shirai, T., "Fault Diagnosis of Rotating Machinery Through Visualization of Sound Waves," *Mechanical Systems and Signal Processing*, Vol. 14, No. 2, March-April 2000, pp. 229-241.

22. Bishop, C.M., **Neural Networks for Pattern Recognitio**n, Oxford University Press, New York, 1998.

23. Simard, P., Le Tavernier, E., "Fractal Approach For Signal Processing and Application to the Diagnosis of Cavitation," *Mechanical Systems and Signal Processing*, Vol. 13, No. 3, May-June 2000, pp. 459-469.

24. Poole, D., Raftery, A.E., "Inference for Deterministic Simulation Models: The Bayesian Melding Approach," *Technical Report No. 346*, Department of Statistics, University of Washington, Seattle, Washington, December 1998. http://www.stat.washington.edu/raftery/.

25. Kennedy, M.C., O'Hagan, A., "Bayesian Calibration of Computer Models," *Journal of the Royal Statistical Society (Series B*), 2001, to be published. http://www.shef.ac.uk/~st1ao/.