# The Good, The Bad, and The Ugly of Predictive Science

*François M. Hemez[1]*

Engineering Sciences and Applications
Los Alamos National Laboratory
Los Alamos, New Mexico 87545, U.S.A.

*Yakov Ben-Haim[2]*

Faculty of Mechanical Engineering
Technion — Israel Institute of Technology
Haifa 32000, Israel

**Abstract:** 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 the assessment of prediction credibility. Credibility, which is demonstrated through the activities of Verification and Validation, quantifies the extent to which simulation results can be analyzed with confidence to represent the phenomenon of interest with accuracy consistent with the intended use of the model. This paper argues that assessing the credibility of a mathematical or numerical model must combine three components: 1) Improving the fidelity to test data; 2) Studying the robustness of prediction-based decisions to variability, uncertainty, and lack-of-knowledge; and 3) Establishing the expected prediction accuracy of the models in situations where test measurements are not available. A recently published Theorem that demonstrates the irrevocable trade-offs between *"The Good, The Bad, and The Ugly,"* or robustness-to-uncertainty, fidelity-to-data, and confidence-in-prediction, is summarized. The main implication is that high-fidelity models cannot, at the same time, be made robust to uncertainty and lack-of-knowledge. Similarly, equally robust models do not provide consistent predictions, hence reducing confidence-in-prediction. The conclusion of the theoretical investigation is that, in assessing the predictive accuracy of numerical models, one should never focus on a single aspect. Instead, the trade-offs between fidelity-to-data, robustness-to-uncertainty, and confidence-in-prediction should be explored. The discussion is illustrated with an engineering application that consists in modeling and predicting the propagation of an impact through a layer of hyper-foam material. A novel definition of sensitivity coefficients is suggested from the slopes of robustness-to-uncertainty curves. Such definition makes it possible to define the sensitivity of a performance metric to arbitrary uncertainty, whether it is represented with probability laws or any other information theory. *This publication has been approved for unlimited, public release on November 18, 2003 (LA-UR-03-8492, Unclassified).*

**Keywords:** Confidence, decision-making, fidelity-to-data, prediction, 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 Verification and Validation (V&V), quantifies the extent to which simulation results can be analyzed with **confidence** to represent the phenomenon of interest with a degree of accuracy consistent with the intended use of the model [1].

The paper argues that assessing the credibility of a mathematical or numerical model must combine three components: 1) Improve the fidelity, $R$, to test data; 2) Study the robustness, $a^*$, of predictions to uncertainty and lack-of-knowledge; and 3) Establish the "prediction looseness," $\lambda_Y$, of the model. Prediction looseness here refers to the range of predictions expected from a model or family of models. Its

---

[1] Technical Staff Member, ESA-WR; E-mail: hemez@lanl.gov.

[2] Professor, Yitzhak Moda'i Chair in Technology and Economics; E-mail: yakov@technion.il.

importance stems from the fact that, to predict with confidence, there should be little difference (or small looseness $\lambda_Y$) between the predictions of equally credible models.

The discussion presented in this manuscript is kept at a conceptual level for the most part. Section 2 defines the notations used in the paper. Section 3 briefly discusses the concepts of fidelity-to-data and robustness-to-uncertainty. A Theorem, that unveils the antagonistic relationships between fidelity-to-data, robustness-to-uncertainty, and confidence-in-prediction, is summarized in Section 4. Finally, Section 5 suggests a novel definition of sensitivity coefficients, obtained from the slopes of robustness-to-uncertainty curves. Such definition makes it possible to define the sensitivity of a performance metric to arbitrary uncertainty, whether it is represented with probability laws or any other information theory.

## 2. BACKGROUND: FIDELITY, ROBUSTNESS, AND CONFIDENCE

Even though the conventional activities of model V&V are generally restricted to improving the fidelity-to-data through the correlation of test and simulation results, and the calibration of model parameters [2-3], the other two components are equally important. The main 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 [4-5]. This means that small variations in the setting of model parameters, or small errors in the knowledge of the functional form of the models, can lead to an actual fidelity that is significantly poorer than the one demonstrated through calibration.

Clearly, fidelity-to-data matters because no analyst will trust a numerical simulation that does not reproduce the measurements of past experiments or the information contained in historical databases. Robustness-to-uncertainty is equally critical to minimize the vulnerability of decisions to uncertainty and lack-of-knowledge. It may be argued, however, that the most important aspect of credibility is the assessment of confidence-in-prediction, which is generally not addressed in the literature. Assessing the confidence in prediction here refers to an assessment of prediction error 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 prediction.

The concepts of fidelity-to-data, robustness-to-uncertainty, and prediction



*Figure 1. Illustration of the concepts of fidelity-to-data, robustness-to-uncertainty, and prediction accuracy (range, "looseness," or confidence).*

confidence are illustrated in Figure 1. It is emphasized that, because this is work-in-progress to a great extent, the concept of prediction accuracy denoted in Figure 1 by the symbol $\lambda_Y$ is somewhat broad. It is analogous to a range of predictions, or "looseness." Clearly, predicting a range of values relates to the notion of confidence-in-predictions. It is believed that future research will narrow down this definition, but a standard accepted throughout the scientific community is not, to the best of the author's knowledge, currently available.
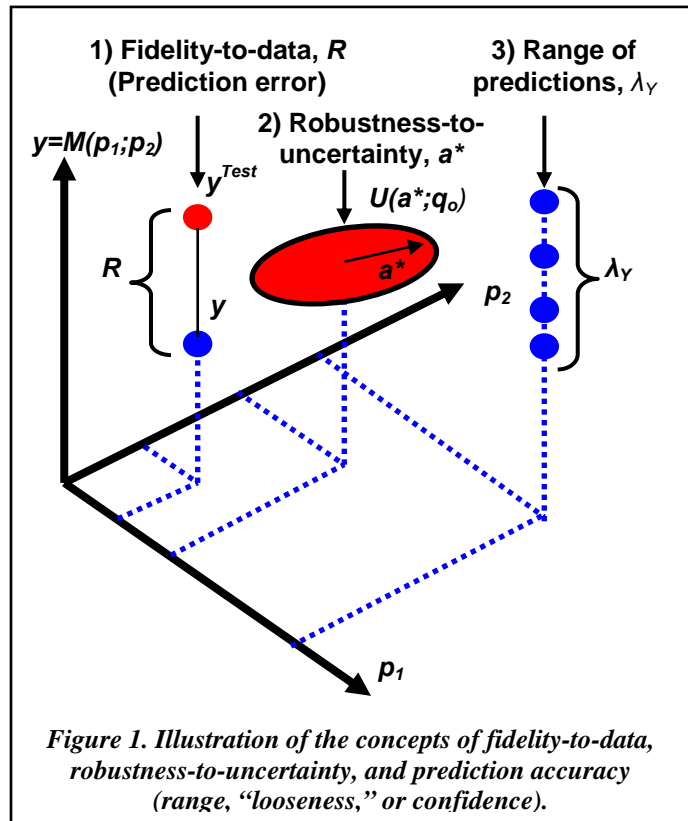
## 2.1 Definitions and Notations

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

- The quantity *y* represents the observable outputs. They can be scalar quantities—which is the case assumed here for simplicity—or vector quantities. These model outputs are usually features extracted from a large-order or large-dimensional response.

- The quantity *p* denotes control parameters of the numerical simulation. These inputs include the control parameters that characterize the experimental configuration. Generally, there will be more than a single input parameter. Inputs *p* represent settings such as, for example, the angle of attack and flow velocity of an aero-elastic simulation that predicts a coefficient of lift $y=C_L$.

- The quantity *q* represents parameters that specify the structure and coefficients of the family of models developed to represent the physical phenomenon of interest. The inputs *q* can include discrete and continuous parameters that control the functional form of the model. Various models can be functions of different subsets of parameters *q*.

In the general case, the model is represented as:

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

where $q_o$ denotes nominal settings for the parameters *q*. In the following, the subscript *( )$_o$* represents the nominal condition of a quantity.

A domain denoted by $D_P$ represents the design space over which predictions must be obtained. This implies that the prediction accuracy must be established for all settings *p* in the design domain $D_P$. In the case, for example, of a two-dimensional operational space where *p = (p$_1$;p$_2$)*, like the one pictured in Figure 1, the prediction accuracy of the model *y = M(p;q$_o$)* must be studied for all combinations *(p$_1$;p$_2$)* that belongs to $D_P = [p_1^{(min)};p_1^{(max)}] \times [p_2^{(min)};p_2^{(max)}]$.

The quantity *M(p;q)* is used to denote alternative possible physical models. This notation is introduced to recognize that some of the model parameters *q* may be subjected to parametric variability. Others may be uncertain, or represent an epistemic lack-of-knowledge about the functional form of the model. For example, the behavior of a particular material under a fast transient load may not be known with certainty. Having to choose between, say, a linear elastic model, a model of perfect plasticity, or a visco-elastic model with hardening represents an epistemic uncertainty denoted by *q*. In the absence of epistemic uncertainty, no alternative to *M(p;q$_o$)* would be feasible. As the horizon of modeling uncertainty increases, more and more alternative models become candidates. The family of predictive models can therefore be represented in a generic sense by the equation:

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

where the symbol *a* and norm || || are left undefined for now. It suffices to say that *a* is a positive scalar quantity that represents the ***horizon-of-uncertainty***. The meaning of definition (2) is that the family of models *U(a;q$_o$)* becomes increasingly inclusive as the parameters *q* are allowed to differ from their nominal settings *q$_o$*. Note that these definitions are purposely broad to encompass a wide range of models and uncertainties.

Measurements are denoted by the symbol $y^{Test}$. Measurements are made at specific experimental configurations controlled by the parameters *p*. The notation used throughout this paper is that replicate measurements made to estimate the environmental variability are collected in the same vector or matrix quantity $y^{Test}$. Measurements made, on the other hand, for different configurations $p_1 \ldots p_m$ will be indexed as $y^{Test\,(1)} \ldots y^{Test\,(m)}$.

## 2.2 Fidelity, Robustness, and Confidence

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 setting $(p;q_o)$:

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

Fidelity-to-data is pictured in Figure 1 as the vertical distance between a measurement $y^{Test}$ and a prediction $y$ for the physical experiment and numerical simulation performed at $(p_1;p_2)$.

Robustness-to-uncertainty refers to the range of settings $q$ that provide no more than a given level of prediction error $R_{Max}$. The concept of robustness is illustrated in Figure 1 by showing a subset $U(a^*;q_o)$ of the design domain $[p_1^{(min)};p_1^{(max)}]$ x $[p_2^{(min)};p_2^{(max)}]$. The significance of the concept of robustness-to-uncertainty is that all predictions made for settings $q$ chosen inside the domain $U(a^*;q_o)$ are guaranteed not to exceed the error level $R_{Max}$. The $a$-parameter represents the "size" of the domain $U(a;q_o)$. The definitions of the sizing parameter and corresponding domain are arbitrary at this point because the purpose of this discussion is to introduce concepts. The only constraint to satisfy is that increasing values of the sizing parameter $a$ must define nested domains $U(a;q_o)$. Reference [5] defines the families of domains as convex sub-spaces. This choice allows the analyst to accommodate a wide variety of uncertainty and lack-of-knowledge models.[3] Clearly, a large robustness-to-uncertainty $(a^*)$ is more desirable than a small one $(a')$ because the former subspace will encompass all events defined in the latter one, or $U(a';q_o) \subset U(a^*;q_o)$. A large robustness indicates that potentially large uncertainty does not deteriorate the prediction error by more than $R_{Max}$. Generally, a trade-off must be decided upon between the robustness-to-uncertainty $(a^*)$ and prediction error $(R_{Max})$, or fidelity-to-data. Studying such trade-off is the basic concept of the information-gap theory for decision-making under severe uncertainty [4-5].

Finally, the symbol $\lambda_Y$ in Figure 1 refers to the range of predictions made by a family of potentially different models. The importance of $\lambda_Y$ stems from the fact that, to have confidence in predictions, there should be as much consistency as possible between the predictions provided by equally credible sources of information. Confidence is generally increased when different sources of evidence all reach the same conclusion. The concept of confidence-in-prediction is illustrated in Figure 1 by showing a range $\lambda_Y$ of predictions obtained when different models are exercised to make predictions at a setting $(p_1;p_2)$ where no test data are available. The ultimate goal of model V&V is to establish predictive confidence by estimating the range of predictions $\lambda_Y$ (or, equivalently, the lack-of-consistency) provided by equally credible sources of information.

Note that the terminology "model" is here defined in a broad sense. In any realistic application, sources of evidence include expert judgment, back-of-the-envelope calculations, measurements, and predictions obtained from phenomenological models or high-fidelity simulations. These can all be considered as models because they define a relationship between the inputs $p$ and the outputs $y$. All available sources of information must be taken into account to assess the credibility of numerical simulations. It is equally important to understand, quantify, and eventually combine the uncertainty associated with each source of information. The integration of uncertainty and definition of total uncertainty metrics are areas of active research at Los Alamos National Laboratory, although not addressed in this publication [6].

---

[3] A first example is a probabilistic model of variability where the values of coefficients in the covariance matrix are controlled by the parameter $a$. A second example is a possibility structure defined to represent a lack-of-knowledge, where the size of intervals is proportional to the parameter $a$. 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 uncertainty parameter $a$.

# 3. THE TRADE-OFF OF FIDELITY AND ROBUSTNESS

A family of models $U(a;q_o)$ such as equation (2) 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 [4-5]. This is a significant departure from other representations of uncertainty, such as the probability theory, that attempt to model the randomness itself. Doing so requires strong assumptions that might not be justifiable in the case of lack-of-knowledge.[4]

## 3.1 Fidelity-optimal Strategy for Model Selection

Let $M(p;q)$ be any physical model in the information-gap domain $U(a;q_o)$. The fidelity-to-data, $R$, of this particular model can be estimated by calculating a test-analysis correlation metric (3). Clearly, many models can potentially be included in the family $U(a;q_o)$, some with higher fidelity-to-data (or smaller $R$ values) than others. For a given horizon-of-uncertainty $a$, an analyst might have the choice between several models to make predictions, and a natural strategy would be to use the model that exhibits the best fidelity to the existing data, $R^*$. In mathematical terms, this approach consists of solving an optimization problem defined as:

$$R^* = \min_{M \in U(a;\, q_o)} R \qquad (4)$$

Solving the optimization (4) provides the set of optimal parameters $q^*$ (or optimal model) for which the predictions $M(p;q^*)$ lead to the best fidelity-to-data $R^*$. This is precisely the problem that parameter calibration solves.

The inappropriateness of this strategy for choosing a model or making decisions comes from the fact that the horizon-of-uncertainty, $a$, is generally unknown. An example in mechanical engineering is the definition of a friction coefficient between two materials. A value of the friction coefficient might be available from the literature, but the extent of the variability is typically unknown. What it 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. These un-doubtfully involve stick-and-slip and complex micro-mechanics that the Coulomb friction can only approximate. The extent to which this model is in error compared to the "true-but-unknown" behavior is generally unknown. Therefore, any uncertainty model that aims at representing the lack-of-knowledge associated with the friction model would have to be associated to an unknown horizon-of-uncertainty.

## 3.2 Robustness-optimal Strategy for Model Selection

The main point of this discussion is that a natural trade-off arises between fidelity-to-data and robustness-to-uncertainty. Instead of fixing the horizon-of-uncertainty—which is practiced all the time, for example, in probabilistic analysis when standard deviation or total entropy are initialized and kept constant—and optimizing the fidelity-to-data, the robustness-to-uncertainty can be maximized for a given ***aspiration*** of fidelity-to-data. The fidelity aspiration is denoted as $R_{Max}$, and it represents a value of prediction error not to be exceeded. This means that a model is rejected if its fidelity-to-data is poorer than the aspiration, or $R > R_{Max}$. It could also happen that models are found that outperform our original fidelity aspiration, which would indeed be good news. In mathematical terms, this approach consists of solving the following embedded optimization problems:

---

[4] 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 derive a probability density function with confidence. In extreme cases, only ranges of values can be obtained. Similar difficulties are encountered with physics-based models, historical databases, and expert opinion. Defining a specific model of uncertainty—such as probabilities, possibilities, or a fuzzy structure—might require assumptions that the available evidence simply does not support.

$$a* = \max_{a \geq 0} \left\{ \min_{M \in U(a; q_o)} R \leq R_{Max} \right\} \tag{5}$$

where $a*$ denotes the robustness-to-uncertainty, or largest amount of uncertainty that can be tolerated in our knowledge of the model and its parameters, while guaranteeing a fidelity-to-data at least equal to $R_{Max}$. As pointed out earlier, it could happen that the robust-optimal model features a better fidelity-to-data, or $R<R_{Max}$, a situation referred to in Reference [5] as opportunity from uncertainty.

## 3.3 Trade-off Between Fidelity and Robustness

Just like the fidelity-optimal strategy for model selection defines an ordering preference where the model $M(p;q*)$ is preferred to the model $M(p;q)$ if $R*<R$, the robustness-optimal strategy defines an ordering preference where the model $M(p;q*)$ is preferred to the model $M(p;q)$ if it is more robust to the uncertainty, that is, $a*>a$. As mentioned previously, a large robustness is more desirable than a small robustness because it indicates that potentially large sources of uncertainty do not deteriorate the prediction error by more than $R_{Max}$.

We are not advocating that fidelity-optimality, as a decision strategy for building and validating models, be systematically replaced with robustness-optimality. Instead, investigating the trade-off between the aspiration of fidelity-to-data $R_{Max}$ and robustness-to-uncertainty $a*$ should be the basis for building and validating models. One significant advantage gained in doing so is that information-gap models can encompass a wide range of uncertainty: probabilistic or non-probabilistic, from parametric uncertainty to linguistic ambiguity and modeling lack-of-knowledge, etc. One practical limitation is the amount of calculation involved in the saddle-point optimization problem (5).

# 4. THE TRADE-OFFS OF FIDELITY, ROBUSTNESS, AND LOOSENESS

In this section we explore the "looseness" of model prediction: the range of predicted values deriving from models which all satisfy a specified fidelity requirement. The notion of prediction looseness (or range of predictions as it is also referred to below) is important because it relates to the confidence that one has in the predictions of equally credible models. We prove a Theorem whose meaning is that a change in the model that enhances fidelity-robustness to modeling error also increases the looseness of the model prediction. In other words, fidelity-robustness and prediction-looseness are antagonistic attributes of any modeling effort.

## 4.1 More Definitions

Let $a*$ be the robustness-to-uncertainty of model $M(p_k;q)$ at the experimental configuration $p_k$, as defined in equation (5). Let $U*=U(a*;q_o)$ denote the set of models whose fidelities are no worse than the aspiration $R_{Max}$, for the $k^{th}$ experiment defined by parameters $p_k$. Note that both $a*$ an $U*$ depend upon the model specification, $q$. We have no reason to reject any model $M(p_k;q)$ in $U*$ if fidelity-to-data is used as the measure of merit. This is because all models $M(p_k;q)$ included in the family $U*$ satisfy, by definition, the aspiration of fidelity-to-data, $R<R_{Max}$. The "best" model is therefore non-unique, which is a well-established result in inverse problem mathematics. As discussed previously, some of these models may be more robust to the uncertainty than others, up to the upper limit $a=a*$. An alternative model selection strategy is to identify models associated with the largest robustness $a$.[5]

If $a*$ is large, then $U*$ contains a wide range of models. The predictive looseness of the family of models, $M(p_k;q)$, that belong to $U*$ is simply defined as the range of predictions in $U*$:

---

[5] Even with robustness-optimal model selection, uniqueness of the solution is not guaranteed. Because the family of models $U(a;q_o)$ is constructed as a family of increasingly-including—or nested—convex subsets, models that would possess the same fidelity-to-data, $R$, **and** the same robustness-to-uncertainty, $a$, can still occur on the convex hull of the domain $U(a;q_o)$.

$$\lambda_Y = \max_{M \in U(a^*;q_o)} M(p_k;q) - \min_{M \in U(a^*;q_o)} M(p_k;q) \tag{6}$$

The definition of a range (6) may not be the most appropriate to represent the concept of confidence-in-prediction provided by a family of models $U^*$, but it is clearly related. This choice is motivated by the fact that no standard is currently available in the scientific community to express confidence. It also offers simplicity, both conceptually and computationally.

Large robustness, $a^*$, and small range of predictions, $\lambda_Y$, are both desirable. We will say that robustness and range are **sympathetic** if a change in input variables or model form parameters $q$ improves them both; otherwise they are **antagonistic**:

$$\text{Sympathetic}: \left(\frac{\partial a^*}{\partial q}\right)\left(\frac{\partial \lambda_Y}{\partial q}\right) \leq 0, \quad \text{Antagonistic}: \left(\frac{\partial a^*}{\partial q}\right)\left(\frac{\partial \lambda_Y}{\partial q}\right) \geq 0 \tag{7}$$

The Theorem enounced in Section 4.2 shows that, under fairly weak conditions, robustness and range are always antagonistic. The axioms of nesting and translation are needed to support the main result. 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, $a$, only differ in their center points. The axioms are not further discussed because they can easily be realized with arbitrary models of uncertainty.

## 4.2 Theoretical Result of the Robustness-range Trade-off

The main Theorem is given as follows. Let $U(a;q_o)$ be an information-gap family of models that obeys the axioms of nesting and translation, and let $a=a(p_k;q_o;R_{Max})$–versus–$R_{Max}$ be its robustness function. Consider two initial models, $q_o^*$ and $q_o'$. If $a(p_k;q_o^*;R_{Max}) \geq a(p_k;q_o';R_{Max})$, then $\lambda_Y(q_o^*) \geq \lambda_Y(q_o')$. That is, robustness and range (or prediction looseness) are antagonistic.

A proof of the Theorem can be found in Reference [7]. The proof relies on the information-gap description of ignorance. No restrictive assumption is made regarding the type of models, sources or types of uncertainty, and their mathematical representations. This makes the theory applicable to a wide range of situations.

## 4.3 Exploring the Trade-offs of Predictive Accuracy

Three quantities are central to the information-gap analysis of modeling and forecasting: fidelity of the model to the data, $R$; robustness-to-uncertainty, $a^*$; and the range of predictions, $\lambda_Y$, from models of comparable fidelity. Combining the results of References [5] and [7] yields several trade-offs briefly discussed in the remainder.

First, the relationship between robustness ($a$) and looseness ($\lambda_Y$) can be written as:

$$\frac{\partial \lambda_Y}{\partial a} \geq 0 \tag{8}$$

which means that a revision of the model, with the purpose of enhancing robustness to modeling error, also increases the looseness of predictions. In other words, robustness and prediction looseness are antagonistic attributes of any model. Extending this result to the following three inequalities is trivial:

$$\frac{\partial a}{\partial R} \geq 0, \quad \frac{\partial \lambda_Y}{\partial a} \geq 0, \quad \frac{\partial \lambda_Y}{\partial R} \geq 0 \tag{9}$$

The inequalities (9) express the trade-offs between fidelity, robustness, and looseness:

- **Robustness decreases as fidelity improves.** The robustness-to-uncertainty gets larger if the prediction error increases. Numerical simulations made to better reproduce the available test data become more vulnerable to 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.*** The prediction looseness gets larger if the robustness-to-uncertainty increases. Numerical simulations that are more immune to uncertainty and modeling errors provide a wider, hence less consistent, range of predictions.

- ***Looseness decreases as fidelity improves.*** The range of predictions gets larger if the prediction error gets larger. Numerical simulations made to better reproduce the available test data provide more consistent predictions. Although intuitive, this result is not necessarily a good thing when the models are employed to analyze configurations of the system that are very different from those tested.

These trade-offs imply that it is not possible to have, simultaneously, high fidelity, large robustness, and small prediction looseness. High fidelity (small $R$) implies that the model is true to the measurements, which adds warrant to the model. Large robustness (large $a$) strengthens belief in the validity of the model or family of models. Small looseness (small $\lambda_Y$) implies that all the models that are equivalent in terms of fidelity, also agree in their predictions of the system behavior. The conflict between robustness, fidelity and prediction looseness is reminiscent of Hume's critique of empirical induction. Our analysis shows that past measurements, accompanied by incomplete understanding of the measured process, cannot unequivocally establish true predictions of the behavior of the system.

# 5. AN APPLICATION TO PREDICTABILITY IN ENGINEERING

The theoretical results discussed in Sections 3 and 4 are illustrated with an engineering application. The numerical simulation of the propagation of an impact wave through a layer of non-linear, crushable foam material is compared to physical measurements [8]. Details about the experimental set-up, finite element modeling, and sources of uncertainty can be obtained from Reference [9]. In the following, the main source of uncertainty analyzed in this work is briefly described, and analysis results are discussed to illustrate the trade-offs between fidelity-to-data, robustness-to-uncertainty, and prediction looseness. Final comments are made about the concept of sensitivity in the case where calculating partial derivatives or variance coefficients is not mathematically feasible.

## 5.1 Sources of Uncertainty

The main source of uncertainty considered in this analysis arises from a lack-of-knowledge about the material behavior of the crushable foam. It is illustrated in Figure 2 that shows data obtained from four physical experiments. The center curve (second curve from the left, in blue) represents the nominal model. The other three curves are also acceptable realizations of a material behavior. This uncertainty matters greatly because selecting a constitutive law that describes how the material behaves is a critical step of building the numerical simulation.

Because Figure 2 represents a rather severe lack-of-knowledge about the material, the predictions made from numerical simulations must account for this uncertainty. A pre-requisite is to obtain a mathematical representation of this uncertainty. Clearly, deriving a probability law based solely on the evidence captured by Figure 2 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 2 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.
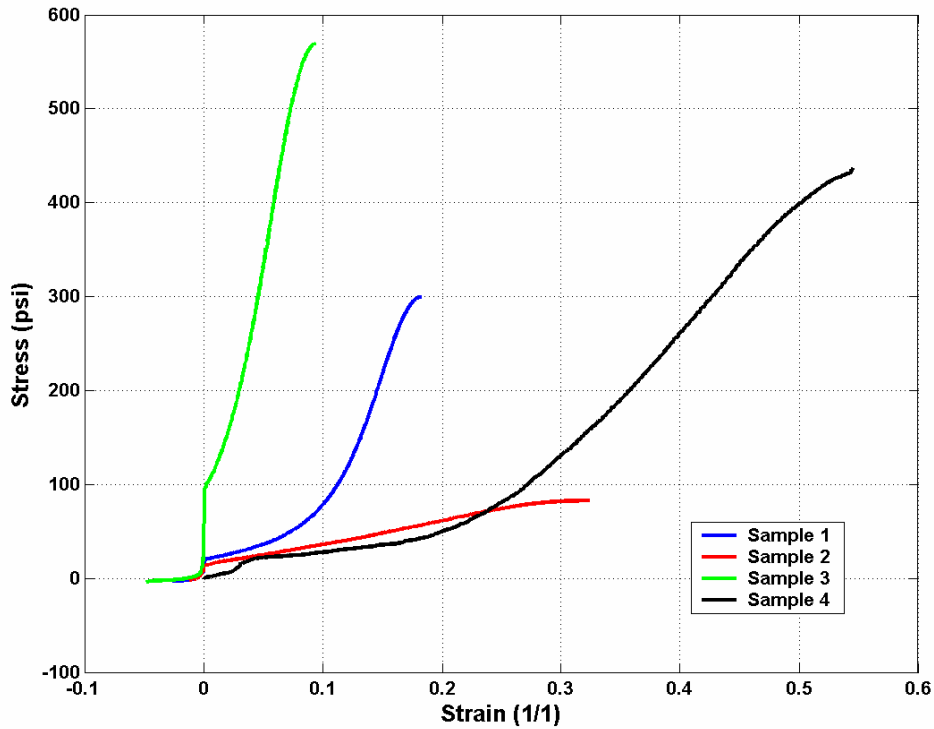
*Figure 2. Four test samples of material behavior for the crushable foam.*

In this study, the uncertainty pictured in Figure 2 is represented mathematically by a family of convex sets $U(a;q_o)$. The nominal material model $q_o$ represents the baseline curve shown in blue color (second from the left). The horizon-of-uncertainty $a$ measures the distance between any curve $q$ and the nominal curve $q_o$ using the Root Mean Square (RMS) metric. The convex domain $U(a;q_o)$ defined in equation (2) therefore becomes the set of predictions made by all possible numerical simulations $M(p;q)$ for which the material model $q$ does not "deviate" from the nominal material model $q_o$ by more than a RMS distance equal to $a$.

It is emphasized that such description of lack-of-knowledge can accommodate a wide variety of practical situations. Here, the symbol $q$ represents variations of material coefficients for linear, bi-linear, or cubic material models. These variations can be made as arbitrary as desired, to the extent where the material models, $q$, do not violate the basic laws of physics. This procedure lets the analyst use his/her prior knowledge to restrict the variations of $q$ around $q_o$ to the ones that make sense. It equally is critical to realize that, in our analysis, the extent of the horizon-of-uncertainty, $a$, is unknown. The family of models $U(a;q_o)$ becomes increasingly inclusive of material models as the uncertainty represented by the parameter $a$ increases. This is a significant departure from the representation of uncertainty provided by probability theory, and a key aspect of the analysis results presented next.

## 5.2 Consistency of Predictions

Figure 3 illustrates the ranges of predictions (maximum minus minimum values) obtained at different horizon-of-uncertainty levels. The horizontal axis represents ranges of predictions, in our case, ranges of peak acceleration of the impact wave as it propagates through the crushable foam material. The vertical axis represents the horizon-of-uncertainty. An expected trend is observed: As the horizon-of-uncertainty increases, the range of predictions made also grows. It expresses that the models included in a given family $U(a;q_o)$ make increasingly less consistent predictions. Discontinuities could be encountered, but

each curve is monotonically increasing because the domains $U(a;q_o)$ are nested within one another. Figure 3 illustrates the antagonistic nature of robustness-to-uncertainty and confidence-in-prediction.
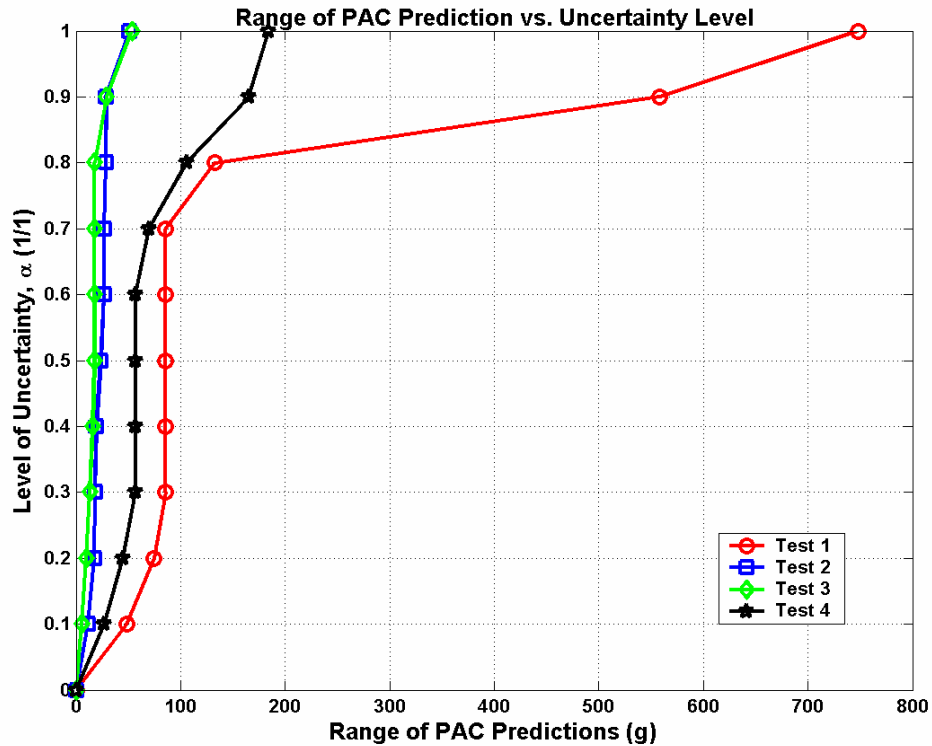


*Figure 3. Ranges of predictions obtained at increasing levels of horizon-of-uncertainty.*

The four curves shown in Figure 3 present results for four different settings of the system. Test 1 (in red color with circle symbols) simulates the propagation of a low impact through a thin layer of material. Test 2 (in blue color with square symbols) simulates the propagation of a low impact through a thick layer of material. Test 3 (in green color with diamond symbols) simulates the propagation of a high impact through a thin layer of material. Lastly, Test 4 (in black color with star symbols) simulates the propagation of a high impact through a thick layer of material. It can be observed that the family of models $U(a;q_o)$ provides more consistent predictions at any given level of uncertainty, $a$, for the second and third configurations of the system. If anything, this would tend to increase the credibility of the numerical simulation for predicting testing conditions similar to those of Tests 2 and 3. It does not, however, address prediction accuracy, which is analyzed next.

## 5.3 Accuracy of Predictions

The results of fidelity-to-data and robustness-to-uncertainty are examined. Figure 4 shows the robustness and opportunity curves obtained with the four configurations of the system (Tests 1 through 4), as a function of increasing levels of horizon-of-uncertainty. The horizontal axis represents test-analysis correlation errors, in our case, the error are expressed as percent between measured peak acceleration values $y^{Test}$ and predicted values $y$. The vertical axis represents the horizon-of-uncertainty.

For each configuration of low/high impact and thin/thick foam, two curves are shown. The robustness curves, in solid lines, are the worst possible test-analysis correlation errors that can be obtained from a family of models at a given horizon-of-uncertainty. As expected, fidelity-to-data worsens with uncertainty, hence, illustrating the antagonistic nature between truthfulness to data and robustness-to-uncertainty.
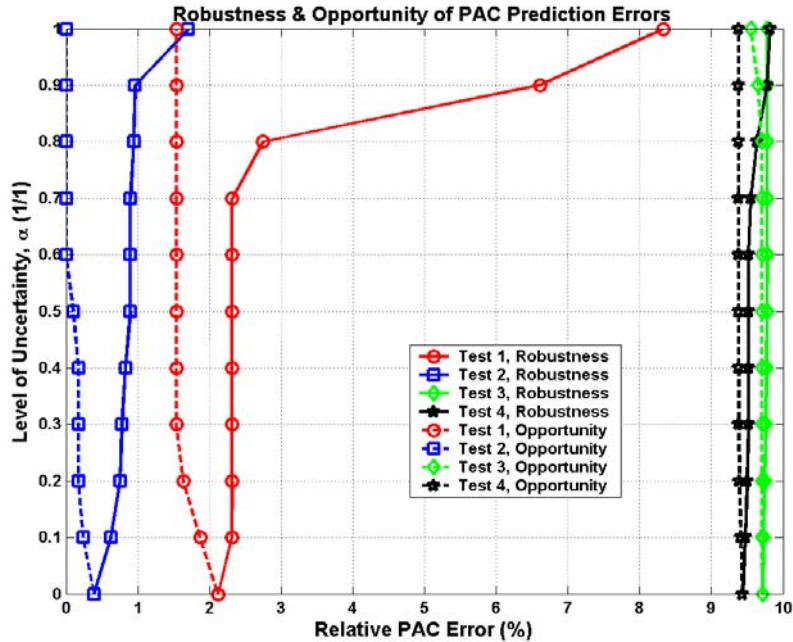
*Figure 4. Test-analysis errors obtained at increasing levels of horizon-of-uncertainty.*

The opportunity curves, shown in dashed lines, are the best possible test-analysis correlation errors that can be obtained from a family of models at a given horizon-of-uncertainty. Opportunity curves illustrate that uncertainty can sometimes be taken advantage of. For Tests 1 and 2, Figure 4 shows that tolerating a small level of uncertainty can yield material models that are more truthful to the test data.[6] As mentioned previously, the robustness and opportunity curves can be discontinuous, but their monotonically increasing (for robustness) or decreasing (for opportunity) natures are guaranteed by the fact that increasing horizon-of-uncertainty levels, $a$, generate nested domains $U(a;q_o)$.

When the robustness and opportunity curves are considered together, Figure 4 shows the ranges of predictive accuracy that can be obtained for each configuration of the system. For example, the prediction error for Test 1 is guaranteed within [1.5%; 2.7%] at the horizon-of-uncertainty level of $a=0.8$. No matter which material model is selected from the family $U(a;q_o)$, its predictive accuracy of Test 1 will be no worst than 2.7% but no better than 1.5%. A slight increase in uncertainty from $a=0.8$ to $a=0.9$ results in a potential deterioration of predictive accuracy from the interval [1.5%; 2.7%] to [1.5%; 6.6%]. Such analysis that examines the trade-off between fidelity-to-data and robustness-to-uncertainty provides useful guidelines to estimate a tolerable level of uncertainty.

Similarly, combining Figures 3 and 4 shows that trade-offs must be made between fidelity-to-data, robustness-to-uncertainty, and looseness-in-prediction. Tolerating increased levels of uncertainty in the knowledge of the material behavior offers the potential benefit of slightly improving the fidelity-to-data while, at the same time, reducing the consistency of predictions made from the family of models.

## 5.4 Sensitivity Analysis

The last question examined is sensitivity analysis (SA). SA provides valuable information to assess the effect of a parametric change or an uncertainty on a performance metric. Local methods generally rely on small perturbation theory, while global methods rely on statistical analysis. Examples are finite

---

[6] Figure 4 shows that the prediction error for Test 1 is equal to 2% with the nominal material model $q_o$ (when $a=0$, no uncertainty). If the material model is allowed to "deviate" from the nominal model up to the horizon of-uncertainty of $a=0.2$, models can be found that yield no more than 1.5% test-analysis error.

differences in the former case, and variance-based effect analysis in the latter case. The question asked is how to handle a non-probabilistic, non-parametric uncertainty such as the material lack-of-knowledge described in Figure 2.

The performance metric of interest in this work is the fidelity-to-data, $R$. Lack-of-knowledge is represented by an information-gap horizon-of-uncertainty parameter, $a$. Sensitivity coefficients are therefore defined as:

$$S_a = \frac{\partial R}{\partial a} \qquad (10)$$

We already know from the theoretical investigation that an increase in robustness worsens the fidelity-to-data, which means that $S_a > 0$. Nevertheless, slopes $S_a$ of the robustness-versus-fidelity curves are of great interest because they indicate "how fast" robustness can be lost. A visual illustration is provided in Figure 4, where it can be observed that the numerical simulation is least sensitive to the lack-of-knowledge for predicting the third and forth configurations of the system (Tests 3 and 4). A qualitative SA is summarized in Table 1. The sensitivity coefficients are estimated from Figure 4 by averaging the slopes of the four robustness curves over the entire range of fidelity-to-data values $R$.

*Table 1. Qualitative sensitivity analysis.*

|  | **Thin layer of material** | **Thick layer of material** |
|---|---|---|
| **Low impact condition** | **Test 1: High ($S_a \sim 6$)** | **Test 3: Low ($S_a < 1$)** |
| **High impact condition** | **Test 2: Medium ($S_a \sim 1$)** | **Test 4: Low ($S_a < 1$)** |

SA results are useful to understand the vulnerability of the simulation to the uncertainty. Clearly, our analysis shows that the model is most vulnerable to uncertainty for Test 1. This means that a significant deterioration in predictive accuracy could result from a small increase in lack-of-knowledge. Caution should be exercised when the numerical simulation is employed to analyze configurations similar to the low impact/thin foam pad settings of Test 1. It is of interest to mention that the measured data provided by Test 1 were the only data used in this study to calibrate the material models. The fact that the average sensitivity coefficient $S_a$ for Test 1 is the highest should not come as a surprise: It is, yet, another manifestation of the theoretical result that ***calibrated models have the least robustness-to-uncertainty*** [5].

Although the sensitivity coefficients $S_a$ do not convey any physical meaning, their main advantage is that they can be estimated based on the robustness-versus-performance functions (such as shown in Figure 4), for any type of uncertainty or lack-of-knowledge. It is emphasized that the information-gap representation of ignorance adopted throughout this work is not exclusive of other representations of uncertainty. Reference [8], for example, presents results where a probabilistic description of uncertainty is handled within the information-gap framework.

## 6. CONCLUSION

This work studies the relationship between several aspects of prediction accuracy. The main conclusion is that, in assessing the predictive 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 models selected for their robustness to uncertainty will tend to make inconsistent predictions. This finding seems discouraging because one would like to make accurate predictions while being robust to the sources of uncertainty and lack-of-knowledge. It is, however, a fundamental limitation of predictive science that scientists and engineers

should not loose sight of. The trade-off simply expresses that obtaining accurate predictions is conditioned by the assumptions upon which the models are built.

# REFERENCES

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

[2] 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.

[3] 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 (ISMA)*, Leuven, Belgium, September 13-15, 2000, pp. 403-413.

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

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

[6] 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 (U)," *14th Biennial Nuclear Explosives Design Physics Conference*, Los Alamos National Laboratory, Los Alamos, New Mexico, October 20-24, 2003. *Unclassified*.

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

[8] Hemez, F.M., Ben-Haim, Y., Cogan, S., "Information-gap Robustness for the Test-analysis Correlation of a Non-linear Transient Simulation," *9th AIAA/ISSMO Symposium on Multi-disciplinary Analysis and Optimization*, AIAA-2002-5420, Atlanta, Georgia, September 4-6, 2002.

[9] 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.