LA-UR- 04-6829

Approved for public release; distribution is unlimited.

Title: THE MYTH OF SCIENCE-BASED PREDICTIVE MODELING (U)

Author(s): François M. Hemez, Los Alamos National Laboratory, ESA-WR

Submitted to: Foundations ’04 Workshop for Verification, Validation, and Accreditation (VV&A) in the 21st Century, October 13-15, 2004, Arizona State University, Tempe, Arizona

The Myth of Science-based Predictive Modeling

François M. Hemez[1]

Engineering Sciences and Applications, ESA-WR
Los Alamos National Laboratory

Abstract: A key aspect of science-based predictive modeling is the assessment of prediction credibility. This publication argues that the credibility of a family of models and their predictions must combine three components: 1) the fidelity of predictions to test data; 2) the robustness of predictions to variability, uncertainty, and lack-of-knowledge; and 3) the prediction accuracy of models in cases where measurements are not available [1]. Unfortunately, these three objectives are antagonistic. A recently published Theorem that demonstrates the irrevocable trade-offs between fidelity-to-data, robustness-to-uncertainty, and confidence in prediction is summarized. High-fidelity models cannot be made increasingly robust to uncertainty and lack-of-knowledge. Similarly, robustness-to-uncertainty can only be improved at the cost of reducing the confidence in prediction. The concept of confidence in prediction relies on a metric for total uncertainty, capable of aggregating different representations of uncertainty (probabilistic or not). The discussion is illustrated with an engineering application where a family of models is developed to predict the acceleration levels obtained when impacts of varying levels propagate through layers of crushable hyper-foam material of varying thicknesses. Convex modeling is invoked to represent a severe lack-of-knowledge about the constitutive material behavior. The analysis produces intervals of performance metrics from which the total uncertainty and confidence levels are estimated. Finally, performance, robustness and confidence are extrapolated throughout the validation domain to assess the predictive power of the family of models away from tested configurations.

Keywords: Confidence, prediction, validation, fidelity-to-data, robustness, uncertainty.

1. Introduction

In computational physics and engineering, numerical models are developed to predict the behavior of a system whose response cannot be measured experimentally. A key aspect of science-based predictive modeling is to assess the ***credibility*** of predictions. Credibility, which is demonstrated through the activities of model Verification and Validation (V&V), quantifies the extent to which simulations can be analyzed with ***confidence*** to represent the phenomenon of interest with a degree of accuracy consistent with the intended use of the model [2].

The paper argues that assessing the credibility of a mathematical or numerical model must combine three components: 1) Improving the fidelity of predictions to test data; 2) Studying the robustness of predictions to variability, uncertainty, and lack-of-knowledge; and 3) Establishing the degree of confidence in model predictions in situations where test measurements are not available. A Theorem has recently been established that demonstrates the irrevocable trade-off between fidelity-to-data, robustness-to-uncertainty, and confidence in prediction [3, 4, 5].

---

[1] Technical staff member and ESA-WR Validation Methods team leader. Mailing address: Los Alamos National Laboratory, ESA-WR, Mail Stop T001, Los Alamos, New Mexico 87545, U.S.A. Phone: (+1) 505-663-5204. Fax: (+1) 505-663-5225. E-mail: hemez@lanl.gov. *This publication is a revised version of Reference [1]. Approved for unlimited, public release on Nov-18-2003, LA-UR-04-6829,* ***Unclassified****.*

Clearly, fidelity-to-data matters because no analyst will trust models and simulations that do not reproduce measurements collected during past experiments. Fidelity-to-data has dominated the concept of prediction accuracy in most scientific activities. This paradigm has resulted in the development of calibration techniques to improve the ability of models to reproduce test data. An example in Structural Dynamics is finite element model updating [6, 7]. The calibration paradigm, however, does not address V&V and the fundamental question of prediction accuracy especially when physical experiments are not available. It is therefore argued that calibration such as finite element model updating is useful but insufficient to reach simulation credibility.

The problem of assessing the prediction accuracy of numerical simulations can be thought of as decision-making under uncertainty where a "best" solution is sought in a family of models. Here, uncertainty should be taken as a broad concept that includes environmental variability; lack-of-knowledge of material behavior, initial conditions, boundary conditions, and loadings; modeling assumptions; model parameter variability; ambiguous or conflicting expert opinion. Likewise, a ***family of models*** includes all models consistent with the sources of uncertainty.[2] The discussion of model validation proposed in this publication emanates from the perspective of making decisions in the context of uncertainty, where the decision is a statement about the prediction accuracy of a family of models, possibly for conditions that have not been tested experimentally, and uncertainty generally arises from competing modeling assumptions.

Model calibration belongs to the class of decision-making strategies that advocate choosing decisions that optimize target performance metrics.[3] Another strategy is to choose decisions that optimize the robustness to uncertainty and lack-of-knowledge [8, 9]. This strategy consists in ***satisficing*** performance, or ensuring that models reproduce the available test data with a level of accuracy that is just good enough. Clearly, the difference with the optimal performance approach is to seek sufficiency, not performance optimality. This frees a degree of freedom in the search for the "most valid" model. Robustness-to-uncertainty can then be optimized. The robust-optimal model not only reproduces the test data up to a given level of accuracy, but it guarantees that the prediction accuracy will be least vulnerable to the uncertainty considered in the analysis. Reference [9] proves the antagonism between performance optimality and robustness optimality, and the two concepts are further discussed and illustrated in Section 2.

The third class of decision-making strategies that can be adapted to model validation is the assessment of prediction accuracy based on nominal predictions to which safety factors are added relative to confidence and uncertainty levels. References [10, 11, 12] provide examples in the context of the accreditation and certification of complex engineered systems. This approach is not further discussed because the definition of safety factors is application-specific to a great extent, although recent attempts have been made at interpreting margins and safety factors in terms of probabilistic reliability [13].

Figure 1 illustrates the application of three broad classes of decision-making strategies, namely, reliability analysis, robustness analysis, and margin analysis, to the problem of model validation. Reliability consists of optimizing performance metrics given a probabilistic, convex, or other, description of uncertainty. In terms of model validation, this leads to the concept of

---

[2] For example, if a coefficient of an ordinary differential equation is unknown, the corresponding family of models would be the sequence of equations obtained as the coefficient is varied. A family of models may also include competing models based on different modeling assumptions, spatial resolutions, temporal discretizations, variational principles, etc

[3] In the case of finite element model updating, for example, "decision" refers to the choice of model parameters and "performance" refers to the ability of predictions to match measurements. The optimal decision optimizes performance, meaning that the "best" model parameters are those that lead to finite element predictions that best reproduce the test data.

calibration previously discussed where the correlation between simulated and measured results is optimized. Figure 1 shows that, of the three models *A*, *B*, and *C*, model *B* is performance-optimal because its prediction, y_B, is closest to the measurement, y^{Test}. Note that the concepts illustrated in Figure 1 can be generalized to account for experimental uncertainty by defining the test-analysis correlation metrics in terms of statistical tests.

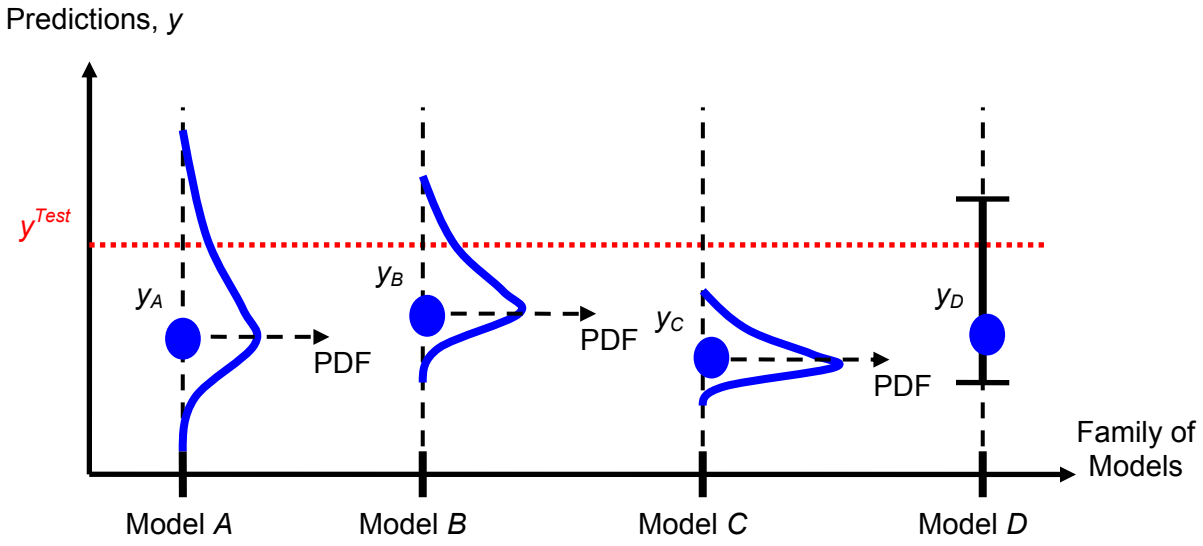


Figure 1. Conceptual illustration of decision-making strategies applied to model validation.

The second class of decision-making methods (robustness) minimizes the vulnerability of performance metrics to the uncertainty. The robust-optimal solution for model validation is the model whose prediction accuracy changes the least given the sources of uncertainty and lack-of-knowledge considered. Of course, the cost to pay is the sub-optimal prediction accuracy. The robust-optimal model of Figure 1 is model *C* because it is the one whose predictions change the least due to modeling uncertainty. This implies that the prediction accuracy of model *C* is least deteriorated even if some of its modeling assumptions are erroneous. In contract, model *A* may be more predictive than model *C*, but its prediction accuracy could be worse than expected should its modeling assumptions be not exactly correct.

The third class of decision-making methods (margin analysis) requires an assessment of prediction uncertainty that combines all sources of uncertainty that can reasonably be estimated such as experimental variability, mesh convergence errors, prediction uncertainty due to model parameter variability and modeling lack-of-knowledge. Safety factors can be added to guarantee that the effect of any source of uncertainty not accounted for in the analysis is included. Note that, here, the concept of optimization does not really apply. All models whose predictions fall within the prediction uncertainty bounds are acceptable. Model *D* shown in Figure 1 is therefore validated based on the fact that its prediction, y_D, falls within the acceptable range.

Clearly, fidelity-to-data and robustness-to-uncertainty are important attributes of any family of models. It may be argued, however, that the most important aspect of prediction credibility is the assessment of confidence in prediction, which is generally not addressed in the literature. Confidence in prediction here refers to an assessment of prediction accuracy away from settings where physical experiments have been performed, which must include a rigorous quantification of the sources of variability, uncertainty, and lack-of-knowledge, and their effects on model-based predictions. Unfortunately, these three attributes are antagonistic [4, 5],

meaning that improving two of them comes to the detriment of the third one. This suggests a decision-making strategy in situations where knowledge is severely lacking: studying the trade-offs between fidelity-to-data, robustness-to-uncertainty, and confidence in predictions.

In Section 2, conceptual illustrations and rigorous definitions are proposed for the concepts of fidelity-to-data R, robustness-to-uncertainty α^*, and prediction "looseness" λ_Y. Looseness refers to the range of predictions made by a family of equally robust models, and this definition is needed only to prove the main Theorem [4, 5] in Section 3. The main contribution of this publication is to link the prediction looseness λ_Y to a confidence level C_F via the concept of total uncertainty TU defined in Section 5. It results an inverse relationship between confidence and looseness (confidence decreases when looseness increases), from which the antagonism between fidelity-to-data, robustness-to-uncertainty, and confidence in prediction is derived.

These concepts are illustrated in Sections 4 and 7 with an engineering application where a family of models is developed to predict the acceleration levels obtained when impacts of varying levels propagate through layers of crushable hyper-foam material of varying thicknesses [14, 15]. Convex modeling is invoked to represent a severe lack-of-knowledge about the constitutive material behavior. The analysis produces intervals of performance metrics from which the total uncertainty and confidence levels are estimated. Finally, performance, robustness and confidence are extrapolated throughout the validation domain to assess the predictive power of the family of models away from tested configurations.

2. Fidelity, Robustness, and Prediction Looseness

Even though the conventional activities of model V&V are generally restricted to improving fidelity-to-data through the correlation of test and simulation results, and the calibration of model parameters, the other two components are equally critical. The reason is that optimal models, in the sense of models that minimize the prediction errors with respect to the available test data, possess exactly ***zero robustness*** to uncertainty and lack-of-knowledge [9]. This means that small variations in the setting of model parameters, or small inaccuracies in the knowledge of the functional form of the models, can lead to an actual fidelity that is significantly worse than the one demonstrated through calibration.

In this Section, conceptual illustrations and rigorous definitions are proposed for the fidelity-to-data R, robustness-to-uncertainty α^*, and prediction looseness λ_Y. The theoretical results that establish the relative sensitivities of R, α^*, and λ_Y are then summarized in Section 3.

2.1 Validation Domain and Uncertainty

Throughout the manuscript, the numerical simulation is represented as a "black-box" input-output relationship between inputs p and q and outputs y:

$$y = M(p; q_o) \tag{1}$$

where the subscript $(\,)_o$ represents the nominal condition of a quantity or state, and:

- The quantity y denotes the observable outputs. They can be scalar quantities, which is the case assumed here for simplicity, or vectors. The model outputs are usually features extracted from a large-order or large-dimensional response.
- The quantity p denotes the control parameters of the numerical simulation and physical experiments. These inputs define the validation domain, as explained below. An example is settings such as the angle of attack, flow velocity of an aero-elastic simulation whose purpose is to predicts a coefficient of lift $y = C_L$.
- The quantity q denotes parameters that specify the structure and coefficients of the family of models developed to represent the physical phenomenon of interest. These inputs

define the uncertainty domain, as explained below. The inputs *q* can include discrete and continuous variables that define modeling assumptions and functional forms.

The most important distinction between inputs *p* and *q* is that the control parameters *p* define the validation domain while calibration parameters *q* define the sources of uncertainty. The validation domain, denoted by D_P, represents the design space over which predictions are made and physical experiments performed. The prediction accuracy must be established for all settings *p* in the design domain D_P. Figure 2 illustrates a two-dimensional design space $(p_1;p_2)$ where predictions or experiments are made. The uncertainty domain, denoted by U_α, represents the sources of modeling uncertainty that must be propagated through the numerical simulations. The Probability Density Function (PDF) shown in Figure 2 illustrates the uncertainty of predicting the response *y*, when uncertainty is propagated through the numerical simulations from the inputs *q* that vary within the domain U_α to the output *y*. A key distinction between the validation (D_P) and uncertainty (U_α) domains is that different modeling choices may change the uncertainty domain, while the validation domain never changes.[4]

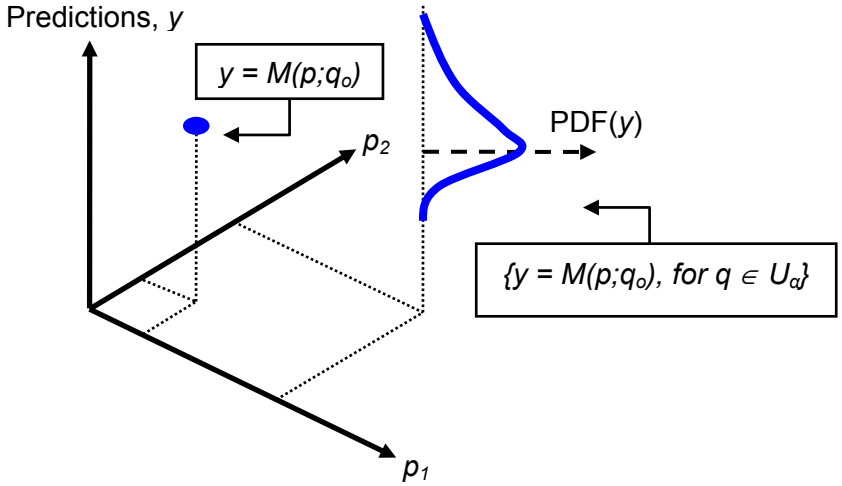


Figure 2. Validation domain, D_P, and propagation of uncertainty from U_α to y.

The family of predictive models is represented in a generic sense by the equation:

$$U_\alpha = \left\{ M(p;\, q) \;\middle|\; \|q - q_o\| \le \alpha \right\}, \quad \text{for} \quad \alpha \ge 0 \tag{2}$$

where the ***horizon-of-uncertainty*** is denoted by a positive scalar α. A member U_α of the family of models, at a given horizon-of-uncertainty α, includes all models consistent with the definition of uncertainty in equation (2). For example, the behavior of a material may not be known with certainty. Having to choose between, say, linear elastic, perfectly plastic, or visco-elastic models represents a lack-of-knowledge. Parameters *q* in equation (2) may include a flag that takes the values "linear", "plastic", and "visco-elastic", in addition to the unknown coefficients (modulus of

[4] An example is modeling the propagation of a transient through a structure. This is fundamentally a wave propagation problem that can be modeled with the continuous wave equation. In this case, unknowns represented by U_α are the speeds of sound of various materials and coefficients of wave reflection at the interface between materials. On the other hand, the same problem can be approximated using a discrete method such as finite element modeling. Unknowns represented by U_α are then the material properties, damping properties, and numerical coefficients of the spatial-temporal discretization. Definitions and dimensions of the two uncertainty domains differ. In both cases, however, the validation domain remains the same. It is defined, for example, by the range of loading conditions that must be simulated.

elasticity, yield stress, etc.) that define these models. In the absence of epistemic uncertainty, no alternative to the nominal model $y = M(p;q_o)$ would be feasible.

It is emphasized that the members U_α of the family of models become increasingly inclusive as the parameters q are allowed to differ from their nominal settings q_o. As the horizon of modeling uncertainty increases, more and more alternative models become candidates and are included in U_α. Note that these definitions are purposely broad to encompass a wide range of models and uncertainties.

2.2 Fidelity-to-data, *R*

Fidelity-to-data represents the distance R, assessed with the appropriate metrics, possibly a statistical test if probabilistic information is involved, between physical measurements y^{Test} and simulation predictions y at a given setting p:

$$R = \left\| y^{Test} - y \right\| \tag{3}$$

The symbol y^{Test} denotes physical measurements. Measurements are made at specific experimental configurations controlled by the parameters p. The norm || || defines the test-analysis correlation metric and is application-specific. Note that the fidelity metric R needs not be necessarily defined in terms of a difference between measured and predicted response features. Correlation coefficients as well as statistical tests of consistency between populations of values $\{y^{Test}\}$ and $\{y\}$ are admissible.

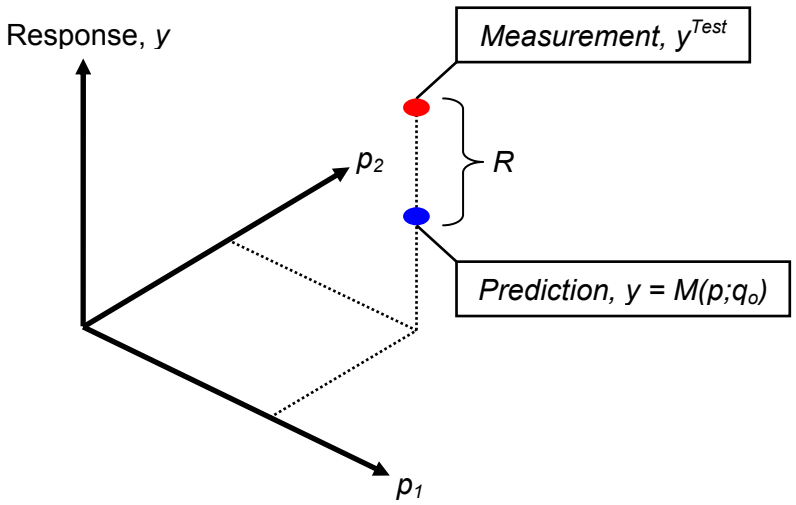


Figure 3. Measurement, prediction, and fidelity-to-data metric R.

Fidelity-to-data is pictured in Figure 3 as the vertical distance between a measurement y^{Test} and a prediction y for the physical experiment and numerical simulation performed at a given setting $(p_1;p_2)$. It is assumed in Figure 3 that the calibration variables q are kept constant and equal to their nominal values, $q = q_o$, because no "spread" of predictions (y) is shown.

2.3 Robustness-to-uncertainty, *α**

Robustness-to-uncertainty refers to the range of parameters q that provides no more than a given level R_{Max} of prediction error. The symbol R_{Max} denotes the ***aspiration*** of fidelity-to-data for all models considered. It represents a value of prediction error not to be exceeded. This means that a model is rejected during the robustness analysis if its fidelity-to-data is poorer than the aspiration, or $R > R_{Max}$.

The robustness α^* of the family of uncertainty models $\{U_\alpha\}$ for all values of the horizon-of-uncertainty ($\alpha > 0$) is defined mathematically by solving the following embedded optimization:

$$\alpha^* = \max_{\alpha \geq 0} \left\{ \min_{q \in U_\alpha} \{ R \leq R_{Max} \} \right\} \tag{4}$$

Equation (4) defines the robustness α^* as the largest amount of uncertainty that can be tolerated in the knowledge of the model and its parameters, while guaranteeing a fidelity-to-data at least equal to R_{Max}. It could happen that the robust-optimal model features a better fidelity-to-data, or $R < R_{Max}$, a situation referred to as ***opportunity*** from uncertainty [9]. Figure 4 illustrates the key point of robustness: the horizon-of-uncertainty α^* solution of equation (4) is the largest amount of uncertainty that can be tolerated while guaranteeing that all models included in the family U_{α^*} satisfy the aspiration of fidelity-to-data R_{Max}.

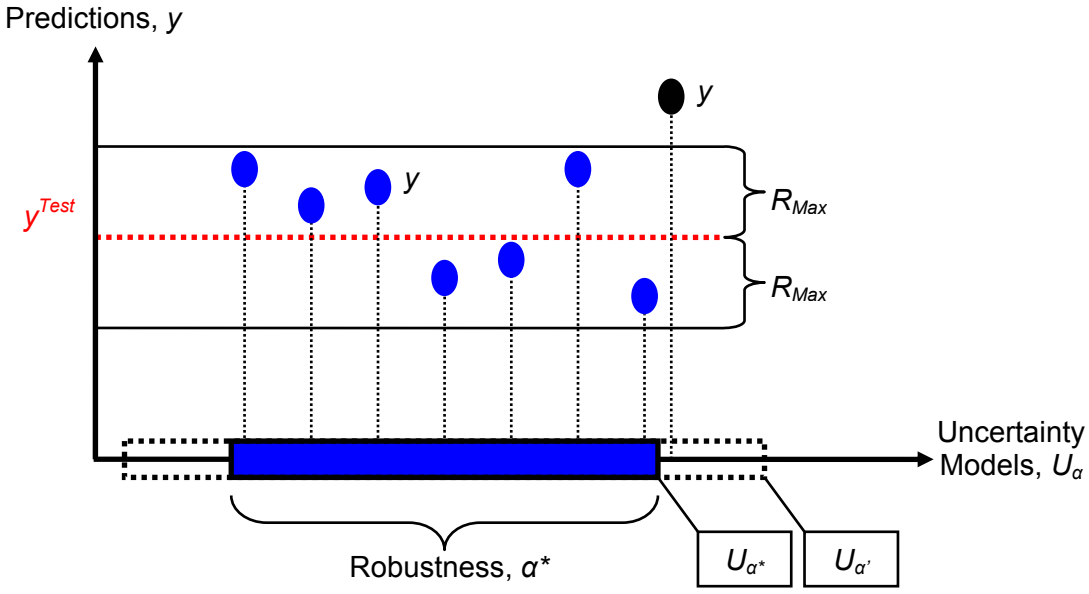


Figure 4. Robustness to uncertainty α^* at the aspiration of fidelity-to-data R_{Max}.

In Figure 4, the uncertainty domain U_α represented on the horizontal axis should not be confused with the validation domain shown in Figures 2 and 3. The validation domain D_F is two-dimensional and described by the pair $(p_1;p_2)$. The uncertainty domain U_α is one-dimensional, which means that the sources of epistemic uncertainty of the numerical simulation are described by a single scalar quantity q. Figure 4 illustrates that all models inside the uncertainty domain of size α^* provide prediction accuracies equal to or better than R_{Max}. However, robustness cannot be extended beyond the value α^* shown because the "next" family of models $U_{\alpha'}$ contains a member whose prediction accuracy is worse than R_{Max}.

The formalism developed through the concepts of horizon-of-uncertainty and family $\{U_\alpha\}$ accommodates a wide variety of uncertainty and lack-of-knowledge models. The only property upon which the definition of robustness (4) relies is the concept of ***structural nesting***. It simply means that increasing values of the horizon-of-uncertainty parameter α must result in nested domains U_α. Reference [9] gives examples of convex models that satisfy the nesting property. Clearly, a large robustness (α^*) is more desirable than a small one (α') because the family U_{α^*} encompasses all events defined in the family $U_{\alpha'}$, or $U_{\alpha'} \subset U_{\alpha^*}$. A large robustness indicates that

potentially large uncertainty or lack-of-knowledge does not deteriorate the prediction accuracy by more than R_{Max}.

2.4 Prediction Looseness, λ_Y

In this section we explore the "looseness" of model prediction: the range of predicted values deriving from equally robust models. The importance of prediction looseness stems from the fact that, to predict with confidence, there should be little difference (or small looseness λ_Y) between the predictions of equally robust models. Section 6 further explores the relationship between prediction looseness λ_Y, total uncertainty TU, and confidence in prediction C_F.[5]

As before, α^* denotes the robustness-to-uncertainty of models $y = M(p;q)$ and U_{α^*} denotes the family of models whose prediction accuracies are no worse than the aspiration R_{Max}. If the robustness α^* is large, then U_{α^*} contains a wide range of models. The prediction looseness of the family of models is defined as the range of predictions in U_{α^*}:

$$\lambda_Y = \max_{q \in U_{\alpha^*}} M(p;q) \; - \; \min_{q \in U_{\alpha^*}} M(p;q) \qquad (5)$$

Figure 5 illustrates prediction looseness by showing the range λ_Y of predictions obtained from all models included in the family U_{α^*} of robustness α^* and aspiration of fidelity-to-data R_{Max}.

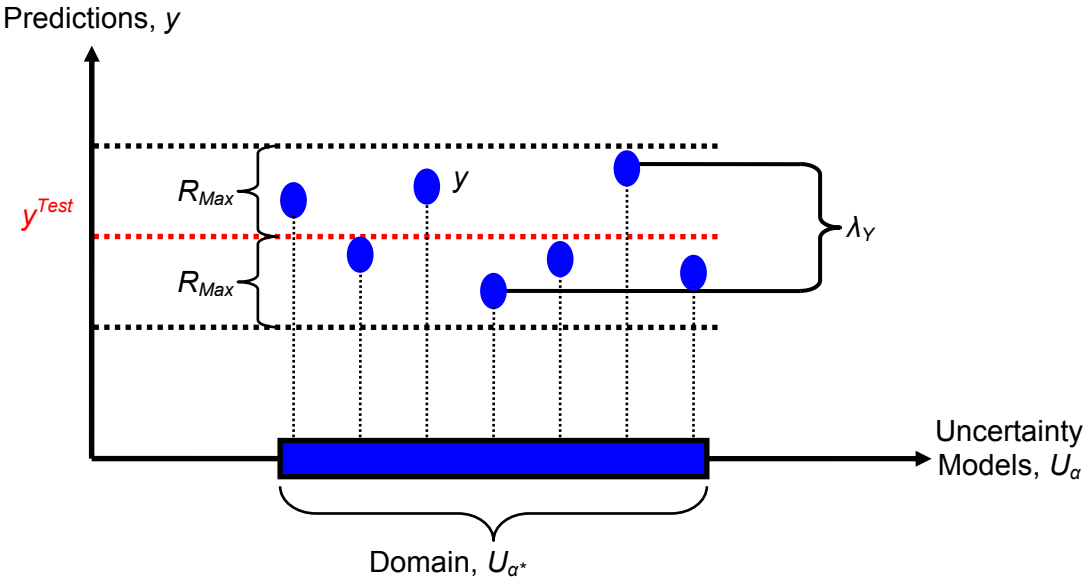


Figure 5. Prediction looseness λ_Y at robustness and fidelity aspiration (α^*;R_{Max}).

Defining the range of predictions (5) over a family of models is needed to achieve the main theoretical results derived in References [4, 5]. It may not be the most appropriate to represent the concept of confidence in prediction, but looseness and confidence are clearly related. The connection between λ_Y and C_F is that confidence generally increases when different sources of evidence reach the same conclusion. Confidence arises from consistency, which is intuitively connected to the notion of prediction looseness λ_Y over the family U_{α^*} of predictive models. This is important for model validation because one of the goals of V&V is to establish confidence in

[5] It is noted that a definition of "confidence" for science-based prediction has yet to be proposed. The only concept that is remotely connected is the notion of confidence interval in probability and statistics. The definition of confidence in Section 6 offers the advantages of conceptual simplicity and computational efficiency, but it is recognized that this is work in progress to a great extent.

science-based predictions by estimating the consistency, or lack thereof, provided by equally credible modeling techniques.

2.5 Modeling and Aggregating Uncertainty

The formalism developed through the concepts of horizon-of-uncertainty α and family of models $\{U_\alpha\}$ accommodates a wide variety of uncertainty modeling.[6] In fact, a family of models such as shown in equation (2) that, in addition, satisfies the nesting property (see Section 2.3) defines a model of ***information-gap***. In the theory of information-gap for decision-making, the difference between what is currently known and what needs to be known to make a decision is modeled. Models of ignorance are hence associated to gaps in knowledge [8, 9]. This is a significant departure from other representations of uncertainty, such as probability theory, that attempt to model randomness itself. Doing so requires strong assumptions that might not be justifiable in the case of severe lack-of-knowledge.[7]

It is not advocated here that epistemic uncertainty should be represented using information-gap. Our opinion is that there is no such thing as a "best" theory to represent uncertainty. Some theories, such as probability theory, are based on well-accepted axioms and they offer powerful algebraic rules [16]. Others, such as the Dempster-Shafer theory of plausibility and belief, can accommodate ambiguity and irrational reasoning [17]. Others yet, such as the theory of random sets, offer a degree of generality that is advantageous to derive models of uncertainty for sparse experimental data sets [18]. What should ultimately dictate the choice of a theory for modeling uncertainty is the purpose of the analysis and the amount of evidence available to justify the assumptions upon which the theory is based, or lack thereof.

If one accepts potentially different theories to represent uncertainty, what then becomes essential is the ability to aggregate, or combine, the uncertainty. The integration of uncertainty and definition of total uncertainty metrics are marginally addressed in Section 5. The short discussion presented in this publication does not communicate the depth of the research and development undertaken at Los Alamos National Laboratory in these areas. To read more on these topics, the reader is referred to References [19, 20].

3. Foundational Theorems of the "Myth" of Predictive Modeling

In this Section, the theoretical foundations needed to discuss the "myth" of science-based predictive modeling are developed. The main two theorems are summarized for completeness. The first one, which derives from a more general formulation in Reference [9], establishes the antagonism between fidelity-to-data and robustness-to-uncertainty. The second one establishes the antagonism between robustness-to-uncertainty and looseness in prediction, as shown in References [4, 5]. Section 6 extends these findings to the notion of confidence in prediction in the particular case of an interval-valued estimation of output uncertainty. At this point, closure is

---

[6] A first example is a probabilistic model of variability where standard deviation and covariance values are controlled by the parameter α. A second example is a possibility structure defined to represent a lack-of-knowledge, where the size of intervals is proportional to the parameter α. A third example is a family of fuzzy membership functions defined to represent expert judgment and linguistic ambiguity, where the membership functions are parameterized by the parameter α.

[7] In probability theory, for example, the frequency of occurrence of random events needs to be assessed. Enough measurements and observations might not be available to confidently derive a probability density function. In extreme cases, only ranges of values can be obtained. Defining a model of uncertainty such as probabilities, possibilities, or a fuzzy structure, might require assumptions that the available evidence simply does not support.

brought to the theoretical foundation and discussion of a strategy for science-based predictive modeling can start.

3.1 Strategies for Model Selection

We start by discussing strategies for model selection, where the question asked is the following one: *"Given two or more competing models, which one is most appropriate to solve a particular problem?"* Although V&V should not be reduced to a model selection problem, such question is of great interest for model validation because it constantly arises when analysts are faced with equally credible alternative modeling choices.

The conventional paradigm for model selection is, not surprisingly, to optimize goodness-of-fit or fidelity-to-data. We have seen that such strategy entails choosing the member of a family of models according to the optimal fidelity-to-data criterion. Fidelity-to-data defines a preference ordering where model $y = M(p;q_A)$ is preferred to model $y = M(p;q_B)$ if $R_A < R_B$. The model that provides the best goodness-of-fit or fidelity-to-data, R^*, is selected where R^* is the smallest of values for all models included in the uncertainty domain (U_α) up to the horizon-of-uncertainty α.

The inappropriateness of this strategy for choosing a model comes from the fact that the horizon-of-uncertainty, α, is generally unknown.[8] Not knowing the values of some parameters with certainty, or what the structural form of the model should be, represents a first level of epistemic uncertainty. A second level is to ignore how far from our best educated guess the solution should be searched, or whether that matters at all. Capturing the complexity of these two levels is to investigate the extent to which the best model performance is vulnerable to the lack-of-knowledge. This means that, in addition to searching for the best model within a family U_α, the robustness of its performance R^* to increasing levels of uncertainty should be examined.

Just like the fidelity-optimal strategy for model selection defines an ordering preference where the model $y = M(p;q_A)$ is preferred to the model $y = M(p;q_B)$ if $R_A < R_B$, the robustness-optimal strategy defines an ordering preference where the model $y = M(p;q_A)$ is preferred to the model $y = M(p;q_B)$ if the former is more robust to uncertainty, that is, $\alpha_A > \alpha_B$, at the common aspiration of fidelity-to-data R_{Max}. As mentioned previously, a large robustness is more desirable than a small robustness because it indicates that potentially large sources of lack-of-knowledge do not deteriorate the prediction accuracy by more than R_{Max}. An alternative model selection strategy is therefore to identify models associated with the largest robustness-to-uncertainty.

3.2 Theoretical Results

The first Theorem summarized below for completeness establishes that a trade-off arises between fidelity-to-data and robustness-to-uncertainty. Instead of fixing, somewhat arbitrarily, the level of lack-of-knowledge represented by the symbol α and optimizing the fidelity-to-data, robustness-to-uncertainty α^* can be maximized for a given aspiration of accuracy R_{Max}.

Theorem 1: Let $\{U_\alpha\}$ denote an information-gap family of models that obeys the axiom of nesting. Its fidelity and robustness functions are denoted by R and $\alpha^* = \alpha^*(q_o;R_{Max})$, respectively. Consider two requirements of fidelity, $R_{A,Max}$ and $R_{B,Max}$. If $R_{A,Max} \geq R_{B,Max}$, then $\alpha^*(q_o;R_{A,Max}) \geq \alpha^*(q_o;R_{B,Max})$.

[8] An example in mechanical engineering is the definition of a friction coefficient between two materials. A value may be available from the literature, but the extent of the variability is typically unknown. What is even more difficult to assess is the suitability of the Coulomb friction model, for which a friction coefficient is sought, to represent the mechanics of friction. Friction un-doubtfully involve stick-and-slip and complex micro-mechanics that the Coulomb model only approximates. The accuracy of the model compared to the "true-but-unknown" behavior is generally unknown.

Theorem 1 states that robustness-to-uncertainty increases monotonically as the minimal required aspiration of fidelity-to-data increases. Recall that R is defined in equation (3) as a test-analysis correlation error. It implies that increasing R_{Max} corresponds to a decrease in prediction accuracy. The Theorem therefore states that prediction accuracy and robustness-to-uncertainty are antagonistic attributes of the family of models.[9] A proof can be found in Reference [9].

The second Theorem establishes that a trade-off arises between robustness-to-uncertainty and looseness in prediction, or the range of predicted values obtained from models which all satisfy a specified aspiration of fidelity-to-data. The notion of prediction looseness is important because it relates to the confidence that one has in predictions of equally credible models. The Theorem implies that maximizing robustness-to-uncertainty may have detrimental effects on the ability of equally credible models to make consistent predictions.

Theorem 2: Let $\{U_\alpha\}$ denote an information-gap family of models that obeys the axioms of nesting and translation.[10] Its robustness and looseness functions are denoted by $\alpha^* = \alpha^*(q_o;R_{Max})$ and $\lambda_Y = \lambda_Y(q_o;R_{Max})$, respectively. Consider two initial models, $q_{A,o}$ and $q_{B,o}$. If $\alpha^*(q_{A,o};R_{Max}) \geq \alpha^*(q_{B.o};R_{Max})$, then $\lambda_Y(q_{A,o};R_{Max}) \geq \lambda_Y(q_{B.o};R_{Max})$.

Theorem 2 states that looseness in prediction increases as the robustness-to-uncertainty increases. Greater looseness means that the models included in the family U_{α^*} (centered about the "nominal" model $q_{A,o}$, with robustness $\alpha^*(q_{A,o};R_{Max})$, and aspiration of fidelity-to-data R_{Max}) tend to make predictions inconsistent with one another. It ultimately translates into a lesser degree of confidence in the prediction accuracy of the family of models, as shown in Section 6. The Theorem therefore states that robustness-to-uncertainty and looseness in prediction are antagonistic attributes of the family of models.[11] A proof can be found in Reference [7].

The proofs provided in References [7, 9] rely on a description of uncertainty that uses the theory of information-gap. The main advantage is that no restrictive assumption is made regarding the source of uncertainty and type of mathematical representation. Information-gap models can be built to encompass a wide range of uncertainty: probabilistic, non-probabilistic, linguistic ambiguity, modeling lack-of-knowledge, etc. Conventional descriptions of uncertainty, such as probabilistic models, can be viewed as models of information-gap as long as families of nested convex sets can be defined. This makes the results of Theorems 1 and 2 applicable to a wide range of situations. It is noted, however, that a practical limitation of robustness analyses is the amount of calculations needed to solve the saddle-point optimization of equation (4).

---

[9] Note that the antagonism between accuracy and robustness does not mean that it is impossible to find a high-fidelity model that is, at the same time, very robust to the uncertainty considered when defining the family $\{U_\alpha\}$. It may be possible to derive a model that is true to the data and robust to the uncertainty. The Theorem simply states that, if such model exists, increasing its fidelity-to-data even more will result in a degradation of robustness. Similarly, increasing its robustness even more will deteriorate its accuracy.

[10] Nesting expresses that, as the horizon-of-uncertainty increases, the family of models includes all previously included models, plus new ones. Translation expresses that two families of models that share the same horizon-of-uncertainty only differ in their center points. A simple example of information-gap model that satisfies these two properties is a scalar q that varies in an interval $[-\alpha; +\alpha]$. The nominal value of q is $q_o = 0$, and increasing values of α define a family of nested intervals. These two technical points are needed to prove the Theorem, but it can be verified that they do not restrict its applicability.

[11] As noted previously for the antagonism between accuracy and robustness, it may be possible to define a highly robust family of models that make highly consistent predictions. The Theorem simply states that increasing its robustness even more will result in a loss in prediction consistency. Similarly, "tightening" the range of predictions made (or reducing the looseness) will come at the cost of loosing robustness.

3.3 The Trade-offs of Prediction Credibility

It is our contention that three quantities are central to the discussion of science-based prediction credibility: fidelity-to-data of the family of models, R_{Max}; robustness-to-uncertainty, α^*; and looseness in prediction, or range of predictions, λ_Y. The trade-off of Theorem 1 between fidelity and robustness can be best expressed by the compact inequality:

$$\frac{\partial \alpha^*}{\partial R_{Max}} \geq 0 \tag{6}$$

which means that an increase in robustness comes at the cost of relaxing the aspiration of prediction accuracy or, equivalently, increasing the maximum authorized error R_{Max}. Likewise, the trade-off of Theorem 2 between robustness and looseness can be expressed as:

$$\frac{\partial \lambda_Y}{\partial \alpha^*} \geq 0 \tag{7}$$

which means that an increase in robustness comes at the cost of loosing consistency between the predictions made by all models included in the family, up to the fidelity and robustness levels R_{Max} and α^*, respectively.

Clearly, multiplying equations (6) and (7) provides a third inequality that expresses that λ_Y increases when R_{Max} is increased. Hence the discussion:

- ***Robustness decreases as fidelity improves.*** Numerical simulations or models made to better reproduce the available test data become more vulnerable to potential errors in modeling assumptions, errors in the functional form of the model, and uncertainty and variability in the model parameters.
- ***Looseness increases as robustness improves.*** Numerical simulations or models that are more immune to uncertainty and modeling errors provide a wider range of predictions. This translates into less consistency between the predictions of models that belong to the same family, which decreases confidence in our ability to forecast configurations, settings or environments that have not been tested experimentally.
- ***Looseness decreases as fidelity improves.*** Numerical simulations or models made to better reproduce the available test data provide more consistent predictions when asked to forecast settings that have not been tested experimentally. Although intuitive, this is not necessarily a good thing when modeling is extrapolated to configurations very different from those tested. It may lead to a false sense of confidence achieved artificially through excessive calibration of the models.

These trade-offs imply that it is not possible to ***improve***, simultaneously, fidelity-to-data, robustness-to-uncertainty, and consistency in predictions. High fidelity (small R_{Max}) implies that the models are true to the measurements, which adds warrant to the family of models. Large robustness (large α^*) strengthens belief in the validity of the family of models because its members are less vulnerable to epistemic uncertainty. High consistency in predictions (small λ_Y) implies that all the models that are equivalent in terms of fidelity, also agree in their predictions when forecasting new, potentially not yet observed, behaviors. Our analysis shows that past measurements, accompanied by incomplete understanding of the measured process, cannot unequivocally establish true prediction of the behavior of the system.

One may now ask what these theoretical results entail for the problem of model selection. It has been argued that the commonly encountered paradigm of optimizing goodness-of-fit is not appropriate because it leads to non-unique solutions. In addition, this strategy negates the fact that the magnitude of our ignorance is often unknown. Worse, fidelity-optimal models provide no

robustness-to-uncertainty, which means that small errors in modeling assumptions or parameter values can lead to large prediction errors [9]. We are not advocating that fidelity-optimality, as a decision strategy for selecting and validating models, be systematically replaced by robustness-optimality. Nevertheless, the basis for selecting and validating models should be to understand the trade-off between the aspiration of fidelity-to-data, R_{Max}, and robustness-to-uncertainty, α^*, for a given requirement of confidence in prediction, C_F. In Section 6, a connection is proposed between looseness (λ_Y) and confidence (C_F) via a total uncertainty metric defined for intervals.

4. An Application to Predictive Modeling in Engineering

The theory discussed in Section 3 is illustrated with results from an engineering application. The purpose is to validate the predictions of numerical models that simulate the behavior of a non-linear crushable foam material when traversed by a short-duration impact wave. Details of the experimental set-up and modeling uncertainty can be obtained from References [14, 15]. In the following, the modeling and main source of uncertainty are briefly described. Numerical predictions are compared to measurements collected during four experiments to establish the accuracy of each model. Trade-offs between fidelity-to-data and robustness-to-uncertainty are illustrated. Extension to the third dimension of confidence in prediction is proposed in Section 7, after a connection between looseness and confidence has been derived.

4.1 Numerical Modeling of the Crushable Foam Impact Experiment

In References [14, 15], the behavior of the layers of foam material subjected to impacts are simulated via finite element modeling and analysis. Here, the numerical models are based on single degree-of-freedom oscillators that obey the following equation of vibration:

$$m\left(\ddot{x}(t)-\gamma(t)\right)+c\dot{x}(t)+kx(t)+F_{NL}(t)=0 \tag{8}$$

where m, c, and k denote the mass, viscous damping, and linear stiffness coefficients and F_{NL} represents the contribution of a non-linear internal force. The oscillator is initially at rest and the equation of motion is integrated over a period of time $[t_0;t_F]$. The signal $\gamma(t)$ is specified by the analyst; it represents the acceleration record that results from the impact applied at the base of the layer of foam material.

To integrate the equation of motion, a sub-model of internal force must be specified by the analyst. Three examples are the bi-linear, quadratic, and cubic non-linearity models illustrated in equations (9), (10), and (11), respectively:

$$F_{NL}(t)=\begin{cases} F_o+k_A x(t), & \text{if } |x|\le x_L \\ F_o+k_B x(t), & \text{if } |x|\ge x_L \end{cases} \tag{9}$$

$$F_{NL}(t)=F_o+k_2\left(x(t)\right)^2 \tag{10}$$

$$F_{NL}(t)=F_o+k_3\left(x(t)\right)^3 \tag{11}$$

The physical meaning of coefficients such as $(F_o;k_A;k_B)$ for the bi-linear model is unclear because the equation is a mathematical idealization, not based on a fundamental understanding of how the crushable foam behaves. Likewise, the values of coefficients $(m;c;k)$ of equation (8) cannot be directly inferred from dimensional or material analysis of the foam material. These sub-models and values of coefficients are unknown, yet, the analyst must select them to make predictions. Convex models of uncertainty are developed to account for the lack-of-knowledge. Robustness analyses assess the vulnerability of prediction accuracy to the choice of forcing function sub-model and coefficient values.

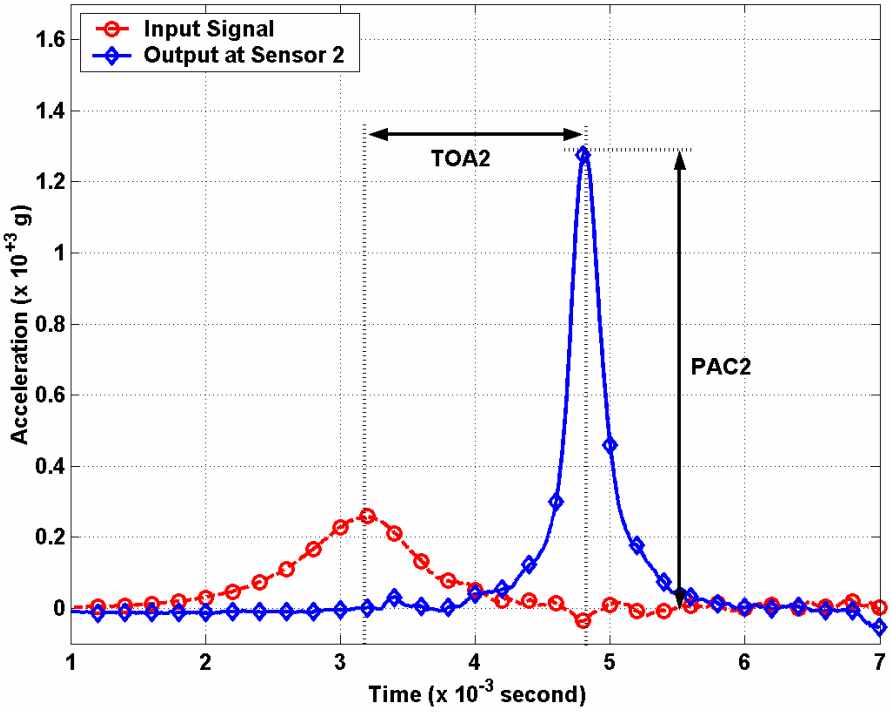


Figure 6. Definition of the peak acceleration (PAC) and time-of-arrival (TOA) features.

The numerical simulation provides an acceleration signal from which two response features are extracted. They are the peak acceleration and time-of-arrival, that is, the time it takes to the peak acceleration to travel from one side of the foam layer to the other. The response features are denoted by the symbols *PAC* and *TOA*, respectively. Figure 6 shows an input acceleration signal $\gamma(t)$ in red, dashed line; the corresponding output $x(t)$ in blue, solid line; and how the *PAC* and *TOA* features are calculated. The prediction accuracy of a model is quantified by calculating a test-analysis correlation metric between the measured and predicted values of *(PAC;TOA)*.

4.2 Sources of Lack-of-knowledge and Models of Uncertainty

The main source of uncertainty considered in this analysis arises from not understanding the constitutive behavior of the crushable foam. Figure 7 shows data obtained from four physical experiments. The center curve, labeled *"Sample 1"* and shown in blue solid line, represents the nominal strain-stress curve. The other curves are acceptable realizations of material behavior. Such uncertainty matters greatly because selecting a constitutive law that describes how the material behaves or, equivalently, selecting an internal force sub-model $F_{NL}(t)$, is a critical step of building the numerical simulation.

Figure 7 illustrates a rather severe lack-of-knowledge about the material, which raises the question of how to represent such uncertainty mathematically. Clearly, deriving a probability law based solely on the evidence captured by Figure 7 would be nothing short of crystal-ball reading. For the same reason, we are not confident postulating a possibility structure, basic Dempster-Shafer probability assignments, or fuzzy membership functions, to name only a few. It is recognized that more testing could be performed, and formal expert elicitation techniques are available to help capture knowledge. The merits of acquiring more knowledge, in one form or another, can never be over-stated. Nevertheless, Figure 7 illustrates a practical reality where

decisions must often be made in the context of severe uncertainty because of constraints such as timetables, budgets, staffing, and lack of testing.

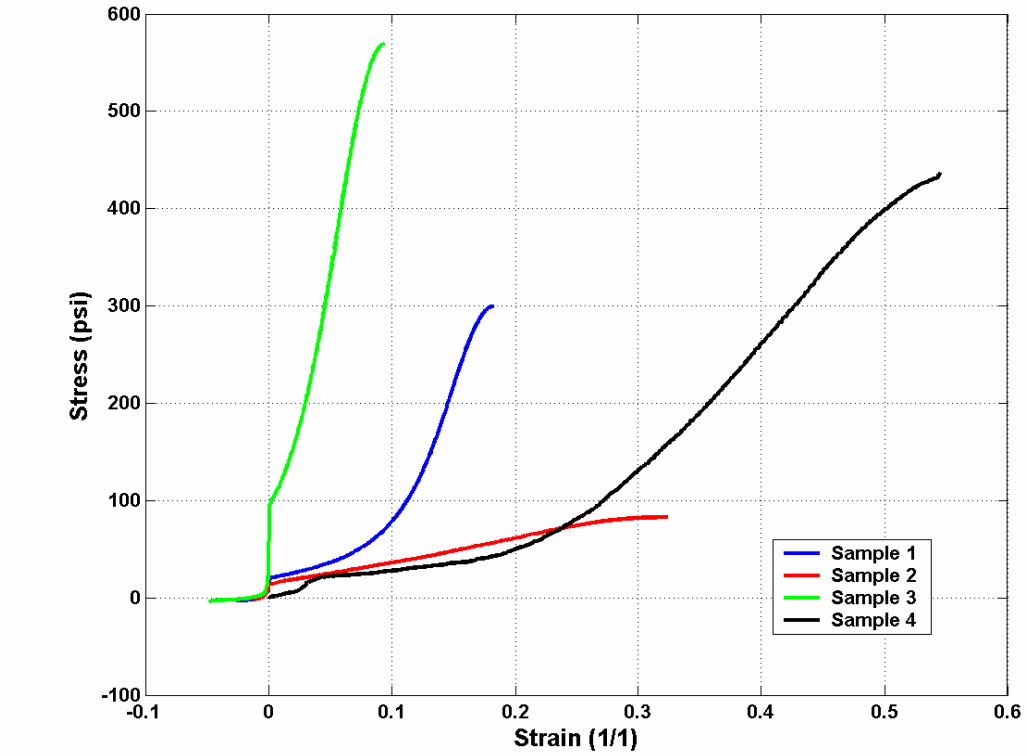


Figure 7. Four test samples of strain-stress curve for the crushable foam material.

In this study, the material behavior uncertainty is represented by a family of convex sets U_α. The nominal stress-strain curve is denoted by q_o. It represents the best available knowledge, here, the curve labeled *"Sample 1"* in Figure 7. The horizon-of-uncertainty α is a parameter that has no explicit physical meaning but measures how far away from the nominal knowledge one is willing to define a strain-stress curve, denoted by q. A curve is defined by selecting the bi-linear model (9), quadratic model (10), or cubic model (11), and its appropriate coefficients. For simplicity, the analysis is restricted to the bi-linear, quadratic, and cubic models, but nothing would prevent us from including other types of strain-stress curves or non-parametric models. Distance between a strain-stress curve, q, used for the numerical simulation and the nominal curve, q_o, is quantified with the Root Mean Square (RMS) metric:

$$U_\alpha = \left\{ \text{Strain} - \text{stress Curves "q" such that } \sqrt{(q-q_o)^T (q-q_o)} \le \alpha \right\} \qquad (12)$$

Definition (12) indicates that an internal force sub-model (9), (10), or (11) is included in the uncertainty domain U_α if the corresponding strain-stress curve, q, does not deviate from the nominal material model, q_o, by more than the RMS distance α. Figure 8 illustrates conceptually the domains U_α. Three domains are shown for increasing values of the horizon-of-uncertainty, $\alpha_1 < \alpha_2 < \alpha_3$. It can be verified that definition (12) satisfies the nesting property. All strain-stress curves included in the domain U_α are automatically included in the domains defined by larger values of the horizon-of-uncertainty. The family $\{U_\alpha, \alpha>0\}$ becomes increasingly inclusive of material models as the uncertainty represented by the parameter α increases.

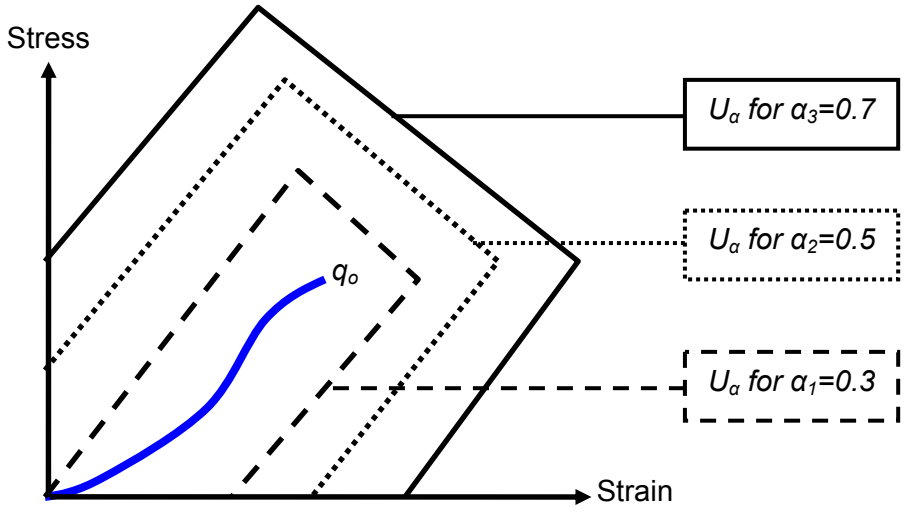


Figure 8. Family of convex models of uncertainty $\{U_\alpha, \alpha>0\}$ for the non-linear forcing function.

Another key departure between information-gap analysis of robustness and other strategies for decision-making, such as reliability, is that the former accommodates unbounded horizons of uncertainty. We do not know how far the "true" strain-stress curve of the foam material is from our nominal best-guess. Therefore, the amount of uncertainty should not be determined a priori. The horizon-of-uncertainty, α, is an unknown of the robustness analysis.

Finally, it is mentioned that epistemic uncertainty is restricted to the constitutive behavior of the crushable foam. Other unknowns include the coefficients of the mass, viscous damping, and linear stiffness contributions to equation (8). They are kept constant and equal to calibrated values obtained by searching for the linear oscillator whose predictions best match the test data.

4.3 Physical Experiments and Domain of Validation

Physical experiments are performed to measure the response of the crushable foam to impact loading. Impacts are generated using a drop table launched from various heights. In addition, foam pads of different thicknesses are used during the experiments. Four settings are tested by combining low and high impact heights with thin and thick foam pads [14, 15].

During each physical experiment, the input and output acceleration signals, $\gamma(t)$ and $x(t)$, are recorded and response features $y^{Test} = (PAC;TOA)$ are extracted from the measurements. In addition, each impact is repeated several times (ten replicates with low drop heights, five with high drop heights) to estimate the effects of environmental variability and unknown conditions of the experiment that cannot be controlled. Populations of measured features y^{Test} are collected from which statistics can be estimated, such as the mean vector of response features and the matrix of variance and covariance coefficients.

The concept of validation domain introduced in Section 2.1 needs to be defined for the crushable foam impact application. Control parameters of the physical experiments, drop height and foam pad thickness, define the two-dimensional domain within which predictions must be obtained. One reason for wanting to substitute numerical predictions to physical experiments is that experimenting with all combinations of drop height and foam pad thickness would be too expensive and time consuming. Analyzing the equation of motion (8), on the other hand, is very efficient, provided that prediction accuracy can be guaranteed. The two parameters, drop height

and foam pad thickness, therefore define the operational space of interest. Validation is achieved when the prediction accuracy of the numerical simulation and, consequently, the sub-model of internal force that idealizes the constitutive behavior of the crushable foam, have been assessed within the operational domain.

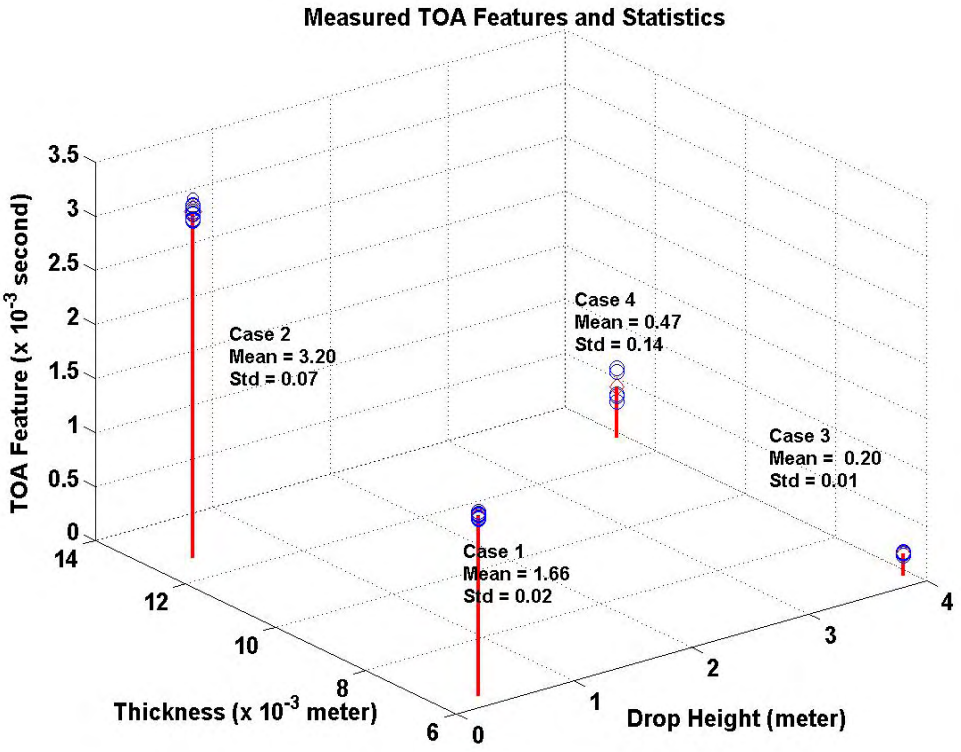


Figure 9. Validation domain, four configurations tested, and TOA measurements.

Figure 9 shows the two-dimensional domain of validation and measured values of response feature *TOA*. According to the nomenclature defined in Section 2.1, the two control parameters are denoted by $p = (p_1;p_2)$, and they represent the drop height and foam pad thickness. The replicate measurements of the time-of-arrival feature are shown for each configuration tested.

4.4 Quantitative Metric of Test-analysis Correlation

The metric defined to quantify the prediction accuracy of the numerical simulation (8), or test-analysis correlation error, is a weighted L^2 norm of the prediction error:

$$y(p;q) = \begin{Bmatrix} PAC \\ TOA \end{Bmatrix}, \qquad R = \left(y^{Test} - y(p;q)\right)^T W_{yy}^{-1} \left(y^{Test} - y(p;q)\right) \tag{13}$$

where $y(p;q)$ represents the two-feature vector predicted by the numerical simulation and y^{Test} is the mean vector of measurements. The weighting W_{yy} is a constant matrix that eliminates the dimensional difference between the units of *PAC* and *TOA*. It is initialized using variance and covariance coefficients estimated from the population of replicate measurements. The notation $y(p;q)$ emphasizes that predictions are made for a given configuration of drop height and foam pad thickness (denoted by p) and given internal force sub-model (denoted by q).

The weighted L^2 norm defined in equation (13) accumulates prediction error that originates from both features, *PAC* and *TOA*, without any power of discrimination. In Section 4.5, relative differences are also calculated to establish the accuracy of *PAC* predictions separate from the accuracy of *TOA* predictions:

$$R = 100\left|\frac{y^{Test} - y(p;q)}{y^{Test}}\right|\% \qquad (14)$$

where $y(p;q)$ and y^{Test} now represent a single feature, either *PAC* or *TOA*. In equation (14), the prediction error R is expressed in percent of the mean measured value.

4.5 Trade-offs Between Prediction Accuracy and Robustness-to-uncertainty

The results of an information-gap analysis of robustness are now examined to study the trade-offs between fidelity-to-data and robustness-to-uncertainty. Figure 10 shows the best and worst prediction errors obtained for increasing values of the horizon-of-uncertainty parameter. Results are reported for a single configuration of the system where impact tests are performed with the low drop height and thin foam pad. The horizontal axis represents prediction errors (14) for the peak acceleration *PAC* feature. The vertical axis represents the horizon-of-uncertainty, α.

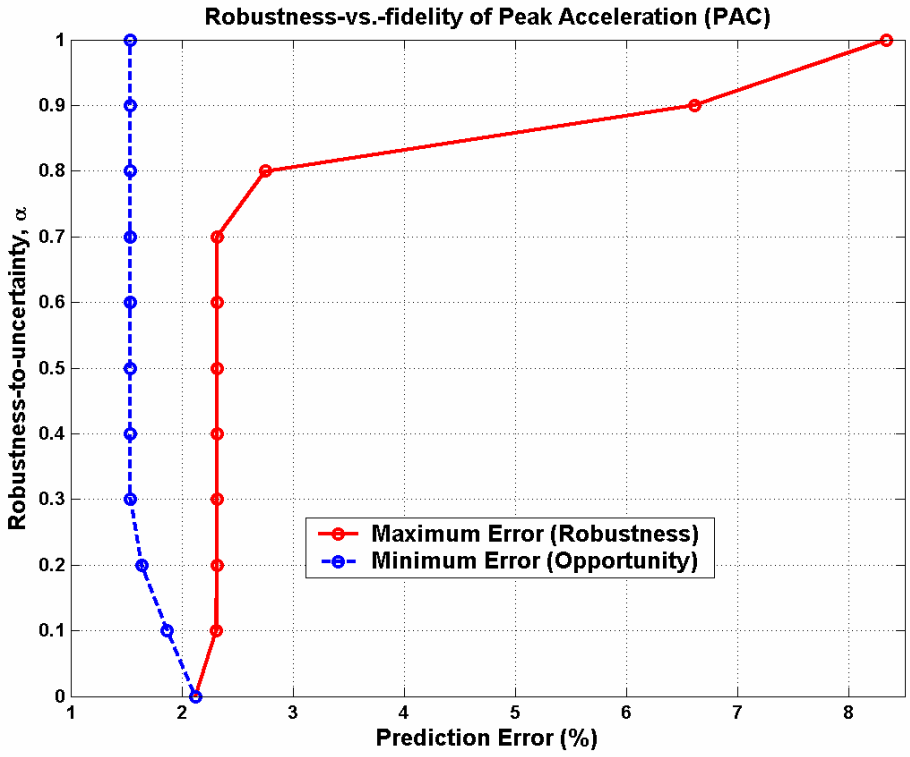


Figure 10. Robustness and opportunity for the configuration (low impact; thin foam pad).

The procedure to obtain the curves shown in Figure 10 is briefly summarized. At a given horizon-of-uncertainty, say α = 0.1, a domain U_α is defined according to equation (12). U_α sets bounds on the family of strain-stress curves, q, that are considered legitimate alternatives to our current model, q_o. An optimization problem is solved to search, within U_α, for the sub-model of internal force that yields the worst prediction error, that is, the maximum value of R defined in equation (14). The solution is the point R = 2.3% shown on the robustness curve (red, solid line) at α = 0.1. A second optimization problem is then solved to search, within U_α, for the sub-model

of internal force that yields the least prediction error. The solution is the point R = 1.9% shown on the opportunity curve (blue, dashed line) at α = 0.1. The main constraint of both optimization problems is that strain-strain curves must be realized within the bounds imposed by the domain U_α. The procedure is iterated for increasing values of the horizon-of-uncertainty, α.

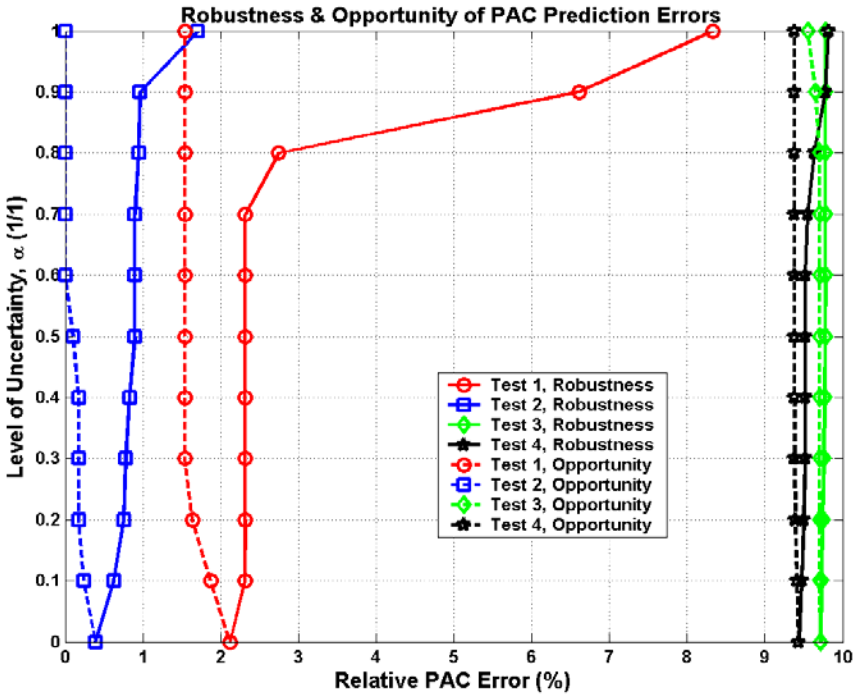
The robustness curve (red, solid line) shows the worst prediction accuracy obtained at any level of uncertainty. The opportunity curve (blue, dashed line) shows the best possible accuracy. Opportunity illustrates that uncertainty can sometimes be taken advantage of to obtain better than expected fidelity-to-data. The robustness and opportunity curves can be discontinuous but the fact that increasing horizon-of-uncertainty levels generate nested domains U_α guarantees their monotonic natures, increasing for robustness and decreasing for opportunity.

When the robustness and opportunity curves are considered together, Figure 10 shows the ranges of prediction accuracy that can be obtained at any level of uncertainty. For example, the prediction error is guaranteed within [1.5%; 2.7%] at the horizon-of-uncertainty level of α = 0.8. No matter which material model is selected from the family U_α, its prediction accuracy for Test 1 will be no worst than 2.7% but no better than 1.5%. A slight increase of uncertainty to α = 0.9 results in a potential deterioration of prediction accuracy to [1.5%; 6.6%]. The robustness-to-uncertainty, α^*, is defined according to equation (4) as the maximum level of uncertainty that can be tolerated while guaranteeing a minimum requirement of fidelity-to-data, R_{Max}. Figure 10 shows that the robustness of the family of models is α^* = 0.8 at the requirement R_{Max} = 3.0%. As Theorem 1 indicates, fidelity-to-data worsens as lack-of-knowledge increases, hence, illustrating the antagonistic nature between truthfulness to data and robustness-to-uncertainty.

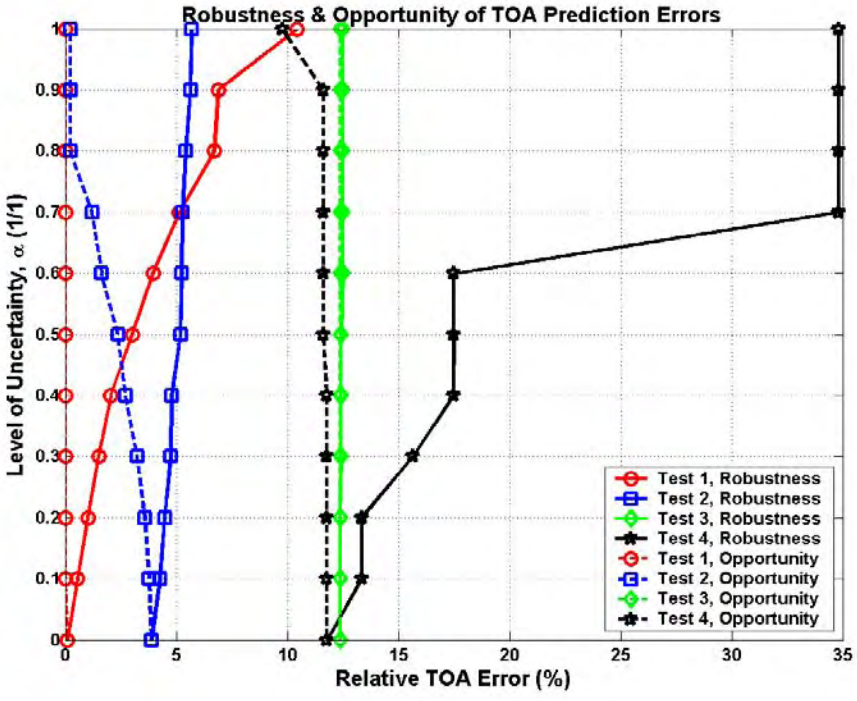
Naturally, the analysis can be repeated for each configuration of the system. Figures 11-a and 11-b show the robustness and opportunity curves obtained with the four configurations for predicting features *PAC* and *TOA*, respectively. Robustness curves, that is, the worst prediction accuracy obtained at any level of uncertainty, are shown with solid lines. Opportunity curves, that is, the best prediction accuracy obtained at any level of uncertainty, are shown with dashed lines. Figure 11-a indicates that the family of models is very robust to the modeling uncertainty when predicting *PAC* for Tests 3 and 4 (high drop height) because the prediction error does not change significantly even at high levels of lack-of-knowledge. Figure 11-b indicates that this is not the case for prediction of the *TOA* response feature.

Finally, the results pictured in Figure 11 are used to assess fidelity-to-data and robustness-to-uncertainty throughout the two-dimensional validation domain. Validation is parameterized by the drop height and foam thickness because the purpose of modeling is to develop a predictive capability to exercise different combinations of the two design variables $(p_1;p_2)$. Extrapolation must be employed here because measurements are not available to calculate a test-analysis correlation metric other than at the settings $(p_1;p_2)$ that have been tested. A simple polynomial-based extrapolation is performed to obtain the results shown in Figures 12 and 13. Figure 12 shows the robustness versus fidelity for predicting the *PAC* feature at six discrete uncertainty levels, α = 0.1, 0.3, 0.6, 0.8, 0.9, and 1.0. Likewise, Figure 13 shows the results for predicting the *TOA* feature. The two-dimensional surfaces indicate the expected prediction error at settings $(p_1;p_2)$ that have or have not been tested.[12] Extrapolation also makes it possible to estimate the overall robustness α^* of the family of models to analyze any combination of drop height and foam thickness. Such information is extremely useful, for example, to decide where to allocate resources for the next round of physical experiments.

[12] This is, of course, conditioned upon the assumption that the extrapolation is correct. In this work, several polynomials have been developed and the solutions presented in Figures 12 and 13 optimize goodness-of-fit while presenting no evidence of over-fitting the data.

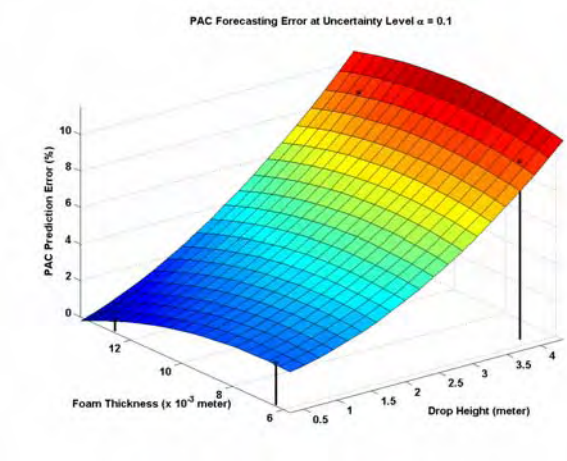


(11-a) Robustness and opportunity versus accuracy for PAC predictions.

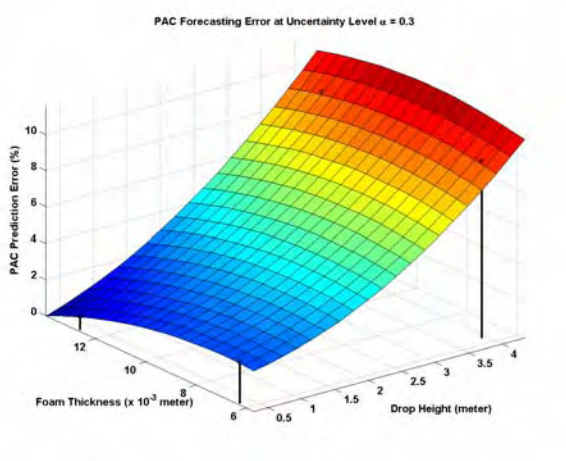


(11-b) Robustness and opportunity versus accuracy for TOA predictions.

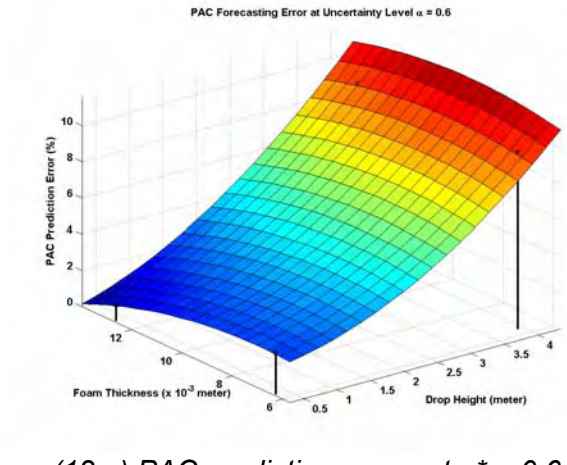
Figure 11. Robustness and opportunity analyses of the four configurations of the system.



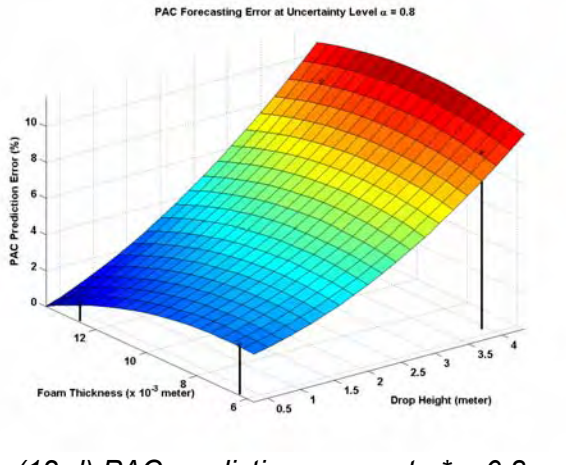
(12-a) PAC prediction errors at α = 0.1.*



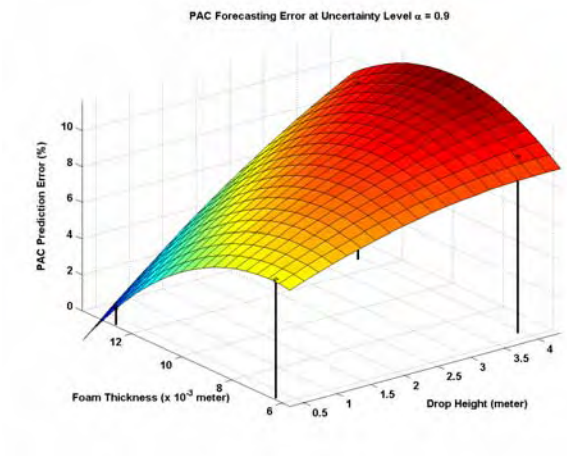
(12-b) PAC prediction errors at α = 0.3.*



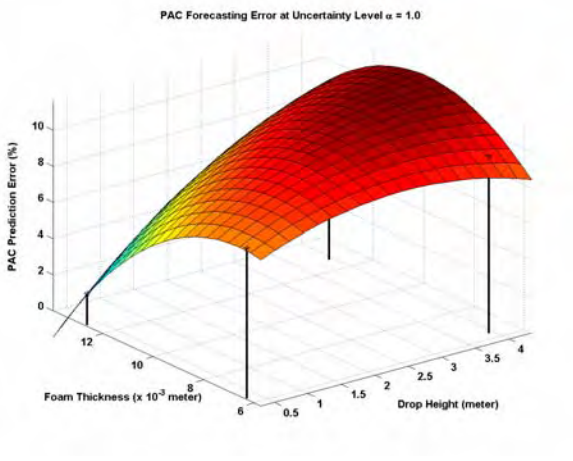
(12-c) PAC prediction errors at α = 0.6.*



(12-d) PAC prediction errors at α = 0.8.*

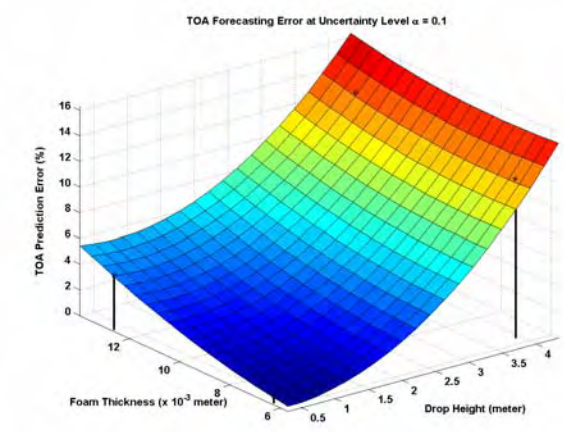


(12-e) PAC prediction errors at α = 0.9.*

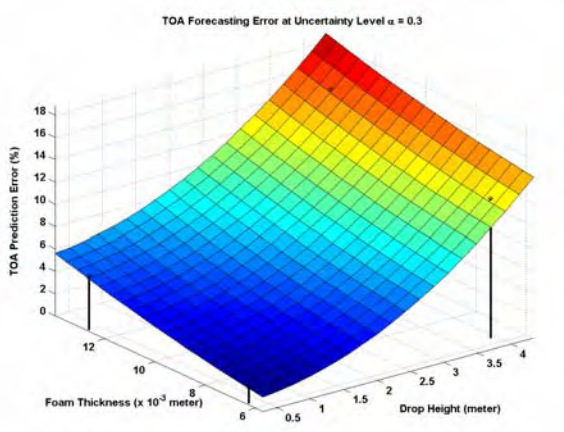


(12-f) PAC prediction errors at α = 1.0.*

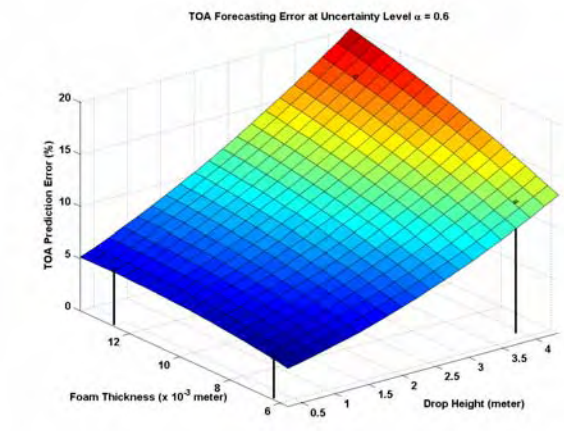
Figure 12. Extrapolated robustness-vs.-fidelity surfaces for the PAC feature.



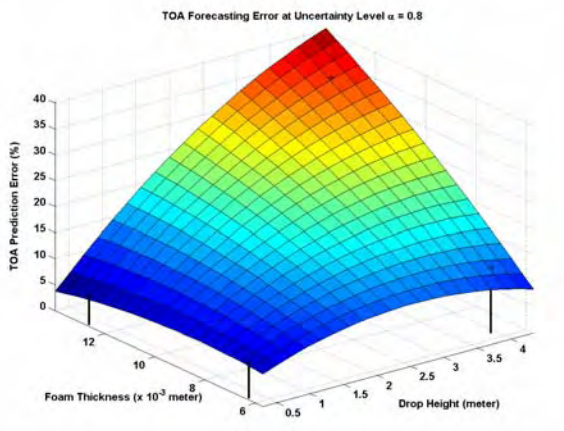
(13-a) TOA prediction errors at α = 0.1.*



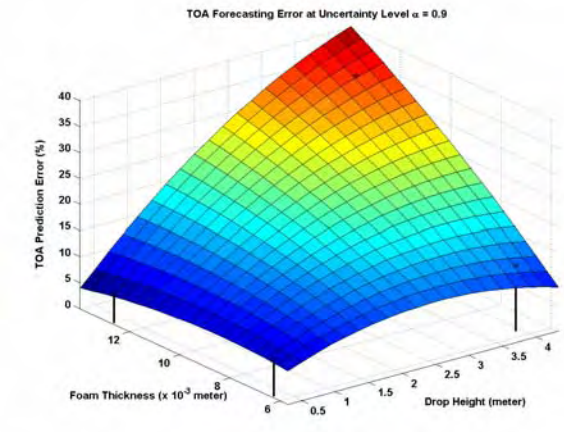
(13-b) TOA prediction errors at α = 0.3.*



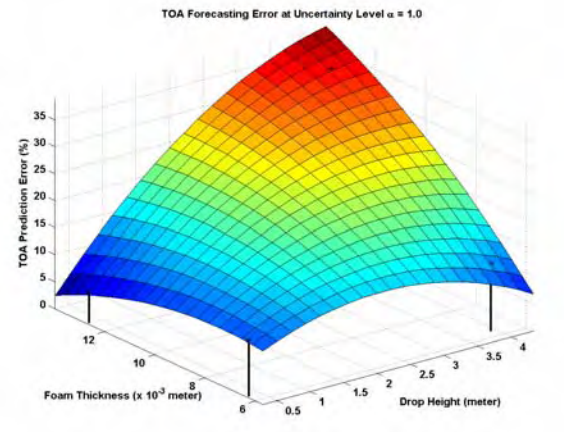
(13-c) TOA prediction errors at α = 0.6.*



(13-d) TOA prediction errors at α = 0.8.*



(13-e) TOA prediction errors at α = 0.9.*



(13-f) TOA prediction errors at α = 1.0.*

Figure 13. Extrapolated robustness-vs.-fidelity surfaces for the TOA feature.

5. Aggregating Uncertainty for Model Validation

In this Section, the concept of total uncertainty, *TU*, introduced in References [19, 20] is briefly summarized. It is then extended to interval-valued uncertainty models. Although not central to the discussion of science-based predictive modeling, a metric of total uncertainty is needed as a practical means to get to the notion of confidence. Section 6 proposes a definition of confidence based on the *TU* metric. The step from prediction looseness (λ_Y) to confidence (C_F) via the *TU* metric is essential to connect fidelity R_{Max}, robustness α^*, and confidence C_F. For completeness, it is mentioned that other applications where the aggregation of uncertainty plays a prominent role include V&V, reliability analysis, and system certification [10, 11, 12, 20].

5.1 The Total Uncertainty Metric

This Section summarizes the on-going development at Los Alamos National Laboratory of a metric to aggregate the various forms of uncertainty. Total uncertainty, denoted by the symbol *TU*, is defined as the combination of the two general types of uncertainty: natural variability and lack of specific information.[13]

A first requirement of the *TU* metric is that it must be able to aggregate different kinds of uncertainties represented by the collection of available mathematical theories. These include measure-based theories such as evidence theory, possibility theory, fuzzy set theory, random intervals, imprecise probabilities, and probability theory, collectively known as the Generalized Information Theories (GIT) [21, 22]. A second requirement is to develop a metric for model validation. The metric would aggregate uncertainty estimates coming from physical experiments (measurements), numerical simulations (calculations), and expert judgment (opinions). A third requirement of the *TU* metric is that it must generalize to multivariate or multi-dimensional comparisons when several response features or time series are considered. The fourth and final requirement is to develop a scaled metric to facilitate the relative comparison between several values. A natural, although arbitrary, choice is that the value $TU = 0$ represents the case of complete certainty while the value $TU = 1$ (or $TU = +\infty$) represents the case of total uncertainty.

In the original development of the *TU* metric, emphasis is given to two of the more prevalent theories of uncertainty for test data and model choice: probability and possibility, respectively [19, 20]. The fundamental difference is that probability theory concentrates evidence on the singletons of a universe of information, whereas possibility theory locates evidence on collections of nested sets within the universe of information. These differences in mathematical properties of the two theories make each one suitable for modeling various types of uncertainty and less suitable for modeling others.[14]

The development of the total uncertainty metric begins by, first, defining an information matrix, $H_{m,2}$, that collects the possibility and probability distributions in each of its two columns. The distributions characterize the uncertainty associated with the estimation of a scalar-valued response feature *y*. Data whose uncertainty is represented by probabilities may come from a collection of N_{Test} replicate measurements, $\{y^{Test\ (k)},\ k = 1 \ldots N_{Test}\}$, while data whose uncertainty is

[13] Variability is, by definition, irreducible. It cannot be reduced, but only quantified. Lack-of-knowledge, on the other hand can be reduced with the acquisition of more information. The taxonomy that uses the terms aleatoric (irreducible) and epistemic (lack-of-knowledge or reducible) is often encountered. This taxonomy specifically distinguishes statistical variation from other, reducible forms of uncertainty.

[14] For example, probability theory is ideal for formalizing uncertainty in situations where event frequencies are known or evidence is based on outcomes of a large number of independent and repeatable trials. Possibility theory, by contrast, is ideal for formalizing incomplete information expressed in terms of vague or ambiguous terms, or where evidence supports conflicting events.

represented by the possibility distribution may originate from the analysis of N_{Model} simulations, $\{y^{(k)}, k = 1 \ldots N_{Model}\}$. The information matrix is defined as:

$$H_{m,2} = \begin{bmatrix} \pi_1 & p_1 \\ \vdots & \vdots \\ \pi_i & p_i \\ \vdots & \vdots \\ \pi_m & p_m \end{bmatrix} \tag{15}$$

where π_i and p_i are the possibility and probability, respectively, of the i^{th} discrete estimate of the value of the (unknown) response feature y. Note that, while the number of physical experiments (N_{Test}) may be different from the number of models analyzed (N_{Model}), the populations of y values must be discretized into the same number of m bins in equation (15). The calculation of the *TU* metric relies on a Singular Value Decomposition (SVD) of the information matrix $H_{m,2}$ [23]. The SVD provides orthogonal sets of left and right singular vectors, and a set of singular values:

$$H_{m,2} = U \Sigma V^T, \quad U^T U = Id, \quad V^T V = Id, \quad \Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \tag{16}$$

The meaning of the singular vectors contained in the columns of matrices U and V is still under investigation in the context of a decomposition of information matrix. The singular values σ_1 and σ_2 can be viewed as measuring the amount and consistency of information provided by the possibility, $\{\pi_i\}$, and probability, $\{p_i\}$, distributions. An analogy with the analysis of time series in Structural Dynamics is that the singular values represent the amount of energy contained in the signals [24]. For the characterization of uncertainty, the singular values represent the ***total energy*** of the uncertainty in $H_{m,2}$, or simply the total uncertainty. The *TU* metric is given by:

$$TU = 2\left(\frac{\sigma_1^2 + \sigma_2^2}{\pi_{max}^2 + p_{max}^2} - 1 \right) \tag{17}$$

where π_{max} is the largest possibility value in the first column of $H_{m,2}$, and p_{max} is the largest probability value in the second column of $H_{m,2}$.

An important property of definition (17) is that the values of the *TU* metric are, by definition, scaled between zero and an upper bound that only depends on m, the number of bins of the information matrix $H_{m,2}$:

$$0 \le TU \le 2(m-1) \tag{18}$$

The total uncertainty metric is guaranteed to provide values between two extreme conditions on uncertainty, that is, between the case of no uncertainty, $TU = 0$, and the case of maximum uncertainty, $TU = 2(m-1)$. Normalizing *TU* values by the upper bound $2(m-1)$ provides positive numbers between zero and one, which makes it possible to compare them efficiently and also eliminates dependency on the number of bins, m. It is recognized, however, that this scaling is arbitrary and that numerical values between zero and one have no physical meaning.

5.2 Extension to Multivariate Analysis and Other General Information Theories

The framework developed by collecting uncertainty distributions in the columns of an information matrix, and processing it using the SVD, makes it possible to extend the definitions (15-18) to multivariate analysis and/or the inclusion of other GIT representations of uncertainty. Even though this is still work in progress at Los Alamos National Laboratory, the main steps are briefly summarized for completeness.

The information matrix is first expanded to include other columns, representing other types of uncertainties. Research is on-going to include uncertainty models such as Dempster-Schafer belief functions [17], random intervals and sets [18], fuzzy membership functions [25], imprecise probabilities [26], all of which play important roles in the kinds of uncertainties often experienced in a V&V process. The information matrix is denoted by $H_{m,N}$ where m is the number of discretization bins (number of rows) and N is the number of distributions (number of columns):

$$H_{m,N} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,N} \\ \vdots & \vdots & \cdots & \vdots \\ h_{i,1} & h_{i,2} & \cdots & h_{i,N} \\ \vdots & \vdots & \cdots & \vdots \\ h_{m,1} & h_{m,2} & \cdots & h_{m,N} \end{bmatrix} \tag{19}$$

The column $\{h_{i,j},\ i=1\ldots m\}$ represents the j^{th} uncertainty distribution of a response feature y. The matrix may collect uncertainty distributions expressed with different theories for a single feature; or distributions expressed with the same theory for different features; or a combination of the previous two. Mixing different uncertainty representations and features is possible here because the SVD provides an automatic normalization. Consequently, the decomposition does not suffer from the adverse effects of ill-conditioning that could result from collecting information about mathematically different theories or physically different features.

The second step is, as before, to perform the SVD of the information matrix:

$$H_{m,N} = U\,\Sigma\,V^{T}, \qquad \Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_N \end{bmatrix} \tag{20}$$

where $\sigma_1, \sigma_2, \ldots, \sigma_N$ are the N singular values that characterize the amount of uncertainty and consistency between distributions. The third step is the calculation of the *TU* metric, defined as:

$$TU = N\left(\frac{\sum_{j=1}^{N} \sigma_j^2}{\sum_{j=1}^{N} h_{\max,\,j}^2} - 1 \right) \tag{21}$$

where $h_{max,j}$ denotes the maximum value of the j^{th} column, that is, $h_{max,j} = max\{h_{i,j},\ i=1\ldots m\}$. The definition (21) produces values scaled between zero and an upper bound that only depends on the dimensions of the information matrix:

$$0 \le TU \le N(m-1) \tag{22}$$

As mentioned previously, normalizing *TU* values by the upper bound $N(m-1)$ provides positive numbers between zero and one and eliminates dependency on the size of $H_{m,N}$. Future work will investigate this approach to estimate total uncertainty in the context of the GIT.

5.3 TU Metric for Interval-valued Uncertainty

Following the same procedure outlined in equations (15-18), a definition of the *TU* metric is proposed in the case where multivariate uncertainty is represented by intervals with crisp end-points. It is assumed that N response features, denoted by $y_{i,j}$ for $j=1\ldots N$, are analyzed.

For the crushable foam impact discussed in Section 4, two response features ($N = 2$) are considered: $y_{i,1}$ is the Peak Acceleration (*PAC*) and $y_{i,2}$ is the Time-of-arrival (*TOA*). What

motivates the development of a *TU* metric in the case of an interval-valued representation of uncertainty is that ranges of *PAC* and *TOA* features are obtained during the robustness and opportunity analyses. A range is an interval $[y_{min,j};y_{max,j}]$. The minimum bound, $y_{min,j}$, is estimated by solving a minimization problem given bounds within which the input parameters vary. Likewise, the maximum bound, $y_{max,j}$, is estimated by solving a maximization problem given bounds for the input parameters. The estimates of $y_{min,j}$ and $y_{max,j}$ can alternatively be estimated from a Monte Carlo simulation if probability distributions were available to characterize the input uncertainty. At any horizon-of-uncertainty α, the two optimization problems provide a rigorous propagation of interval-valued uncertainty from inputs to output responses of the simulation.

Because the only information available is the range $[y_{min,j};y_{max,j}]$ for each feature of interest, the information matrix becomes a two-row, *N*-column matrix:

$$H_{2,N} = \begin{bmatrix} y_{\min,1} & y_{\min,2} \cdots & y_{\min,N} \\ y_{\max,1} & y_{\max,2} \cdots & y_{\max,N} \end{bmatrix} \tag{23}$$

The information matrix $H_{2,N}$ is factorized using the SVD in equation (16) and *TU* is defined as:

$$TU = N\left(\frac{\sum_{j=1}^{N} \sigma_j^2}{\sum_{j=1}^{N} y_{\max,\,j}^2} - 1 \right), \qquad 0 \le TU \le N \tag{24}$$

where values are bounded between $TU = 0$ (case of absolute certainty) and $TU = N$ (case of complete uncertainty). Note that the value $TU = 0$ is obtained if and only if the lower and upper bounds of each feature are equal, that is, $y_{min,j} = y_{max,j}$ for $j=1…N$. It means that the analysis of the numerical simulation is deterministic and the intervals of uncertainty collapse down to a single point, which is consistent with absolute certainty.

For the crushable foam impact discussed in Section 4 where $y_{i,1}$ is the peak acceleration and $y_{i,2}$ is the time-of-arrival, equation (24) becomes:

$$TU = 2\left(\frac{\sigma_1^2 + \sigma_2^2}{PAC_{\max}^{\,2} + TOA_{\max}^{\,2}} - 1 \right), \qquad 0 \le TU \le 2 \tag{25}$$

The calculations of confidence in prediction in Section 7 are based on the *TU* metric (25) where the two features of the crushable foam impact simulations are the *PAC* and *TOA*.

6. Foundations of Confidence in Prediction

We have argued that three quantities are central to the discussion of science-based prediction credibility: fidelity-to-data of the family of models, R_{Max}; robustness-to-uncertainty, α^*; and looseness in prediction, or range of predictions, λ_Y. Our purpose, however, is to ultimately estimate confidence in predicting behavior that may or may not have already been observed. In this Section, a connection is established between looseness λ_Y and confidence C_F using the total uncertainty metric defined in Section 5. The theoretical results of Section 3 are extended from $(R_{Max};\alpha^*;\lambda_Y)$ to the triplet $(R_{Max};\alpha^*;C_F)$. Section 7 illustrates the trade-offs between fidelity, robustness, and confidence with numerical results from the crushable foam.

6.1 Confidence in prediction

The reason why the concept of confidence in prediction is of interest is that it is often central to the accreditation or certification of complex engineered systems. Certification is here defined as the assessment of the overall system performance and its ability to meet design, safety and

other requirements. An example of certification exercise is the mission of stockpile stewardship of the Los Alamos National Laboratory that provides an assessment of the performance and reliability of the physics package of nuclear weapons. Credibility of the decision-making process clearly depends on the levels of scientific rigor and confidence that one places in the evidence assembled to support the decisions. Confidence in prediction is especially critical when science-based simulations, such as provided by the ASC Program at Los Alamos, are developed to predict environments or behaviors that may never have been tested experimentally [27].

Our discussion of confidence in prediction is largely paraphrases comments made in Reference [20] where the authors note that, if predictions are made on the response of a system and the level of uncertainty expressed in these predictions are close to the extreme of no uncertainty, then credibility in that prediction exists. On the other hand, if the uncertainty is closer to the case of maximum uncertainty, then less credibility exists. It is therefore important to develop a metric of "credibility" or "confidence" that scales monotonically with the quantified level of uncertainty and, in a mathematical sense, measures the degree of closeness. Jane Booker and her co-authors also note that:

> *"Confidence is a commonly used term whose definitions include words like trust, belief, reliance, and certitude. It is the state of feeling sure [28]. Note, however, that even the great Greek philosophers were unable to precisely, or mathematically, define what is meant by confidence. Outside of the statistical context, there is no definition for the mathematical meaning or quantification of confidence.[15] Therefore, we discourage its use in V&V and uncertainty quantification studies unless defined using the statistical definitions. However, we are willing to note that confidence seems to have an inverse relationship to uncertainty."*

The previous discussion gives rise to three basic requirements that a metric developed to quantify confidence in prediction, and denoted by the symbol C_F here, must satisfy:

- The confidence in prediction metric, C_F, is a positive number.
- Numerical values must be bounded between a minimum, $C_F = 0$, that expresses a complete lack of confidence and a maximum, $C_F = 1$ or $C_F = +\infty$, that expresses total confidence.
- Confidence has an inverse relationship to uncertainty.

It is noted that the first two requirements are somewhat arbitrary. Others, such as requiring that C_F varies monotonically or linearly with the level of uncertainty or requiring continuity and differentiability, could be added to the list. It is, however, not our intent to propose an axiomatic definition for confidence metrics. Instead, the simplest possible expression is defined:

$$C_F = 1 - \left(\frac{TU}{TU_{max}} \right), \qquad 0 \le C_F \le 1 \tag{26}$$

[15] Also quoted from Reference [20]: *"In statistical sciences, confidence has a specific meaning when referring to a confidence interval for an unknown parameter. The interpretation of a confidence interval is often misused. It refers to a sampling process and the calculation of multiple confidence intervals for multiple repeated samples. For example, if one were to take a hundred samples from a population and calculate a hundred times the 95% confidence intervals for an unknown parameter, such as the mean, then 95 of those confidence intervals would contain the true value of the mean. Another statistically based confidence definition is in common use. The so-called confidence level is defined as the complement of a significance level in statistical hypothesis testing. The confidence level is (1-α), where α is the significance level or the Type I error. The Type I error is a controlled error in statistical inference, it refers to the chance (e.g., 5%) that a null hypothesis is rejected when it is true and should not have been rejected."*

where TU_{max} denotes the upper bound of the *TU* metric, defined by equation (22) in the general case. Clearly, the definition satisfies the above three requirements.[16] It can also be verified that complete lack of confidence leads to the value of $C_F = 0$. Likewise, total confidence corresponds to the value of $C_F = 1$. For the interval-valued uncertainty of the crushable foam application, the *TU* metric from equation (25) is used and calculations of C_F values are based on:

$$C_F = 2 - \left(\frac{\sigma_1^2 + \sigma_2^2}{PAC_{max}^2 + TOA_{max}^2} \right), \qquad 0 \le C_F \le 1 \tag{27}$$

It is emphasized that the definition proposed here of a quantitative metric for confidence in prediction is not meant to be final. It is likely to be revised in future work. Nevertheless, the C_F metric captures in the simplest possible way trends expected to be observed between prediction looseness from a family of models, uncertainty of the predictions, and confidence.

For completeness, other attempts at defining confidence during a V&V process are briefly mentioned. In References [29, 30], statistical confidence interval estimation is considered to define validation metrics in cases where replicate measurements are performed and multiple predictions are obtained from simulation. These metrics remain within the realm of probabilistic representation of uncertainty and, therefore, take advantage of the narrow-but-solid foundation of confidence in statistical sciences. Another promising approach is the QRC metric developed at the Lawrence Livermore National Laboratory [11, 12]. QRC stands for Quantifying Reliability at Confidence and it provides a risk-based methodology for analyzing the reliability of complex systems and validating predictions against test data.[17] By recognizing that a reliability number depends on the level of confidence with which the estimate can be obtained, Roger Logan and his co-authors implicitly define a framework to explore the trade-offs of the performance-confidence pair $(R;C_F)$, which is not unlike the one proposed here for the triplet $(R_{Max};\alpha^*;C_F)$.

6.2 Relationship Between Looseness λ_Y and Confidence C_F

The relationship between looseness λ_Y and confidence C_F is now investigated in the context of the interval-valued uncertainty model discussed in Section 5.3. To extend the Theorems of Section 3 from $(R_{Max};\alpha^*;\lambda_Y)$ to the triplet $(R_{Max};\alpha^*;C_F)$, one must understand how the uncertainty and confidence metrics *TU* and C_F vary when looseness λ_Y is increased or decreased. Because the definition of confidence in equation (26) depends on the *TU* metric, derivations are made specific to the interval-valued representation of uncertainty proposed to analyze the crushable foam application.

The main result is intuitive: An increase in looseness λ_Y, that is, less consistent predictions of a family of models or less consistency in a body of evidence collected to support a decision, translates into more uncertainty (*TU* increases) and less confidence (C_F decreases). The main

---

[16] Definition (26) is based on the mathematical relationship $y=1-x$ to express that the quantities x and y are inversely related. Alternatives include $y=1/x$, $y=e^{-x}$, etc. It is argued that such choice is somewhat arbitrary. It also depends on the range of numerical values, $C_F \in [0;1]$ or $C_F \in [0;+\infty]$. Our preference, at this point, goes to a C_F metric that varies between zero and one. This is to prevent translation of the value "$C_F=+\infty$" into a statement of "infinite confidence". Although appealing in the context of common language, it is our opinion that there is no such thing as a scientific statement or mathematical theory in which we can place infinite confidence.

[17] Reliability is defined in a broad sense. In the context of conventional reliability analysis, it is obtained as $R = 1-P_F$, or the probability of not obtaining failure. In the context of test-analysis correlation for V&V, it is the result of statistical testing that assesses the probability that measurements and predictions come from the same parent population.

contribution of this Section is a formal proof in the particular case of the interval-valued model of uncertainty proposed by equations (23-24).

Starting from the definition (23), the information matrix $H_{2,2}$ that collects uncertainty about predictions made for two features $y_{i,1}$ and $y_{i,2}$ is written as:

$$H_{2,2} = \begin{bmatrix} y_{\min,1} & y_{\min,2} \\ y_{\max,1} & y_{\max,2} \end{bmatrix} \tag{28}$$

where uncertainty about the true-but-unknown value of the j^{th} feature, $j = 1$ or 2, is represented not by a distribution, but an interval $[y_{min,j};y_{max,j}]$. The analysis can be performed for any feature of the response although peak acceleration and time-of-arrival are considered for the crushable foam application. In addition, the symbol $\lambda_{Y,j}$ denotes the prediction looseness of the j^{th} feature:

$$\lambda_{Y,j} = \left(y_{\max,j} - y_{\min,j}\right) \tag{29}$$

The proof starts by calculating the singular values of matrix $H_{2,2}$. By definition, the singular values, σ_i, are equal to the eigen-values of the squared matrix:

$$\left(H_{2,2}\right)^T H_{2,2} = \begin{bmatrix} a & c \\ c & b \end{bmatrix} \tag{30}$$

where:

$$a = y_{\min,1}^2 + y_{\max,1}^2, \qquad b = y_{\min,2}^2 + y_{\max,2}^2, \qquad c = y_{\min,1}y_{\min,2} + y_{\max,1}y_{\max,2} \tag{31}$$

The eigen-values of matrix (30) are calculated by solving the second-order equation:

$$\sigma^2 - \sigma(a+b) + ab - c^2 = 0 \tag{32}$$

that admits two solutions:

$$\sigma_1 = \frac{a+b}{2} + \sqrt{\left(\frac{a+b}{2}\right)^2 + c^2 - ab}, \quad \sigma_2 = \frac{a+b}{2} - \sqrt{\left(\frac{a+b}{2}\right)^2 + c^2 - ab} \tag{33}$$

Finally, it can be verified that the sum of squares of the two singular values is equal to:

$$\sigma_1^2 + \sigma_2^2 = a^2 + b^2 + 2c^2 \tag{34}$$

With these intermediate results established, the sensitivity of the *TU* metric to changes in the looseness $\lambda_{Y,j}$ can be studied. A positive partial derivative of *TU* with respect to $\lambda_{Y,j}$ indicates that *TU* and $\lambda_{Y,j}$ vary sympathetically. A negative partial derivative indicates that *TU* and $\lambda_{Y,j}$ vary antagonistically. Extending the results to the confidence metric is trivial since $C_F = 1-(TU/TU_{max})$.

The *TU* metric derived in equation (25) for the crushable foam application depends on the sum of squares of the two singular values. Substituting equation (34) leads to:

$$\begin{aligned} \frac{\partial TU}{\partial \lambda_{Y,j}} &= \frac{2}{\left(y_{\max,1}^2 + y_{\max,2}^2\right)} \frac{\partial\left(\sigma_1^2 + \sigma_2^2\right)}{\partial \lambda_{Y,j}} \\ &= \frac{2}{\left(y_{\max,1}^2 + y_{\max,2}^2\right)} \frac{\partial\left(a^2 + b^2 + 2c^2\right)}{\partial \lambda_{Y,j}} \end{aligned} \tag{35}$$

Partial derivatives with respect to $\lambda_{Y,j}$ are calculated, without any loss of generality, by assuming that the upper bounds $y_{max,i}$ do not vary and substituting $(y_{max,i} - \lambda_{Y,i})$ for $y_{min,i}$. It results that:

$$\frac{\partial TU}{\partial \lambda_{Y,1}} = \frac{8\left(a y_{\min,1} + c y_{\min,2}\right)}{\left(y_{\max,1}^2 + y_{\max,2}^2\right)}, \qquad \frac{\partial TU}{\partial \lambda_{Y,2}} = \frac{8\left(b y_{\min,2} + c y_{\min,1}\right)}{\left(y_{\max,1}^2 + y_{\max,2}^2\right)} \tag{36}$$

The two features of interest are $y_{i,1} = PAC_i$ for the peak acceleration and $y_{i,2} = TOA_i$ for the time-of-arrival. Their values are positive numbers, and it follows that the sign of the partial derivatives shown in equation (36) is always positive:

$$\frac{\partial \mathrm{TU}}{\partial \lambda_{\mathrm{Y}}} \geq 0 \tag{37}$$

where the subscript $(\,)_i$ identifying the feature is dropped for simplicity. It follows that:

$$\frac{\partial \mathrm{C_F}}{\partial \lambda_{\mathrm{Y}}} = \frac{\partial}{\partial \lambda_{\mathrm{Y}}}\left(1 - \frac{\mathrm{TU}}{\mathrm{TU_{max}}}\right) = -\frac{1}{\mathrm{TU_{max}}}\frac{\partial \mathrm{TU}}{\partial \lambda_{\mathrm{Y}}} \leq 0 \quad \blacksquare \tag{38}$$

As suspected, confidence in prediction C_F is decreased when the range of predictions made by a family of models, or the lack of consistency of a body of evidence, increases. The proof has been proposed for the information matrix $H_{2,2}$ defined by equation (28), and it is easy to verify that it can be extended to any information matrix $H_{2,N}$ such as defined in equation (23).[18] It has not been verified that this result can be extended to cases such as the possibility-probability model of equation (15) or the other GIT representations of equation (19).

6.3 Discussing the "Myth" of Predictive Modeling

The Theorems of Section 3 have established that an increase in robustness-to-uncertainty comes at the cost of relaxing the aspiration of prediction accuracy or, equivalently, increasing the maximum authorized error R_{Max}. Likewise, prediction looseness increases with robustness. These results are expressed compactly by the following two inequalities:

$$\frac{\partial \alpha^*}{\partial \mathrm{R_{Max}}} \geq 0\,, \qquad \frac{\partial \lambda_{\mathrm{Y}}}{\partial \alpha^*} \geq 0 \tag{39}$$

Extension to confidence in predictions is straightforward by combining equations (38) and (39):

$$\frac{\partial \mathrm{C_F}}{\partial \alpha^*} = \underbrace{\frac{\partial \mathrm{C_F}}{\partial \lambda_{\mathrm{Y}}}}_{\text{Negative}} \overbrace{\frac{\partial \lambda_{\mathrm{Y}}}{\partial \alpha^*}}^{\text{Positive}} \leq 0 \tag{40}$$

The final trade-offs between aspiration of fidelity-to-data R_{Max}, robustness-to-uncertainty α^*, and confidence in prediction C_F, are expressed by the following two inequalities:

$$\frac{\partial \alpha^*}{\partial \mathrm{R_{Max}}} \geq 0\,, \qquad \frac{\partial \mathrm{C_F}}{\partial \alpha^*} \leq 0 \tag{41}$$

In conclusion, it is not possible to improve the fidelity-to-data and, simultaneously, make the family of models more robust to epistemic uncertainty. Likewise, it is not possible to improve robustness-to-uncertainty and, simultaneously, increase confidence that the models will predict

[18] The argumentation that the proof given using matrix $H_{2,N}$ is valid for any other matrix $H_{2,N}$ relies on two technicalities. The main reason is that a two-row, N-column matrix such as $H_{2,N}$ admits only two non-zero singular values, no matter how many columns $N \geq 2$. Fundamentally, the case of matrix $H_{2,N}$ is therefore no different, with the exception that derivations such as equation (33) will change. A second technicality is to study the sign of expressions given in equation (36). A negative sign could be obtained if the response features take negative values, which then results in sympathetic variations of the pair $(C_F;\lambda_Y)$. In the event of negative feature values, entries in matrix $H_{2,N}$ can be altered by shifting them to guarantee that they are positive. Shifting the values is not a concern, as long as minima $y_{min,j}$ and maxima $y_{max,j}$ are shifted by the same amounts, because what matters are the ranges $\lambda_{Y,j} = (y_{max,j} - y_{min,j})$, not the values themselves.

environments or behaviors that have not been observed experimentally. A word of caution is necessary to remind the reader that the proof has only been provided in the special case of a two-feature prediction with interval-valued representation of uncertainty. Note, however, that no restrictive assumption is made about the models of input uncertainty, as long as they define a family of convex nested sets when the horizon-of-uncertainty parameter, α, increases.

It is expected that, in the case of a general information matrix $H_{m,N}$ defined in equation (19), the existence of an antagonism between robustness, α^*, and confidence, C_F, will depend on the nature and representation of the evidence collected in matrix $H_{m,N}$. Special cases may exist where robustness-to-uncertainty and confidence in prediction turn out to be sympathetic.

Our interpretation of these results is ***not*** that achieving science-based predictive credibility is a myth. Clearly, this work suggests that it is impossible to find models that match the available test data "perfectly" while being "highly" robust to the lack-of-knowledge and providing "infinite" confidence in their forecasting ability or predictive power. The good news is that a framework has been outlined to study which requirements of fidelity, robustness, and confidence are attainable given the current limitation of our knowledge, and which combinations of $(R_{Max};\alpha^*;C_F)$ are not feasible. Understanding that science-based prediction has limits should not be pretext for not developing much needed predictive modeling tools. Instead, understanding the trade-offs of $(R_{Max};\alpha^*;C_F)$ should be the mechanism through which prediction credibility is established.

7. An Application to Quantifying Confidence in Prediction

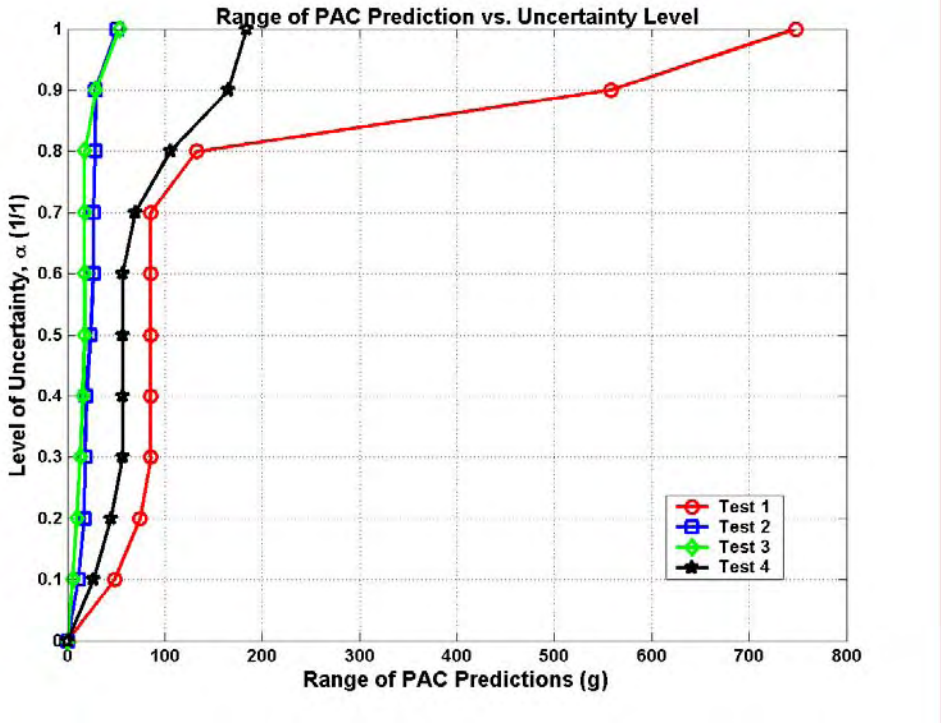
In this Section, the antagonism between robustness-to-uncertainty, α^*, and confidence in prediction, C_F, is illustrated for the crushable foam impact problem. Results shown here apply to the prediction of peak acceleration *PAC* and time-of-arrival *TOA* features.

The main source of lack-of-knowledge is the behavior of the crushable foam material. It is modeled using a family of convex nested sets $\{U_\alpha,\ \alpha>0\}$ that become increasingly inclusive of strain-stress curves as the horizon-of-uncertainty, α, increases. Members of the sets U_α are the sub-models of internal force described in equations (9-11) and needed to solve the equation of motion (8). Section 4.2 describes the models of input uncertainty and Section 4.5 discusses how they are propagated through the simulation.

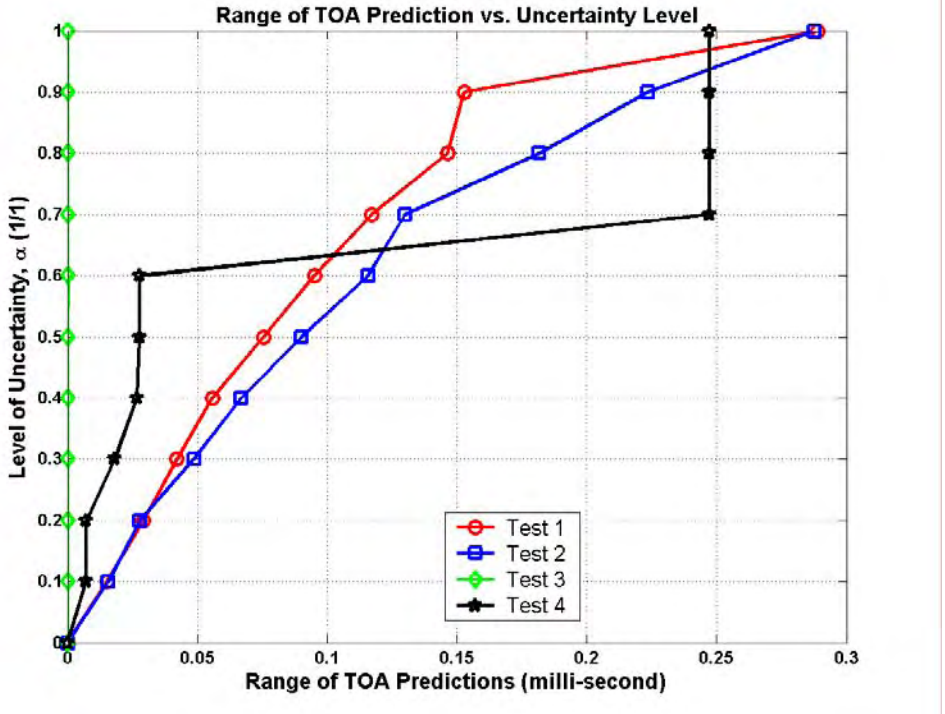
Output uncertainty, that is, uncertainty about the values of *PAC* and *TOA*, is represented by an interval $[y_{min};y_{max}]$ for each feature, as explained in Section 5.3. Intervals are collected in an information matrix and the total uncertainty framework is implemented according to the steps (23-25). Confidence in prediction is quantified using the definition (27).

7.1 Looseness vs. Robustness for the Crushable Foam Modeling

First, the relationship between looseness in prediction, $\lambda_Y = y_{max} - y_{min}$, and robustness-to-uncertainty, α^*, is illustrated. Theorem 2 states that looseness increases with robustness. This is verified in Figure 14 that pictures looseness as a function of robustness. Each curve represents one of the four configurations of the system tested experimentally. In Figure 14-a, the looseness of *PAC* prediction increases with robustness, especially for Tests 1 and 4. A similar trend can be observed for the *TOA* feature in Figure 14-b. This expresses that the models included in a given family U_α, up to the robustness level α^* and aspiration of fidelity-to-data R_{Max}, make increasingly less consistent predictions. Discontinuities are seen, but each curve is monotonically increasing because the domains U_α are nested within one another.



(14-a) PAC prediction looseness versus modeling lack-of-knowledge.



(14-b) TOA prediction looseness versus modeling lack-of-knowledge.

Figure 14. Prediction looseness of the four configurations of the system.

7.2 Confidence vs. Robustness for the Crushable Foam Modeling

Figure 15 translates the information conveyed by Figure 14 into an estimate of confidence in prediction. The ranges of *PAC* and *TOA* predictions, made by all numerical simulations based on sub-models that belong to U_α, are combined in the information matrix $H_{2,2}$ of equation (28). Total uncertainty, *TU*, and confidence, C_F, are then quantified according to equations (25) and (27), respectively. As before, each curve shows results for one of the tested configurations. The antagonistic nature of robustness and confidence can be observed from Figure 15.

It is our opinion that such information is of great value to investigate the strengths and weaknesses of numerical simulations. For example, if one is asked to meet the requirement of C_F = 60% confidence when predicting *PAC* and *TOA* for all four configurations, Figure 15 shows that no more than α = 0.3 uncertainty can be tolerated. If evidence is available to suggest that our lack-of-knowledge is greater than this level of uncertainty, then the confidence requirement cannot be met. Corrective actions can be suggested to reduce the modeling uncertainty, such as performing validation experiments, collecting more data, or re-thinking some of the modeling assumptions.

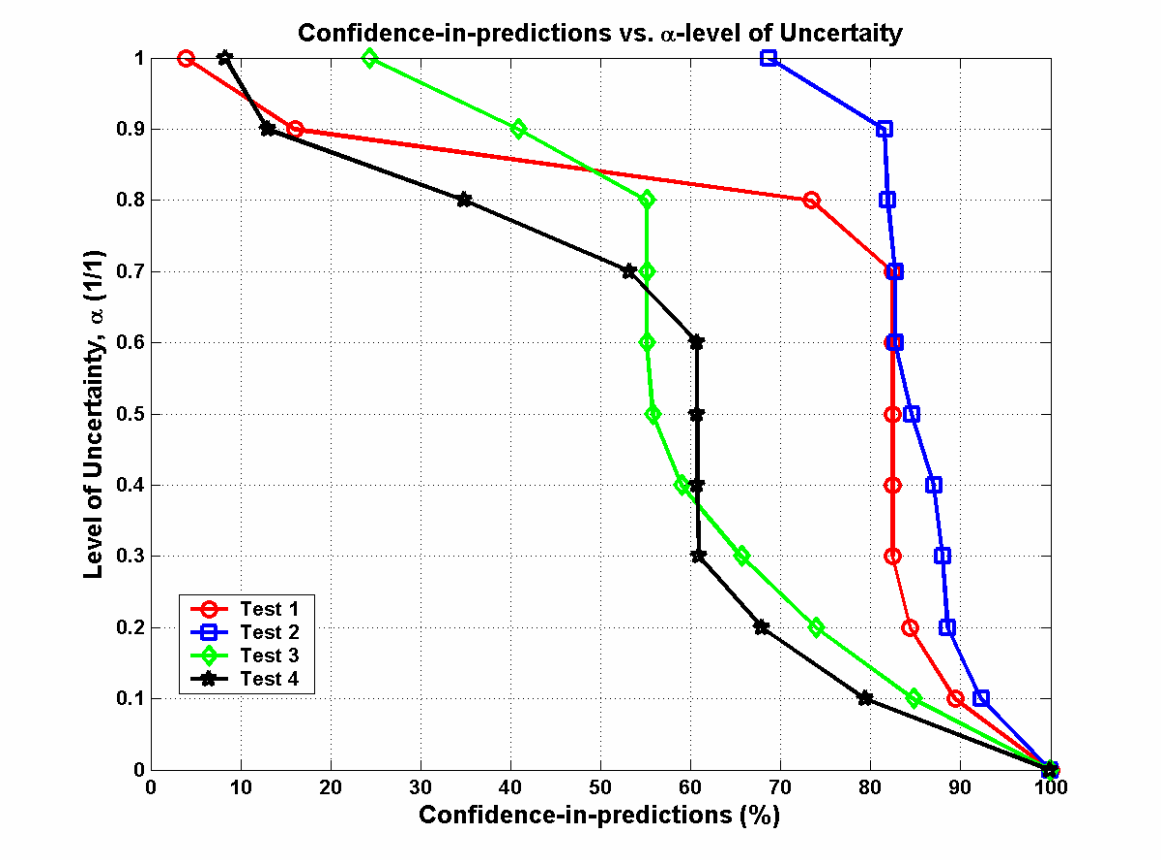
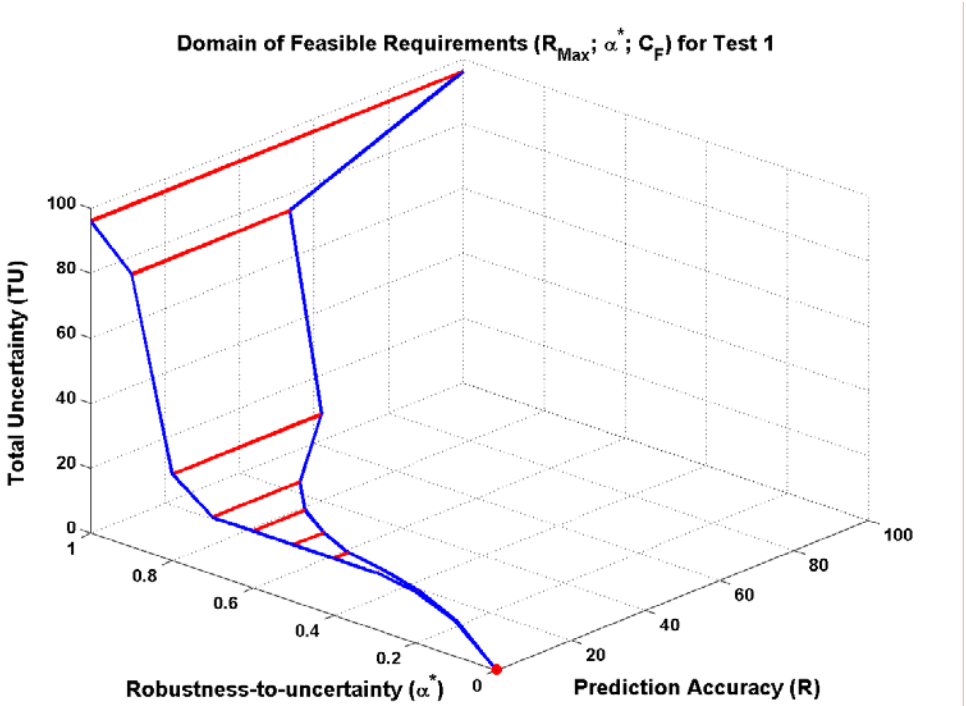
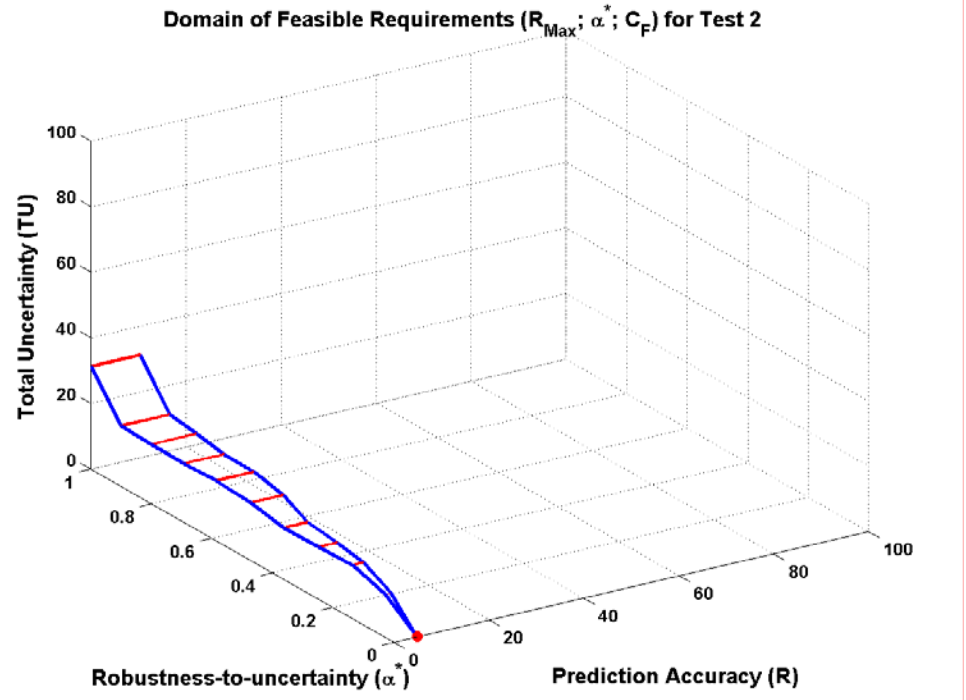


Figure 15. Confidence in prediction, C_F, for the four configurations of the system.

The reader will understand that the curves shown in Figure 15 are projections of the three-dimensional triplet $(R_{Max};\alpha^*;C_F)$ in the plane defined by confidence and robustness. To each point shown on the curves of Figure 15 corresponds a range of prediction accuracy, from best attainable to worst possible. The domains of feasible requirements defined by the triplets of values $(R_{Max};\alpha^*;C_F)$ are examined next.

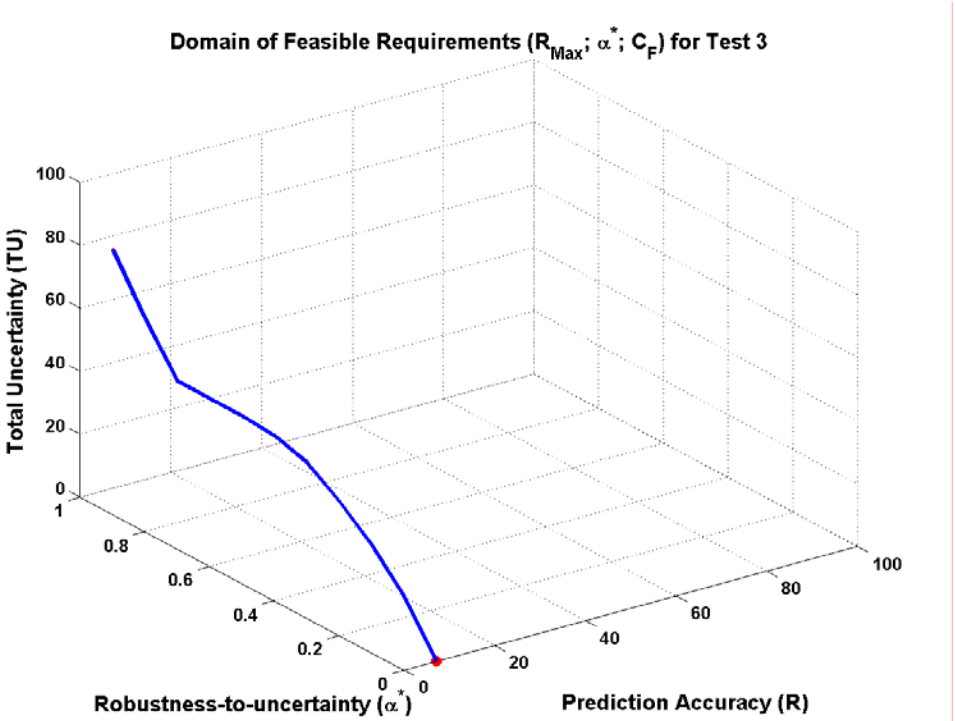


(16-a) Feasible requirements for Test 1, low drop height and thin foam pad.

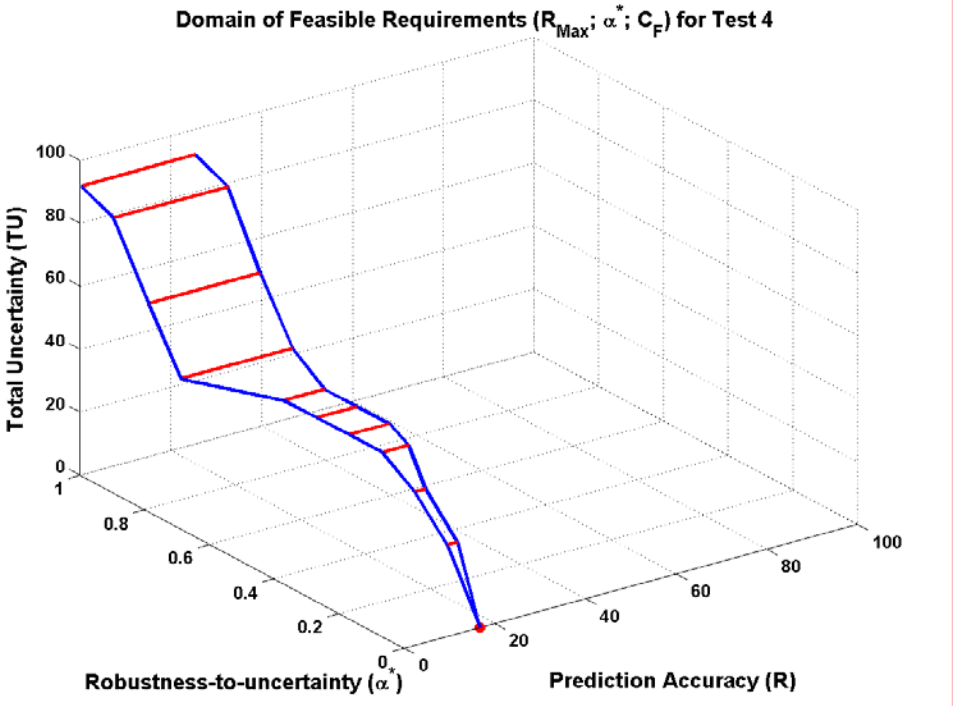


(16-b) Feasible requirements for Test 2, low drop height and thick foam pad.

Figure 16. Feasible requirements (R_{Max};α^*;C_F) for the four configurations of the system.

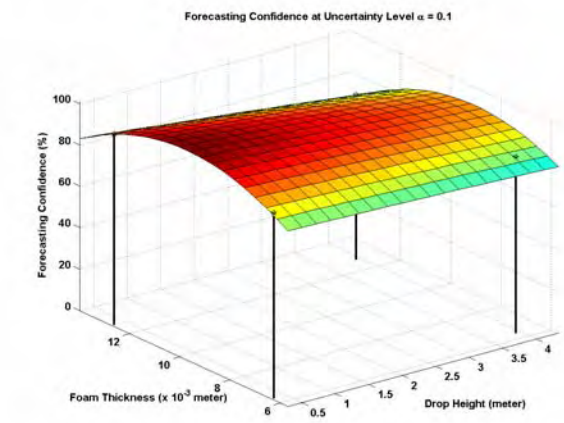


(16-c) Feasible requirements for Test 3, high drop height and thin foam pad.

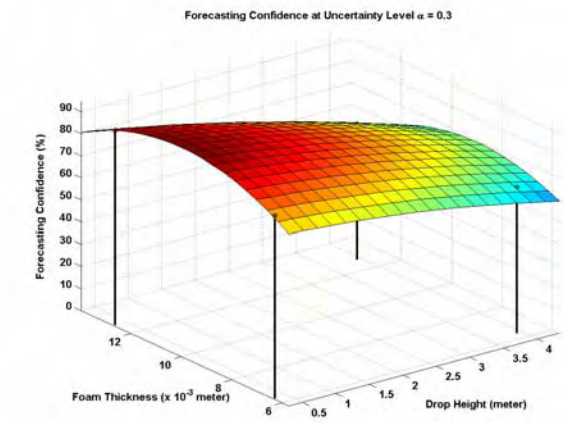


(16-d) Feasible requirements for Test 4, high drop height and thick foam pad.

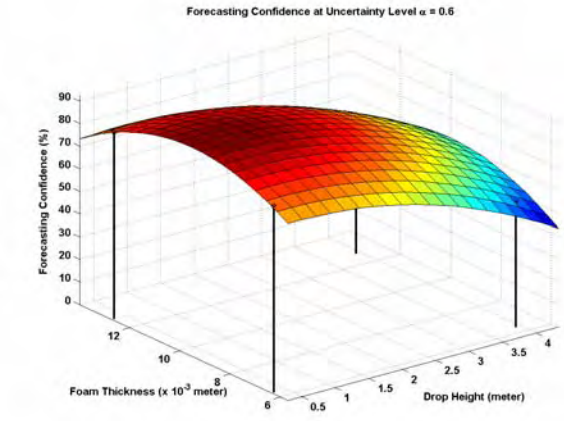
Figure 16. Feasible requirements (R_{Max};α^*;C_F) for the four configurations of the system.



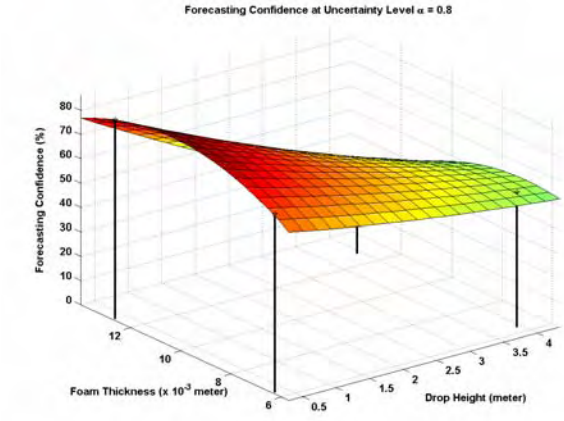
(17-a) Confidence C_F at α = 0.1.*



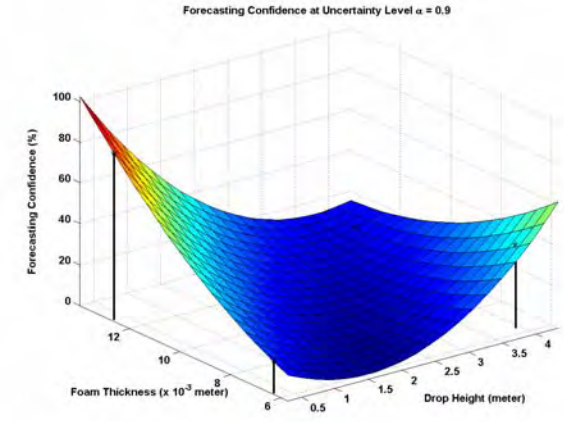
(17-b) Confidence C_F at α = 0.3.*



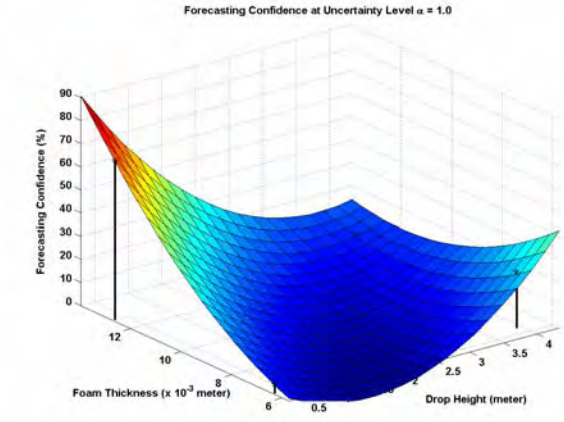
(17-c) Confidence C_F at α = 0.6.*



(17-d) Confidence C_F at α = 0.8.*



(17-e) Confidence C_F at α = 0.9.*



(17-f) Confidence C_F at α = 1.0.*

Figure 17. Confidence-vs.-robustness surfaces extrapolated throughout the validation domain.

7.3 What Constitute Feasible Requirements for the Crushable Foam Modeling?

Figures 16 show the three-dimensional domain defined by all combinations of fidelity, robustness, and confidence $(R_{Max};\alpha^*;C_F)$ requirements. To provide an overall assessment of prediction accuracy, test-analysis correlation R values are defined by the Mahalanobis distance of equation (13). The total uncertainty TU metric, normalized between zero and one, is shown instead of the confidence C_F metric. Scaling of axis values is kept constant to make it easier to compare across the four figures.[19] The ranges of prediction accuracy, from best attainable to worst possible, are illustrated by the vertical lines that connect within each horizontal plane $(\alpha^*;C_F)$ the opportunity curve (or best accuracy) to the robustness curve (or worst accuracy). The discontinuous appearance of these "tornado-looking" plots is an artifact of using only ten discrete horizon-of-uncertainty values, α.

Ideally, we would like the family of models to provide low prediction error (or $R\sim0$), high robustness (or $\alpha^*\sim1$), and high confidence in prediction (or $TU\sim0$), which is the region near the lower left corner of each cube in Figures 16. For Tests 1 and 4, Figures 16-a and 16-d show that the family of models can satisfy quite stringent requirements $(R_{Max};\alpha^*;C_F)$, up to the level of robustness of $\alpha^* = 0.8$, approximately. Predictions made for Test 3, on the other hand, are relatively insensitive to the modeling uncertainty. The only noticeable trend of Figure 16-c is that confidence decreases with robustness, which illustrates the Theorem 2 and equation (41). The family of models appears very appropriate to predict Test 2. This simple example illustrates the methodology proposed to study the "myth" of science-based predictive modeling.

7.4 Extrapolation of Confidence vs. Robustness Throughout the Validation Domain

Finally, Figure 17 extrapolates the confidence-versus-robustness curves throughout the validation domain, that is, to all settings of drop height and foam thickness values that have not been tested experimentally. The procedure is similar to the extrapolation of robustness-versus-fidelity curves of Figures 12 and 13 in Section 4.5. The overall trend, that increasing robustness decreases confidence throughout the validation domain, is in agreement with the theory and sensitivity results shown in equation (41).

8. Thoughts About the Department of Energy's ASC Program

In this Section, the theory developed is discussed in the context of the Department of Energy's Advanced Simulation and Computing (ASC) Program. ASC has been tasked with the development of massively parallel computing platforms and high-fidelity, physics-based codes for weapon applications at Los Alamos and other national laboratories.

The rationale for investing in this technology is that, in the absence of full-scale testing, predictions made by ASC codes will eventually become more credible than predictions assembled by analyzing legacy codes and eliciting expert opinion. The notions of "credibility", "confidence", and "predictive capability" are central to the mission of ASC, as indicated by these quotes extracted from several documents:

> *"The development of three-dimensional, high-fidelity applications for execution on massively parallel computers is required to properly steward the enduring stockpile and maintain a* ***credible*** *deterrent."* From Reference [31].

[19] The fidelity-to-data values shown on the X-axis are expressed as percentages of the maximum prediction error obtained using the Mahalanobis metric for Test 1. The values of the X-axis have therefore no physical meaning. The robustness values shown on the Y-axis, and likewise the total uncertainty values shown on the Z-axis, are scaled between zero and one.

> *"Advanced physics and material models, and the coupling of such models to these applications, are required to create a* ***predictive capability*** *for the modeling of nuclear weapons as our stockpile continues to age."* From Reference [31].

> *"An essential task of the weapons program has always been to determine, with* ***confidence****, the performance of stockpile weapons. Today we face the additional challenge of accomplishing this task without nuclear testing."* From Reference [32].

One important characteristic of legacy codes is that they have been calibrated to match the observations collected from a large number of nuclear and non-nuclear diagnostics and experiments. Theorem 1 suggests that the price to pay for increased fidelity is vulnerability to epistemic uncertainty. Examples of uncertainty include the value of discretization parameters, material properties, nuclear properties, numerical settings that control the solution algorithms, and the degree to which different physics are coupled.

High-fidelity codes, on the other hand, are based on first-principle physics. The basic idea is to better understand the physics, materials, and environments; more accurately describe the mechanical, thermal, and hydro-dynamical states; and couple different physics such as phase transformation, fission, radiation, and thermonuclear burn. This approach aims at substituting first-principle physics to simplifying assumptions, hence, pushing back the boundaries of our ignorance. In doing so, what is gained is robustness. Because ASC codes are general-purpose, as opposed to specific to a system or series of tests, and because they can simulate a wide range of physical behaviors, they are made less vulnerable to what may still be unknown about the fundamental physics. Theorem 1 suggests that the price to pay for increased robustness is a lesser ability to match the available test data. In addition, Theorem 2 and its extension to confidence suggest that a deteriorated ability to predict with confidence should be expected.

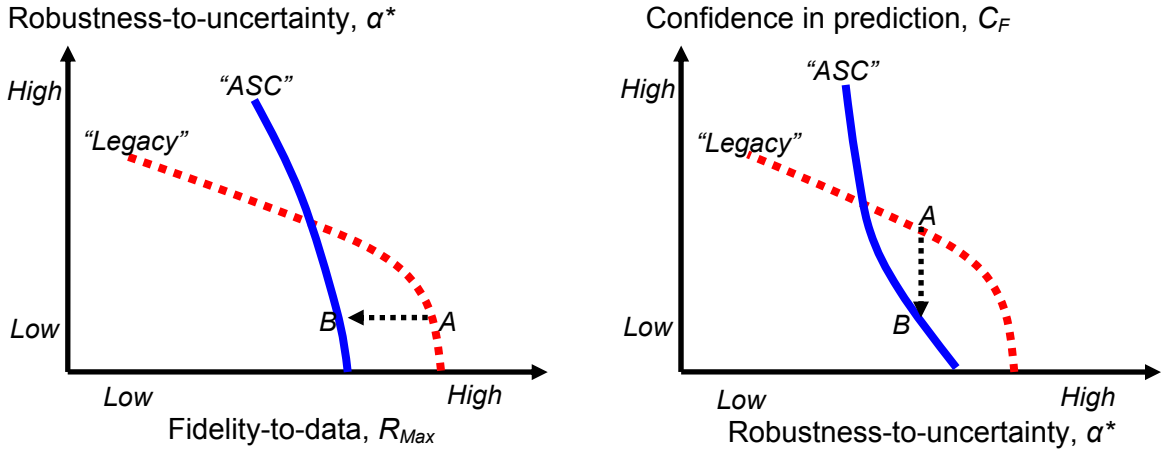


Figure 18. Conceptual illustration of a fidelity-robustness-confidence curves (R_{Max};α;C_F).*

This is conceptually illustrated in Figure 18. The trade-offs between fidelity and robustness are shown on the left; the trade-offs between robustness and confidence are shown on the right. Performance of legacy codes is illustrated by the red dashed lines; performance of ASC codes is illustrated by the blue solid lines. It is emphasized that the figures are notional only. The reader will understand that what should really be compared are the triplets *(R_{Max};α*;C_F)* for each family of codes, projected here in two dimensions for simplicity.

Figure 18 suggests that switching from one family of codes to the next, that is, moving from points *A* to points *B*, tends to initially degrade fidelity and confidence. However, moving to the

points *B* means that trade-offs of fidelity-robustness-confidence requirements have changed. They no longer are dictated by the red dashed curves; the trade-offs move instead along the blue solid curves. After having learned the new tools and perfected the modeling rules, confidence is gained and better matches to test data are obtained. Although the transition may, at first, appear to be a poor decision, the enhanced robustness of the new family of codes should eventually yield better fidelity-to-data and more confidence in prediction.

The message should not be that, after over a decade of investment by Congress and the Department of Energy, the capability to match past experiments and predict with confidence is deteriorating. Indeed, lesser fidelity-to-data and lesser confidence are likely to be observed, at least initially. However, what matters to demonstrate prediction credibility is not the starting point on a fidelity-robustness-confidence curve $(R_{Max};\alpha^*;C_F)$. What ultimately matters is the overall shape of the curve that defines constraints between R_{Max}, α^*, and C_F. Such message may be difficult to communicate to high-level managers, but our opinion is that the role of scientists is to educate them to the sometimes harsh reality of predictive modeling.

9. Conclusion

This work studies the relationship between several aspects of prediction accuracy. The main conclusion is that, in assessing the prediction accuracy of numerical models, one should never focus on a single aspect only. Instead, the trade-offs between fidelity-to-data, robustness-to-uncertainty, and confidence in prediction should be explored. One consequence that cannot be emphasized enough is that the calibration of numerical models—which focuses solely on the fidelity-to-data aspect—is not a sound strategy for selecting models capable of making accurate predictions. Calibration leaves models vulnerable to modeling uncertainty.

It is further established that predictive models selected for their robustness-to-uncertainty tend to make inconsistent predictions, hence, decreasing confidence in prediction. A practical definition of confidence is proposed using the concept of total uncertainty. Total uncertainty is a general framework to aggregate different types of uncertainty and their mathematical modeling. Although preliminary, this definition of confidence is useful to illustrate the antagonism between prediction accuracy, robustness to modeling assumptions, and confidence that one places in the ability of a family of models to forecast environments that have not been tested experimentally.

Disclaimer and Acknowledgements

Opinions expressed in this publication are the author's own. They do not necessarily reflect official policy of the U.S. Department of Energy, National Nuclear Security Administration, Los Alamos National Laboratory, or the Advanced Simulation and Computing (ASC) Program. The support of the ASC Program for engineering analysis verification and validation at Los Alamos is acknowledged. The author is grateful to Mark Anderson, ESA-WR, Jane Booker, ESA-WR, and Scott Doebling, ESA-WR, for their continuing support.

References

[1] Ben-Haim, Y., Hemez, F.M., "Robustness-to-uncertainty, Fidelity-to-data, and Prediction Looseness of Models," *22nd SEM International Modal Analysis Conference*, January 26-29, 2004, Dearborn, Michigan.

[2] Doebling, S.W., "Structural Dynamics Model Validation: Pushing the Envelope," *First International Conference on Structural Dynamics Modeling: Test, Analysis, Correlation and Validation*, Madeira Island, Portugal, June 3-5, 2002.

[3] Hemez, F.M., Ben-Haim, Y., "The Good, the Bad, and the Ugly of Predictive Science," *4th International Conference on Sensitivity Analysis of Model Output*, Santa Fe, New Mexico, March 8-11, 2004.

[4] Ben-Haim, Y., Hemez, F.M., "Robustness, Fidelity, and Prediction-looseness of Models," *7th Biennial Conference on Engineering Systems Design and Analysis*, Manchester, United Kingdom, July 19-22, 2004.

[5] Ben-Haim, Y., Hemez, F.M., "Robustness, Fidelity, and Prediction-looseness of Models," *Physical Transactions of the Royal Society*, April 2004, *in preparation*.

[6] 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.

[7] Hemez, F.M., Doebling, S.W., "Inversion of Structural Dynamics Simulations: State-of-the-art and Orientations of the Research," *25th International Conference on Noise and Vibration Engineering*, Leuven, Belgium, September 13-15, 2000, pp. 403-413.

[8] Ben-Haim, Y., "Robust Rationality and Decisions Under Severe Uncertainty," *Journal of the Franklin Institute*, Vol. 337, 2000, pp. 171-199.

[9] Ben-Haim, Y., **Information-Gap Decision Theory: Decisions Under Severe Uncertainty**, Academic Press, 2001.

[10] Sharp, D.H., Wood-Schultz, M.M., "QMU and Nuclear Weapons Certification: What's Under the Hood?," *Los Alamos Science*, Number 28, June 2003, pp. 47-53. (Also available as report LA-UR-03-5704, Los Alamos National Laboratory, Los Alamos, New Mexico.)

[11] Logan, R.W., Nitta, C.K., "Verification & Validation Methodology and Quantitative Reliability at Confidence (QRC): Basis for an Investment Strategy", *Report UCRL-ID-150874*, Lawrence Livermore National Laboratory, Livermore, California, November 2002.

[12] Nitta, C.K., Logan, R.W., "Qualitative and Quantitative Linkages from V&V to Adequacy, Certification, Risk, and Benefit/Cost Ratio," *Foundations '04 Workshop for Verification, Validation, and Accreditation in the 21st Century*, Arizona State University, Tempe, Arizona, October 13-15, 2004. (Also available as report UCRL-TR-205809, Lawrence Livermore National Laboratory, Livermore, California.)

[13] Elishakoff, I., **Safety Factors and Reliability: Friends or Foes?**, Kluwer Academic Publishers, 2004.

[14] 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.

[15] Hemez, F.M., Ben-Haim, Y., "Info-gap Robustness for the Correlation of Tests and Simulations of a Non-linear Transient," *Mechanical Systems and Signal Processing*, Vol. 18, 2004, pp. 1443-1467.

[16] Kolmogorov, A.N., **Foundations of the Theory of Probability**, Chelsea, New York, 1956.

[17] Dempster, A.P., "Upper and Lower Probabilities Generated by a Random Interval," *Annals of Mathematical Statistics*, Volume 39, Number 3, 1968, pp. 957-966.

[18] Joslyn, C., "Aggregation and Completion of Random Sets with Distributional Fuzzy Measures," *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, Volume 4, Number 4, 1996, pp. 307-329.

[19] Anderson, M.C., Booker, J., Hemez, F.M., Joslyn, C., Reardon, B., Ross, T., "Quantifying Total Uncertainty in a Validation Assessment Using Different Mathematical Theories," *14th Biennial Nuclear Explosives Design Physics Conference*, Los Alamos National Laboratory, Los Alamos, New Mexico, October 20-24, 2003. (Also available as report LA-UR-03-9001, Los Alamos National Laboratory, Los Alamos, New Mexico.)

[20] Booker, J.M., Ross, T.J., Rutherford, A.C., Reardon, B.J., Hemez, F.M., Anderson, M.C., Doebling, S.W., Joslyn, C.A., "An Engineering Perspective on Uncertainty Quantification for Validation, Reliability, and Certification," *Foundations '04 Workshop for Verification, Validation, and Accreditation in the 21st Century*, Arizona State University, Tempe, Arizona, October 13-15, 2004.

[21] Klir, G., Wierman, M., **Uncertainty-based Information: Elements of Generalized Information Theory**, Physica-Verlag/Springer-Verlag, 1999.

[22] Zadeh, L.A., "Towards a Unified Theory of Uncertainty," *International Conference on Information Processing and Management of Uncertainty*, Perugia, Italy, July 4-7, 2004.

[23] Klema, V.C., Laub, A.J., "The Singular Value Decomposition: Its Computation and Some Applications," *IEEE Transactions on Automatic Control*, Volume AC-25, 1980, pp. 164-176.

[24] Lenaerts, V., Kerschen, G., Golinval, J.-C., "Proper Orthogonal Decomposition for Model Updating of Non-linear Mechanical Systems," *Mechanical Systems and Signal Processing*, Volume 15, 2001, pp. 31-43.

[25] Klir, G.J., Yuan, B., **Fuzzy Sets and Fuzzy Logic: Theory and Applications**, Prentice Hall, 1995.

[26] Walley, P, **Statistical Reasoning With Imprecise Probabilities**, Chapman and Hall, 1990.

[27] Doebling, S.W., "Overview of Los Alamos National Laboratory Simulation Validation Program," *ASC Alliance/Laboratory Verification & Validation Workshop*, La Jolla, California, July 13-14, 2004.

[28] Gove, P.B., Editor, **Webster's Third New International Dictionary**, Merriam-Webster Inc., 1986.

[29] Trucano, T.G., Piltch, M., Oberkampf, W.L., "General Concepts for Experimental Validation of ASCI Code Applications," *Report SAND-2002-0341*, Sandia National Laboratories, Albuquerque, New Mexico, 2002.

[30] Schwer, L., "Guide for the Verification and Validation of Computational Solid Mechanics," *Report 2004-03-22*, ASME PTC-60 Committee on Verification and Validation in Computational Solid Mechanics, American Society of Mechanical Engineers, August 2004.

[31] **ASCI Technology Prospectus, Simulation and Computational Science**, *Publication DOE/DP/ASC-ATP-001*, Defense Programs, National Nuclear Security Administration, Department of Energy, Washington, D.C., July 2001.

[32] Benjamin, R.F., "Validation Experiments in Support of the Nuclear Weapons Stockpile," *Nuclear Weapons Journal*, Winter 2004, pp. 16-21. (Also available as report LALP-04-013, Los Alamos National Laboratory, Los Alamos, New Mexico.)