LA-UR-10- 02003

Approved for public release; distribution is unlimited.

Title:	DEMONSTRATING THE IMPROVEMENT OF PREDICTIVE MATURITY OF A COMPUTATIONAL MODEL *(Manuscript)*
Author(s):	François M. Hemez, Los Alamos National Laboratory, XCP-1 Sezer Atamturktur, Civil Engineering Department, Clemson University Cetin Unal, Los Alamos National Laboratory, CCS-DO
Intended for:	51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and 12th AIAA Non-Deterministic Approaches Conference, Orlando, Florida, April 12-15, 2010

Form 836 (7/06)

This page is left blank intentionally.

12th AIAA Non-Deterministic Approaches (NDA) Conference
Orlando, Florida, April 12-15, 2010, AIAA Tracking Number 213426

Demonstrating the Improvement of Predictive Maturity of a Computational Model

François Hemez*
Los Alamos National Laboratory
XCP-Division (XCP-1)
Los Alamos, New Mexico

Sezer Atamturktur†
Clemson University
Civil Engineering Department
Clemson, South Carolina

Cetin Unal&
Los Alamos National Laboratory
CCS-Division (CCS-DO)
Los Alamos, New Mexico

Abstract: We demonstrate an improvement of predictive capability brought to a non-linear material model using a combination of test data, sensitivity analysis, uncertainty quantification, and calibration. A model that captures increasingly complicated phenomena, such as plasticity, temperature and strain rate effects, is analyzed. Predictive maturity is defined, here, as the accuracy of the model to predict multiple Hopkinson bar experiments. A statistical discrepancy quantifies the systematic disagreement (bias) between measurements and predictions. Our hypothesis is that improving the predictive capability of a model should translate into better agreement between measurements and predictions. This agreement, in turn, should lead to a smaller discrepancy. We have recently proposed to use discrepancy and coverage, that is, the extent to which the physical experiments used for calibration populate the regime of applicability of the model, as basis to define a Predictive Maturity Index (PMI). It was shown that predictive maturity could be improved when additional physical tests are made available to increase coverage of the regime of applicability. This contribution illustrates how the PMI changes as "better" physics are implemented in the model. The application is the non-linear Preston-Tonks-Wallace (PTW) strength model applied to Beryllium metal. We demonstrate that our framework tracks the evolution of maturity of the PTW model. Robustness of the PMI with respect to the selection of coefficients needed in its definition is also studied. *(Approved for unlimited, public release on March-xx-2010, LA-UR-10-xxxx.)*

1. Introduction

The ever-increasing role that numerical models and computer simulations play in decision-making motivates the need to quantify ***"predictive maturity."*** This concept is needed to assess the extent to which a modeling capability may be applied successfully to a specific problem. Tracking progress in the development of a predictive capability is another important motivation. Code developers and their customers have a need to understand the degree of improvement in

* Technical staff member in XCP-Division and manager of the Advanced Scientific Computing (ASC) code verification project. Mailing address: Los Alamos National Laboratory, XCP-1, Mail Stop F644, Los Alamos, New Mexico 87545. E-mail: hemez@lanl.gov. AIAA Senior Member.

† Associate Professor of civil engineering at Clemson University. Mailing address: Clemson University, Department of Civil Engineering, Lowry Hall, Box 340911, Clemson, South Carolina 29634-0911. E-mail: sez@clemson.edu.

& Technical staff member in CCS-Division and manager of the Nuclear Energy Advanced Modeling and Simulation (NEAMS) program. Mailing address: Los Alamos National Laboratory, CCS-DO, Mail Stop B297, Los Alamos, New Mexico 87545. E-mail: cu@lanl.gov.

predictive capability potentially realized by implementing a new algorithm or performing a new experimental campaign. This cannot easily be achieved without tracking progress with a metric.

Whether it is to make a statement about the predictive maturity of a modeling capability or track progress, an important question is whether predictive maturity can even be achieved in the case of multi-scale, multi-physics models. Our working hypothesis is that, if one could conduct a large-enough number of physical experiments, then one could reasonably envision reducing the prediction uncertainty down to the level of uncontrolled, natural variability. Therefore, it can be postulated that, as new experimental information becomes available, the disagreement between simulation predictions and physical measurements should be reduced down to a "constant" level of systematic bias. Such statement should also apply vis-à-vis the physical complexity or degree of sophistication of models implemented in simulation codes. As models needed to represent the relevant physics are progressively developed and implemented in a computer code, the disagreement between simulation predictions and physical measurements should be reduced.

In Reference [1], we have proposed to use this concept of ***"stabilization"*** of the systematic discrepancy between numerical predictions and physical measurements to assess the maturity of a model or simulation. Our approach is in contrast to other efforts to define predictive maturity through the combination of qualitative assessments and quantitative scores. An institution that leads these efforts is the National Aeronautics and Space Administration (specifically, NASA Langley Research Center), as documented in References [2-3]. The authors of References [4-5] develop yet another approach based on a goodness-of-fit metric to correlate predictions and measurements. The metric includes the concept of novelty of information to quantify whether an additional physical test is independent from the others in terms of exploring the physics regimes that previous experiments have stayed away from.

What can be observed from literature available on the subject is that very few investigations venture away from the concept of goodness-of-fit, that is, the extent to which predictions from models or simulations correlates well with physical measurements. We argue that goodness-of-fit metrics are only part of the puzzle because they do not address the domain of applicability of a modeling capability. It is our opinion that predictive maturity should include a statement about the ability of models to deliver accurate predictions over a ***"range"*** of settings of the domain of applicability. These observations motivated the proposal, in Reference [1], of a metric based on several attributes to go beyond the simple concept of goodness-of-fit. The necessary attributes include, as mentioned previously, a measure of discrepancy between physical observations and numerical predictions and, also, a measure of complexity of the model and the extent to which physical experiments used for calibration cover the domain of applicability of the model.

In this work, we use the non-linear Preston-Tonks-Wallace (PTW) material model of plastic deformation for illustration. Physical experiments are performed on samples of Beryllium metal to measure its strain-stress behavior as a function of various settings of temperature and strain-rate. These experiments are Hopkinson bar tests. A statistical procedure is then applied to infer the joint probability distribution of eight coefficients of the PTW model such that its predictions match the measured strain-stress curves over the regime of temperature and strain-rate settings of interest. Three model variants are investigated, ranging from a moderately predictive version to the original PTW equation. The accuracy of model predictions is tracked as new experiments are provided to the analysis and better models are analyzed. It is shown that the PMI is capable to track progress, as more physical experiments are provided to the analysis, and discriminate the moderately predictive model from better variants.

These results suggest that the approach can be useful to claim completion of a calibration process, deliver model predictions with quantified uncertainty and bias, and provide insight into the predictive maturity of a model or numerical simulation. These results also open the door to

designing physical experiments such that a minimal resource can be spent on performing tests while guaranteeing a given level of predictive maturity. This direction is one that will be pursued in future research. What this work does not yet address, that will also be pursued in future work, is to understand the extent to which these concepts can be applied to the verification, validation and uncertainty quantification of multi-space, multi-physics models or simulations.

The manuscript is organized as follows. Section 2 summarizes the PMI metric developed in Reference [1]. The PTW model and datasets of Hopkinson bar measurements are presented in section 3. Section 4 discusses the results obtained by analyzing several variants of the model and progressively adding measurements obtained at different settings of temperature and strain rate. Finally, conclusions and directions of future research are briefly discussed in section 5.

2. The Predictive Maturity Index Metric

In this work, the terms "model," "code" and "simulation" are used inter-changeably to denote an analytical or numerical model. They are collectively referred to as the "model" that, in short, builds a functional relation between input variables and output responses:

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

where the pair (p; θ) denotes inputs to the model and y is the output. Variables (p; θ) and y can be scalar-valued or multi-dimensional. The symbol p refers to inputs that define the domain of applicability while θ denotes other inputs such as ancillary variables or calibration variables. In the application to the PTW model, input variable p refers to the pair (temperature; strain rate) of control variables while θ refers to eight, material-dependent coefficients.

The terms "data" and "dataset," likewise, refer to physical measurements or observations collected by performing experimental tests. Measurements are denoted by the symbol y^{Test}, a quantity that, again, can be either scalar-valued or multi-dimensional. Verification and Validation (V&V) investigations typically assess the extent to which predictions y and measurements y^{Test} are correlated, and quantify the sensitivity and uncertainty of these predictions [6-8].

2.1 The Attributes of Predictive Maturity

Reference [1] postulates that, at least, three attributes are essential to define predictive maturity: (A-1) the extent to which datasets available "cover" the domain of applicability; (A-2) the "complexity" of the model; and (A-3) the level of accuracy of model predictions. Clearly, the list is not exhaustive and other attributes could be added to define predictive maturity. Examples include the robustness of model predictions to assumptions, and time-to-solution.

Time-to-solution is important because getting the answer from a computer code that runs in one minute as opposed to another simulation that runs in one hour, everything else being equal, matters greatly. Robustness is essential, as stressed in References [9-10]. Robustness refers to the extent to which model predictions vary or, to the contrary, are insensitive or ***"robust,"*** when assumptions upon which the model is derived are modified. Achieving robustness means that predictions do not change significantly and, therefore, can be trusted with a higher degree of confidence, even if some of the assumptions of the simulation are incorrect. Robustness can be dealt with by quantifying the extent to which the PMI is sensitive to modeling assumptions.

2.2 Attribute (A-1): Coverage of the Domain of Applicability

Coverage of the domain of applicability $\{\Omega_V\}$ refers to the location of physical experiments performed. Figure 1 illustrates this concept for a notional 2D domain parameterized by a pair $(p_1; p_2)$ of control parameters. In Reference [1], coverage is measured by assessing the extent to which the physical experiments performed fill the space. One can, for example, compute the

convex hull $\{\Omega_{CH}\}$ of physical experiments and measure it relative to the total "volume" of the multi-dimensional domain of applicability $\{\Omega_V\}$.

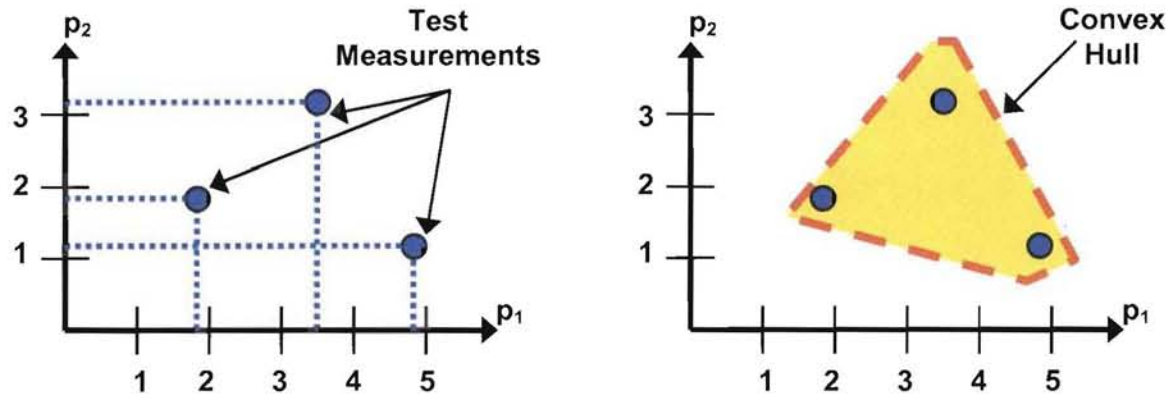

Figure 1. Definition of a 2D domain of applicability (left) and coverage by test data (right).

The convex hull is, by definition, the smallest domain that, while remaining convex, includes all physical experiments. The ratio between the volume of the convex hull $\{\Omega_{CH}\}$ and that of the domain of applicability $\{\Omega_V\}$ defines the coverage metric:

$$\eta_C = \frac{Volume(\Omega_{CH})}{Volume(\Omega_V)}, \tag{2}$$

where Volume(•) denotes a function that calculates the N-dimensional volume of a region of $\Re^N$. For the domain of applicability $\{\Omega_V\}$ = [1; 5] x [1; 3] shown in Figure 1, coverage is equal to the area highlighted in orange divided by the total area, that is, η_C = Area(Ω_{CH})/Area(Ω_V).

2.3 Attribute (A-2): Complexity of a Model or Numerical Simulation

Defining the complexity of a model can be an extremely difficult task. It may involve making a statement about the sophistication of physical principles that are modeled; the complexity of mathematical spaces where the continuous or discrete solutions are constructed; the degree to which different physics are coupled; how many sub-models, algorithms or numerical methods are implemented; how many lines of codes are written; etc.

To define complexity, one metric is chosen that, while remaining simple, cuts across most of these aspects: the number of calibration ***"knobs,"*** or ancillary variables θ_k, of the model. This choice is guided by the principle that, in general, more sophisticated models possess larger numbers of ancillary variables. The number of calibration knobs is denoted by the symbol N_K.

2.4 Attribute (A-3): Level of Accuracy of Model Predictions

The ability of a model or simulation to accurately reproduce the datasets is defined, not so much in terms of goodness-of-fit, but through a ***"discrepancy"*** as proposed in Reference [11]:

$$y^{Test}(p) = y(p;\theta) + \delta(p) + \varepsilon^{Test}, \tag{3}$$

where symbols $y^{Test}(p)$, $y(p;\theta)$, $\delta(p)$ and ε^{Test} denote, respectively, the physical measurements, model predictions, discrepancy term, and measurement error. The measurement error takes the form of a zero-mean, Gaussian process, $\varepsilon^{Test} \sim N(0; \sigma^{Test})$, obtained from replicate experiments.

The role played by the discrepancy term $\delta(p)$ is to capture residual differences between predictions and measurements that cannot be accounted for when calibration parameters θ are varied or calibrated. Discrepancy is a statistical process that represents model form error, as

opposed to parametric uncertainty captured by ancillary variables θ. The theory behind equation (3) comes from Reference [11]. It is implemented at the Los Alamos National Laboratory in the Gaussian Process Modeling for Simulation Analysis (GPM/SA) software package [12-13].

Once the statistics of discrepancy δ(p) are estimated, the metric of accuracy is defined as:

$$\delta_S = \frac{\sigma_{Max}(\delta(p))}{\sigma_{Max}(y(p;\theta))}, \tag{4}$$

where the operator σ(•) denotes the maximum singular value computed when all realizations of either discrepancy, that is, δ(p), or prediction, that is, y(p;θ), are analyzed. The Singular Value Decomposition (SVD) algorithm is used in these computations. (See Reference [1] for details.) The maximum singular value is a convenient metric because it exhibits the same physical units as those of measurements and predictions analyzed, while being independent of the number of realizations analyzed. The resulting scaled discrepancy metric δ_S exhibits no unit.

2.5 Definition of the Predictive Maturity Index (PMI) Metric

The PMI metric defined in Reference [1] depends on coverage η_C, number of knobs N_K, and goodness-of-fit δ_S. Its values are, without loss of generality, bounded in the interval $0 \leq PMI \leq 1$ for intuitive interpretation. PMI = 0 means that the model has no maturity what-so-ever. PMI = 1 implies, on the other hand, perfect maturity over the entire domain of applicability. Clearly, these two cases are asymptotes that cannot be reached with a finite number of physical experiments.

In Reference [1], several mathematical and asymptotic properties of predictive maturity are proposed that constrain the definition of the metric. The first property is that the level of maturity increases when coverage of the domain of applicability increases. Conversely, maturity should decrease when the number of knobs increases. It means that simpler models, that tend to have fewer knobs, are more mature than complicated models defined with more knobs, at equivalent levels of coverage and discrepancy. Thirdly, maturity decreases when discrepancy increases.

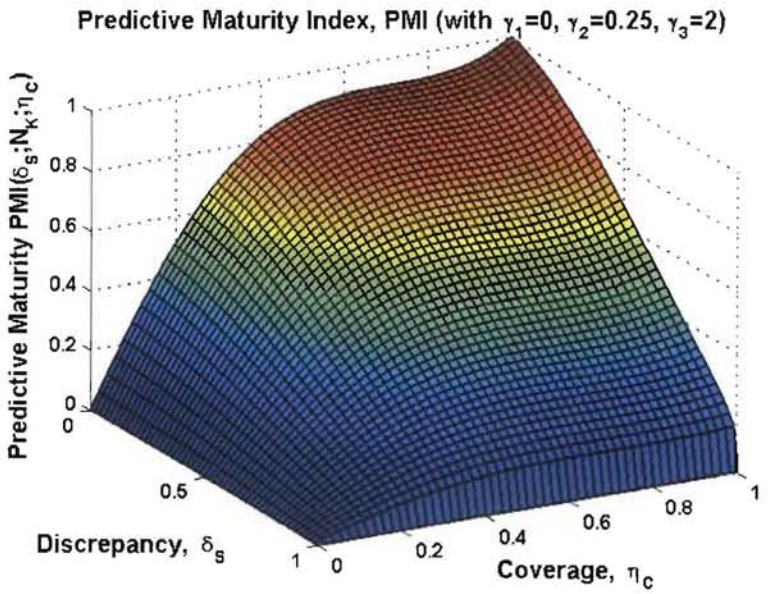

Figure 2. Predictive maturity $PMI(\delta_S;N_K;\eta_C)$ with coefficients $\gamma_1 = 0$, $\gamma_2 = ¼$ and $\gamma_2 = 2$.

Four asymptotic properties of predictive maturity are also defined. The first limit case is that maturity tends to zero as predictions of the model become increasingly inaccurate, or $\delta_S \rightarrow 1$.

The second case is that maturity deteriorates as $N_K \rightarrow \infty$. This expresses that it becomes easier to match physical measurements if a large number of knobs is available for calibration. The third case is that maturity deteriorates if coverage tends to zero, or $\eta_C \rightarrow 0$. Finally, we postulate that "perfect" maturity, or PMI = 1, can only be reached to the extent that the model delivers "perfect" accuracy, or $\delta_S \rightarrow 0$, ***and*** the physical experiments used for correlation and calibration provide "infinite" coverage of the domain of applicability, or $\eta_C \rightarrow 1$.

Reference [1] proposes the following PMI metric that verifies the above-stated properties:

$$PMI(\delta_S;N_K;\eta_C)=\eta_C\times\left(\frac{N_R}{N_K}\right)^{\gamma_1}\times(1-\delta_S)^{\gamma_2}\times e^{\left(1-\eta_C^2\right)^{\gamma_3}-\delta_S^2}, \tag{5}$$

where γ_1, γ_2 and γ_3 are strictly positive, user-defined coefficients used to weight the effects of various contributions relative to the first one. The symbol N_R denotes a "reference" number of knobs. It is a characteristic number of calibration variables that one would expect to encounter in a class of similar models. The ratio (N_R/N_K) defines a non-dimensional number of knobs.

Figure 2 illustrates a "2D slice" of the PMI function obtained when the number of knobs is kept constant. Coefficients $\gamma_1 = 0$, $\gamma_2 = ¼$ and $\gamma_3 = 2$ are used to show the combined effect of discrepancy δ_S and coverage η_C. The PMI reaches "perfect" predictability, or PMI = 1, only when $(\delta_S; \eta_C) \rightarrow (0; 1)$. It can also be observed that maturity decreases as either coverage reduces or discrepancy increases. The triplet (γ_1; γ_2; γ_3) is kept constant for consistency with Reference [1].

3. Application of the Preston-Tonks-Wallace Model to Beryllium Metal

In this section, the Preston-Tonks-Wallace (PTW) model of plastic deformation documented in Reference [14] is briefly discussed. It is applied to the prediction of strain-stress curves for the light-weight, high-strength Beryllium (Be) metal. Two variants of the original PTW equations are also defined to assess, in section 4, the evolution of the PMI as "better" physics are analyzed.

3.1 The Preston-Tonks-Wallace (PTW) Model of Strength and Plastic Deformation

The PTW model of strength and plasticity describes strain-stress curves obtained at various regimes of strain rate and temperature. It models the plastic flow of metals and is usually though of as suitable to simulate the material response to fast transients such as those from explosive loading or high velocity impacts. The main equations from Reference [14] are:

$$\sigma_S = s_0 - \frac{2\left(s_0 - s_\infty\right)}{\sqrt{\pi}}\int_0^{\kappa T\log(\gamma\dot{\varepsilon}/\theta)} e^{-\lambda^2}\,d\lambda \quad \text{and} \quad \sigma_Y = y_0 - \frac{2\left(y_0 - y_\infty\right)}{\sqrt{\pi}}\int_0^{\kappa T\log(\gamma\dot{\varepsilon}/\theta)} e^{-\lambda^2}\,d\lambda, \tag{6}$$

where symbols σ_S and σ_Y denote the dimensionless work hardening saturation stress and yield stress, respectively. Control parameters that define the two-dimensional domain of applicability are the strain rate ($d\varepsilon/dt$) and temperature (T) of equation (6). Symbols θ, κ, γ, s_0, s_∞, y_0 and y_∞ are seven, also dimensionless, calibration variables that depend on the material analyzed.

Table 1. Definition of variants of the PTW model analyzed.

Identifier	Description
Variant-0	It is the original PTW model of equation (6).
Variant-1	The "erf" function replaced with an exponential "err" function.
Variant-2	The stress hardening is turned off (with PTW parameter p = 0).

In addition to the original PTW model of equation (6), two variants are analyzed to observe how the PMI metric changes as "better" physics are implemented to describe the behavior of the

material. These variants are defined in Table 1. The first version is the original PTW model. The second version, labeled "Variant-1," is an implementation where "erf" functions of equation (6) are replaced by simpler, exponential "err" functions. This implementation is expected to perform less accurately than the original model. A third version, labeled "Variant-2," is an implementation where stress hardening is turned off. Because this implementation is incapable to account for an important phenomenology, it is expected to be the least mature of the variants considered.

3.2 Application to the High-strength Beryllium (Be) Metal

The application considered is the development of a material model for the Be metal over a range of temperatures and strain rates likely to be encountered in an application of interest that could, for example, be the numerical simulation of the mechanical and thermal responses of fuel rods subjected to irradiation in a nuclear reactor. Hopkinson bar experiments are performed on Be samples to collect the strain-stress curves shown in Figure 3. The pairs (T; dε/dt) of control parameters that define the Hopkinson bar tests are listed in Table 2.

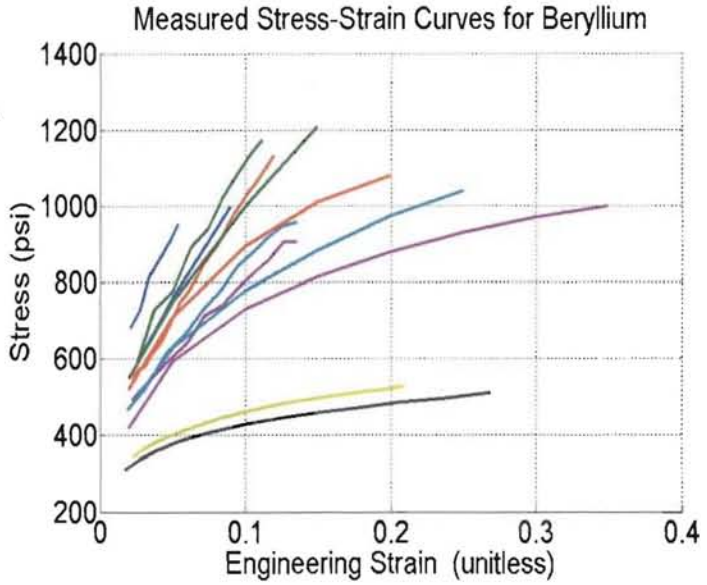

Figure 3. Strain-stress curves measured from Hopkinson bar tests for the Be metal.

Table 2. Definition of settings for experiments performed on Be samples.

Dataset	Maximum Strain, ε_{Max}	Temperature, T (°K)	Strain-rate, dε/dt (sec.$^{-1}$)
1	0.0539	77.0 °K	3,000.0 sec.$^{-1}$
2	0.1118	223.0 °K	3,500.0 sec.$^{-1}$
3	0.1202	298.0 °K	3,500.0 sec.$^{-1}$
4	0.1355	473.0 °K	3,700.0 sec.$^{-1}$
5	0.1360	573.0 °K	3,900.0 sec.$^{-1}$
6	0.2100	573.0 °K	1.0 sec.$^{-1}$
7	0.2689	473.0 °K	0.001 sec.$^{-1}$
8	0.090	293.0 °K	2,000 sec.$^{-1}$
9	0.150	293.0 °K	950.0 sec.$^{-1}$
10	0.200	293.0 °K	2.0 sec.$^{-1}$
11	0.250	293.0 °K	0.02 sec.$^{-1}$
12	0.350	293.0 °K	0.0001 sec.$^{-1}$

Figure 3 illustrates the range of stain and stress values, as well as the variety of shapes, that the material model is expected to reproduce over the domain of applicability. Measurement error is modeled as a Gaussian process with zero mean and 2.5% variance, as described by the experimentalists.

The seven ancillary variables (θ; κ; γ; s_0; s_∞; y_0; y_∞) of the PTW model and its variants are defined in Table 3, together with the lower and upper bounds within which they are believed to vary for the Be metal. Ideally, one could calibrate these variables to improve the goodness-of-fit between strain-stress curves predicted by the model and those measured. Attempting, however, to find through numerical optimization a unique set of values that reproduce the measurements is impossible because of intrinsic grain size and texture variability of the metal. Also, performing Hopkinson bar tests at various regimes of temperature and strain rate exercise different-enough effects that it is unrealistic to envision that all could be represented by a unique set of calibration variables (θ; κ; γ; s_0; s_∞; y_0; y_∞).

Table 3. Definition of calibration variables of the PTW model for the Be metal.

Symbol	Description	Minimum	Maximum
θ	Initial strain hardening rate	0.009979	0.0480590
κ	Temperature dependence of thermal activation energy	0.013516	0.4901500
γ	Strain rate dependence of thermal activation energy	-22.15299	7.4708000
y_0	Minimum yield stress (at T = 0 °K)	0.001054	0.0021643
y_∞	Maximum yield stress (at T ≈ melting)	0.000194	0.0016100
s_0	Minimum saturation stress (at T = 0 °K)	0.002493	0.0480680
s_∞	Maximum saturation stress (at T ≈ melting)	0.000599	0.0080031

Instead, our procedure searches for the joint probability distribution of calibration variables (θ; κ; γ; s_0; s_∞; y_0; y_∞) such that model predictions are statistically consistent with measurements over the two-dimensional domain of applicability $\{\Omega_v\}$. It is performed with the GPM/SA software that also infers from the comparison between predictions and measurements a statistical model of the discrepancy term $\delta(p)$ of equation (3). GPM/SA explores the joint probability distribution with a Markov-chain random walk that is based on a simple but effective principle: predictions that better match the measurements originate from combinations of calibration variables that tend to be visited more frequently by the random walk. After performing a sufficient number of iterations, selected to be 10,000 here, the statistics of calibration variables visited are computed to estimate the (unknown) joint probability distribution that represents our modeling uncertainty.

3.3 Definition of the Analysis Performed

The Markov-chain exploration of the posterior probability distribution of calibration variables (θ; κ; γ; s_0; s_∞; y_0; y_∞), and estimation of the discrepancy term $\delta(p)$, is repeated for each variant of the PTW model (see previous Table 1) and different combinations of physical measurements. The combinations of physical tests used in each case are defined in Table 4.

Table 4. Definition of the nine sets of Hopkinson bar experiments analyzed.

Case	List of Experiments	Case	List of Experiments
1	1, 2	6	1, 2, 3, 4, 5, 6, 7
2	1, 2, 3	7	1, 2, 3, 4, 5, 6, 7 and 10
3	1, 2, 3, 4	8	1, 2, 3, 4, 5, 6, 7 and 10, 11
4	1, 2, 3, 4, 5	9	1, 2, 3, 4, 5, 6, 7 and 10, 11, 12
5	1, 2, 3, 4, 5, 6		

Nine separate cases are defined, as opposed to a single analysis with twelve experiments, to assess the effect that increasing the number of physical tests has on predictive maturity. With the analysis of three variants of the PTW model, the effect of implementing "better" physics is also studied in section 4. Our hypothesis is that the PMI metric of Reference [1] will be sensitive to both increasing the number of physical tests available for analysis and improving the model.

4. Assessment of Predictive Maturity of the PTW Model and Variants

Figure 4 illustrates the location of physical experiments in the 2D domain (T; dε/dt). Blue, square symbols represent the Hopkinson bar experiments defined in Table 2. This domain is a two-dimensional hyper-cube defined as $\{\Omega_V\} = [10^{-4}; 4.10^{+3}] \times [70; 600]$ °K.sec.$^{-1}$. The convex hull $\{\Omega_{CH}\}$ of all twelve experiments is shown with a red, dashed line.

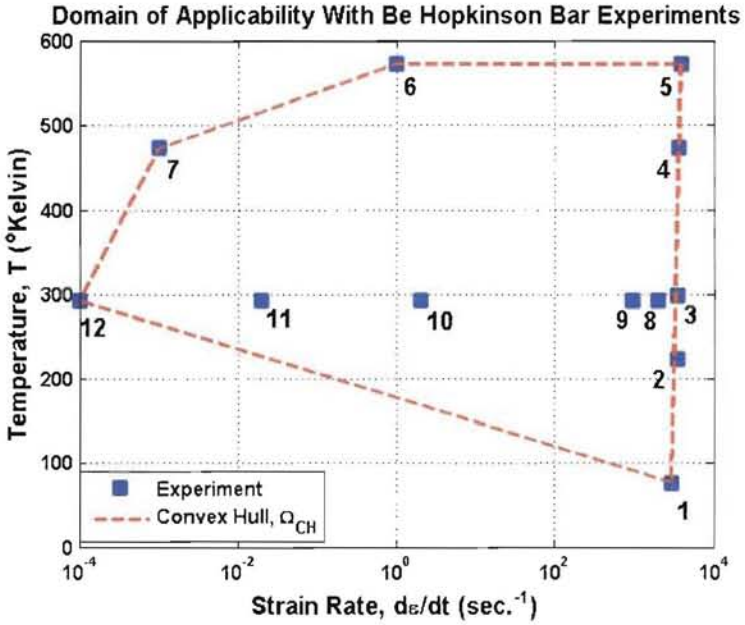

Figure 4. Domain of applicability and coverage of Hopkinson bar experiments.

Results are presented by, first, briefly discussing the analysis of the original PTW model, labeled "Variant-0" in Table 1. Results obtained with the other two variants are discussed next. The discussion concludes with an illustration of the robustness of predictive maturity to unknown modeling choices, such as those of the PTW model or coefficients (γ_1; γ_2; γ_3) of the PMI metric.

4.1 Analysis of Predictive Maturity of the PTW Model

Results shown in this section have been published in Reference [1]. They are summarized here for completeness. The values of coverage η_C, number of calibration variables N_K, scaled discrepancy δ_S and PMI are listed in Table 5 for each one of the nine cases defined in Table 4. The PMI is computed from equation (6) with $\gamma_1 = ½$, $\gamma_2 = ¼$, $\gamma_3 = 2$ and $N_R = 5$. A reference of five calibration variables, that is, $N_R = 5$, is used because this number is typical of material models in solid mechanics such as, for example, Johnson-Cooke or Zerilli-Amstrong.

It can be seen in Table 5 that coverage increases from 3.11% to nearly 75% of the domain of applicability. Discrepancy starts at 7.11% and decreases to 0.63% when test 5 is added to the analysis. Tests 5 and 6 are experiments performed at high temperatures; they are essential to the predictability of the PTW model, which explains the improvement of accuracy obtained when they are included in the analysis. It then becomes increasingly more difficult to match the

variety of strain-stress curves obtained when new experiments are added at low strain rates. It results that discrepancy increases from 0.63% (case 4) to nearly 17% (case 9). With PMI values above 70%, it is nevertheless assessed that maturity of the PTW model is acceptable.

Table 5. Predictive maturity for the nine sets of experiments (PTW model of Be metal).

Case	Coverage, η_C (%)	Number of Knobs, N_K	Discrepancy, δ_S (%)	PMI Metric (%)
1	3.11%	7	7.11%	6.96%
2	4.62%	7	7.21%	10.32%
3	8.17%	7	1.55%	18.44%
4	9.68%	7	0.63%	21.80%
5	34.16%	7	9.72%	60.83%
6	55.55%	7	11.82%	72.36%
7	55.55%	7	11.23%	72.58%
8	62.74%	7	12.01%	73.12%
9	74.25%	7	16.85%	71.24%

Table 5 also conveys the "stabilization" of the PMI metric as the level of coverage provided by physical experiments is increased. (This can also be observed in Figure 5 where PMI values for the "Variant-0" model are identical to those of Table 5.) The fact that the predictive maturity metric stabilizes after enough physical tests have been provided to the analysis illustrates our hypothesis that ***"maturity can be reached if enough experiments are analyzed."***

4.2 Effect on Predictive Maturity of Varying the Degree of Fidelity of the PTW Model

The analysis summarized above is repeated with the "Variant-1" and "Variant-2" versions of the PTW equation, as defined in Table 1. ("Variant-0" is the original PTW model of section 4.1.) Results are summarized in Figure 5 that compares PMI values obtained for these three variants.

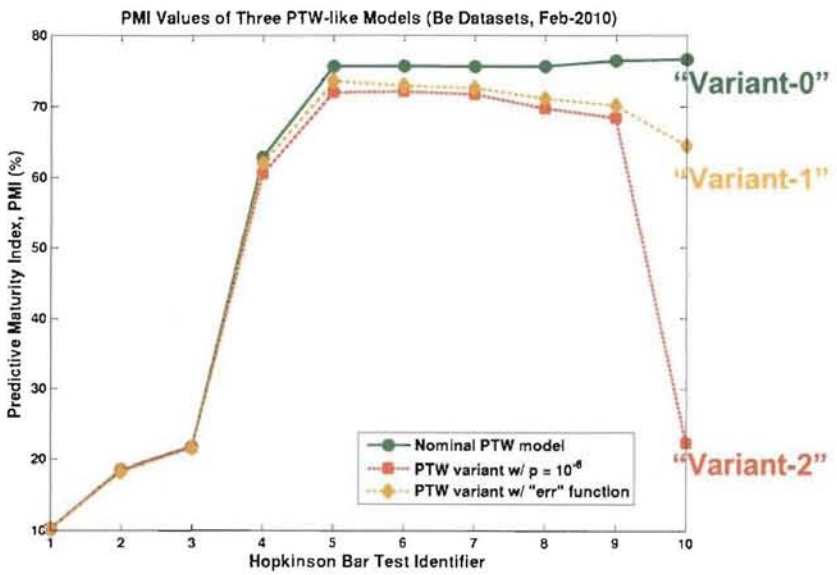

Figure 5. Predictive maturity values for three variants of the PTW model.

Results obtained with the "Variant-1" model are shown with an orange, dashed line and diamond symbols. Results of "Variant-2" are shown with a red, dashed line and square symbols.

Figure 5 illustrates a degradation of PMI values as one progresses from the nominal model to "Variant-1" and "Variant-2." This observation is consistent with expectation because "Variant-1" implements an exponential function in lieu of the "erf" function of equation (6). This perturbation of the original PTW equation is thought to be somewhat less severe than the implementation of "Variant-2," where stress hardening is completely turned off.

This application, while it does not constitute a formal proof, confirms our hypothesis that the PMI metric of Reference [1] is capable of tracking progress as "better" physics become available in a model or code. These results open the door to designing physical experiments such that a minimal resource can be spent on collecting measurements while guaranteeing a given level of predictive maturity. The availability of a quantitative metric of predictive maturity can also guide studies where the benefits of improving the fidelity of a phenomenon are traded against those of performing more physical experiments. Figure 5, for example, clearly illustrates that the benefits of adding "better" physics to progress from "Variant-1" to "Variant-0" are small relative to the benefits of including in the analysis the high-temperature Hopkinson bar tests 5 and 6.

4.3 Illustration of the Robustness of Predictive Maturity

In this last section, we attempt to address one criticism often encountered when attempting to define metrics for model validation or, in our case, predictive maturity. The concern is that the definition of a metric usually relies on arbitrary coefficients. One can, appropriately-so, question whether the value of the metric is sensitive to these arbitrary choices. The PMI of equation (5), for example, introduces a triplet of weighting factors (γ_1; γ_2; γ_3) to account for the contributions of discrepancy, coverage and complexity. These arbitrary coefficients can be seen as strength for their ability to add flexibility and account for expert judgment. But they constitute a weakness if PMI values happen to be overly sensitive to the choice of arbitrary coefficients (γ_1; γ_2; γ_3).

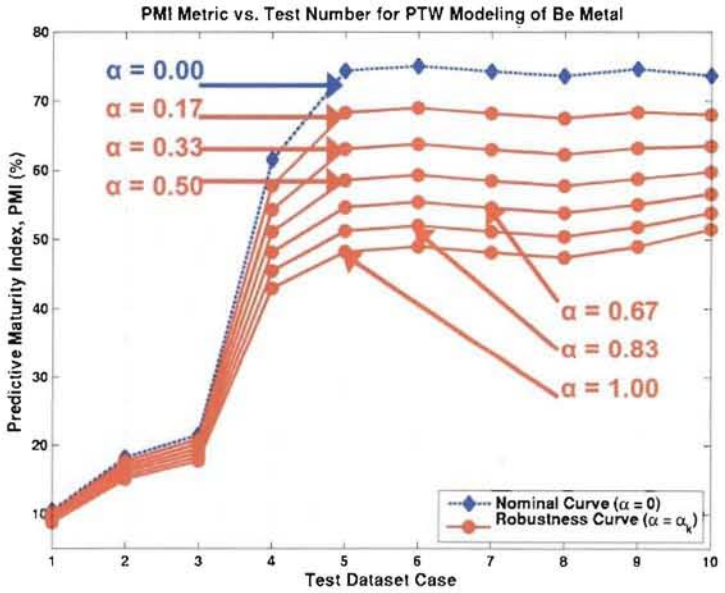

Figure 6. Robustness of the PMI with respect to the choice of coefficients (γ_1; γ_2; γ_3).

Our contention is that the question is ***not*** to justify the choice of these arbitrary coefficients because their introduction in the definition of a metric is somewhat unavoidable. Instead, we can make the question go away by demonstrating that it is irrelevant since values of the PMI metric are somewhat robust, or insensitive, to these arbitrary choices.

To do so, a robustness analysis is performed for the "Variant-0," or original, PTW equation. The analysis consists of varying the coefficients (γ_1; γ_2; γ_3) and searching for the worst-possible PMI, or smallest value, by solving an optimization problem. Gamma coefficients γ_k are varied up to a ***"horizon-of-uncertainty,"*** denoted by the symbol α, that defines how far away from their nominal values $\gamma_k^{(0)}$ one is willing to vary the arbitrary coefficients (γ_1; γ_2; γ_3):

$$(1-\alpha) \le \frac{\gamma_k}{\gamma_k^{(0)}} \le (1+\alpha), \tag{7}$$

where nominal values are $\gamma_1^{(0)} = \frac{1}{2}$, $\gamma_2^{(0)} = \frac{1}{4}$ and $\gamma_3^{(0)} = 2$, as used previously. The horizon-of-uncertainty parameter α is then increased progressively to examine by how much the smallest PMI value changes as one ventures further away from the nominal triplet ($\gamma_1^{(0)}$; $\gamma_2^{(0)}$; $\gamma_3^{(0)}$). Note that this procedure involves solving a constrained optimization problem, where constraints are defined by three instances of equation (7), one for each coefficient γ_k. Reference [15] discusses the theory and application of information-gap robustness that has inspired the present analysis.

Figure 6 compares the nominal PMI curve, obtained with triplet ($\gamma_1 = \frac{1}{2}$; $\gamma_2 = \frac{1}{4}$; $\gamma_3 = 2$), to worst-case PMI curves obtained when coefficients (γ_1; γ_2; γ_3) are allowed to vary up to the level of uncertainty α. The analysis is performed at seven discrete levels α = 0, 0.17, 0.33, 0.50, 0.67, 0.83 and 1. Even though a deterioration of maturity is clearly visible, the overall trend remains unchanged as more physical experiments are provided to the analysis. This result translates the fact that the PMI metric is robust to the choice of arbitrary coefficients γ_k. A practical implication is that one should not worry too much about the choice of coefficients γ_k: the trends identified by the PMI metric remain unchanged no matter which values of coefficients (γ_1; γ_2; γ_3) are used.

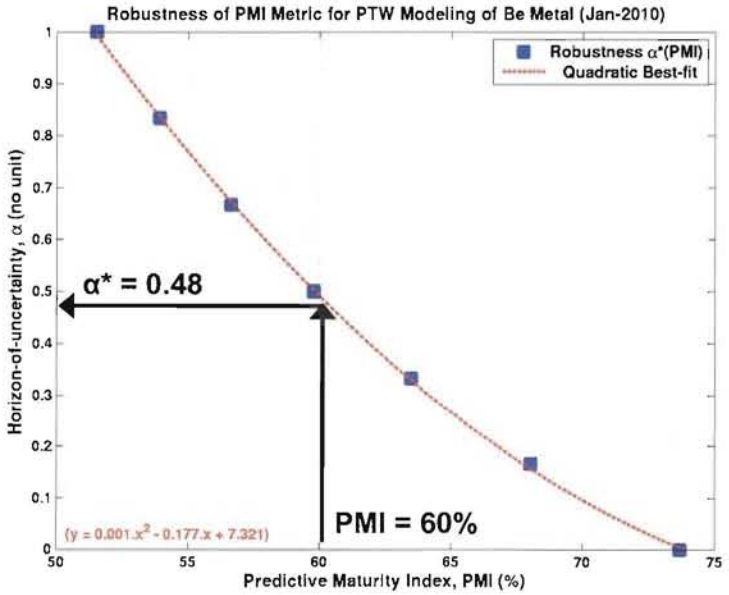

Figure 7. Robustness of the Case-9 PMI metric as a function of α-parameter.

Figure 7 shows the worst-case PMI metric as a function of increasing horizon-of-uncertainty parameter α. This result is produced for case 10 where the ten tests 1, 2, 3, 4, 5, 6, 7, 10, 11 and 12 are provided to the test-analysis correlation and calibration procedure. The figure shows that the analysis can be performed at a small number of discrete α-values. A simple polynomial curve-fit can then be performed to estimate the PMI metric at other (non-analyzed) α-values.

Figure 7 also illustrates how to interpret the analysis of robustness if the objective is, for example, to guarantee a minimum level of predictive maturity equal to PMI = 60%. The figure shows that 60% maturity corresponds to a maximum horizon-of-uncertainty equal to 0.48. The implication is that, for the dataset of case 10 considered, our analysis of predictive maturity of the PTW model has a level of robustness of $\alpha^* = 0.48$. Stated simply, it means that the PMI is ***guaranteed to be, at least, equal to 60%*** as long as the analysis is performed with a triplet of coefficients (γ_1; γ_2; γ_3) that does not deviate from nominal values $\gamma_k^{(0)}$ by more than:

$$\underbrace{(1-\alpha^*)}_{=0.52}\gamma_k^{(0)} \le \gamma_k \le \underbrace{(1+\alpha^*)}_{=1.48}\gamma_k^{(0)}. \tag{8}$$

Note that, in the particular case of the PMI metric, an analytical expression is available from equation (5) that can be used to obtain sensitivities ($\partial PMI/\partial\gamma_k$). These sensitivities can be used, in a crude sense, to assess the robustness of PMI values to coefficients γ_k. Even though it offers another possibility, the robustness analysis is preferred because it is applicable to all situations, especially, those where obtaining a closed-form solution is not an option.

5. Conclusion

In this publication, we pursue the development, started in Reference [1], of a quantitative metric to assess the predictive maturity of a model or numerical simulation. The metric proposed is based on a statistical discrepancy term that quantifies the systematic disagreement, or bias, between measurements and predictions. It also accounts for coverage, or the degree to which physical experiments cover the domain of applicability of the model or code. A third attribute of the metric is the level of complexity of the model analyzed.

Our hypothesis is that improving the predictive capability of a model should translate into better agreement between measurements and predictions. This agreement, in turn, should lead to a smaller discrepancy term. The hypothesis is illustrated with the non-linear Preston-Tonks-Wallace strength model applied to Hopkinson bar experiments performed on Beryllium metal. It is shown that predictive maturity improves when additional physical tests are made available to increase coverage of the domain of applicability. It is also shown that the maturity metric tracks progress as "better" physics are implemented in the model. Finally, robustness of the metric with respect to the choice of arbitrary coefficients needed in its definition is demonstrated.

These results, while preliminary, open the door to designing physical experiments such that a minimal resource can be spent on collecting measurements while guaranteeing a given level of predictive maturity. The availability of a quantitative metric of maturity can also guide studies where the benefits of improving the fidelity of a phenomenon, for example, by implementing a "better" model, are traded against those of performing more physical experiments.

Acknowledgements

This work is performed under the auspices of the fuels modeling program element of the Nuclear Energy Advanced Modeling and Simulation (NEAMS) program at Los Alamos National Laboratory (LANL). The first two authors are grateful to Cetin Unal, NEAMS program manager, for his support and technical leadership. Support with the statistical inference software GPM/SA provided by Brian Williams is acknowledged and greatly appreciated. LANL is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396.

References

1. Hemez, F.M., Atamturktur, S.H., "Prediction with Quantified Uncertainty of Temperature and Rate Dependent Material Behavior," *11th AIAA Non-Deterministic Approaches Conference*, Palm Springs, California, May 4-7, 2009.
2. Zang, T.A., "Perspectives on Uncertainties (and Margins) in NASA Engineering Decisions," *10th AIAA Non-Deterministic Approaches Conference*, Schaumburg, Illinois, April 7-10, 2008.
3. NASA, **Standard for Models and Simulation**, *Technical Standard NASA-STD-7009*, National Aeronautics and Space Administration, Washington, D.C., November 2007.
4. Sornette, D., Davis, A.B., Kamm, J.R., Ide, K., "A General Strategy for Physics-based Model Validation Illustrated with Earthquake Phenomenology, Atmospheric Radiative Transfer and Computational Fluid Dynamics," *Proceedings of the Lawrence Livermore National Laboratory Workshop on Computational Methods in Radiation and Particle Transport*, Lake Tahoe, California, Springer-Verlag Editors, September 9-14, 2006.
5. Sornette, D., Davis, A.B., Ide, K., Kamm, J.M., "Theory and Examples of a New Approach to Constructive Model Validation," *NATO RTA AVT-147 Symposium on Computational Uncertainty*, Athens, Greece, December 3-6, 2007. (Los Alamos Report LA-UR-07-7013.)
6. Roache, P.J., **Verification in Computational Science and Engineering**, Hermosa Publishers, Albuquerque, New Mexico, 1998.
7. Hemez, F.M., Doebling, S.W., Anderson, M.C., "A Brief Tutorial on Verification and Validation," *22nd SEM International Modal Analysis Conference*, Dearborn, Michigan, January 26-29, 2004.
8. ASME, **Guide for Verification and Validation in Computational Solid Mechanics**, *Publication V&V-10-2006*, American Society of Mechanical Engineers, 2006.
9. Hemez, F.M., "Answering the Question of Sufficiency: How Much Uncertainty is Enough?," *1st International Conference on Uncertainty in Structural Dynamics*, University of Sheffield, United Kingdom, June 11-13, 2007. (Los Alamos Report LA-UR-07-3575.)
10. 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, March 2004, pp. 1443-1467. (Los Alamos Report LA-UR-02-3538.)
11. Kennedy, M., O'Hagan, A., "Predicting the Output from a Complex Computer Code When Fast Approximations are Available," *Biometrika*, Vol. 87, 2000, pp. 1-13.
12. Higdon, D., Gattiker, J., Williams, B., Rightley, M., "Computer Model Calibration Using High-Dimensional Output," *Journal of the American Statistical Association*, Vol. 103, No. 482, June 2008, pp. 570-583.
13. Higdon, D., Nakhleh, C., Gattiker, J., Williams, B., "A Bayesian Calibration Approach to the Thermal Problem," *Computational Methods in Applied Mechanics*, Vol. 197, 2008, pp. 2431-2441.
14. Preston, D.L., Tonks, D.L., Wallace, D.C., "Model of Plastic Deformation for Extreme Loading Conditions," *Journal of Applied Physics*, Vol. 93, No. 1, January 2003, pp. 221-220.
15. Ben-Haim, Y., **Information-gap Decision Theory**, Academic Press, Elsevier Publishers, 2nd edition, 2006.