ABSTRACT

A key aspect of science-based predictive modeling is to assess the credibility of predictions. To gain confidence in predictions, one should demonstrate consistency between physical observations, expert judgments, and the predictions of equally credible models. This suggests a relationship between fidelity-to-data, robustness-to-uncertainty, and confidence in prediction. The purpose of this work is to explore the interaction between these three aspects of predictive modeling. The concepts of fidelity, robustness, and confidence are first defined in a broad sense. A Theorem is then proven that establishes that these three objectives are antagonistic. This means that high-fidelity models cannot, at the same time, be made robust to uncertainty and lack-of-knowledge. Similarly, equally robust models cannot provide consistent predictions, hence reducing confidence. The conclusion of this theoretical investigation 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.

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 usually demonstrated through the activities of model 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 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 theoretical level for the most part. A Theorem, that we believe is important to explore the relationship between fidelity-to-data, robustness-to-uncertainty, and prediction looseness, is enounced and proven. The objective of this publication is to establish the main theoretical result upon which future applications will be based. Section 2 provides background information in non-technical terms, and defines the notations used in the remainder. Section 3 discusses briefly the concepts of fidelity-to-data and robustness-to-uncertainty. The main theoretical result is given in Section 4, where the ability to make consistent predictions with a family of equally credible models is investigated. Finally, the trade-offs between fidelity, robustness, and looseness are briefly explored in Section 5.
2. BACKGROUND

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 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 that one has in the ability to make accurate predictions. The notion of accuracy $\lambda_Y$ is further discussed below. 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 our 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 basic 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. Examples include resonant frequencies, peak stress or acceleration responses, shock response spectra, etc. These outputs are usually features extracted from the structural 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 $y=C_L$. 

Figure 1. Illustration of the concepts of fidelity-to-data, robustness-to-uncertainty, and prediction accuracy (range, “looseness,” or confidence).
Another example is the response of a building to Earthquakes. The inputs might represent the amplitude and frequency contents of the excitation, and the output prediction \( y \) might be a peak level of structural stress occurring in the structure.

- The quantity \( q \) represents parameters that specify the structure and coefficients of the family of models developed to represent the physical phenomenon of interest. For example, a material model of plasticity can be parameterized with quantities such as a linear modulus of elasticity and yield stress. It is emphasized that the inputs \( q \) can include discrete and continuous parameters that control the functional form of the model. Also, various models can be functions of different subsets of parameters \( q \).

In the general case, the model is represented as:

\[
y = M(p; q_o)
\]

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

A domain denoted by \( D_p = \{p^{(\text{min})} \times p^{(\text{max})}\} \) 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 \( D_p = [p_1^{(\text{min})}; p_1^{(\text{max})}] \times [p_2^{(\text{min})}; p_2^{(\text{max})}] \).

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:

\[
U(a; q_o) = \left\{ M(p; q) \left| \| q - q_o \| \leq a \right. \right\}, \quad \text{for } a \geq 0
\]

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 \( y^{\text{Test}} \). 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 \( y^{\text{Test}} \). Measurements made, on the other hand, for different configurations \( p_1 \ldots p_m \) will be indexed as \( y^{\text{Test}(1)} \ldots y^{\text{Test}(m)} \).

2.2 Concepts of 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 \( y^{\text{Test}} \) and simulation predictions \( y \) at a setting \( (p; q_o) \):

\[
R = \left\| y^{\text{Test}} - y \right\|
\]

Fidelity-to-data is pictured in Figure 1 as the vertical distance between a measurement \( y^{\text{Test}} \) and a prediction \( y \) for the physical experiment and numerical simulation performed at the setting \( (p_1; p_2) \). Examples of fidelity metrics are the conventional Root Mean Square (RMS) error between tests and predictions, and the Mahalanobis multivariate test that can account for experimental variability:
Robustness-to-uncertainty refers to the range of settings \( q \) that provide no more than a given level of prediction error \( R_{\text{Max}} \). The concept of robustness is illustrated in Figure 1 by showing a subset \( U(a;q_o) \) of the design domain \([p_{1}\text{ (min)};p_{1}\text{ (max)}] \times [p_{2}\text{ (min)};p_{2}\text{ (max)}]\). 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 \( R_{\text{Max}} \). 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 \( U(a;q_o) \), as shown in Figure 2. 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 \( R_{\text{Max}} \). Generally, a trade-off must be decided upon between the robustness-to-uncertainty \( (a) \) and prediction error \( (R_{\text{Max}}) \), 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 validation is to establish predictive confidence by estimating the range of predictions \( \lambda_Y \) (or, equivalently, the lack-of-consistency)

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 \( a \).
provided by equally credible sources of information. The range of predictions is related to the notion of confidence through, for example, the use of statistical testing.

Figure 3. Sources of evidence for predicting the peak acceleration of an impact propagating through a crushable foam material.

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. An example of what is meant by “model” is illustrated in Figure 3. It shows the prediction of peak acceleration levels, given control parameters \((p_1; p_2)\) of the physical experiment, obtained from physical tests, expert judgment, and several physics-based models (some very crude, like a single degree-of-freedom shock response equation; some more sophisticated, like a three-dimensional finite element simulation). These can all be considered as models because they define a relationship between the inputs \((p_1; p_2)\) and the peak acceleration output, \(y\). 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.

3. FIDELITY-TO-DATA AND ROBUSTNESS-TO-UNCERTAINTY

A family of models \(U(a;p,q)\) such as defined in 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 uncertainty 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 modeling lack-of-knowledge.\(^4\)

---

\(^4\) 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.
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:

\[
R^* = \min_{M \in U(a; q_o)} R
\]

the result of which is 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 or finite element model updating solve.

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 \( R_{\text{Max}} \), 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 \( R > R_{\text{Max}} \). 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:

\[
a^* = \max_{a \geq 0} \left\{ \min_{M \in U(a; q_o)} R \leq R_{\text{Max}} \right\}
\]

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 \( R_{\text{Max}} \). As pointed out earlier, it could happen that the robust-optimal model features a better fidelity-to-data, or \( R < R_{\text{Max}} \), 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 and lack-of-knowledge do not deteriorate the prediction error by more than \( R_{\text{Max}} \).

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 \( R_{\text{Max}} \) 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 (6).

Figure 4. Investigation of the trade-off between fidelity-to-data and robustness.

Figure 4 illustrates the trade-off between the aspiration of fidelity and robustness. It is borrowed from a study documented in Reference [6], where a simulation of the propagation of an impact through a layer of non-linear, crushable foam is compared to physical measurements. The experimental set-up, finite element modeling, and sources of uncertainty are discussed in Reference [7].

The figure shows the degradation in prediction accuracy that can result from accepting increasingly more uncertainty about the value of model parameters. The uncertainty is defined as the gap of knowledge between values used in the simulation and the “true-but-unknown” parameters. A total of four parameters are considered: a bolt preload, two angles, and a calibration scaling. The robustness-versus-fidelity curve of the bolt preload variable, shown with black diamonds in Figure 4, suggests that the prediction accuracy of the finite element simulation is highly sensitive to uncertainty in the knowledge of the “true” preload value. On the other hand, large uncertainty in the values of the angles has almost no effect on the fidelity metric. From this analysis, it is learned that controlling the bolt preload is more important to the predictive accuracy of the model than controlling the angles. Conversely, the study shows the worst prediction error that can be expected if a measurement system cannot provide values of a parameters within, say, $a^*=20\%$ accuracy. Note that performing a calibration of the parameters would not provide as much insight into the understanding of which factors are important to control the predictive accuracy. Calibration would only identify the optimal parameter values that provide no more than 14% prediction error (see Figure 4).

4. LOOSENESS OF MODEL PREDICTIONS

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 Definitions and Set-up of the Problem

Let \(a^*\) be the robustness-to-uncertainty of model \(M(p_k;q)\) at the experimental configuration \(p_k\), as defined in equation (6). Let \(U^* = U(a^*;q_0)\) 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 \(a^*\) and \(U^*\) 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 \(a\).

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 all belong to \(U^*\) is simply defined as the range of model predictions in \(U^*\):

\[
\lambda_Y = \max_{M(p_k;q) \in U^*} M(p_k;q) - \min_{M(p_k;q) \in U^*} M(p_k;q)
\]

(7)

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:

\[
\text{Sympathetic: } \frac{\partial a^*}{\partial q} \cdot \frac{\partial \lambda_Y}{\partial q} \leq 0, \quad \text{Antagonistic: } \frac{\partial a^*}{\partial q} \cdot \frac{\partial \lambda_Y}{\partial q} \geq 0
\]

(8)

The Theorem enounced in Section 4.2 shows that, under fairly weak conditions, robustness and range are always antagonistic. Two basic axioms are needed to support the main result. They are:

Axiom 1, Nesting: \(a < a'\) implies that \(U \subseteq U'\), where the compact notation is used: \(U = U(a;q_0)\) and \(U' = U(a';q_0)\). Axiom 1 expresses that, as the horizon-of-uncertainty increases, the family of models includes all previously included models, plus new ones. This is a direct consequence of the property of convexity.

Axiom 2, Translation: \(U(a;q_0) = U(a;q_0') - M(p_k;q_0) + M(p_k;q_0)\). Axiom 2 expresses that two families of models that share the same horizon-of-uncertainty, \(a\), only differ in their center points. One can move from one to the other through a simple translation. Addition of an element \(M(p_k;q_0)\) to the set \(U(a;q_0)\) means that the value \(M(p_k;q_0)\) is added to each element of \(U(a;q_0')\).

4.2 Theoretical Result of the Robustness-range Trade-off

| Theorem: Let \(U(a;q_0)\) be an info-gap family of models that obeys the axioms of nesting and translation, and let \(a = a(p_k;q_0;R_{Max})\) versus \(-R_{Max}\) be its robustness function. Consider two initial settings of model parameters, \(q_0^*\) and \(q_0'\). If \(a(p_k;q_0^*;R_{Max}) \geq a(p_k;q_0';R_{Max})\), then \(\lambda_Y(q_0^*) \geq \lambda_Y(q_0')\). That is, robustness and range (or prediction looseness) are antagonistic. |

---

5 Even with robustness-optimal model selection, uniqueness of the solution is not guaranteed. Because the family of models \(U(a;q_0)\) 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 \(U(a;q_0)\). This points to an interesting contradiction. Even though it is theoretically possible to find two models with the same fidelity and robustness, the probability of occurrence of such event is zero because the mathematical measure (which is how a probability is defined) of a convex hull is null. This contradicts the common understanding of possibility and probability: “All that is possible has a non-zero probability of occurring.”
The proof employs the following concise notation: \( M^*=M(p_0; q_0^*), \) \( M'=M(p_0; q_0'), \) \( a^*=a(p_0; q_0^*; R_{Max}), \) and \( a'=a(p_0; q_0'; R_{Max}). \) Starting by assuming \( a^* \geq a' \), and using the property of nesting (Axiom 1) leads to \( U(a'; q_0^*) \subset U(a^*; q_0^*). \) The translation property (Axiom 2) is used next:

\[
U(a^*; q_0^* \cap M^* + M^*) \subset U(a^*; q_0^*)
\]

Hence the following two expressions hold:

\[
\begin{align*}
\max_{M \in U(a^*; q_0^*)} M(p_k; q) & \geq \max_{M \in U(a^*; q_0^*)} M(p_k; q) = \max_{M \in U(a^*; q_0^*)} (M(p_k; q) - M^* + M^*) \\
\min_{M \in U(a^*; q_0^*)} M(p_k; q) & \leq \min_{M \in U(a^*; q_0^*)} M(p_k; q) = \min_{M \in U(a^*; q_0^*)} (M(p_k; q) - M^* + M^*)
\end{align*}
\]

From these relations and the definition of looseness in equation (7) we find:

\[
\begin{align*}
\lambda_Y(q_0^*) \geq & \lambda_Y(q_0^*) \\
\lambda_Y(q_0^*) \geq & \lambda_Y(q_0^*)
\end{align*}
\]

which completes the proof. ■

It is emphasized that the previous proof relies on the information-gap description of uncertainty. 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.

5. EXPLORING THE TRADE-OFFS OF PREDICTION ACCURACY

Three quantities are central to the info-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. The results of Reference [5], combined with our Theorem, establish several trade-offs briefly discussed below.

The Theorem that explores the relationship between robustness \( a \) and prediction looseness \( \lambda_Y \) can be summarized by:

\[
\frac{\partial \lambda_Y}{\partial a} \geq 0
\]

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 the Theorem to the following inequalities is trivial:

\[
\begin{align*}
\frac{\partial a}{\partial R} \geq 0, \quad \frac{\partial \lambda_Y}{\partial a} \geq 0, \quad \frac{\partial \lambda_Y}{\partial R} \geq 0
\end{align*}
\]

Inequalities (13) express the three 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.

6. CONCLUSION

This work studies the relationship between several aspects of prediction accuracy. The main conclusion of this theoretical investigation 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 lose sight of. The trade-off simply expresses that obtaining accurate predictions is conditioned by the assumptions upon which the models are built.

One may ask what the practical finality of this work is. The answer is that the robustness-to-uncertainty and range of predictions should be extended to future model predictions (or forecasts), as opposed to the “post-diction” of past experiments. In mathematical terms, the robustness and range should be made functions of the control parameters that define the validation domain, or $a=a(p)$ and $\lambda_Y=\lambda_Y(p)$. Being able to study the trade-off between robustness $a(p)$ and looseness $\lambda_Y(p)$ at settings or configurations $p$ that have not been tested yet is the ultimate goal of prediction accuracy assessment. This is a difficult problem because we have virtually no fundamental understanding of how robustness and prediction looseness should vary with the setting or configuration $p$. Consequently, any meta-model $p \rightarrow a(p)$ or $p \rightarrow \lambda_Y(p)$ developed will face significant structural uncertainty that may require the integration of various information theories [8]. We nevertheless expect that forecasting robustness and looseness will tend to be antagonistic. An application is currently being considered to demonstrate this concept for assessing the prediction accuracy of a numerical model.

7. REFERENCES


