6-2012

Resource Allocation Framework: Validation of Numerical Models of Complex Engineering Systems against Physical Experiments

Joshua Hegenderfer
Clemson University, jhegend@yahoo.com

RESOURCE ALLOCATION FRAMEWORK: VALIDATION OF NUMERICAL MODELS OF COMPLEX ENGINEERING SYSTEMS AGAINST PHYSICAL EXPERIMENTS

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Civil Engineering

---

by
Joshua L. Hegenderfer
June 2012

Accepted by:
Dr. Sezer Atamturktur, Committee Chair
Dr. C. Hsein Juang
Dr. Abdul A. Khan
Dr. Lonny Thompson

ABSTRACT

An increasing reliance on complex numerical simulations for high consequence decision making is the motivation for experiment-based validation and uncertainty quantification to assess, and when needed, to improve the predictive capabilities of numerical models. Uncertainties and biases in model predictions can be reduced by taking two distinct actions: (i) *increasing the number of experiments in the model calibration process*, and/or (ii) *improving the physics sophistication of the numerical model*. Therefore, decision makers must select between further code development and experimentation while allocating the finite amount of available resources. This dissertation presents a novel framework to assist in this selection between experimentation and code development for model validation strictly from the perspective of predictive capability. The reduction and convergence of *discrepancy bias* between model prediction and observation, computed using a suitable convergence metric, play a key role in the conceptual formulation of the framework. The proposed framework is demonstrated using two non-trivial case study applications on the Preston-Tonks-Wallace (PTW) code, which is a continuum-based plasticity approach to modeling metals, and the ViscoPlastic Self-Consistent (VPSC) code which is a mesoscopic plasticity approach to modeling crystalline materials. Results show that the developed resource allocation framework is effective and efficient in path selection (i.e. experimentation and/or code development) resulting in a reduction in both model uncertainties and discrepancy bias.

The framework developed herein goes beyond path selection in the validation of numerical models by providing a methodology for the prioritization of optimal experimental settings and an algorithm for prioritization of code development.

If the path selection algorithm selects the experimental path, optimal selection of the settings at which these physical experiments are conducted as well as the sequence of these experiments is vital to maximize the gain in predictive capability of a model. The Batch Sequential Design (BSD) is a methodology utilized in this work to achieve the goal of selecting the optimal experimental settings. A new BSD selection criterion, *Coverage Augmented Expected Improvement for Predictive Stability* (C-EIPS), is developed to minimize the maximum reduction in the model discrepancy bias and coverage of the experiments within the domain of applicability. The functional form of the new criterion, C-EIPS, is demonstrated to outperform its predecessor, the EIPS criterion, and the distance-based criterion when discrepancy bias is high and coverage is low, while exhibiting a comparable performance to the distance-based criterion in efficiently maximizing the predictive capability of the VPSC model as discrepancy decreases and coverage increases.

If the path selection algorithm selects the code development path, the developed framework provides an algorithm for the prioritization of code development efforts. In coupled systems, the predictive accuracy of the simulation hinges on the accuracy of individual constituent models. Potential improvement in the predictive accuracy of the simulation that can be gained through improving a constituent model depends not only on the relative importance, but also on the inherent uncertainty and inaccuracy of that

particular constituent. As such, a unique and quantitative code prioritization index (CPI) is proposed to accomplish the task of prioritizing code development efforts, and its application is demonstrated on a case study of a steel frame with semi-rigid connections. Findings show that the CPI is effective in identifying the most critical constituent of the coupled system, whose improvement leads to the highest overall enhancement of the predictive capability of the coupled model.

ACKNOWLEDGMENTS

My journey to this point in my academic career has been by no means a solo one. Only through the encouragement, guidance, and assistance provided by committee members, instructors, colleagues, and fellow students have I been able to conduct the research that is assembled in this dissertation proposal. Many thanks go to the United State Department of Energy: Nuclear Energy University Programs (Project: Predictive Maturity of Multi-Scale Simulation Models for Fuel Performance; Contract Number: 00101999) for providing the funding for this project. I would like to thank my committee chair, Dr. Atamturktur, for the countless hours she has spent training, guiding, and teaching me how to be a high quality researcher. I would also like to thank Dr. C. Hsein Juang, Dr. Abdul A. Khan, and Dr. Lonny Thompson for their willingness to mentor and serve on my committee throughout the remainder of my time here at Clemson. Many thanks go to my mentors from Los Alamos National Laboratories: Dr. Brian Williams, Dr. Francois Hemez, Dr. Ricardo Lebensohn, and Dr. Cetin Unal for going out of their way to corroborate with me during my education. Special thanks also go to Dr. Karma Yonten for his continued assistance in coding efforts. Finally, I would like to thank Daniel Metz for constructing our steel models and Austin Gillen and Andrew Lisicki for their assistance with the experimental campaigns.

TABLE OF CONTENTS

Table of Contents (Continued)

Page

LIST OF TABLES

Table Page

List of Tables (Continued)

Table Page

LIST OF FIGURES

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

CHAPTER 1: INTRODUCTION

1.1. Motivation

With advances in computer technology, the traditional design-build-test strategy for the certification of complex engineering systems is being replaced with physics-based complex numerical models. Such models can greatly reduce the reliance on and need for extensive experimental campaigns; however, one must bear in mind that these models are only approximations of reality and their predictions contain a level of inaccuracy and imprecision. Applying such models in high-consequence decision making, especially in scientific and engineering fields involving the safety and security of human life, requires a rigorous assessment of the deficiencies of the model through efforts collectively referred to as *experimental-based model validation*. Many questions arise during experimental-based model validation, including but not limited to: Are sufficient quality and quantity of experiments being collected? Are the engineering or science principles modeled with sufficient detail? With many vital questions to be answered, it is crucial to have a systematic, formal experimental-based model validation framework.

A framework for experimental-based model validation must not only apply rigorous methods for quantifying the model inaccuracies and imprecisions, but also provide decision makers with a tool for determining what actions to take when the model accuracy or precision is found insufficient. Two obvious paths available to the decision

maker are: (i) conducting additional experiments and (ii) improving the model (or code[1]) itself; while the selection of one path over the other is less obvious. Faced with time, budget, and model performance requirements, decision makers need tools to choose the most efficient path toward model validation. Thus, a model validation framework incorporating such decision making tools is essential for the science and engineering community. The focus of this study is to provide such a framework.

1.2. Background Overview

Over the past several decades, many studies have been devoted to the verification and validation of numerical models. To gain a full understanding of these previous research efforts in the pertinent literature, a consistent definition of the terms *validation* and *verification* is necessary. The Department of Defense (DoD) defines validation as "the process of determining the degree to which a model is an accurate representation of the real-world from the perspective of the intended uses of the model" while verification is defined as "the process of determining that a model implementation accurately represents the developer's conceptual description and specifications" (DoD, 2007). Balci et al. (1998) provides a similar definition where validation "is substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with modeling and simulation objectives" and verification is "substantiating that the model is transformed from one form into another, as intended, with sufficient accuracy."

[1] Herein *codes* are defined as the mathematical algorithms developed to compute physical quantities and phenomena. *Models* incorporate codes for particular applications (material, geometry, etc.). When referring to improving the physics sophistication of the numerical model, the terms *model* and *code* will be used interchangeably as model development and code development can be used as a means of physics sophistication.

Balci summarizes these terms in stating: "Model verification deals with building the model *right,* and model validation deals with building the *right* model" (Balci et al., 1998).

It is also necessary to accompany definitions of validation and verification with a formal definition for the *accuracy* and *precision*. The *accuracy* of a model represents the fundamental inability to reproduce reality due to missing physics principles, idealizations, or unsuitable assumptions. A direct comparison of model predictions to experimental results can quantify these inaccuracies and is defined herein as *model form error* (or *systematic bias)*. If one can conduct an infinite number of experiments, the model form error would be known throughout the entire domain of applicability; however, given resource limitations, an independent estimate of model form error throughout the domain of applicability, referred to herein as the *discrepancy bias*, must be evaluated.

The assessment of predictive ability must go beyond deterministic approaches that measure accuracy but fail to account for imprecisions of the model predictions. In many science and engineering problems, the imprecisions stem from aspects of the model that are not known with absolute certainty including, but not limited to, material property variability, geometric or material property parameter values, statistical limitations in estimating parameter values, predictions related to time (past or future), and human error (Bulleit, 2008). Such model imprecisions are herein referred to as model uncertainties. *Uncertainty quantification* is the process by which the effect that these imprecisions have on model predictions is quantified (Hemez, 2007). In fact, the National Aeronautics and Space Administration (NASA), in response to a lack of uncertainty quantification and

risk assessment in the Columbia accident in 2003, created a model validation and verification framework that includes rigorous uncertainty quantification in the development of the Uncertainty Structure (Green, 2008).

Many such frameworks have been developed in academic institutions (Leye et al., 2009; Jung, 2011; Sargent, 2010; Bayarri et al., 2007; Sornette et al., 2007; Jiang and Mahadevan, 2007), industry (Balci et al., 2002), National Laboratories (Oberkampf et al., 2007; Unal et al., 2011), and NASA (Green et al., 2008). Frameworks such as the Uncertainty Structure developed at NASA (Green et al., 2008) or the Predictive Capability Maturity Model from Sandia National Laboratories (Oberkampf et al., 2007) seek to assess the overall predictive capability of a modeling and simulation campaign through the use of qualitative, expert-judgment based ranking systems that include assessment of model verification, validation, and uncertainty quantification. Balci et al. (1998) has developed an online credibility assessment tool to improve the communication and overall effectiveness of such qualitative approaches. General high-level approaches to assessing model validity against experimental data (see for instance Leye et al., 2009 and Sargent, 2010) include sensitivity analysis, experimentation, and a qualitative or quantitative comparison of model predictions to experimental observations; however, in these mainly qualitative high-level approaches, the rigorous, quantitative assessment metrics for model validation and uncertainty quantification are not explicitly provided.

To provide such rigorous and quantitative assessments, Jung (2011), Bayarri et al. (2007), Jiang and Mahadevan (2007), and Unal et al. (2011) provide model validation frameworks with similar general structures including: initial modeling efforts,

experimentation, model calibration, assessment against an evaluation criterion, and a decision of whether or not the model is fit for use in its intended domain of applicability.

Model calibration is the process by which uncertainties in model input parameters can be reduced by inferring improved estimates of these input parameters through a comparison of model predictions to experimental data. Model calibration can be conducted in such a way that model uncertainties can be integrally quantified through the use of Bayesian calibration techniques (Kennedy and O'Hagan, 2000) as implemented in Unal et al. (2011) and Bayarri et al. (2007) or the intrinsically random maximum likelihood principal as employed in Jung (2011). Additionally, within the frameworks provided in Unal et al. (2011) and Bayarri et al. (2007), the Bayesian calibration approach can account not only for model uncertainties but can also integrally account for both experimental uncertainties and model accuracy in the training of a discrepancy bias surrogate model. This treatment allows for a more rigorous and systematic separation and quantification of model accuracy and precision.

During model calibration, *Evaluation and Assessment Criterion* (also known as validation metrics) can take on many forms depending on the goals and needs of the project. For example, Jung (2011) directly compares calibrated model output to experimental measurements using statistical approaches such as confidence intervals and hypothesis testing. In Bayarri et al. (2007) discrepancy bias-corrected model predictions under uncertainty are compared to experimental measurements also accounting for experimental uncertainty. A forecasting metric developed in Atamturktur et al. (2011) seeks to assess not only the model form error at known experimental locations but also

assesses the disagreement between the estimated discrepancy bias at untested settings and the true model form error. Moving beyond validation metrics solely based on model accuracy, the validation metric used in Sornette et al. (2007) combines a measure of uncertain model predictions to experimental measurements in addition to a novelty term incorporating how well experimentation has covered the domain of applicability. Another metric, the Predictive Maturity Index (PMI), quantitatively combines model accuracy, coverage of the domain of applicability by experimentation, model robustness, and model complexity to provide a comprehensive validation metric for use in the model validation process (Hemez et al., 2010; Stull et al., 2011).

1.3. Dissertation overview

Frameworks mentioned in the previous section include a step for *the decision of whether or not the model is fit for use in its intended domain of applicability.* Although these earlier frameworks acknowledge the paths to improvement, i.e. (i) collecting additional experimental data and/or (ii) improving the mathematical description of underlying physics principles of the model, none provide a quantitative strategy for choosing the appropriate path. This dissertation addresses this need for such a decision making algorithm.

Furthermore, if the experimentation path is selected, the validation experiments must be conducted at optimal settings. The study of optimal experimental design provides metrics and algorithms suitable for making such selections. One such experimental design methodology is Batch Sequential Design (BSD) (Williams et al., 2011; Atamturktur et al., 2012). BSD optimizes the experimental settings for a budget-driven,

user-chosen batch size and number of batches. This approach operates on the trained discrepancy bias model and uses a modified Federov exchange algorithm to exchange candidate design settings for optimal settings that maximize/minimize the design criterion. In published literature, design criterion utilized in BSD operate directly on a measure of the model accuracy and neglect additional aspects important to model validation such as the overall coverage of the domain of applicability by experiments, an attribute emphasized in the validation metric PMI by Hemez et al. 2011. Therefore, this dissertation develops a tool for prioritization of experiments taking into account not only the effects of discrepancy bias but also coverage of the domain of applicability.

In a similar fashion, if the physics sophistication path is chosen, the model constituents of a coupled numerical system that most effectively reduce model form error should be chosen for improvement. Qualitative ranking systems such as the Phenomenon Identification and Ranking Table (Boyack, 1989) used in the nuclear engineering field and quantitative approaches to optimize software fault detection (Korel, 2009) and parallel processing of finite element codes (Zeyao and Lianxiang, 2004) have been developed in their respective fields. However, a quantitative ranking system for selecting the optimal constituent for code development has not been addressed in the literature. In response to this need, this dissertation develops a code prioritization metric for ranking constituents of models based upon an analysis of each constituent's importance, uncertainty, and estimated initial inaccuracy.

In summary, the goal of this dissertation is to utilize and expand on previously developed model validation frameworks to provide a more comprehensive framework

and its associated validation metrics that incorporate and directly addresses model uncertainties, experimental uncertainties, and model inaccuracies. This validation framework also provides a path selection algorithm for choosing either an augmented experimental campaign or code development to improve the modeling and simulation campaign, an algorithm for selecting optimal experimental design settings, and a metric for determining the most efficient code prioritization efforts.

1.4. Main Dissertation Contributions

This dissertation contributes to the knowledge in the model validation field in *three* distinct ways. First, the main contribution of this dissertation is to extend the validation framework proposed by Unal et al. (2011) for the licensing of nuclear reactors into a generally applicable validation framework that can be used for certification/licensing in many technology areas including defense, energy and environment, risk assessment, national security, and basic and applied engineering and science projects. The proposed framework includes a path selection algorithm for choosing whether to conduct additional experiments or improve the physics of the model that is not provided previously in the literature. This algorithm uses convergence rate metrics applied to the model accuracy (measured through the discrepancy bias) and model uncertainty (measured through an entropy-based information gain metric) to select the most effective path and is demonstrated in three case studies: a conceptual mathematical example, the Preston-Tonks-Wallace (PTW) material model for plasticity of metals, and the viscoplastic self-consistent (VPSC) material model.

Second, this dissertation inserts BSD into the developed model validation framework to be used when the experimentation path is chosen. Furthermore, a new optimization criterion, C-EIPS, is developed that seeks to select experimental settings based not only on the discrepancy bias but also on the coverage of the domain of applicability. The C-EIPS combines a measure of the expected improvement in the trained discrepancy bias with a computation of the overall coverage of the domain of applicability; then, the BSD algorithm optimizes C-EIPS for potential candidate points for the next batch of experimental settings in the domain. This new criterion (implemented in its functional form) is shown to more efficiently improve PMI when compared to its predecessor.

Finally, to address the need for resource allocation in code development, this dissertation develops a novel metric for the most effective ranking of model constituents for code development of coupled numerical systems and incorporates the metric in the developed model validation framework. This metric, the Code Prioritization Index (CPI), quantitatively assesses code constituents' importance, uncertainty, and estimated model form error and subsequently ranks constituents accordingly. The importance and uncertainty components of the CPI metric can be assessed using sensitivity analysis and uncertainty assessment techniques available in literature; however, the analysis of estimated model form error (referred to as error analysis (EA) herein) of CPI is specifically developed in this dissertation to assess the estimated initial inaccuracy of model constituents of coupled systems. The CPI metric is demonstrated using a steel frame with semi-rigid connections in which model constituents are identified as the semi-

rigid beam-to-column connections. Model predictions are compared against experimentally collected static deflections and natural frequencies. The CPI metric requires obtaining model predictions with perturbed input parameters, and as such, as input parameters are perturbed the order of simulated vibratory modes of the steel frame changes (also known as mode-swapping). Therefore, a mode-pairing algorithm is developed to automate the process of matching mode shapes with corresponding natural frequencies. Each constituent's sensitivity, uncertainty, and error analysis are then evaluated, and the CPI is computed. The CPI chosen ranking of model constituents is compared against all other possible combinations of code development and shown to provide the most efficient reduction in model form error. These three major contributions are summarized and depicted in the model validation flow chart (Figure 1.1). Through the development and application of each contribution on real-life, non-trivial case studies, the following chapters demonstrate the unique contributions of this dissertation to the model validation community.

1.5. Dissertation Organization

This dissertation[2] begins with a description of the model validation framework and the incorporation of the path selection algorithm. Chapter 2 concludes by demonstrating the path selection algorithm for a case where experimentation and code development are considered equally demanding on resources on both a conceptual mathematical model and the PTW model for plastic deformation of metals. This chapter

[2] Chapters 2-5 in this dissertation serve as stand-alone journal publications, and as such, some level of conceptual overlap is unavoidable.

is currently under review for publication. Chapter 3 then extends the use of the framework to a case where experimentation is considered much less demanding on resources than code development. This chapter also analyzes the convergence of both model form error at known experimental settings and the information gain metric as demonstrated on the VPSC model. The content of Chapter 3 is currently under review. Next, the optimization of experimental settings is incorporated into the model validation framework through the use of the BSD algorithm. Chapter 4 also includes a full description of the PMI metric and the demonstration of the effectiveness of the newly developed BSD criterion, C-EIPS, on the VPSC model. Chapter 4 is currently under review as a journal publication. The final integral component of the validation framework, the CPI metric, is derived in Chapter 5. This chapter discusses the foundations of the CPI metric and coupled numerical models and demonstrates the developed CPI metric on a steel frame with semi-rigid connections. This chapter has been submitted for publication and is currently under review. Finally, conclusions, limitations and scope of the current research, and necessary future work are presented in Chapter 6.

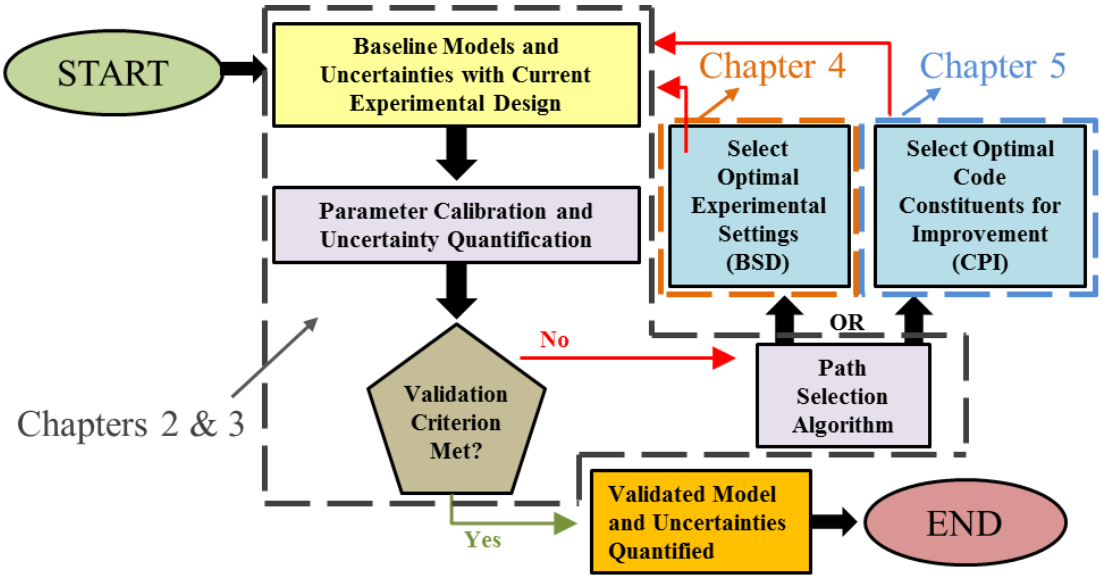

Figure 1.1: General Model Validation Framework

1.6. References

Atamturktur, S., Williams, B.J., Hamutcuoglu, M.O., & Unal, C. (2012), Batch sequential calibration for improved predictive maturity in physics-based modeling, *International Journal for Numerical Methods in Engineering*, (under review).

Atamturktur, S., Hemez, F., Williams, B., Tomé, C., & Unal, C. (2011), A forecasting metric for predictive modeling, *Computers and Structures*, 89, 2377-87.

Balci, O. (1998), Verification, validation, and accreditation, Proceedings of the 1998 Winter Simulation Conference. ed. D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, Piscataway, NJ: IEEE, 41-48.

Balci, O., Adams, R.J., Myers, D.S., & Nance, R.E. (2002), Credibility assessment: a collaborative evaluation for credibility assessment of modeling and simulation applications, In Proceedings of the 34th winter simulation conference: exploring new frontiers, San Diego, California, 214-20.

Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C.H., & Tu, J. (2007), A framework for validation of computer models, *Technometrics*, 49(2), 138-54.

Boyack, B.E., Duffey, R.B., & Griffith, P., et al. (1989), Quantifying reactor safety margins: application of code scaling, applicability, and uncertainty methodology to a large-break, loss-of-coolant accident. US NRC Reprt, NUREG/CR-5249.

Bulleit, W.M. (2008), Uncertainty in structural engineering, *Practice Periodical on Structural Design and Construction (ASCE)*, 13, 24-30.

Department of Defense (2007), Directive number 5000.59, August 8.

Green, L.L., Blattnig, S.R., Hemsch, M.J., Luckring, J.M., & Tripathi, R.K. (2008), An uncertainty structure matrix for models and simulations, American Institute of Aeronautics and Astronautics, AIAA-2008-2154.

Hemez, F., Atamturktur, S., & Unal, C. (2010), Defining predictive maturity for validated numerical simulations, *Computers and Structures*, 88, 497-505.

Hemez, F.M. (2007), A course on the validation of computational models, Los Alamos National Laboratories Report, LA-UR-07-8454.

Jiang, X. & Mahadevan, S. (2007), Bayesian risk-based decision method for model validation under uncertainty, *Reliability Engineering and System Safety*, 92, 707-18.

Jung, B. (2011), A hierarchical framework for statistical model validation of engineered systems, Dissertation: University of Maryland.

Kennedy, M. & O'Hagan, A. (2000), Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87, 1-13.

Korel, B. (2009), Experimental comparison of code-based and model-based test prioritization, IEEE International Conference on Software Testing Verification and Validation Workshops, Denver, CO, United States; April 1-4, 77-84.

Leye, S., Himmelspach, J., & Uhrmacher, A.M. (2009), A discussion on experimental model validation, UKSim 2009: 11th International Conference on Computer Modelling and Simulation, 161-7.

Oberkampf, W.L., Pilch, M., & Trucano, T.G. (2007), Predictive capability maturity model for computational modeling and simulation, Sandia National Laboratories Report, SAND2007-5948.

Sargent, R.G. (2010), Verification and validation of simulation models, Proceedings of the 2010 Winter Simulation Conference. ed. B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesann, 166-83.

Sornette, D., Davis, A.B., Ide, K., Vixie, K.R., Pisarenko, V., & Kamm, J.R. (2007), Algorithm for model validation: theory and applications, Proceedings of the National Academy of Sciences of the United States of America (PNAS), 104(16), 6562-7.

Stull, C.J., Hemez, F., Williams, B.J., Unal, C., & Rogers, M.L. (2011), An improved description of predictive maturity for verification and validation activities. Los Alamos National Laboratories Technical Report, LA-UR-11-05659.

Unal, C., Williams, B., Hemez, F., Atamturktur, S.H., & McClure, P. (2011), Improved best estimate plus uncertainty methodology, including advanced validation concepts, to license evolving nuclear reactors. *Nuclear Engineering and Design*, 241, 1813–1833.

Williams, B.J., Loeppky, J.K., Moore, L.M., & Macklem, M.S. (2011), Batch sequential design to achieve predictive maturity with calibrated computer models. *Reliability Engineering and System Safety*, 96, 1208-19.

Zeyao, M. & Lianxiang, F. (2004), Parallel Flux Sweep Algorithm for Neutron Transport on Unstructured Grid, *The Journal of Supercomputing*, 30, 5-17.

CHAPTER 2: PREDICTIVE CAPABILITY OF NUMERICAL MODELS: BENEFITS OF AUGMENTING EXPERIMENTAL CAMPAIGNS AND IMPROVING PHYSICS SOPHISTICATION

2.1. Introduction

In predictive modeling, the primary purpose of constructing a model is to reduce the number of experiments necessary to gain an understanding of the physical phenomenon of interest, while the purpose of experiments is to ensure the precision of the model parameters and accuracy of model form. As Kennedy and O'Hagan (2000) stresses, model predictions are plagued with systematic bias, referred to herein as model form error. The origin of model form error is the inevitable inexactness and incompleteness of the model caused by missing physics principles, overly crude idealizations, and unsuitable assumptions. Examples of such inexactness and incompleteness include oversimplification of lower length scale phenomena in a higher length scale simulation model, or approximate representation of first principle physics with a purely non-physical calibration parameter. Note that such causes of incomplete and inaccurate physics may also lead to missing input parameters. The inherent inaccuracy of a model could be reduced if a more complete physics[3] description to capture the true response of the phenomenon of interest is incorporated. An improved physics description would reduce the need for experiments to infer the model form error. On the other hand, improving physics sophistication may naturally translate to specifying

[3] Herein, the term physics is used to describe physical laws of the system in question, while the term physics sophistication indicates the extent to which those laws are known and incorporated in the model.

additional physics parameters, which in turn may lead to the need for an increase in experiments to learn about the additional uncertain and influential parameters. These parameters are herein referred to as calibration parameters. As seen, code development to improve physics sophistication (relates to accuracy) and experiments to reduce parametric uncertainty of a model (relates to precision) are intertwined, and must therefore be treated as such.

In a hypothetical scenario, let us assume one could conduct a sufficiently large number of experiments to learn about the imprecisely known, influential model parameters. Then, one could reasonably envision reducing the parametric uncertainty down to the level of uncontrolled, natural variability (i.e. irreducible or aleatory uncertainty). In this scenario, the disagreement between model predictions and experiments can be explained primarily by the systematic bias of the model (Draper, 1995). Therefore, as new experimental information becomes available, the disagreement between simulation predictions and physical measurements can be expected to converge to a consistent level of systematic bias (δ_C). This constitutes the first hypothesis, which is illustrated in Figure 2.1 for an averaged, univariate systematic bias.

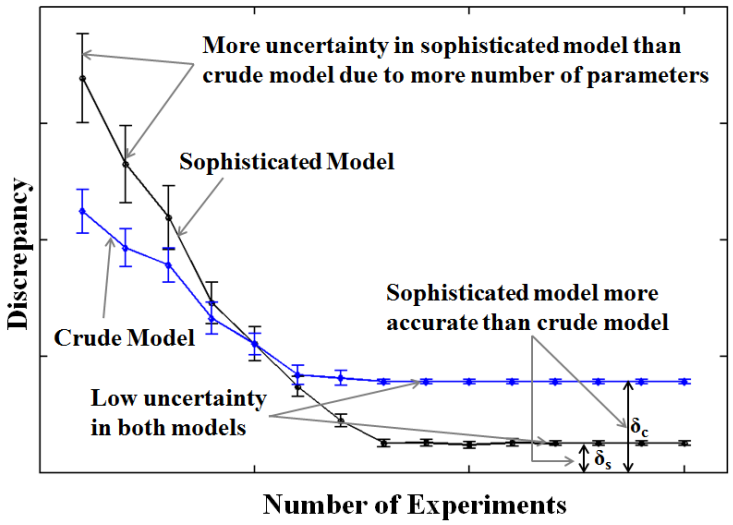

Figure 2.1: The hypotheses: experimental campaign will lead to stabilization of discrepancy, and improving physics will reduce the stabilized discrepancy.

Such a statement applies vis-à-vis to the degree of sophistication of models implemented in simulation codes, which constitutes the second hypothesis – also illustrated in Figure 2.1. As new first-principle-physics are progressively implemented in a simulation code, the assumptions would gradually be replaced with first principle physics and the code would reproduce reality with improved fidelity (δ_S). Such improvement can be expected to be asymptotic if the first-principle-physics are implemented in an order based on both their influence (i.e., sensitivity) and incompleteness, formulated for instance by a Parameter Identification and Ranking Table (PIRT) inventory (Wilson and Boyack, 1998; Diamond, 2006) or the CPI metric proposed later in this dissertation. Note that, improving physics sophistication may naturally introduce new uncertain input parameters to the simulation model, and thus require a larger number of experiments to learn about the increased number of

parameters, as indicated by the larger number of experiments needed for the convergence of the sophisticated model in Figure 2.1. Also note that, for a low number of available experiments, the lack of knowledge due to the increased number of parameters (i.e. reducible or epistemic uncertainty) may contaminate the empirically estimated discrepancy bias of the more sophisticated model, which may in turn appear to be higher than that of the crude model.

The focus herein is on highly complex numerical models for high consequence decision making, development of which may take several years (even decades) and many iterations. Furthermore, it is envisioned that projects involving such models will contain both experimental and model development components. This chapter claims that selection of an optimal route during the model development process, further experimentation or code development, can be achieved by evaluating the convergence of the systematic bias. If sufficient experimental data is available such that the systematic bias has converged *throughout the domain of applicability*, the epistemic portion of parameter uncertainty (reducible uncertainty) can be considered mitigated (remaining uncertainty in this case would be aleatory and thus, irreducible). A converged systematic bias that is sufficiently low for the application of interest demonstrates the predictive capability of the model. On the other hand, systematic bias that fails to converge or that converges to a value that is too high for the application of interest necessitates a more complete physics description. Based on these seemingly simple principles, this chapter casts the problem of resource allocation into an algorithmic framework. The proposed

algorithm utilizes a metric for the convergence of the systematic bias in the optimal path selection that consists of a combination of experimental campaign and code development.

First, a purely mathematical proof-of-concept exercise is demonstrated to establish the aforementioned concepts. An arbitrary function representing "truth" is compared to its degraded variants representing incomplete and inexact simulation models with various levels of "theoretical knowledge." Systematic biases between the truth function (i.e., experiments) and degraded functions (i.e., simulation models) are calculated and are shown to converge as the number of experimental data points is increased; additionally, the corresponding systematic bias is shown to decrease as the sophistication of the simulation model is improved. Furthermore, the inability to know the "truth" is discussed by illustrating the difference between true systematic bias (also known as model form error) and the estimated systematic bias (also known as discrepancy). The proposed path selection algorithm is then applied to this proof-of-concept example. Second, the conceptual discussion is extended to the non-linear Preston-Tonks-Wallace (PTW) material model of plastic deformation (Preston et al., 2003). Physical experiments performed on samples of Tantalum metal, measuring the stress-strain behavior as a function of various settings of temperature and strain-rate, are available. A statistical inference procedure is applied to infer the posterior distribution of seven coefficients of the PTW model such that its predictions better match the measured stress-strain curves over the domain of interest. Two variants of the PTW model are investigated, including a low-fidelity version and the original PTW model. The systematic bias of the two models, as well as their convergence rate as the number of

available experiments increases, is analyzed. The PTW model informed with an increased number of experiments is observed to have reduced random uncertainty, while the model with improved theoretical knowledge, i.e. the original PTW model, is observed to have reduced systematic bias. Finally, the proposed resource allocation algorithm is further demonstrated using the two variants of the PTW model.

2.2. Parameter Calibration and Bias Correction

Herein, the focus lies on the imprecision in the model input parameters, and inaccuracy due to inadequate and incomplete physics; however, note that there are several other sources of uncertainty and error that a model developer faces[4]. It is a routine application, that uncertain and influential[5] model parameters are calibrated against experiments to reduce the uncertainty in the models predictions. These parameters, which are known to the model developer while their precise values or distributions are unknown, are referred to as *calibration parameters*. The systematic bias, on the other hand, represents the incomplete or inadequate physics principles and parameters; thus it is either unknown or insufficiently known to the model developer and is referred to as *model form error*. It is less common to see the treatment of model form error as a bias

[4] Additional sources are numerical uncertainty caused by truncation errors, discretization errors, and round-off errors, and experimental uncertainty in the observations acquired to inform the numerical models. Note that experimental and numerical uncertainties play a similar role in that each adds uncertainty to one side of Equation (2.1) during test-analysis correlation. Moreover, one can also consider emulator uncertainty in cases where a fast-running interpolator (emulator or a surrogate model) is needed to replace a computationally demanding code.

[5] Calibration of insensitive parameters results in inefficient use of resources and therefore, influential input parameters must first be determined through a sensitivity analysis.

integrated into the validation process; however, Higdon et al. (2008), Rebba et al. (2006), and Roy and Oberkampf (2011) provide examples of such treatments.

A model calibration procedure solely focusing on the calibration of input parameters, and thus overlooking the presence of model form error, is prone to converging to a mathematically viable but physically incorrect solution (Kennedy and O'Hagan, 2000; Draper, 1995), i.e. model parameters may converge to values other than their true values to counteract the presence of model form error. What is more, if model form error is not accounted for during calibration, determining the residual difference between the experiments and calibrated model would portray an incorrect picture –often overly optimistic– regarding the model's predictive capability (Kennedy and O'Hagan, 2000). The inference procedures must therefore be conceived to simultaneously address calibration of imprecise parameters and bias correction of inaccurate physics principles. For this purpose, this study utilizes an equality initially proposed by Kennedy and O'Hagan (2000) and later implemented in a fully Bayesian context by Higdon et al. (2008a). The aim is to reduce the random uncertainty of the model parameters and simultaneously correct for the systematic bias. The method relies on the following relations:

$$y(x) = \zeta(x) + \varepsilon(x) \tag{2.1}$$

$$\zeta(x) = \eta(x, \theta) + \psi(x) \tag{2.2}$$

where the pair (x, θ) denotes model input and $\eta(x, \theta)$ denotes the model output. Model input and output can be scalar or multi-dimensional. The symbol x refers to inputs that define the domain of applicability (parameters that can be controlled during experimental

testing); while θ denotes calibration parameters (parameters that cannot be controlled during experimental testing). In the application to the PTW model (discussed later in Section 2.5), input variable x refers to the pair of control variables (temperature; strain rate), while θ refers to seven material-dependent coefficients. Measurements[6] are denoted by the symbol $y(x)$ in Equation (2.1) and are defined as the summation of the truth, $\zeta(x)$, and experimental error, $\varepsilon(x)$. Note that experimental errors affect the model accuracy and precision indirectly through Equation (2.1). In Equation (2.2), the truth is further defined as the summation of model simulations $\eta(x, \theta)$ and model form error, $\psi(x)$. When the systematic bias of the model is known (i.e. model form error), the model predictions can be corrected and the truth can be obtained by directly substituting Equation (2.2) in (2.1). In reality, however, the model form error is unknown and it must be replaced with an 'error model' estimate, i.e. discrepancy bias, $\delta(x)$, resulting in a best-estimate of the observation:

$$\hat{y}(x) = \eta(x, \theta) + \delta(x) + \varepsilon(x) \tag{2.3}$$

Note that the error model, herein referred to as discrepancy bias, is evaluated independently. Equation (2.3) defines the best estimate of observations, $\hat{y}(x)$, over the domain of applicability as the sum of model predictions, discrepancy, and experimental error. The theory behind Equations (2.1)-(2.3) comes from Kennedy and O'Hagan (2000) and applications can be found in Higdon et al. (2008a) and (2008b).

[6] The terms "measurements," "data," and "dataset," likewise, refer to physical measurements or observations collected by performing experiments.

2.3. Convergence of Discrepancy

It is envisioned that the model is executed to predict at varying operational and environmental conditions that define the application domain. The systematic bias of the model will naturally be non-uniform throughout this application domain. Therefore, the empirically defined discrepancy bias should convergence to a non-uniform 'functional form' as new experiments become available. If the true discrepancy or its best estimate is known, convergence can be quantified by a standard distance-based metric whereby the convergence is checked point-wise between the estimated and the true discrepancy (or the best-estimate discrepancy). Alternatively, one could treat convergence in an average sense whereby the area between the true and estimated discrepancy curves provide the 'error'. Both aforementioned approaches measure convergence to the true functional form of discrepancy based on an error calculation; convergence is considered to be reached when the acceptable error level is achieved.

When the true functional form of discrepancy is not known *a priori*, these approaches can still be utilized whereby the relative distance or error measure is computed between the estimated functional forms of current and the preceding discrepancies. These values can then be tracked until there is no significant change between two successive pairs of discrepancies, i.e., until the functional form of discrepancy stabilizes. While tracking the stabilization of discrepancy typically, the mean of the discrepancy normalized with respect to the mean of the predictions is often used. In this chapter, discrepancy is presented by its normalized mean plotted against the number of experiments or physics improvement, as shown in Figure 2.2. Next, the convergence

can be characterized by calculating the difference between the current and previous data points (Figure 2.2) until three consecutive points exhibit stabilization.

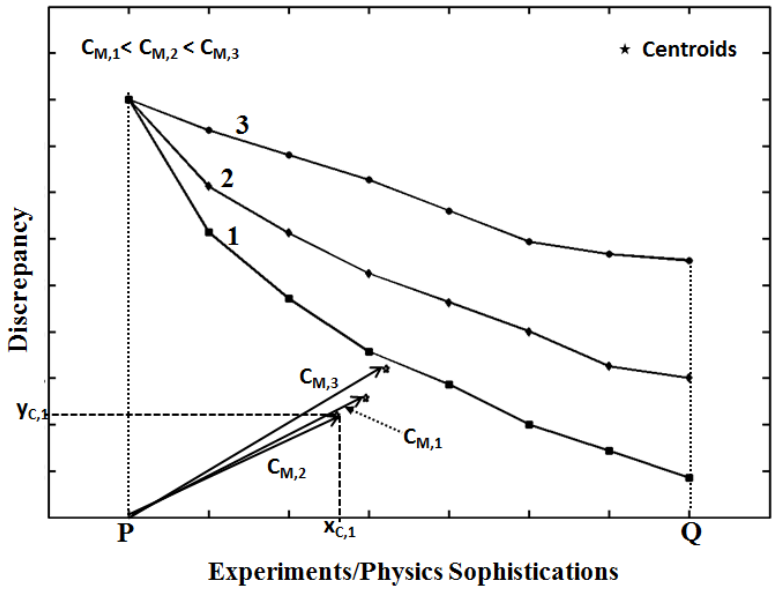

Figure 2.2: The convergence rate can be calculated as the distance between the origin and the centroid of the area under the curve.

To quantify the convergence rate of the discrepancy bias, the metric of Saxena et al. (2010) proposed for comparison of prognostic algorithms is implemented. This metric is defined as the distance between the origin and the centroid of the area under the curve formed by a chosen prediction metric (i.e. discrepancy in this application). For the prognostic application, the rate of convergence is defined in real time, but for the purposes of this chapter the metric is defined in terms of the number of experiments and level of physics sophistications. The rate of convergence is then expressed as:

$$C_M = \sqrt{(x_c - \alpha_P)^2 + y_c^2} \tag{2.4}$$

where α_P represents experiments (E) or codes (C) and the subscript P denotes their value at the origin, (P, 0). The centroid is given by:

$$x_c = \frac{0.5\sum_{i=P}^{Q}(\alpha_{i+1}^2 - \alpha_i^2)\frac{\delta_i}{\delta_{\max}}}{\sum_{i=P}^{Q}(\alpha_{i+1} - \alpha_i)\frac{\delta_i}{\delta_{\max}}}, \text{ and } y_c = \frac{0.5\sum_{i=P}^{Q}(\alpha_{i+1} - \alpha_i)\left(\frac{\delta_i}{\delta_{\max}}\right)^2}{\sum_{i=P}^{Q}(\alpha_{i+1} - \alpha_i)\frac{\delta_i}{\delta_{\max}}} \quad (2.5)$$

It is plausible that for two given curves the initial discrepancy values do not coincide, hindering the direct comparison of convergence characteristic of the addition of new experiments with that of improving physics. Therefore, note that the calculation of C_M values is normalized such that discrepancy in Equation (2.5) is divided by the corresponding maximum discrepancy. While comparing the convergence rate of multiple curves, curves with a smaller C_M value indicate faster convergence (Figure 2.2). Note that the rate of convergence should be considered in light of accuracy or precision of the models. The discrepancy at convergence is assumed to be zero, which can only be reached by a *perfect* model. However, the discrepancy bias of each model eventually converges to its unique discrepancy value (which represents the respective model incompleteness and inexactness).

Note that the convergence rate metric can be used to track convergence by computing the metric at each experiment or physics improvement until the change in metric values is insignificant. The C_M metric for convergence rate assumes that convergence will eventually be achieved as either physics sophistication is improved or more experiments are added. This assumption calls for further discussion on convergence characteristics of discrepancy bias.

Convergence of discrepancy as the number of experiments increases is intuitive and has been demonstrated in Hemez (2007) and Hegenderfer et al. (2012). However, such convergence presents a path-dependency, i.e. the settings of experiments as well as the sequence in which these experiments are conducted influence the rate of convergence. This is demonstrated to be especially true for data-poor situations in Atamturktur et al. (2011). However, for data-rich scenarios the path dependency diminishes (Atamturktur et al., 2011). To assure monotonic convergence of discrepancy bias, one can employ sequential calibration techniques, which accelerate the convergence of discrepancy bias through the selection of optimal experimental settings whereby discrepancy convergence can be achieved by using the least possible number of experiments. Batch Sequential Design (Loeppky et al., 2010; Williams et al., 2011) is one recently proposed efficient approach for the selection of experiments. Other sequential, non-sequential approaches (Davis and Prieditis, 1999) and hybrid approaches (Crombecq et al., 2009) can also be implemented to assure monotonic and rapid convergence of the discrepancy function as the number of experiments increases.

While the concept of *diminishing returns* is intuitive for experiments, the same cannot be said for physics sophistication. Nonetheless, a theoretical breakthrough development in the modeling strategy, which improves the accuracy of the model significantly, can defy the asymptotic convergence expectation. While surprise findings can be expected in theoretical development, the discussion is limited to cases where the model is built based on a systematic inventory of "known" physics principles, such as those provided by the Phenomena Identification and Ranking Table (PIRT). PIRT,

originated as part of the U.S. Nuclear Reactor Commissions' Code Scaling, Applicability and Uncertainty (CSAU) evaluation procedure (Boyack, 1989), is a qualitative approach to effectively gather information from experts on the phenomena and processes of interest, and ranking their importance in a systematic fashion with proper documentation (Wilson and Boyack, 1998). It has been used primarily in nuclear engineering and safety (Diamond, 2006; Olivier and Nowlen, 2008; Tregoning et al, 2009). The relevant phenomena are categorized and ranked based on *importance* (determined through sensitivity analysis) and *knowledge level* (determined through uncertainty quantification). In Chapter 5 of this dissertation the fundamental concepts of PIRT are quantified and amended for the case of coupled numerical models through the developed CPI metric. It is herein claimed that a modeling effort that strictly follows the sequence as ranked by PIRT or CPI (in partitioned analysis of coupled models) would include the physics principles prioritized by reduced importance. Such a model can be argued to converge asymptotically as the physics sophistication increases.

2.4. Proof-of-Concept: Experimentation versus Code Development

This section provides a proof-of-concept illustration of the trade-off between allocating resources for new experimental campaigns versus further code development. First, a simple mathematical function is adapted to represent the "truth." Next, two simpler variants of this truth function are generated to mimic the incomplete and inexact nature of numerical models. The truth function is executed to generate five (synthetic) experiments at five different control parameter settings.

The first numerical model, $y_1(x,\theta)$, is a crude representation of truth. This model includes a single term in the function and a single calibration variable θ_1 that is allowed to vary between -1 and 1 with a uniform probability distribution. An ensemble of model predictions obtained by varying θ_1 is compared to the experiments in Figure 2.3 (Left). Note that $y_1(x,\theta)$ is unable to reproduce the experiments because its mathematical equation is missing essential terms. The discrepancy associated with $y_1(x,\theta)$, $\delta_1(x)$, depicted in Figure 2.4, illustrates the need for a series of *additional terms* to be added to the function of the crude numerical model.

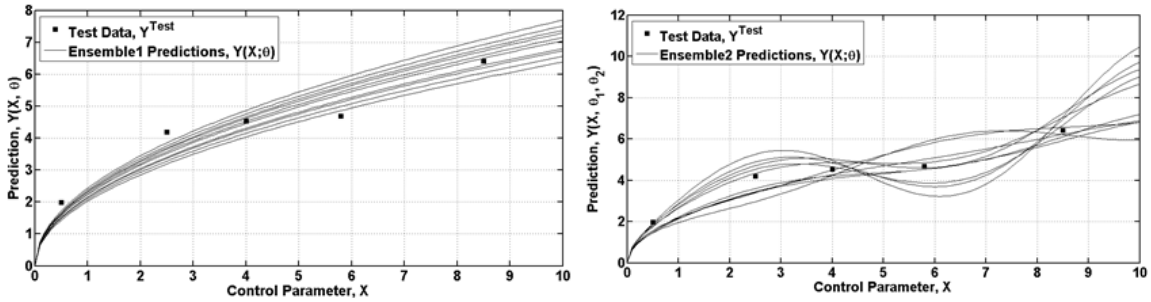

Figure 2.3: (Left) Ensemble of very crude model predictions $y_1(x,\theta)$ and test data. (Right) Ensemble of more sophisticated model predictions $y_2(x,\theta)$ and test data.

The second prediction curve, $y_2(x,\theta)$, represents a more sophisticated model, and includes one of the additional terms of the "truth" function that is absent from the first model, $y_1(x,\theta)$. As a result, this more sophisticated model, $y_2(x,\theta)$, includes two calibration variables θ_1 and θ_2. The newly introduced model input, θ_2, is also allowed to vary between -1 and 1 with uniform probability. Figure 2.3 (Right) presents the ensemble of model predictions generated by varying the two input parameters, θ_1 and θ_2.

One can see that the implementation of a new calibration parameter into $y_2(x,\theta)$ increases the dispersion of the predictions in the ensemble. The mean standard deviation of $y_1(x,\theta)$ is 0.295 and of $y_2(x,\theta)$ is 0.596. This increase in standard deviation indicates an increase in uncertainty due to the addition of a new uncertain parameter, θ_2. However, in comparing the predictions of the two models (see Figure 2.4), one can also see that the mean of the second model prediction more closely tracks the experiments. The discrepancy herein is calculated as the mean of the absolute value between the "truth" and the simulation normalized as a percent of the mean of the "truth". The discrepancy of the first model prediction, $\delta_1(x)$, is 11.40% of the mean "truth" while the corresponding discrepancy of the second model prediction, $\delta_2(x)$, is equal to 5.86% of the mean "truth" (Figure 2.4). As seen in Figure 2.4, the discrepancy of the second model, $y_2(x,\theta)$, is less than that of $y_1(x,\theta)$. This illustrates the second hypothesis that as more physics sophistication is added to the model, the discrepancy bias reduces.

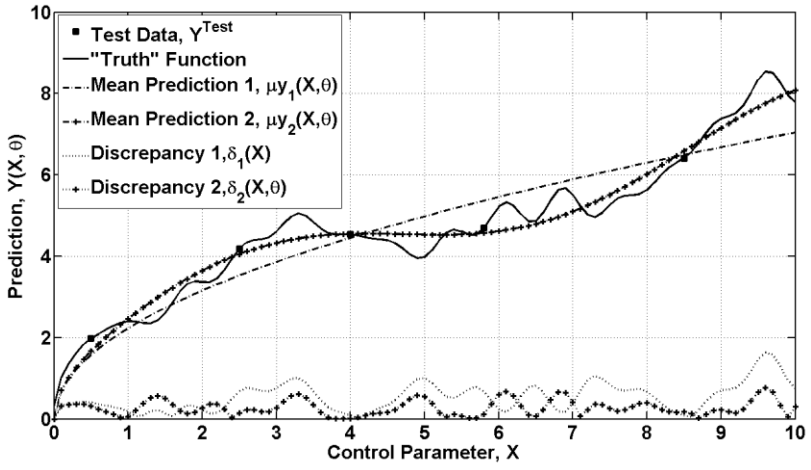

Figure 2.4: Comparison of simulations, "truth function", and associated discrepancy terms.

In the controlled proof-of-concept example herein, the truth function is known and thus, the "true" discrepancy biases for both $y_1(x,\theta)$ and $y_2(x,\theta)$ are known, i.e. model form error can be plotted in Figure 2.4. In reality, the truth is unknown, and therefore the true values of discrepancy are only known at the discrete values of the finite number of available experiments. However, simulation models are routinely executed to predict at settings where experiments are unavailable. Therefore, the discrepancy bias must be estimated at these untested settings. Such estimation can be obtained by training an 'error model' (i.e., an emulator) exploiting the "true" discrepancy values available for a finite number of experiments[7], to represent the discrepancy bias throughout the entire domain of applicability. With a finite number of available experiments, many different discrepancy emulators can be trained, however. In Figure 2.5 (Left), for instance, a plausible discrepancy bias estimate is constructed using a cubic interpolation (the dotted line). The fidelity of this estimated discrepancy to the true discrepancy can be improved by increasing the number of available experiments In Figure 2.5 (Right), the number of experimental data points is increased from five to 15, which yields a discrepancy estimation closer to the true discrepancy bias. The percent error between the estimated discrepancy and the true discrepancy that is 65.5% for Data Set 1, is reduced to 24.3% for Data Set 2 (Figure 2.5 (Right)). As seen, as more experimental data is added to the analysis, the estimated discrepancy bias converges to the true functional form of discrepancy throughout the entire domain. However, if the discrepancy bias has

[7] The experimental uncertainty is neglected in this discussion. Note that the true values of discrepancy are still unknown at the experimental settings due to the presence of experimental uncertainty.

converged and the resulting discrepancy remains unacceptably high, the only route left for improvement would be through improving the physics sophistication of the model.

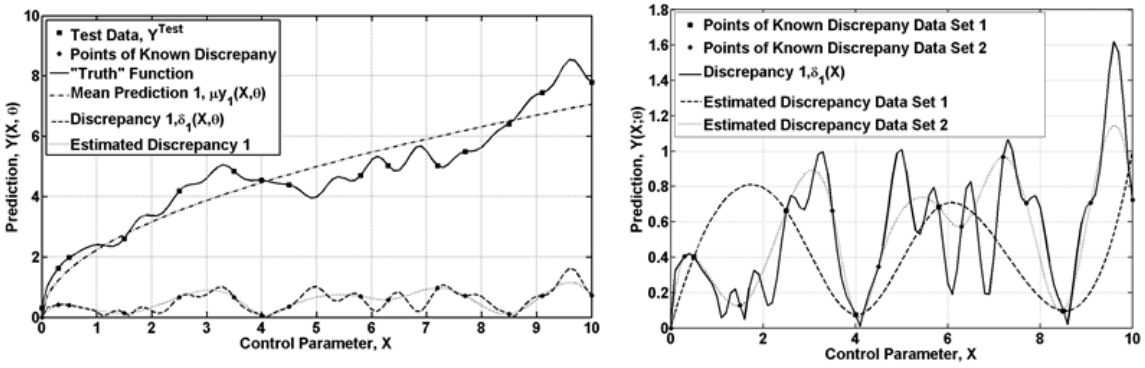

Figure 2.5: (Left) Comparison of estimated discrepancy term and true discrepancy term. (Right) Comparison of "True" discrepancy and estimated discrepancies.

As seen, there is a trade-off between improving physics and increasing the number of experiments. This is of course of great interest for decision makers while allocating a finite amount of available resources. The results of both routes are shown in Figure 2.6 for a total of five simulations of increasingly improved physics with incrementally increased experiments (from two-14). Figure 2.6 summarizes both hypotheses discussed in the previous section. Once discrepancy convergence is achieved, no further reduction in discrepancy is obtained through additional experimentation; therefore, physics sophistication is required to further reduce the discrepancy bias. However, more sophisticated models may naturally need more experiments for the discrepancy to converge.

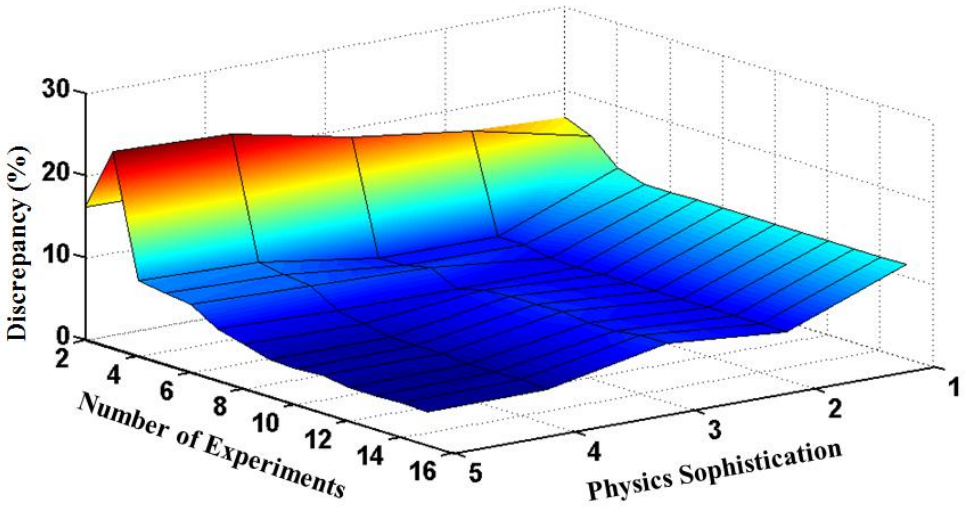

Figure 2.6: Proof-of-concept: relationship between number of experiments, physics sophistication, and discrepancy.

The C_M metric can be used to compare the convergence rate of multiple curves; however, since the main interest is in the rate of convergence to the converged discrepancy of each curve respectively, y_c in Equation (2.4) and δ_i in Equation (2.5) are replaced by $(y_c - \delta_f)$ and $(\delta_i - \delta_f)$, respectively, where δ_f denotes the ultimate value of the discrepancy for each model such that the convergence metric is calculated based upon the unique characteristics of each curve respectively. δ_f is assumed to be the value of discrepancy at the final experiment (at the 14^{th} experiment) for each physics sophistication curve. In Figure 2.6, the fastest convergence as the number of experiments is increased (C_M=1.21), is obtained for the crudest model (see Table 2.1 for the convergence rates of five models with increasing sophistication) and the convergence rate

slows as physics sophistication is increased. However, note that the crudest model has the highest discrepancy bias remaining after convergence.

Table 2.1: Values of convergence rate metric for experiments and physics sophistication

Level of Physics Improvement	1	2	3	4	5
$\mathbf{C_M}$	1.21	1.35	1.37	2.28	2.38

Of course, the point made in Figure 2.6 is not to show that with the benefit of hindsight, the decision maker can determine the optimal route to reach predictively mature models. Rather, Figure 2.6 aims to illustrate that there exists an optimal decision for resource allocation between theoretical advancement and empirical training of the numerical model. This optimal decision depends on various interrelated factors related to the existing accuracy, and precision of the model, and the quality and quantity of available experiments. To seek a path that optimally combines physics improvements and experimental campaign, it is proposed to exploit the previously introduced rate of convergence metric. Note that since the final converged discrepancy in real applications is unknown, the C_M values are calculated according to Equation (2.4). An algorithm can be devised to carry out the experiment-code selection process based on the previous conceptual discussions as follows:

1. Initialize indices i and j for the experiment (E) and code (C): i=2 and j=2, and flag for path selected: pflag=0.
2. Compute discrepancy, $\delta_{i,j}$ at i^{th} experiment and j^{th} code.

3. If (pflag=0), compute convergence rate metrics, $(C_M^E)_i$ and $(C_M^C)_j$, respectively.

   Else if (pflag=1), compute convergence rate metric, $(C_M^E)_i$ for i^{th} experiment;

   Else, compute convergence rate metric, $(C_M^C)_j$ for j^{th} code.

4. Select the path, i.e. further experimentation or further code development

   a. At initial step, i.e, i=2,j=2, and pflag=0

      If $(C_M^E)_i < (C_M^C)_j$, choose further experimentation, and set pflag=1

      Else, choose further code development, and set pflag=2

   b. If (pflag=1), previously on experimentation path

      If $|\delta_{i,j} - \delta_{(i-1),j}| <= \beta \delta_{i,j}$ (tolerance value of 0.1 is used) and $(C_M^E)_i < (C_M^C)_j$, choose further experimentation again, and set pflag=1

      Else, choose further code development, and set pflag=2

   c. If (pflag=2), previously on code development path

      If $|\delta_{i,j} < \delta_{i,(j-1)}| <= \beta \delta_{i,j}$ and $(C_M^C)_j < (C_M^E)_i$, choose further code development again, and set pflag=2

      ELSE, choose further experimentation, and set pflag=1

5. Update indices: If (pflag=1), i=i+1; Else, j=j+1
6. If either of the limit (maximum number of iterations allowed based on the budget) on experiment or code development is reached, proceed with the other by default
7. Termination conditions:

a. If $\delta_k \leq \varepsilon^{\delta}$ (acceptable discrepancy value), check three consecutive points

   If (pflag=1), k=(i,j), (i-1, j), (i-2,j); Converged, Exit

   Else if (pflag=2), k=(i,j), (i, j-1), (i,j-2); Converged, Exit

b. If iteration = limit of experiments + limit of code, Exit

8. Repeat Steps 2 to 7

To demonstrate the use of the proposed path selection algorithm, it is applied to the proof-of-concept example, assuming that at least one code refinement and one additional experiment are available (therefore, the algorithm starts from i=2 and j=2). Two different applications are considered, each with different accuracy requirements; for the high consequence application with high accuracy demands, the acceptable discrepancy value is set to 5% and for the low consequence application with lower accuracy demands, the acceptable discrepancy value is set to 10%.

As shown in Figure 2.7, for the high consequence application, with the proposed selection procedure, the process for the development of an empirically trained, physics-based numerical model goes from point A to L (shown with their respective discrepancy and its corresponding convergence rate, C_M, values (boxed)). It takes five levels of physics improvement and eight experiments to reach the acceptable discrepancy, and the algorithm concludes once three consecutive points remain below the acceptable value (i.e. 10 experiments). The black arrows represent the accepted paths while the gray arrows are the rejected paths. At point A, the C_M values in the physics sophistication path that is obtained using two levels of physics improvement at the second experiment is 1.11 while the C_M values for the experiment path at the second experiment with second

level of physics improvement is 1.15. This indicates that at A the convergence of discrepancy in the physics improvement path is faster than in the experiment direction; thus the physics path is selected. However, this leads to increase in discrepancy (from 18.55 to 20.41) going from A to B due to increases number of uncertain model parameters as more sophisticated model is used indicating the need for more experiments. Furthermore, the C_M value also increases from A to B (from 1.11 to 1.64). Note that a similar pattern is also observed in steps C to D, and F to G. As shown in step 4 of the algorithm, the reduction in discrepancy is considered in addition to the rate of convergence, and since this increase in discrepancy is greater than the set threshold, for the next step from B to C the algorithm chooses the experiment path. The algorithm proceeds as such until the acceptable level of discrepancy is achieved. As seen in Figure 2.7, for the low consequence application, the desired accuracy can be obtained at step C, where the discrepancy value falls below 10%; however, to assure convergence, the algorithm continues until three consecutive discrepancies under 10% are realized (point F – much sooner than for the high consequence case which terminates at point L).

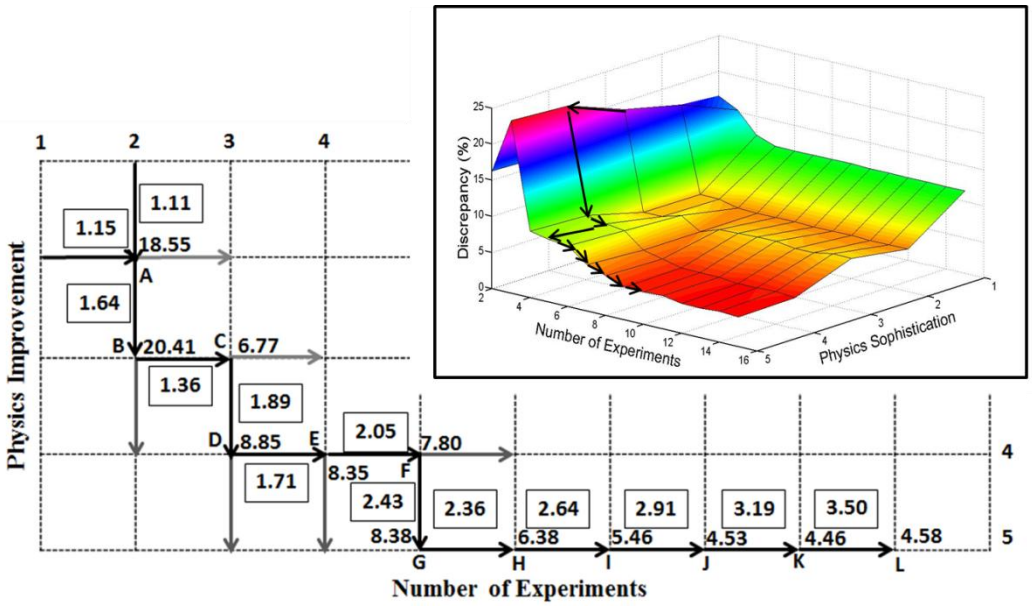

Figure 2.7: Path of optimal combination of physics sophistication and experimental campaign with the corresponding discrepancy values at each step.

2.5. PTW Model: Background and Model Calibration

In this section, the Preston-Tonks-Wallace (PTW) model of plastic deformation documented in Preston et al. (2003) is briefly introduced. It is applied to the prediction of stress-strain curves for light-weight, high-strength Tantalum (Ta) metal. A degraded variant of the original PTW equations is also evaluated to observe the increase in discrepancy as physics sophistication is degraded[8]. These analyses are demonstrative applications of the hypothesis presented herein in the context of an experiment-based model validation framework where code development and experimentation are considered equally costly as shown in Figure 2.8.

[8] The approach taken here, i.e. degrading the physics sophistication of a numerical model, is in the sense opposite of improving the physics of a code, which in some cases can take years to complete.

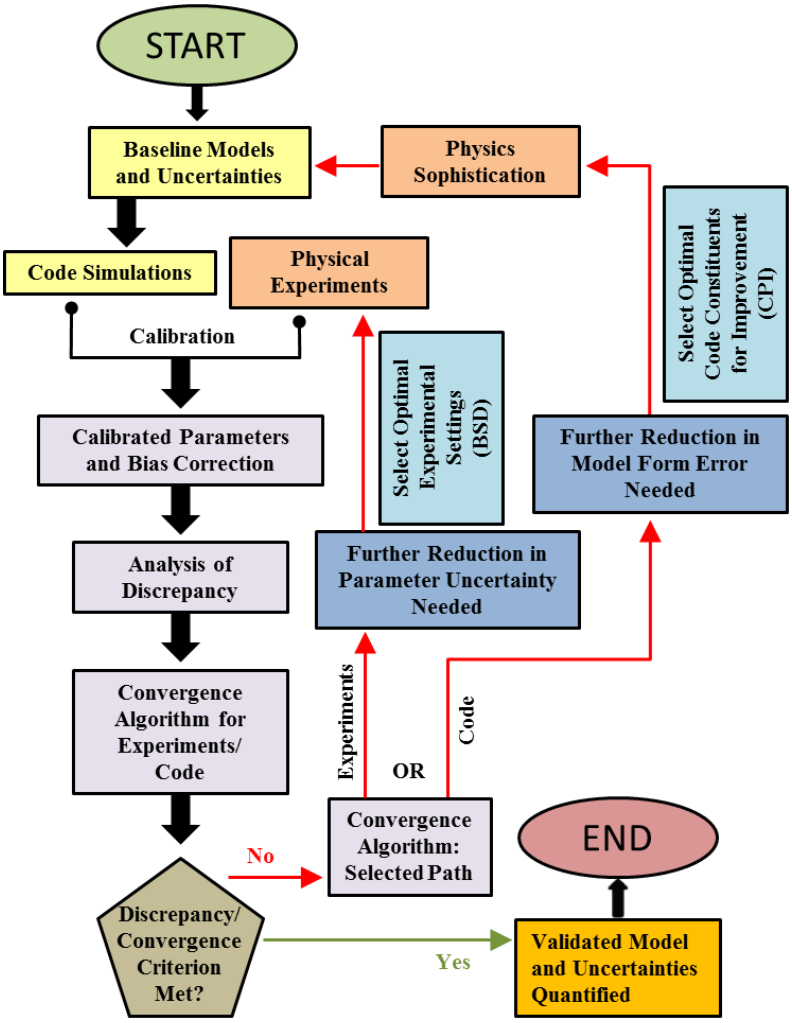

Figure 2.8: Experiment-based model validation framework

2.5.1. 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 suitable to simulate the material response to fast transients such as those from explosive loading or high velocity impacts. The main equations of the PTW model from Preston et al. (2003) are:

$$\sigma = \sigma_S + \frac{1}{p}(s_0 - \sigma_Y)\ln\left\{1-[1-\exp(-pr)]\exp\left[-\frac{p\theta\Delta}{(s_0-\sigma_Y)[\exp(pr)-1]}\right]\right\} \tag{2.6}$$

$$\sigma_S = s_0 - (s_0 - s_\infty)\,erf\left[\kappa T \ln\left(\frac{\gamma\dot{\xi}}{\dot{\Delta}}\right)\right] \tag{2.7}$$

$$\sigma_Y = y_0 - (y_0 - y_\infty)\,erf\left[\kappa T \ln\left(\frac{\gamma\dot{\xi}}{\dot{\Delta}}\right)\right] \tag{2.8}$$

$$r = (\sigma_S - \sigma_Y)/(s_0 - \sigma_Y) \tag{2.9}$$

where σ denotes stress, Δ represents strain, $\dot{\xi}$ denotes the atomic vibration frequency, and 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 ($\dot{\Delta}$) and temperature (T) of Equations (2.6)-(2.9). Symbols θ, κ, γ, s_0, s_∞, y_0 and y_∞ are seven dimensionless calibration variables that depend on the material analyzed.

Table 2.2: Definition of variants of the PTW model analyzed

Identifier	Description	Error Function Used
Variant 0	It is the original PTW model with "erf" function.	$\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$
Variant 1	The "erf" function is replaced with piecewise linear approximation.	$\frac{2}{\sqrt{\pi}}x, x \le \frac{2}{\sqrt{\pi}}; 1, x > \frac{2}{\sqrt{\pi}}$

In addition to the original PTW model of Equation (2.6), a degraded variant of the PTW model is analyzed to monitor the changes in the discrepancy as the physics implemented to describe the behavior of the material is artificially degraded. The characteristics of this degraded variant of the PTW model are defined in Table 2.2. In the original PTW model, the "erf" function is utilized in the formulation of work hardening saturation stress and yield stress in the thermal activation regime in Preston et al. (2003). This function is chosen because it depicts evidence seen in experimental data (Preston et al., 2003). The second version, labeled Variant 1, is an implementation where the "erf" function is replaced by a piecewise linear function. This implementation is expected to perform less accurately than the original model. The difference between the two error functions is shown in Figure 2.9. Compared to Variant 1, the original PTW model supplies a model with "improved" physics sophistication.

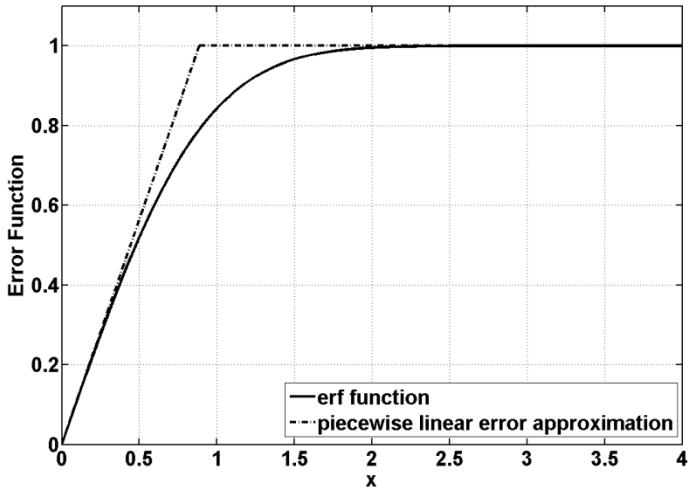

Figure 2.9: Error functions used in the two variants of the PTW model.

2.5.2. PTW model for high-strength Tantalum (Ta) metal

The application considered is the material model for the Ta 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 Ta samples to collect the stress-strain curves shown in Figure 2.10. The pairs (T; dε/dt) of control parameters that define the Hopkinson bar tests are listed in Table 2.3. Figure 2.10 illustrates the range of stain and stress values that the material model is expected to reproduce over the domain of applicability. Measurement error is modeled as a Gaussian distribution with a zero mean and 2.5% variance.

Table 2.3: Definition of settings for experiments performed on Ta samples

Data Set	$\mathbf{n_i}$	**Strain (max)**	**Temperature (K)**	**Strain Rate (s^{-1})**
1	32	0.9757	298	0.001
2	31	0.2218	1073	3900
3	10	0.1057	298	0.1
4	29	0.2433	298	1300
5	27	0.1737	673	2600
6	13	0.1735	77	0.001

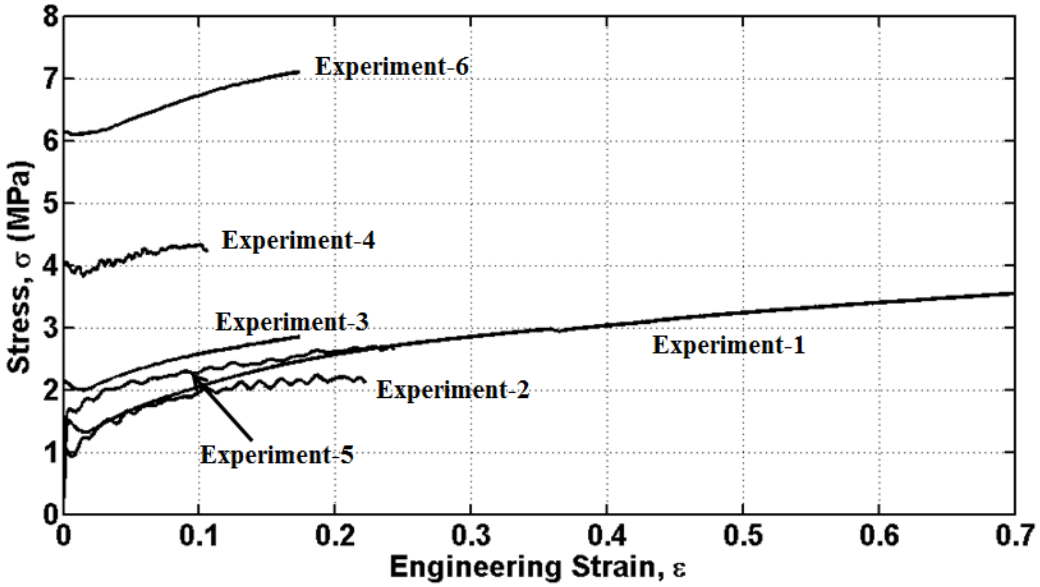

Figure 2.10: Strain-stress curves measured from Hopkinson bar tests for the Ta metal.

The nominal values for the seven ancillary variables (θ, κ, γ, s_0, s_∞, y_0 and y_∞) of the PTW model are given in Table 2.4, together with lower and upper bounds within which they are reported to vary for the Ta metal (Fugate et al, 2006). These parameters are initially assumed to vary uniformly between minimum and maximum values. Ideally, one could calibrate these variables to improve the goodness-of-fit between stress-strain curves predicted by the PTW model and those that are measured. Attempting, however, to find a unique set of values that reproduce the entirety of the measurements is difficult 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 model input variables. Thus, the calibration parameters must be treated as random variables.

Table 2.4: Definition of calibration variables of the PTW model for the Ta metal

	Description	Minimum	Maximum	Nominal
θ	Initial strain hardening rate	8.37×10^{-4}	0.0276	0.01
κ	Temperature dependence of thermal activation energy	0.399	1.078	0.6
$-\ln(\gamma)$	Strain rate dependence of thermal activation energy	7.20	15.508	12.3
y_0	Minimum yield stress (at T = 0 ^{o}K)	0.00653	0.0128	0.00925
y_∞	Maximum yield stress (at T ≈ melting)	7.05×10^{-4}	0.00214	0.00123
s_0	Minimum saturation stress (at T = 0 ^{o}K)	0.00803	0.0344	0.0122
s_∞	Maximum saturation stress (at T ≈ melting)	0.00144	0.00452	0.00375

Therefore, through implementing Bayesian inference, the joint probability distribution of calibration variables is sought such that model predictions are statistically consistent with measurements over the two-dimensional domain of applicability. This approach also infers from the comparison between predictions and measurements a statistical error model of the discrepancy term $\delta(x)$. The joint probability distribution is explored with a Markov chain Monte-Carlo (MCMC) 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 (Metropolis et al., 1953; Hastings, 1970). After performing a sufficient number of iterations, selected to be 20,000, the statistics of calibration variables visited are computed to estimate the (initially unknown) joint probability distribution of the uncertain input parameters of the PTW model.

2.5.3. Definition of the analysis

The MCMC exploration of the posterior probability distribution of calibration variables (θ, κ, γ, s_0, s_∞, y_0 and y_∞), and estimation of the discrepancy term $\delta(x)$, is repeated for the two variants of the PTW model (see Table 2.2) and different combinations of physical measurements (see Table 2.3). The combinations of physical tests used in each case are defined in Table 2.5.

Table 2.5: Definition of the four sets of Hopkinson bar experiments analyzed

Case	**List of Experiments**
1	1, 2, 3
2	1, 2, 3, 4
3	1, 2, 3, 4, 5
4	1, 2, 3, 4, 5, 6

Four separate cases with varying number of experiments are defined to assess the effect that increasing the number of experiments has on the predictive maturity of the PTW model. On the other hand, the analysis of the two variants of the PTW model enables us to study the effect of implementing improved physics sophistication. The

working hypothesis is that the discrepancy will be sensitive to both increasing the number of physical tests available for analysis and improving the physics of the model. This hypothesis and further discussion on the effects of increased numbers of experiments versus improved physics sophistication are provided in the following section.

2.6. PTW Model: Findings and Discussion

For the discussion herein, Variant 1 is treated as the less sophisticated "starting" point for the PTW analysis. The discrepancy of Variant 1 is first analyzed for convergence and fidelity-to-data. The discrepancy of the original and "more sophisticated" PTW model, Variant 0, is then considered to assess the reduction in model form error with more sophisticated physics.

Figure 2.11 shows the discrepancy as a percentage of the mean stress of the model prediction throughout the tested domain for both models plotted against an increasing number of experiments. As seen in Figure 2.11, the discrepancy of Variant 1 starts at approximately 18% and converges to approximately 3.5% after the fifth experiment is added. The first hypothesis illustrates that *the convergence of discrepancy indicates that sufficient experimental data has been utilized to mitigate uncertainty in the discrepancy estimate*. Therefore, a further reduction in discrepancy is achieved through the physics sophistication of the original PTW model, Variant 0. Figure 2.11 shows the normalized discrepancy of the original PTW model starts at 10%, and reduces to approximately 2.7% when the fifth experiment is added. As seen in Figure 2.11, the discrepancy of the original PTW model converges to a lower level than Variant 1. This illustrates the second

hypothesis, that *as the physics sophistication of a model is increased, the discrepancy reduces.*

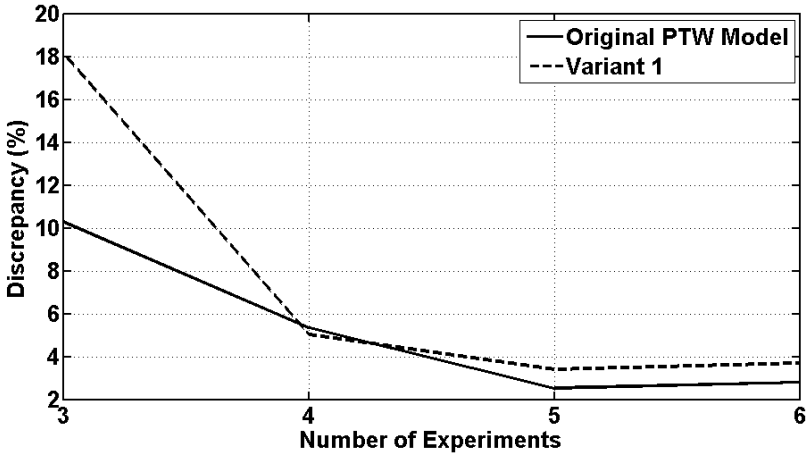

Figure 2.11: Discrepancy convergence with number of experiments for PTW model and Variant 1.

The values of the rate of convergence metric of the discrepancy bias for the original PTW model and the Variant 1 are C_M=3.10 and C_M=2.97, respectively when all six experiments are considered for each variant separately. Thus the discrepancy of Variant 1 converges faster than the original PTW model with respect to number of experiments.

Next, the proposed algorithm for path selection between experiments and physics improvement discussed in Section 2.3 is applied to the PTW model. The results are shown in Figure 2.12. A C_M = 0.94 is obtained for the original PTW model calibrated and bias corrected with four experiments (from three to four experiments). For the physics sophistication path, a C_M = 1.13 is obtained as the physics of the model is improved (from Variant 1 to the original PTW model) both of which are calibrated and bias corrected using four experiments. Based on the C_M values, the algorithm suggests path A

to B in the experimental direction to achieve reduction and convergence of discrepancy. From point B, the algorithm suggests improvement of the model. However, for the purposes of this illustration, note that the discrepancy value is already significantly low at point B (2.53%). Therefore, the analyst may choose to conduct additional experiments to ensure converged discrepancy predictions.

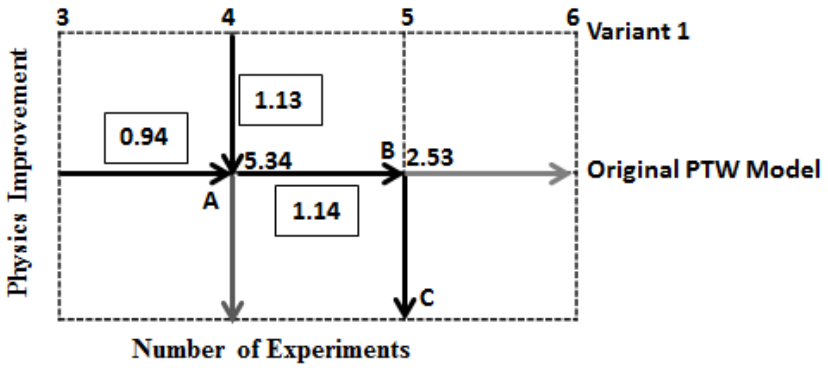

Figure 2.12: Path selection of physics improvement and experiments for PTW model and Variant 1 with the corresponding discrepancy and convergence rate, C_M, (shown in boxes) values at each step.

2.7. Conclusion

A crude model with few parameters is typically inaccurate (perhaps precise due to the low number of uncertain parameters); thus experiments would be needed to infer the systematic bias of the model predictions. On the other hand, a very sophisticated model with many parameters might be imprecise (perhaps accurate due to the sophistication of the physics in the model); thus experiments would be needed to learn about the uncertain and influential parameters. As seen, while exploiting experiments to inform numerical models, the inaccuracy of the model form and imprecision of model parameters are interrelated.

In this chapter, improvement of predictive simulations by (i) mitigating the inaccuracies due to incomplete physics in the model and (ii) uncertainties of the input parameters is investigated. This is achieved by improving the physic sophistication of the model and/or increasing the experimental campaign and is efficiently executed through a framework that considers the relative benefits of the two in terms of the overall cost-effectiveness of resource allocations. Time and cost effectiveness is of course the main driving force in decisions pertaining to resource allocations. The improvement of the predictive capability can be determined by considering the discrepancy bias (representing the systematic disagreement between experimental measurements and numerical simulations). Two hypotheses are developed: first, a progressive reduction in input parameter uncertainty as the number of experiments is increased is claimed. Therefore, due to the reduction in input parameter uncertainty, discrepancy is expected to converge to the model form error as the number of experiments is increased. Once discrepancy has converged, the second hypothesis claims that when the physics sophistication of a model increases, the discrepancy should reduce.

These hypotheses are first illustrated using a purely mathematical exercise. The rate of convergence metric is used to compare the convergence as physic sophistication is improved and as the number of experiments is increased in this example. Next, a practical application of these concepts is presented on the non-linear Preston-Tonks-Wallace strength model applied to Hopkinson bar experiments performed on the Tantalum metal. Although the present chapter does not provide a formal proof, aforementioned applications confirm that the discrepancy for models that are of an acceptable level of

physics sophistication converges as the number of experiments is increased. It is also shown that the discrepancy is reduced when the physics sophistication of the model is improved.

The results presented herein suggest that this stabilization concept 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 simulation model. These results open the door for further study, as the concepts presented herein are generally applicable and are not solely limited to the studies presented.

In the examples considered here, initial stages of code development and calibration are considered where the analysts start with no or very little experimental data and limited number of codes. While this might be true in some cases, in industry or national laboratories where there is a long history of use of simulation-based predictions, rich data collection of experiments as well as multiple versions of codes that have been developed over the years are often available. For these cases, algorithms such as the one proposed here can be employed for the case study of determining the optimal combination of experimental campaign and code development. Such studies using past information can benefit decision making for future steps of model development.

In this chapter, an algorithm is proposed for selection of the path that leads to the optimal combination of physics improvement and experimental campaign, considering only the reduction and convergence of discrepancy. Note that for the results presented herein, it is assumed that the cost of an experimental campaign and code development are the same, which is hardly the case in practice. Therefore, another important factor to be

considered is the cost of code development and experimentation. However, incorporation of the cost factor in the algorithm in a systematic manner is not a trivial matter. Namely, the decision cannot be made solely on cost since it would drive the path only in one direction. However, convergence in discrepancy can be most effectively achieved by considering both directions, i.e. experimental campaign and physics improvement. Therefore, the optimization problem becomes multi-objective requiring compromise between the two driving factors of cost and discrepancy. In future studies, cost factors must be incorporated to make the algorithm more suitable to practical applications.

The convergence of discrepancy indicates a reduction in parameter uncertainty; however, in future studies, an entropy-based information gain metric can be developed to quantify this reduction in uncertainty. As with a physics model that is too crude, an insufficient experimental campaign can yield discrepancy that fails to converge and is not suitable for quantifying the predictive capability of the model. To expedite the convergence of discrepancy, an experiment campaign can be optimized through the selection of optimal design settings. This way, monotonic convergence can be ensured with the least possible number of experiments. For the improvement of a simulation model, a PIRT must be carefully constructed or CPI computed to guide what aspects of the code need further enhancement, and identify the influential parameters.

Additional topics that this work does not yet address and will be pursued in future work, is the understanding of the extent to which these concepts can be applied to the verification, validation and uncertainty quantification of multi-space, multi-physics simulation models.

2.8. References

Atamturktur, S., Bornn, L., & Hemez, F. (2011), Damage detection in masonry vaults by time-domain vibration measurements, *J. Eng. Struct.*, 33, 2472-84.

Boyack, B.E., Duffey, R.B., & Griffith, P., et al. (1989), Quantifying reactor safety margins: application of code scaling, applicability, and uncertainty methodology to a large-break, loss-of-coolant accident, US NRC Reprt, NUREG/CR-5249.

Crombecq, K., Gorissen, D., de Tommasi, L., & Dhaene, T. (2009), A novel sequential design strategy for global surrogate modeling. In: Proceedings of the 41th Conference on Winter Simulation, Austin, Texas, December 2009, 731-742.

Davis, R. & Prieditis, A. (1999), Designing optimal sequential experiments for a Bayesian classifier, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3), 193-201.

Diamond, D.J. (2006), Experience using Phenomena Identification and Ranking Technique (PIRT) for nuclear analysis. BNL-76750-2006-CP.

Draper, D. (1995), Assessment and propagation of model uncertainty, *J. Royal Stat. Soc*., Series B, 57(1), 45-97.

Fugate, M., Williams, B., Higdon, D., Hanson, K.M., Gattiker, J., Chen, S.H., & Unal, C. (2006), Hierarchical Bayesian Analysis and the Preston-Tonks-Wallace Model, Los Alamos National Laboratories Technical Report, LA-UR-06-5205.

Hastings, W.K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57, 97-109.

Hegenderfer, J., Atamturktur, S., Williams, B., Lebensohn, R., Yonten, K., & Unal, C. (2012), Framework for experiment-based validation of numerical models of complex systems: application to VPSC code for metals, *Computer Methods in Applied Mechanics and Engineering,* (submitted for review).

Hemez, F.M. (2007), Answering the question of sufficiency: how much uncertainty is enough? 1st International Conference on Uncertainty in Structural Dynamics. June 11-13, 2007; University of Sheffield, United Kingdom. (Also Los Alamos Technical Report LA-UR-07-3575.)

Higdon, D., Gattiker, J., Williams, B., & Rightley, M. (2008), Computer model calibration using high-dimensional output, *J. A. Stat. Assoc,* 103(482), 570-583.

Higdon, D., Nakhleh, C., Gattiker, J., & Williams, B. (2008), A Bayesian calibration approach to the thermal problem, *Comput. Meth. Appl. Mech.*, 197, 2431-441.

Kennedy, M. & O'Hagan, A. (2000), Predicting the output from a complex computer code when fast approximations are available, *Biometrika*, 87, 1-13.

Loeppky, J.L., Moore, L.M., & Williams, B.J. (2010), Batch sequential designs for computer experiments, *J.Stat. Plan. Infer.,* 140, 1452-1464.

Metropolis, N., Rosenbluth, A., Teller, A., & Teller, E. (1953), Equations of state calculations by fast computing machines, *J. Chem. Phys.*, 21, 1087-91.

Olivier, T.J. & Nowlen, S.P. (2008), A Phenomena Identification and Ranking Table (PIRT) exercise for nuclear power plant fire modeling applications, NUREG/CR-6978 SAND2008-3997P.

Preston, D.L., Tonks, D.L., & Wallace, D.C. (2003), Model of plastic deformation for extreme loading conditions, *J. Appl. Phy*., 93(1), 211-20.

Rebba, R., Mahadevan, S., & Huang, S. (2006), Validation and error estimation of computational models, *Reliab. Eng. Sys. Safety*, 91, 1390–7.

Roy, C.J. & Oberkampf, W.L. (2011), A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing, *Comput. Methods Appl. Mech. Engrg.*, 200, 2131-44.

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010), Metrics for offline evaluation of prognostic performance, *Int. J. Prog. Health Manag.*, 1, 1-20.

Tregoning, R.T., Apps, J.A., Chen, W., Delegard, C.H., Litman, R., & MacDonald, D.D. (2009), Phenomena Identification and Ranking Table evaluation of chemical effects associated with generic safety issue 191, NUREG-1918.

Williams, B.J., Loeppky, J.K., Moore, L.M., & Macklem, M.S. (2011), Batch sequential design to achieve predictive maturity with calibrated computer models, *Reliab. Eng. Sys. Safety*, 96, 1208-19.

Wilson, G.E. & Boyack, B.E. (1998), The role of the PIRT process in experiments, code development and code applications associated with reactor safety analysis, *Nucl. Eng. Design*, 186, 23–37.

CHAPTER 3: FRAMEWORK FOR EXPERIMENT-BASED VALIDATION AND UNCERTAINTY QUANTIFICATION OF NUMERICAL MODELS OF COMPLEX SYSTEMS: APPLICATION TO VPSC CODE FOR METALS

3.1. Introduction

Numerical models[9] are approximate representations of real-world phenomena and thus, simulations invariably suffer from a degree of inaccuracy and imprecision, which can be attributed to a variety of sources, including: (i) incomplete modeling of physics and/or engineering principles, (ii) imprecisely known model parameters, and (iii) numerical uncertainties incurred while solving the mathematical equations.

Incomplete modeling of physics and/or engineering principles, the first factor, refers to the physical phenomena that are either completely unforeseen or that are known, but too complex to incorporate in the model. This inevitable inexactness and incompleteness results in the systematic bias in predictions, also known as model form error. Though imprecise model parameters, the second factor, can be identified by the analyst, their precise values (or distributions) are unknown. These imprecise model parameters are main contributors to the uncertainty in predictions. Numerical uncertainties, the third factor, can be treated by rigorous mathematical analysis of code and solution verification studies, which assures that the equations within the code are being solved correctly (Oberkampf et al., 2003). An exact numerical model with

[9]Herein, the following distinctions are made for the terms code, model, and simulation. Codes are the mathematical algorithms developed to compute physical quantities and phenomena. Models incorporate codes for particular applications (material, geometry, etc.). Simulations constitute the execution of models for particular boundary conditions.

precisely known parameters will still yield inaccurate solutions if the mathematical equations are solved incorrectly (Christie et al., 2005). Hence, verification activities that confront numerical uncertainties must be completed prior to validation activities that confront parameter uncertainty and systematic bias (Thacker et al., 2004). Numerical model verification is considered to be completed prior to validation and thus will be excluded from the scope of the present chapter.

The ultimate objective of an experiment-based validation is then to reduce the parameter uncertainty and systematic bias in model predictions. This reduction is possible by allocating resources to either (i) *experimentation* to increase the number of physical observations used in the model validation process or (ii) *code development* to improve the manner in which physics and/or engineering principles are defined. Of course, the relative benefits of these two approaches, further experimentation versus further code development, vary depending upon available experimental measurements and the existing predictive capability of the numerical model. Therefore, considering the finite resources, effectively choosing one approach over the other becomes an issue of efficient allocation of available resources.

By focusing on the relative benefits of each approach strictly from the perspective of predictive capability, the authors propose a resource allocation framework that aids in the selection between further experimentation and code development. The proposed framework extends the work by Unal et al. (2004) for certification of nuclear fuels by incorporating a resource allocation decision node. Figure 3.1 illustrates the proposed

framework, in which the validation process begins with an initial numerical model[10] and a first set of available physical measurements.

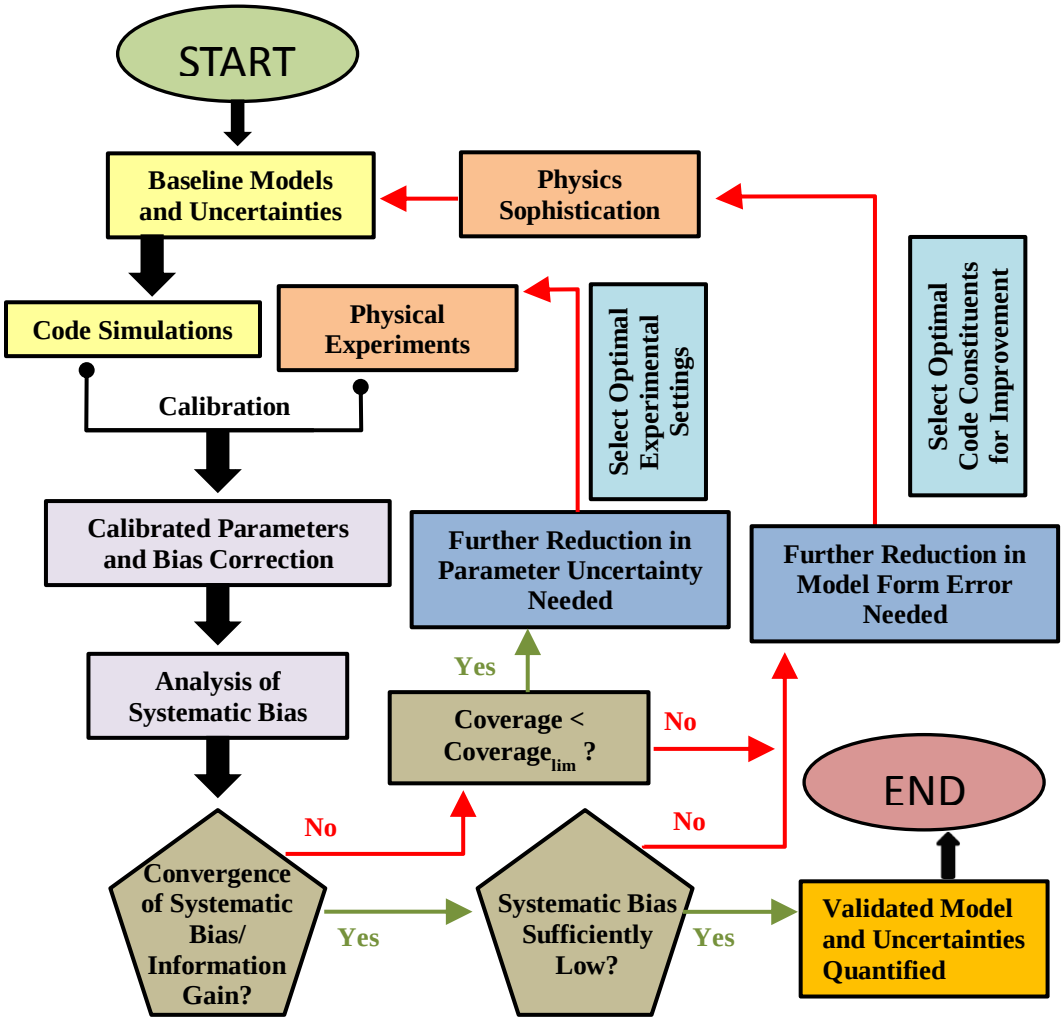

Figure 3.1: Predictive capability framework.

This initial model is imprecise due to parameter uncertainties and inaccurate due to systematic bias and thus must be calibrated against experimental data. Such calibration is possible using a variety of back-calculation techniques. In this study, a statistical

[10]Hereafter, the term "model" is used to mean both codes and numerical models.

inference procedure (Kennedy and O'Hagan, 2001; Higdon et al., 2008a; 2008b) is used to infer these imprecise model parameters. This calibration procedure conditions the probability distributions of input parameters to the experimental evidence resulting in a refined posterior probability distribution of these parameters. Thus, as more physical measurements become available, the uncertainties in the input parameter values are reduced resulting in a reduction in prediction uncertainty. In the present study, this reduction in prediction uncertainty is traced by quantifying the information gained using an entropy-based metric.

As approximations of reality, numerical models have a fundamental inability to reproduce the reality that cannot be remedied by calibrating the input parameters. This fundamental inability results in a systematic bias in prediction that's determination is closely tied with the calibrated parameter values. Thus, in this study, resolution of parameter uncertainty and systematic bias is treated in a simultaneous and intertwined manner. Expanding the experimental campaign by conducting new experiments can reduce parameter uncertainty and produce a more refined estimate of systematic bias. On the other hand, an improvement in the description of physics and/or engineering principles can lead to a reduction in the systematic bias.

At the decision node in the framework, the decision maker must assess the predictive capability of the numerical model by evaluating the stabilization (or convergence) of the systematic bias and the information gain metric throughout the domain of applicability. As the absence of stabilization indicates that the parameter uncertainty has not yet been fully mitigated, systematic bias in the model predictions is

not yet properly defined. In this case, resources must be allocated for experimentation to better infer the parameter values and systematic bias. In some cases, poorly predicting models may fail to exhibit convergence in systematic bias and information gain even after a significant number of experiments are conducted; therefore, at this point in the framework, a measure of how well the current experiments cover the domain of applicability (referred to as coverage) is compared to a case-specific coverage limit. As long as the coverage limit has not been exceeded, resources are allocated for experimentation. However, if the coverage limit is exceeded, resources are to be allocated for code development.

If stabilization is observed, however, then the parameter uncertainty can be expected to be adequately reduced[11], and the inferred systematic bias can be considered to be a proper representation of the incompleteness and inexactness of the model, i.e. model form error. In this case, as further experimentation would only marginally reduce the systematic bias, allocating resources to experiments cannot be justified. The decision maker must then evaluate if the remaining systematic bias is at an acceptably low level[12] for the application of interest. The model is considered valid for the particular application if the systematic bias is acceptably low. On the other hand, if the systematic bias is too high, the physics and/or engineering principles in the model need improvement. It is

[11]With statement comes a caveat, only the epistemic component of the parameter uncertainty that is due to lack of knowledge can be mitigated, and therefore, the parameter uncertainty can only be reduced to the levels of aleatory uncertainty. Of course, if all the uncertain and sensitive parameters are not incorporated in the uncertainty quantification, then there will always be a remaining uncertainty due to the parameters that are not accounted for. Therefore, one cannot estimate the incompleteness of the model in a deterministic manner.

[12]What constitutes "acceptably low level" depends on the particular problem being solved.

important to note that a more sophisticated model may have larger numbers of uncertain parameters and may require a more extensive experimental campaign to mitigate parameter uncertainty. The framework in Figure 3.1 therefore loops through the aforementioned steps until the systematic bias and information gain have converged to the acceptable levels.

The application of this proposed framework is demonstrated on the viscoplastic self-consistent (VPSC) material code for modeling the plastic deformation of polycrystals as described in Section 3.2. The experimental data utilized in the calibration process is described in Section 3.3. The methods used for calibrating model parameters, computing the systematic bias, and calculating information gain are discussed in Section 3.4. Results and discussions are presented in Section 3.5. Concluding remarks, limitations of this framework, and the necessary future work and are provided in Section 3.6.

3.2. VPSC Model

Lebensohn and Tomé developed a viscoplastic self-consistent (VPSC) material model for modeling the plastic deformation of polycrystals (Lebensohn et al., 1993). A polycrystal is modeled by a set of single crystals (grains) with initial crystallographic orientations that represent the initial texture of the aggregate and evolve during plastic deformation. In turn, each grain is treated as an ellipsoidal inclusion with anisotropic viscoplastic properties, deforming in a homogenous equivalent medium that has the *a priori* unknown average properties of the aggregate. This leads to a relation between the strain-rate and stress in each individual grain with the global stress and strain-rate of the aggregate through localization equations. The viscoplastic deformation of the crystals

occurs by dislocation motion and therefore can be modeled in terms of constitutive relations between the deviatoric stress and strain-rate tensors. Viscoplastic deformation will occur when a slip system activates and dislocations move under an applied stress. The final deformation is obtained in the VPSC formulation through imposing a macroscopic strain-rate during each incremental deformation step. The strain-rate and stress from each previous step is used as the starting values for the next step. Stress-strain curves and texture development constitute the typical output of a VPSC calculation. Here, two versions of the VPSC formulation are utilized: the original glide-only (G) model, used for predictions of plasticity of polycrystals; and the climb-and-glide (C&G) version, with improved physics for the prediction of polycrystal response under creep conditions[13].

3.2.1. Glide VPSC

In the G version of the VPSC code, a Schmid-type constitutive behavior is used to describe the dislocation motion in the constituent single-crystals (Lebensohn et al., 1993). As such, dislocations lie and move within the slip plane and are activated by shear stresses, their motion can only accommodate simple shear deformation on this plane. Glide activity in several slip planes is able to accommodate arbitrary deformation applied to the crystal. The constitutive equation at the single crystal level is expressed as:

[13]Prior to performing validation on the VPSC model, verification is conducted (Recall, in Section 3.1 it is stressed that verifying the numerical accuracy of the mathematical codes must be addressed prior to validation endeavors). In (Lebensohn et al., 2004), the VPSC model is verified and numerical errors are quantified by comparison with "exact" full-field formulations.

$$\dot{\varepsilon} = \dot{\gamma}_o \sum_{s=1}^{N_s} m^s \left(\frac{\left| m^s : \sigma \right|}{\tau_o^s} \right)^{n_g} \operatorname{sgn}\left(m^s : \sigma \right) \tag{3.1}$$

where σ is the stress applied to the crystal and $\dot{\varepsilon}$ is the strain-rate, accommodated by glide; m^s and τ_o^s are the Schmid tensor and the critical resolved shear stress associated with glide in the system(s), respectively, The stress exponent, n_g, represents the inverse of rate-sensitivity for glide activity, and $\dot{\gamma}_o$ denotes a normalization factor. The single crystal equation for strain-rate is summed over all active slip systems, N_s.

3.2.2. Climb-and-Glide VPSC

Lebensohn et al.'s (2010) more sophisticated constitutive model for aggregates of single crystals deforming by climb and glide is an improvement to the original VPSC approach that considers deformations by glide only. At temperatures below 50 percent of the melting temperature, glide-controlled creep dominates; however, when higher temperatures are present, local non-equilibrium concentrations of point defects interacting with dislocations allow for dislocations to climb in addition to glide. Dislocation climb becomes very relevant in high-temperature plasticity and irradiation creep. The direction of dislocation motion is determined by the velocity vector made up of two components: the glide velocity (lies in glide plane) and the climb velocity (normal to the glide plane). The glide component depends upon the shear stress component acting on the glide plane while the climb component depends on the full stress tensor. The extension of Equation (3.1) to the C&G case is expressed as:

$$\dot{\varepsilon} = \dot{\gamma}_o \sum_{s=1}^{N_s} \left\{ m^s \left(\frac{\left| m^s : \sigma \right|}{\tau_o^s} \right)^{n_g} \times \mathrm{sgn}\left(m^s : \sigma \right) + c^s \left(\frac{\left| c^s : \sigma \right|}{\sigma_o^s} \right)^{n_c} \times \mathrm{sgn}\left(c^s : \sigma \right) \right\} \tag{3.2}$$

where c^s and σ_o^s are, respectively, the climb tensor and a critical stress associated with climb in the system(s) and n_c is the stress exponent associated with climb. In comparing Equations (3.1) and (3.2), it is evident that the additional physics of the C&G model result in an increased number of parameters over the G model.

It is important to note that n_g and τ_o^s (glide stress exponent and critical resolved shear stress associated with glide) are temperature, strain-rate, and microstructure-dependent (Lebensohn et al., 2010). Likewise, n_c and σ_o^s (climb stress exponent and critical stress associated with climb) are dependent of the same variables. This latter dependency however, is much more complex due to not only the dynamics and interactions of dislocations but also the interactions between point defects and dislocations. Additional versions of the code, including an improved the C&G model (Lebensohn et al., 2011) and an atomistic scale coupled model, are currently being developed and will be added to future analysis.

3.3. Application of the VPSC Code to 5182 Aluminum Alloy

Experiments performed on 5182 aluminum samples with an initial (001) ("cube") texture, deformed in compression are reported in Stout et al. (1998a; 1998b). Stress-strain curves and final textures (in terms of inverse pole figures) are measured for various temperatures (from 22°C to ~550°C) and strain-rates (1 s^{-1} and 10^{-3} s^{-1}). The experiments are performed until the specimens reached a true strain of 0.6. Final textures are available

for seven of the 11 stress-strain curves shown in Figure 3.2. In this study, the stress corresponding to the maximum measured strain of 0.6 and the intensities of the textures corresponding to the (001) and (101) corners of the inverse pole figure are used for validation of the VPSC model. Note that although a complete, quantitative description of crystallographic textures requires, in general, a large number of parameters (e.g. weights associated with a partition of a 3-D orientation space), the final compression textures of the 5182 Al samples can be characterized by, at most, two components with associated intensities, corresponding to a retained (001) "cube" texture and/or a (101) "compression" texture. These chosen features offer a low-dimensional yet highly informative metric for validation as shown by the information gain metric presented later in Section 3.5.

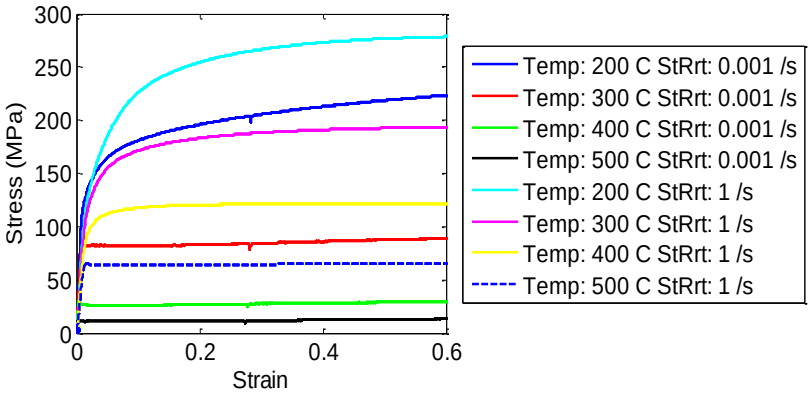

Figure 3.2: 5182 Al stress-strain curves.

3.3.1. Experimental observations

Table 3.1 lists the experimental settings and measured stress and texture features for the experimental campaign performed to explore the domain of applicability defined by temperature and strain-rate. The initial results of the study in Stout et al. (1998a) show typical stress-strain behavior, in which yielding is followed by strain-hardening at low

temperatures. At higher temperatures however, very little work-hardening and lower yield stresses are observed. Additionally, for yield stresses below 50 MPa, negative work hardening occurs with a clear upper/lower yield point; while for stresses above 50 MPa, a yield point is not observable (Stout et al., 1998a). Textures at elevated temperatures are likewise inconsistent with standard glide-only deformation textures, while textures at lower temperatures develop a (101) fiber texture, typical of unixial compression applied to a fcc polycrystal. Experiments run at 500°C and 550°C with a strain-rate of 10^{-3} s^{-1} displayed a (001) cube component (generally thought of as a recrystallization texture not a deformation texture) and almost no (101) deformation component. For 400°C, 500°C, and 550°C with 1 s^{-1} strain-rate experiments both (001) and (101) textures are observed (Stout et al., 1998b). To explain the difference in the combination of textures, Stout et al. (1998b) assume a sharp decrease in rate sensitivity as strain-rate increases.

Table 3.1: Experimental results for 5182 Al

Experiment	Temperature (C°)	Strain-Rate (s^{-1})	Stress (MPa) @ Strain=0.6	Texture Intensity (001)	Texture Intensity (101)
A	200	10^{-3}	226.2	1.00-1.41	4.00-6.00
B	300	10^{-3}	91.4	0.58-0.71	4.00-6.00
C	400	10^{-3}	30.6	NA	NA
D	500	10^{-3}	14.9	4.00-6.00	2.00-2.83
E	200	1	280.0	NA	NA
F	300	1	193.7	NA	NA
G	400	1	121.3	1.41-2.00	4.00-6.00
H	500	1	65.5	2.00-2.83	0.00-0.58
I	550	1	43.0	2.83-4.00	0.58-0.71
J	350	10^{-3}	50.0	2.00-2.83	2.83-4.00
K	550	10^{-3}	7.0	NA	NA

3.3.2. Prior work

Utilizing the VPSC model to make texture predictions on the 5182 aluminum data reported by Stout et al. (1998b), Lebensohn et al. (2010) observes that the (001) "cube" component at high temperatures and low strain-rates are due to an increase in climb activity at those external settings. Seven different VPSC simulations of texture evolution are computed through the analysis of various glide-only and climb-and-glide scenarios, an analysis of which shows that an increase in rate-sensitivity contributes to prevalence of the (001) cube component. The texture simulated using an equal climb-to-glide activity ratio most accurately predicts the experimental texture for the 400°C and 10^{-3} s^{-1} experimental case, however. As proposed in Lebensohn et al. (2010), the final retained (001) "cube" component is achieved by means of an increase in climb activity, which can be explained by the fact that climb mechanism accommodates plastic deformation applied to a single crystal involving a reduced plastic spin (crystal rotation) as compared with glide (see Lebensohn et al. (2010) for details). Note that hereafter, the response features of interest (stress at the maximum measured strain of 0.6, (001) "cube" texture, and (101) "compression" texture) will be referred to as: maximum stress, texture 001, and texture 101, respectively.

3.4. Calibration against Experimental Data

During experiments conducted on 5182 aluminum alloy, temperature and strain-rate are controlled (also known as control parameters). The upper and lower bounds of these control parameters, given in Table 3.2, define the domain of applicability, within which the VPSC model will be executed. The VPSC model studied herein however,

omits temperature-dependency. Therefore, the stress exponents and critical stresses are considered as proxies for temperature to describe the mechanical response at large strains for various temperatures (Lebensohn et al., 2010) (see Section 3.4.1).

As this study is concerned with the ability of the VPSC model to predict the saturation stress, the critical stresses are treated as uncertain model parameters (recall Equation 3.2). Note that the initial strain-hardening is not of interest and thus, the critical stresses are assumed constant throughout the deformation. In addition to these critical stresses, the stress exponents described in Equations (3.1) and (3.2) are identified as highly uncertain parameters. Thus, the two parameters of interest for the G model are then the glide stress exponent, n_g, and the initial critical resolved shear stress for glide, τ_o^s. For the C&G model, in addition to the two parameters corresponding to the glide mechanism, the parameters of interest also include two more parameters, the climb stress exponent, n_c, and critical stress associated with climb, σ_o^s. The plausible ranges for these variables are determined by expert opinion and are listed in Table 3.2.

Table 3.2: Control and uncertain model parameter values

	Parameter	Minimum	Maximum
Control Parameters	Temperature (C°)	180	570
	Strain-Rate (s^{-1})	0.0005	1.05
Uncertain Model Parameters	n_g	1	5
	τ_o^s (MPa)	1.2	1343.4
	n_c	1	5
	σ_o^s (MPa)	1.2	6045.4

3.4.1. Correlation function

The dependency of the uncertain model parameters (the stress exponent(s), *n*, and initial critical shear stress(es) for glide and/or climb, generically referred to as *τ)* upon the control parameters (temperature and strain-rate) makes it implausible to search for a single set of parameter values for the stress exponent and critical stress that will yield satisfactory agreement between the predictions and experiments throughout the domain of applicability. Rather, exploiting the available experimental data, a correlation function is constructed to represent the dependency of the stress exponent and critical stress on control parameters. Using a nonlinear constrained optimization algorithm (Powell, 1978; Waltz et al., 2006), the least-square objective function given in Equation (3.3) is minimized for the experimental settings given in Table I. Note that for the C&G model four variables are considered for optimization while only two variables are needed for the G model. The optimized values of stress exponents and critical stresses are obtained via the objective function given by:

$$x_{optim} = \sqrt{\left(\frac{\left|\sigma_{sim} - \sigma_{exp}\right|}{\sigma_{exp}}\right)^2 + \left(\frac{\left|\psi_{sim}^{(001)} - \psi_{\exp}^{(001)}\right|}{\psi_{\exp}^{(001)}}\right)^2 + \left(\frac{\left|\psi_{sim}^{(101)} - \psi_{\exp}^{(101)}\right|}{\psi_{\exp}^{(101)}}\right)^2} \tag{3.3}$$

where σ_{exp} represents the stress at the maximum measured strain (strain = 0.6), and σ_{sim} represents the corresponding predicted stress. In Equation (3.3), $\psi_{\exp}^{(001)}$ and $\psi_{sim}^{(001)}$ denote the measured and predicted texture 001 intensities, respectively, and $\psi_{\exp}^{(101)}$ and $\psi_{sim}^{(101)}$ represent the texture 101 intensities. Note that for each experimental setting given in Table I, the $\psi_{\exp}^{(001)}$ and $\psi_{\exp}^{(101)}$ values are given as ranges to incorporate the experimental

uncertainty, and thus constraints are imposed on the second and third terms in Equation (3.3) as follows:

$$\left(\frac{\left|\psi_{sim}^{(Tx)}-\psi_{\exp}^{(Tx)}\right|}{\psi_{\exp}^{(Tx)}}\right)^2 = 0 \text{ for } (\psi_{\exp}^{(Tx)})_L \le \psi_{sim}^{(Tx)} \le (\psi_{\exp}^{(Tx)})_U \tag{3.4}$$

$$\left(\frac{\left|\psi_{sim}^{(Tx)}-\psi_{\exp}^{(Tx)}\right|}{\psi_{\exp}^{(Tx)}}\right)^2 = \left(\frac{\left|\psi_{sim}^{(Tx)}-(\psi_{\exp}^{(Tx)})_L\right|}{(\psi_{\exp}^{(Tx)})_L}\right)^2 \text{ for } \psi_{sim}^{(Tx)} < (\psi_{\exp}^{(Tx)})_L \tag{3.5}$$

$$\left(\frac{\left|\psi_{sim}^{(Tx)}-\psi_{\exp}^{(Tx)}\right|}{\psi_{\exp}^{(Tx)}}\right)^2 = \left(\frac{\left|\psi_{sim}^{(Tx)}-(\psi_{\exp}^{(Tx)})_U\right|}{(\psi_{\exp}^{(Tx)})_U}\right)^2 \text{ for } \psi_{sim}^{(Tx)} > (\psi_{\exp}^{(Tx)})_U \tag{3.6}$$

In Equations (3.4)-(3.6), the superscript Tx denotes the texture 001 or 101 and the subscripts "U" and "L" denote the upper and lower bound of the given ranges for texture data in Table I. Because increased stress leads to an increase in glide at the expense of climb, a constraint that $n_c < n_g$ for the entire domain is imposed. The optimized, deterministic point estimates for the uncertain model parameters for the G and C&G model are plotted in Figures 3.3 and 3.4, respectively.

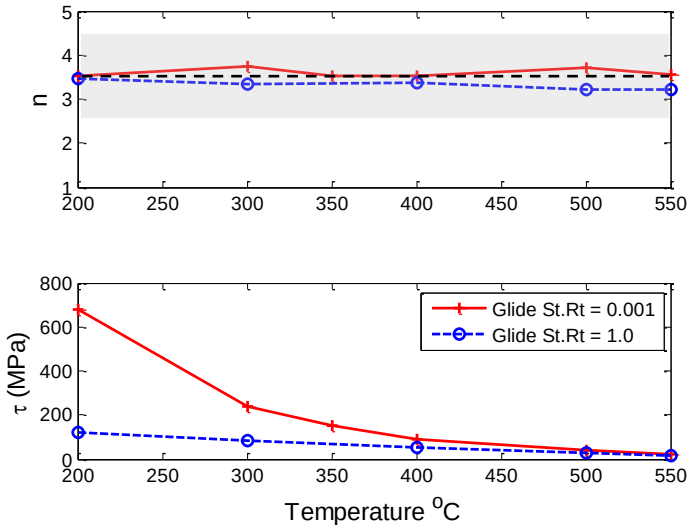

Figure 3.3: Optimized point estimates for rate sensitivity (n_g) (top) and critical stress (τ_0) (bottom) for the G model.

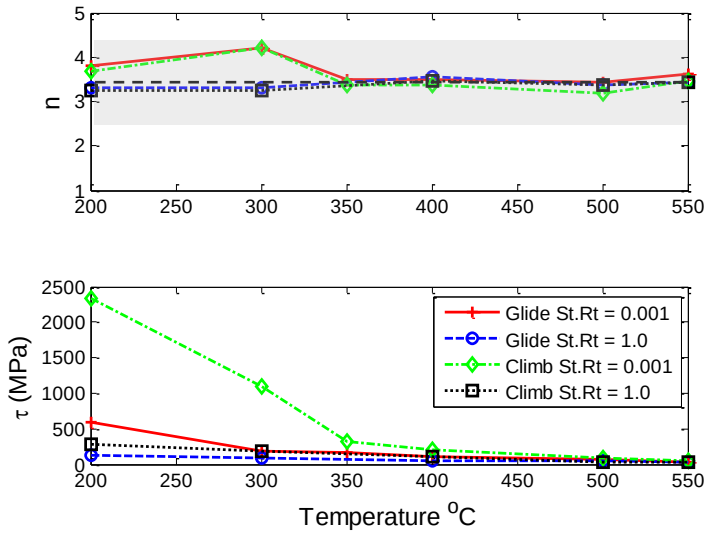

Figure 3.4: Optimized point estimates for rate sensitivities (n_g and n_c) (top) and critical stresses (τ_0 and σ_0) (bottom) for the C&G model.

It can be seen from Figures 3.3 and 3.4 that the optimized stress exponent, *n*, values concentrate around 3.5 without exhibiting a clear dependency on the temperature or strain-rate. Therefore, the parameters n_g and n_c are considered to be independent of temperature and strain-rate. Furthermore, the optimization process indicates that bounds for the *n* parameters can be restricted further (see the gray band on Figures 3.3 and 3.4) from the originally proposed range given by expert judgment in Table 3.2. The further constrained ranges for the *n* parameters are listed in Tables 3.3 and 3.4.

On the other hand, the optimized critical stress, τ , values in Figures 3.3 and 3.4 exhibit an exponential relationship with temperature and a linear relationship with strain-rate. Therefore, suitable correlation functions for the critical stresses require an exponential fit along the temperature domain and a linear correlation along the strain-rate domain, which can be expressed as:

$$\tau = ae^{bT} + \frac{\left(\dot{\varepsilon} - \dot{\varepsilon}_1\right) ce^{dT} - \left(\dot{\varepsilon} - \dot{\varepsilon}_1\right) ae^{bT}}{\left(\dot{\varepsilon}_2 - \dot{\varepsilon}_1\right)} \tag{3.7}$$

where $\dot{\varepsilon}$ is the strain-rate and T is the temperature. The variables *a* and *c* are the leading intercept coefficients for the exponential fit for critical stress for a strain-rate of $\dot{\varepsilon}_1 = 0.001$ and $\dot{\varepsilon}_2 = 1$, respectively. Likewise, *b* and *d* are the decay rate coefficients for the exponential fit. The computed values for the coefficients, shown in Tables 3.3 and 3.4 for the two models, are used as calibration parameters (i.e. uncertain parameters for which posterior distributions will be inferred during the calibration process) in place of critical stresses.

The stress exponents as well as the variables for the exponential fit are accepted as the parameters to be calibrated resulting in a total of five calibration parameters for the G model and ten for the C&G model as listed in Tables 3.3 and 3.4. To account for uncertainties in the experiments, the values of the correlation function coefficients are allowed to vary from their optimized fits. The variance of the coefficients *a* and *c* is set to $\pm 25\%$, and the variance of coefficients *b* and *d* is set to $\pm 20\%$. These ranges, given in Tables 3.3 and 3.4, are selected such that all the available experimental data is encompassed by the family of functions created with varying coefficients.

Table 3.3: Calibration parameters for the G model

Parameter	Optimized/Mean Value	Min	Max
a_g	4577.1	3432.8	5721.4
b_g	-0.01	-0.008	-0.012
c_g	372.48	279.36	465.60
d_g	-0.005	-0.004	-0.006
n_g	3.5	2.5	4.5

Table 3.4: Calibration parameters for the C&G model

Parameter	Optimized/Mean Value	Min	Max
a_g	2970.2	2227.65	3712.75
b_g	-0.008	-0.0064	-0.0096
c_g	281.7	211.275	352.125
d_g	-0.004	-0.0032	-0.0048
n_g	3.5	2.5	4.5
a_c	24727	18545.25	30908.75
b_c	-0.012	-0.0096	-0.0144
c_c	1595.2	1196.4	1994
d_c	-0.008	-0.0064	-0.0096
n_c	3.5	2.5	4.5

3.4.2. Model calibration

For the calibration of the imprecise parameters of the VPSC model, a statistical inference approach is implemented. In particular, a multivariate generalization (Higdon et al., 2008a; 2008b) of the Bayesian calibration method proposed by Kennedy and O'Hagan (2001) is used, in which experimental data are combined with the model predictions to carry out statistical inference while considering the inadequacy of the model. In this framework (Kennedy and O'Hagan, 2001; Higdon et al., 2008a; 2008b), the experimental observation, y(x), is given by:

$$\mathbf{y}(\mathbf{x}) = \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta}) + \boldsymbol{\delta}(\mathbf{x}) + \boldsymbol{\varepsilon}(\mathbf{x}) \tag{3.8}$$

in which $\boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta})$ denotes the model predictions, $\boldsymbol{\delta}(\mathbf{x})$ represents the estimated systematic bias between reality and the predictions, and $\boldsymbol{\varepsilon}(\mathbf{x})$ denotes the experimental error. Here, $\mathbf{x}$

represents settings at which observations are made (i.e. control parameters), and $\boldsymbol{\theta}$ denotes the best values for the calibration parameters **t**.

For many practical problems, numerical models are complex and computationally demanding and thus only a limited number of runs are possible. To mitigate this problem, an inexpensive surrogate (also known as emulator) can be trained to substitute for the numerical model to obtain predictions at untried settings. Here, a Gaussian process (GP) emulator is used to represent the numerical model predictions, $\boldsymbol{\eta}(\mathbf{x},\mathbf{t})$, which is specified by a mean function, $\mu(\mathbf{x},\mathbf{t})$, and a covariance function defined by Higdon et al. (2008a) as:

$$\mathrm{Cov}((\mathbf{x},\mathbf{t}),(\mathbf{x}',\mathbf{t}')) = \frac{1}{\lambda_\eta}\prod_{k=1}^{p_x}\rho_{\eta k}^{4(x_k-x_k')^2}\times\prod_{k=1}^{p_t}(\rho_{\eta k,p_x+k})^{4\left(t_k-x_k'\right)^2} \tag{3.9}$$

where λ_η and the $\rho_{\eta k}$ vector are the so-called hyper-parameters for the GP emulator of the model predictions which control the marginal precision of $\boldsymbol{\eta}(\mathbf{x},\mathbf{t})$ and the dependence strength in the components of the **x** *and* **t** directions**,** respectively; p_x and p_t are the number of control and calibration parameters, respectively. Similarly, for the estimated systematic bias $\boldsymbol{\delta}(\mathbf{x})$, a GP emulator is employed with a zero mean function and a covariance function defined by Higdon et al. (2008a) as:

$$\mathrm{Cov}((\mathbf{x},\mathbf{x}')) = \frac{1}{\lambda_\delta}\prod_{k=1}^{p_x}\rho_{\delta k}^{4(x_k-x_k')^2} \tag{3.10}$$

where λ_δ and $\rho_{\delta k}$ are hyper-parameters for the GP emulator estimating systematic bias which control the marginal precision of $\boldsymbol{\delta}(\mathbf{x})$ and the dependence strength in the

components of the **x** direction**,** respectively. The hyper-parameters ensure a smooth and differentiable form for both $\boldsymbol{\eta}(\mathbf{x},\mathbf{t})$ and $\boldsymbol{\delta}(\mathbf{x})$.

In the Bayesian calibration framework, the true but unknown values of the calibration parameters, $\boldsymbol{\theta}$, are inferred exploiting the availability of the experimental data. Existing knowledge of calibration parameters as well as the hyper-parameters of the GP emulators can be incorporated through prior distributions, which are then updated to obtain the joint posterior distributions based on the experimental data. The posterior distribution conditioned on experimental data is given by:

$$\begin{aligned}&\pi(\boldsymbol{\theta},\mu,\lambda_\eta,\boldsymbol{\rho}_\eta,\lambda_\delta,\boldsymbol{\rho}_\delta\,|\,\mathbf{D})\ \alpha\\&L(\mathbf{D}\,|\,\boldsymbol{\theta},\mu,\lambda_\eta,\boldsymbol{\rho}_\eta,\lambda_\delta,\boldsymbol{\rho}_\delta,\textstyle\sum_y)\times\pi(\boldsymbol{\theta})\times\pi(\mu)\times\pi(\lambda_\eta)\times\pi(\boldsymbol{\rho}_\eta)\times\pi(\lambda_\delta)\times\pi(\boldsymbol{\rho}_\delta)\end{aligned} \tag{3.11}$$

where **D** is the joint vector of experimental data and numerical model outputs, $L(\mathbf{D}\,|\,\boldsymbol{\theta},\cdot)$ is the likelihood function, $\sum_y$ is the observation covariance matrix, and $\pi(\cdot)$ is the prior distributions (for detailed discussion see Higdon et al. (2008a)).

A Markov chain Monte Carlo (MCMC) algorithm, specifically Metropolis-Hasting algorithm, is used to explore the posterior distributions for both the calibration parameters and the aforementioned hyper-parameters. During the MCMC random walk, calibration parameter values that generate predictions with greater fidelity to the experimental data over the domain of applicability are accepted based on the established acceptance criterion for the likelihood function. Upon obtaining the posterior distribution for the calibration parameters and hyper-parameters of the GP models, predictions $\boldsymbol{\eta}(\mathbf{x}^*,\boldsymbol{\theta})$ and the systematic bias $\boldsymbol{\delta}(\mathbf{x}^*)$ can be estimated at untested input settings, $\mathbf{x}^*$, using the GP emulators.

Herein, a 140-run Latin-hypercube (LHS) design is used to construct the GP emulators for the model predictions and systematic bias for the G model, while a 240-run LHS design is used for the C&G model. For the calibration parameters, a uniform prior distribution is assumed between the ranges given in Tables 3.3 and 3.4. 10,000 MCMC iterations are used to estimate the posterior distribution of the calibration parameters. Exercising the GP emulators, predictions are obtained for 500 linearly spaced samples from the posterior distributions of the calibration parameters and GP hyper-parameters. To compare the calibrated model predictions against experimental data, predictions, $\boldsymbol{\eta}(\mathbf{x},\boldsymbol{\theta})$, are generated at experimental settings shown in Table 3.1.

The relative disagreement between predictions and experimental results, i.e. systematic bias, is hypothesized to converge as the number of the experiments used in the calibration process increases. However, the systematic bias is herein treated as a scalar representative value by computing the difference between the average of the sampled model predictions and the experimental observation; it is not a direct quantification of the reduction in prediction uncertainty. Therefore, the information gain metric described in the next section is proposed to quantify the reduction in prediction uncertainty.

3.4.3. Information gain

As the information gain is equal to the amount of uncertainty removed, entropy, defined as a measure of uncertainty, is equivalent to the amount of information (Shannon, 1948). Herein, an information gain metric based on Shannon entropy (Shannon, 1948) is utilized to quantify the prediction uncertainty as additional experiments become available

to condition the posterior distributions. For a discrete random variable, Z, with a probability mass function, p(z), the entropy is expressed as:

$$H(Z) = -\sum_{z \in Z} p(z) \log p(z) \tag{3.12}$$

where the logarithm to base 2 is used to measure entropy in bits. Here Z represents the calibrated model predictions at respective experimental settings (recall Table 3.1), as increasing numbers of experiments are used in the calibration process. The information gain metric, expressed in percentage, is then calculated using the following relationship:

$$\text{Info-Gain(i)(\%)} = \left(\frac{H_{ref} - H_{\text{expt(i)}}}{H_{ref}} \right) * 100 \tag{3.13}$$

where H_{ref} is the entropy calculated for the model predictions obtained with the prior distributions of the calibration parameters. $H_{\text{expt(i)}}$ is the entropy of the calibrated model predictions, in which *i* number of experiments are used in the calibration process.

While the information gain metric is an excellent tool for quantifying the reduction in uncertainty of the model predictions, it does not make any assertions about the systematic bias associated with a particular model and should not be used to assess the accuracy of one model over another. Therefore, both the convergence of systematic bias and information gain must be evaluated while discerning the necessity of additional experiments or further code development.

3.5. Results and Discussion

This section demonstrates the application of the proposed framework and provides a discussion on the computed systematic bias and information gain as the two

versions (G and C&G) of the VPSC model are calibrated using one through 11 available experiments in the sequence shown in Table 3.5. The calibrated VPSC models are executed to predict maximum stress and texture 001 and 101 intensities at different experimental settings.

Table 3.5: Experimental cases (see Table I for the experimental settings)

Case	Experiments
1	A
2	A,B
3	A,B,J
4	A,B,J,D
5	A,B,J,D,G
6	A,B,J,D,G,H
7	A,B,J,D,G,H,I
8	A,B,J,D,G,H,I,C
9	A,B,J,D,G,H,I,C,K
10	A,B,J,D,G,H,I,C,K,E
11	A,B,J,D,G,H,I,C,K,E,F

3.5.1. G model

Figure 3.5 shows a reduction in uncertainty in the calibration parameters where all 11 experiments are used in the calibration process of the G model compared to the case where only one experiment is used (with the exception of d_g[14]). Figure 3.6 shows a box plot of the three predicted features, where the horizontal lines represent the bounds of the

[14]It is important to note that the posterior distributions of insensitive parameters may not narrow as more experimental data is added. If a parameter has little effect (i.e., low sensitivity) relative to other parameters on the output of interest, many alternative values for this parameter could be accepted during the calibration process.

experiments at settings D and I (recall Table 3.1). Overall, reduction in uncertainty in model predictions is observed as the number of experiments increases. Note that the predictions at the other nine experimental settings also show similar trends. Figure 3.6 illustrates that systematic bias decreases and ultimately converges as the number of experiments used in the analysis increases. As evident from Figures 3.6(c-f), a level of systematic bias remains between predictions and experiments.

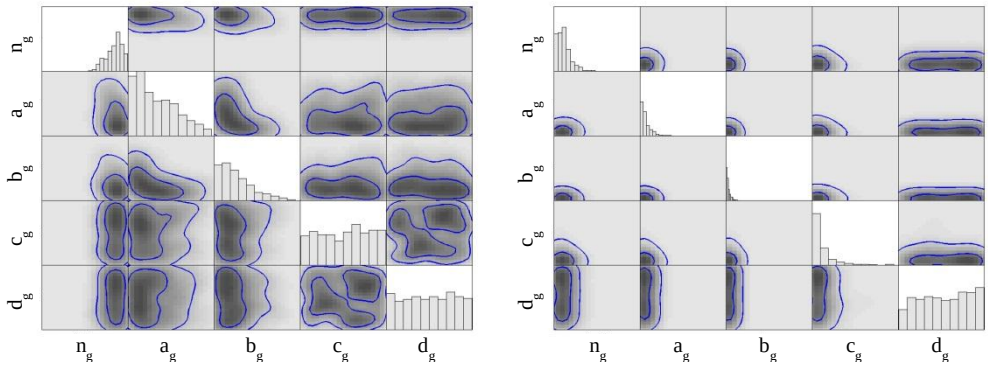

Figure 3.5: Posterior distribution for the G model (Left: one Experiment, Right: 11 Experiments).

The systematic bias first is normalized as a percentage of the experimental observation throughout the domain of applicability and then averaged to obtain the mean[15]. Figures 3.7-3.9 depict the mean systematic bias for the maximum stress, texture 001, and texture 101 predictions at each experimental setting, as a function of the number of experiments used in the calibration process. For the maximum stress output, little convergent behavior is evident; while for the texture outputs, upon addition of the fourth

[15]For each experimental setting, the systematic bias is obtained as a distribution, which is averaged to reduce the dimensionality to a scalar value.

experiment, the systematic bias of the textures for all prediction settings converges as shown in Figures 3.8 and 3.9.

The information gain is computed for the model predictions at settings where experiments are available (Table 3.5). For brevity, however, Figure 3.10 shows the information gain plots for the maximum stress, texture 001, and texture 101 predictions at settings D and I, respectively. Remaining experimental settings show similar trends as the model calibration proceeds consecutively through these 11 experiments. For stress predictions, the information gain monotonically increases and ultimately converges. Slight fluctuations are expected given the stochastic nature of the approach adopted in the study. However, for the texture predictions, a convergent behavior is not observed, indicating the G model's inability to reproduce experimental texture data at certain regions in the domain of applicability.

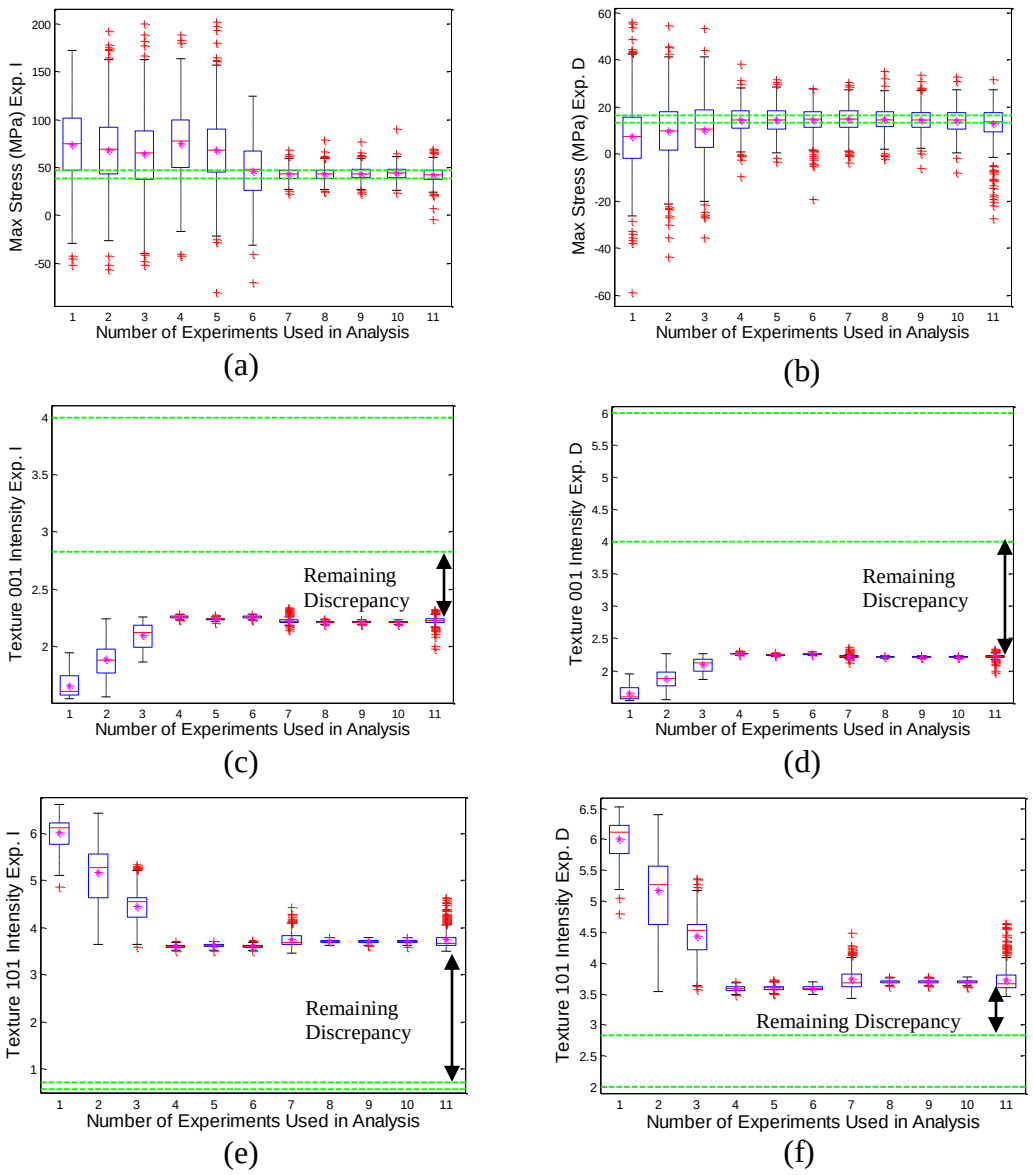

Figure 3.6: Prediction statistics for the G model (Left: Experiment I, Right: Experiment D).

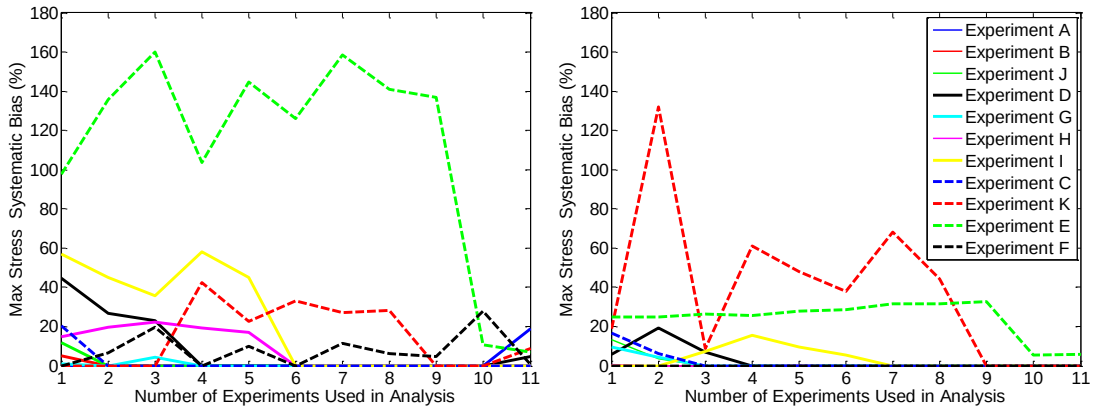

Figure 3.7: Maximum stress systematic bias for each experiment (Left: G model, Right: C&G model).

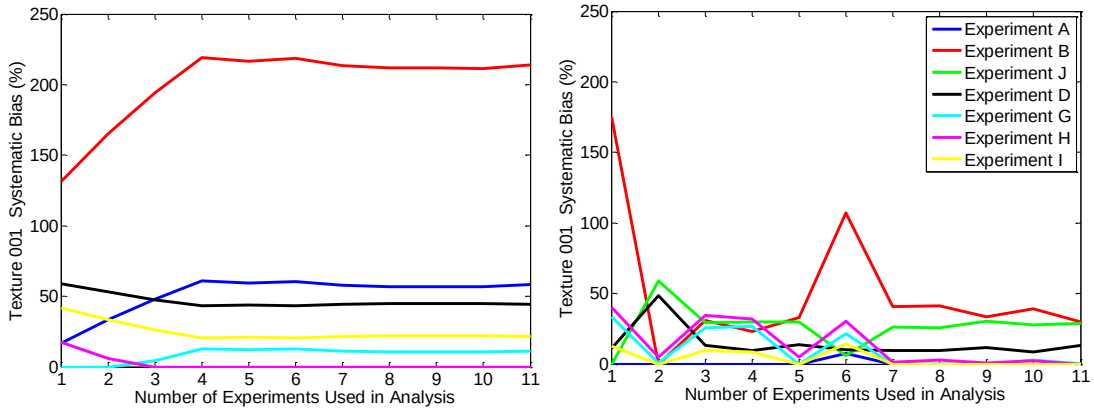

Figure 3.8: Texture 001 systematic bias for each experiment (Left: G model, Right: C&G model).

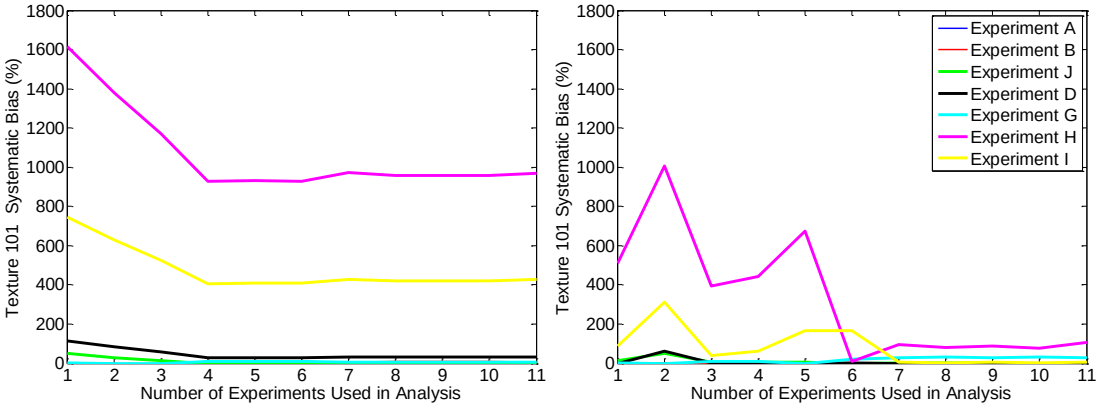

Figure 3.9: Texture 101 systematic bias for each experiment (Left: G model, Right: C&G model).

In Figure 3.10, even though the information gain metric for both texture outputs approaches 100% for some cases, Figures 3.8 and 3.9 depict a systematic bias well in excess of 100% for the texture predictions at settings D and I indicating that uncertainty reduction does not guarantee that the model is making accurate predictions. In fact, Figure 3.9 reveals a systematic bias for texture 101 over 400% for setting I, and as high as 1000% for setting H. For purposes of illustration, the G model's systematic bias is deemed unacceptably high. Since further reduction in parameter uncertainty through conducting more experiments is not feasible, the rigor in which the physics principles are modeled must be improved, which leads to the use of C&G model as it will be discussed below.

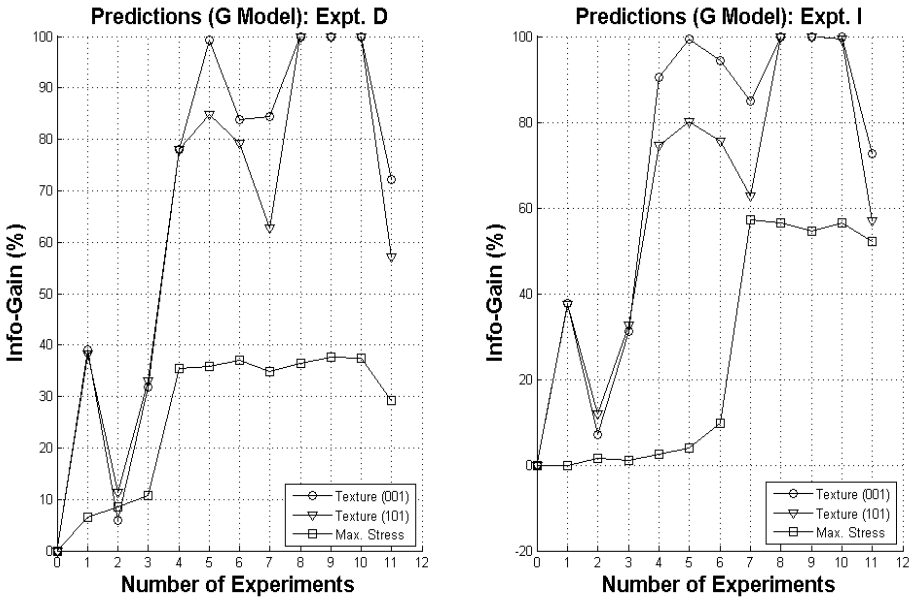

Figure 3.10: Model predictions at Experiment D and Experiment I: Information gain from experiments used for calibration (G model).

3.5.2. C&G model

Figure 3.11 illustrates the uncertainty reduction in the calibration parameters as the number of experiments used for the C&G model calibration is increased through each of the 11 experiments. In Figure 3.12, box plots of the outputs for the experiment settings D and I are shown with the reduction of prediction uncertainty clearly seen. Additionally, the reduction of systematic bias in the more sophisticated model is evident from a comparison of Figures 3.6 and 3.12; significantly lower values of systematic bias for the texture outputs are obtained for the C&G model than its G counterpart. The trends, seen Figure 3.12, are also observed for the remaining experimental settings.

Figure 3.7 shows that the systematic bias for the maximum stress converges at the tenth experiment for all settings. The systematic biases of the texture outputs converge after the addition of the seventh experiment as shown in Figures 3.8 and 3.9. Recall that convergence is achieved with only four experiments for the G model. This is expected, of course, since models with a higher number of parameters necessitate a greater number of experiments to achieve convergence. As the C&G model has twice as many calibration parameters as the G model (see Tables 3.3 and 3.4), more experimental data is required to reduce the parameter uncertainty for the systematic bias to converge.

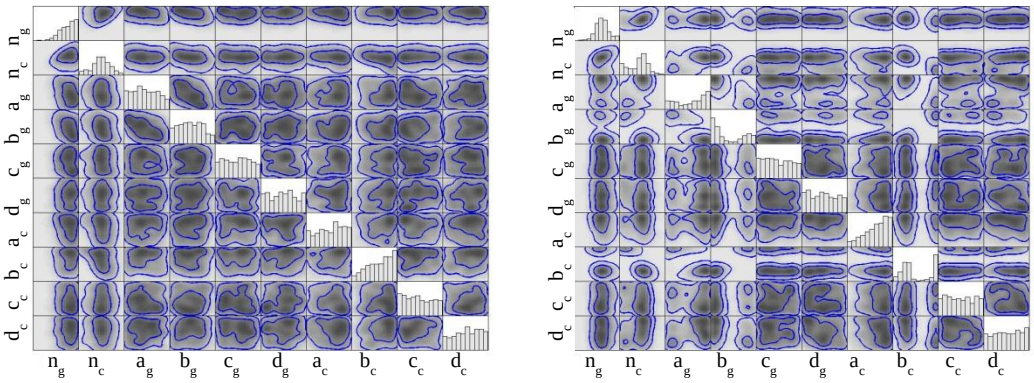

Figure 3.11: Posterior distribution for the C&G model (Left: one Experiment, Right: 11 Experiments).

Figure 3.13 shows information gain for experimental settings D and I. Overall, information gain increases (with only minor fluctuations) as the number of experiments is increased and converges after a sufficient number of experiments are used in the calibration process. The specific level of information gain at which convergence is achieved varies based upon the predicted outputs and the control settings. For instance, the information gain for the maximum stress and texture 001 predictions at experimental setting D are approximately 30% and 43%, respectively while for I the values are approximately 40% and 59%, respectively.

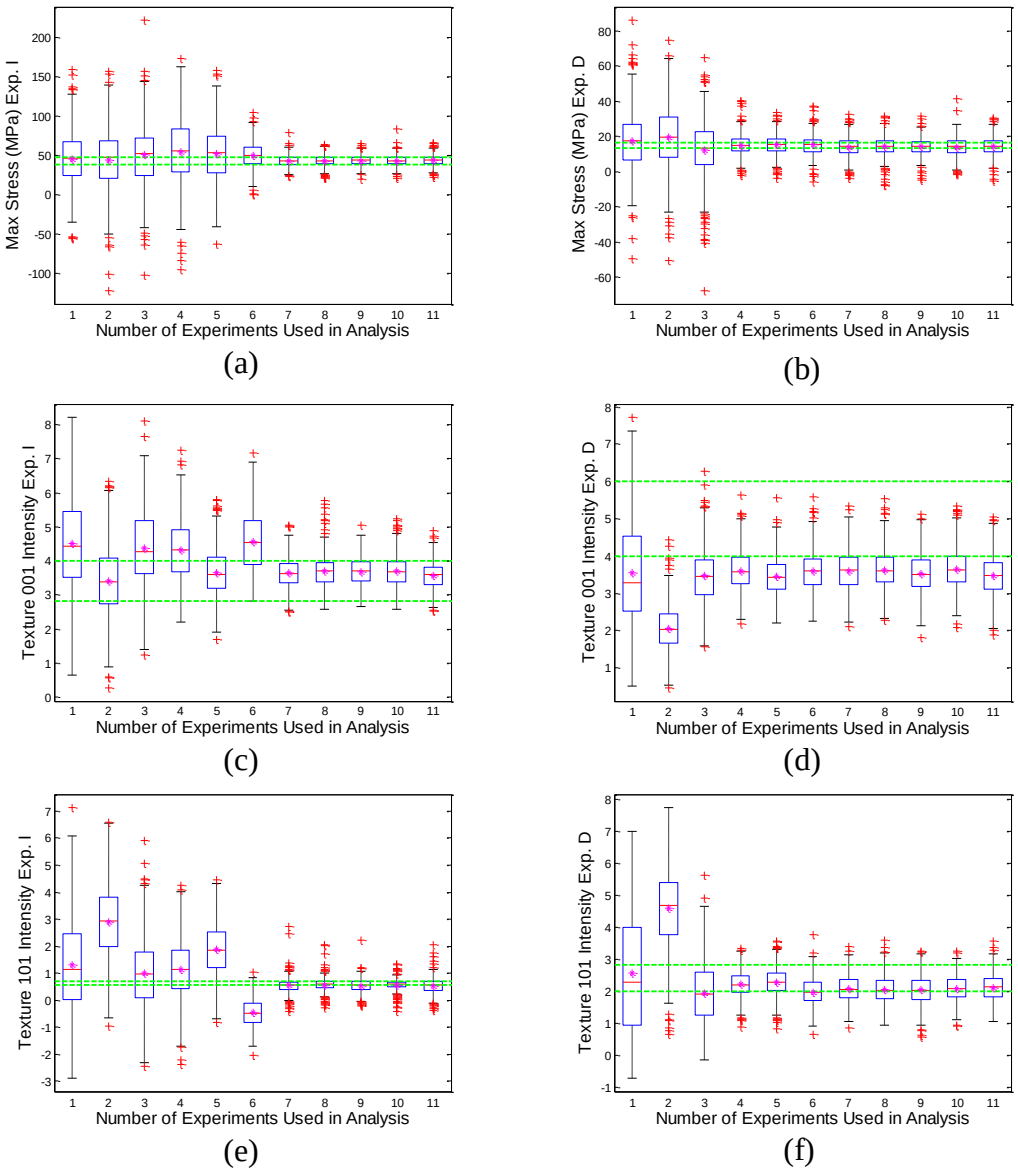

Figure 3.12: Prediction statistics for the C&G model (Left: Experiment I, Right: Experiment D).

The convergence of both the systematic bias and information gain permits analysis of the level of systematic bias. For the C&G model, the systematic bias converges to below 10% for the maximum stress prediction and below 50% for the texture predictions. Should this level of systematic bias be deemed acceptable for the

purposes of simulation, the model may be considered validated. If additional reduction in systematic bias is required, however, a further physics sophistication of the model should be considered.

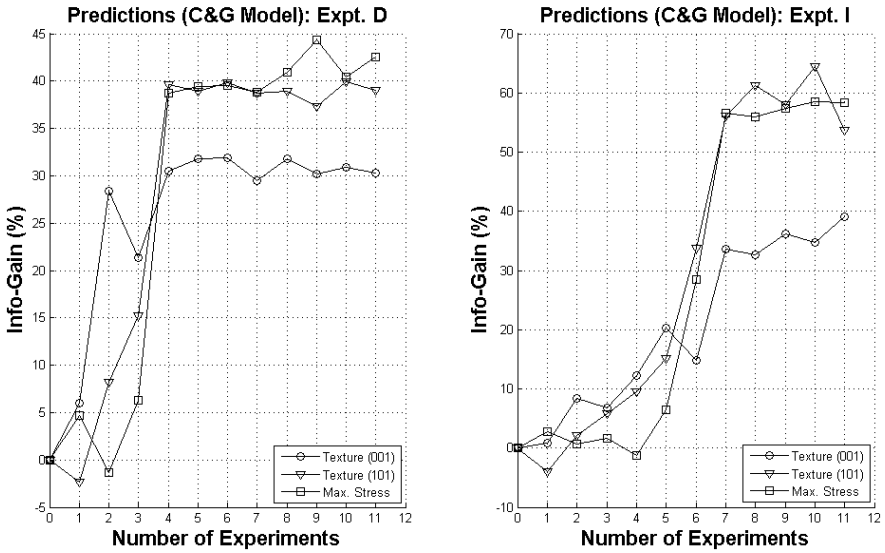

Figure 3.13: Model predictions at Experiment D and Experiment I: Information gain from experiments used for calibration (C&G model).

3.5.3. Comparison of the G and C&G models

Figure 3.14 illustrates the systematic bias, averaged for each feature over each of the experimental settings (from Figures 3.7-3.9). Here, the reduction and ultimate convergence of this averaged systematic bias is observed. Furthermore, Figure 3.14 indicates that the improved physics description, i.e. the C&G model, converges to a lower systematic bias in all three outputs when compared to the G model, confirming that the addition of physics principles in the model reduces systematic bias. Due to a larger number of calibration parameters, however, the C&G model requires increased experiments to reach convergence. For resource allocation problems, therefore, the cost

of additional experimentation for more sophisticated models must be evaluated against the potential reduction in systematic bias.

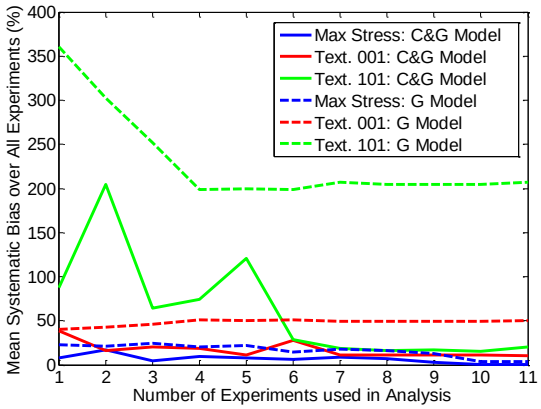

Figure 3.14: Mean systematic bias for each model output computed over each experimental setting from Figures 3.7-3.9.

The G model and C&G model are also compared using the Divergence Information Criterion (DIC), a statistical metric for evaluating preferable models (Spiegelhalter et al., 2002). The DIC operates on the model parameters and their distributions to create a metric that rewards a model for predicting closer to the experiments and penalizes a model for having a larger effective number of parameters. The DIC is computed as shown in Equations (3.14)-(3.16).

$$\tilde{p}_j = -2\int \log[f^{(j)}(y \mid \theta^{(j)})]\pi^{(j)}(\theta^{(j)} \mid y)d\theta^{(j)} + 2\log[f^{(j)}(y \mid \tilde{\theta}^{(j)}(y))] \tag{3.14}$$

$$D_j(\theta^{(j)}) = -2\log[f^{(j)}(y \mid \theta^{(j)})] + 2\log[g^{(j)}(y)] \tag{3.15}$$

$$DIC = D_j(\bar{\theta}^{(j)}) + 2\bar{p}_j \tag{3.16}$$

In Equation (3.14), $f^{(j)}(y \mid \theta^{(j)})$ represents the probability density function for the experimental data y for a given θ and the j-th model, $\pi^{(j)}$ represents the prior distribution

of emulator parameters, and $\tilde{\theta}^{(j)}(y)$ estimates $\theta^{(j)}$ based upon the experimental data y (Williams et al., 2011). In Equation (3.15), $2\log[g^{(j)}(y)]$ is a standardizing term dependent on the observed data. Finally, in Equation (3.16), the DIC is computed for the case where $\bar{p}_j$ represents $\tilde{p}_j$ when $\tilde{\theta}^{(j)}(y)$ is equal to the posterior mean $\bar{\theta}^{(j)}$. When comparing two models, a smaller value for DIC indicates a more preferable model (Williams et al., 2011). As the DIC is used for comparison purposes only, the actual DIC values are only relevant relative to another model.

In Figure 3.15, DIC values are plotted against the number of experiments. The DIC of the C&G model converges to a smaller value than that of the G model indicating that the lower systematic bias of the C&G model outweighs the penalty of having a larger number of parameters. Though the C&G model is preferable over the G model in this instance, it is important to note that until the fourth experiment, the G model has a lower DIC value. This observation again confirms that more complex models require larger numbers of experiments to achieve maturity in their predictive capabilities.

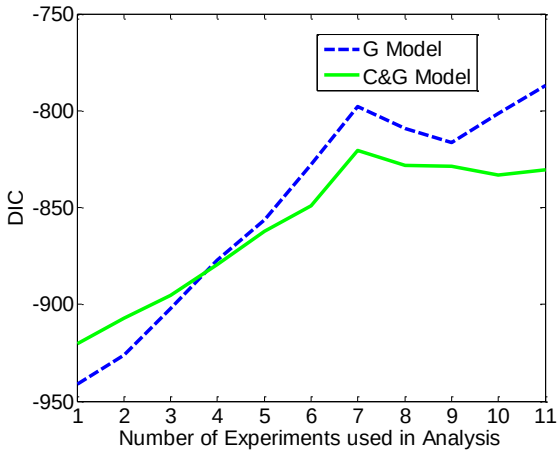

Figure 3.15: DIC comparison of the G and C&G models.

3.6. Conclusions

In this chapter, a framework to guide the allocation of resources for the validation of numerical models is demonstrated. Improvement in the predictive capabilities of a numerical model can be achieved through the reduction of parameter uncertainty and systematic bias. The parameter uncertainty can be reduced through calibration of model parameters against experimental data. An increase in the number of experiments used in the model calibration results in a decrease in parameter uncertainty; this is indicated by the convergence of the systematic bias and the information gain to a stable level throughout the domain. Upon convergence of systematic bias and information gain, parameter uncertainty cannot be reduced further. In this case, improvements to predictive capabilities is only possible by reducing systematic bias, which can be achieved only by improving the physics and/or engineering principles of the given model.

The proposed framework is demonstrated on a non-trivial application of the VPSC code to predict stress and texture behavior of 5182 aluminum alloy. The availability of two versions of the VPSC code, G and C&G models, presents an ideal opportunity to demonstrate the fundamental concepts presented herein. In this example, the systematic bias of the texture predictions of the G model converges as more experimental data is utilized in the calibration process. The analysis of the systematic bias and information gain show that parameter uncertainty cannot be further reduced. As the resulting converged systematic bias for the G model is deemed unacceptably high, the C&G model, with a more sophisticated physics than its G counterpart, is used for predictions. Again, the convergence of the systematic bias and information gain of the

C&G model is observed. Finally, though the converged systematic bias of the more sophisticated C&G model is smaller than the G model for texture predictions, this bias is possible only when calibrated with a higher number of experiments. The results from this case study of the proposed framework provide a science-based, quantifiable, and defendable rationale for allocating resources between coding and experimentation to reduce both parameter uncertainty and systematic bias.

While this work demonstrates the proposed framework for improving the predictive capability of numerical models, certain limitations constrain its effectiveness. Specifically, this framework operates under the assumption that (i) neither resources nor statute can prohibit further experimentation, (ii) that the analyst can improve the code through first-principle knowledge of the missing physics laws; (iii) that the *a priori* values of calibration parameters of the code do not change within the domain of applicability, and (iv) that numerical uncertainty is considered sufficiently small (verification activities are successfully completed *a priori*). A numerical model violating one of these assumptions would require a modification to the framework and provides additional areas topics for further study.

3.7. References

Christie, M.A., Glimm, J., Grove, J.W., Higdon, D.M., Sharp, D.H., & Wood-Schultz, M.M. (2005), Error analysis and simulations of complex phenomena, *Los Alamos Science*, (29), 6-25.

Higdon, D., Gattiker, J., Williams, B., & Rightley, M. (2008a), Computer model calibration using high-dimensional output, *Journal of the American Statistical Association*, 103 (482), 570-83.

Higdon, D., Nakhleh, C., Gattiker, J., & Williams, B. (2008b), A Bayesian calibration approach to the thermal problem, *Computer Methods in Applied Mechanics and Engineering*, 197 (29-32), 2431-41.

Kennedy, M. & O'Hagan A. (2001), Bayesian calibration of computer models (with discussion), *Journal of the Royal Statistical Society Series B*, (68), 425–64.

Lebensohn, R.A., Hartley, C.S., Tomé, C.N., & Castelnau, O. (2010), Modeling the mechanical response of polycrystals deforming by climb and glide, *Philosophical Magazine*, 90 (5), 567-83.

Lebensohn, R.A., Holt, R.A., Caro, J.A., Alankar, A., & Tomé, C.N. (2011), Improved constitutive description of single crystal viscoplastic deformation by dislocation climb, *Comptes Rendus Mecanique*, (submitted for review).

Lebensohn, R.A., Liu, Y., & Castañeda, P.P. (2004) On the accuracy of the self-consistent approximation for polycrystals: comparison with full-field numerical simulations, *Acta Materialia*, 52, 5347-61.

Lebensohn, R.A. & Tomé, C.N. (1993), A self-consistent anisotropic approach for the simulation of plastic deformation and texture development of polycrystals: application to zirconium alloys, *Acta metal. Mater*, 41 (9), 2611-23.

Oberkampf, W.L., Trucano, T.G., & Hirsch, C. (2003), Verification, validation, and predictive capability in computational engineering and physics, Sandia National Laboratories Report, SAND2003-3769 2003.

Powell, M.J.D. (1978), A fast algorithm for nonlinearly constrained optimization calculations, *Numerical Analysis*, G.A. Watson (Eds.), Lecture Notes in Mathematics, Springer Verlag, 630.

Shannon, C.E. (1948), A mathematical theory of communication, *The Bell System Technical Journal*, 27, 623-56.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde, A. (2002), Bayesian measure of model complexity and fit (with discussion), *Journal of the Royal Statistical Society Series B*, 64, 583-639.

Stout, M.G., Chen, S.R., Kocks, U.F., Schwartz, A.J., MacEwen, S.R., & Beaudoin, A.J. (1998a), Constitutive modeling of a 5182 aluminum as a function of strain rate and temperature, *Hot Deformation of Aluminum Alloys II*, Bieler, T., Lalli, T.A., MacEwen, L.A., SR(Eds.), TMS: Warrendale, PA, 1009, 205-16.

Stout, M.G., Chen, S.R., Kocks, U.F., Schwartz, A.J., MacEwen, S.R., & Beaudoin, A.J. (1998b), Mechanisms responsible for texture development in a 5182 aluminum alloy deformed at elevated temperature, *Hot Deformation of Aluminum Alloys II*, Bieler, T., Lalli, T.A., MacEwen, L.A., SR(Eds.), TMS: Warrendale, PA, 1009, 243-54.

Thacker, B.H., Doebling, S.W., Hemez, F.M., Anderson, M.C., Pepin, J.E., & Rodriguez, E.A. (2004), Concepts of model verification and validation, Los Alamos National Laboratories Report, NM, LA-14167-MS 2004.

Unal, C., Williams, B., Hemez, F., Atamturktur, S.H., & McClure, P. (2011), Improved best estimate plus uncertainty methodology, including advanced validation concepts, to license evolving nuclear reactors, *Nuclear Engineering and Design*, 241, 1813–33.

Waltz, R.A., Morales, J.L., Nocedal, J., Orban, D. (2006), An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming*, 107(3), 391–408.

Williams, B.J., Picard, R., & Swiler, L. (2011), Multiple model inference with application to model selection for the reactor code R7, Los Alamos National Laboratories Technical Report, NM, LA-UR-11-05625.

CHAPTER 4: A SELECTION CRITERION BASED ON EXPLORATION-EXPLOITATION APPROACH FOR BATCH SEQUENTIAL DESIGN

4.1. Introduction

Numerical models are increasingly relied upon to make predictions about engineering systems in lieu of physical experiments. Such models can greatly reduce the reliance on extensive experimental campaigns. However, the need for experiments cannot be entirely eliminated, since experiments are vital to ensure the validity of model predictions. As numerical models are executed at different settings within the operational domain of the engineering system being represented, experiments are needed to validate model predictions at different settings within this domain of interest.

Improvement in the predictive capability of a model can be accomplished through a reduction in two principal quantities: (i) the fundamental inability of the model to reproduce experiments, *model form error*, and (ii) the *uncertainties* associated with model predictions. By calibrating the model against physical experiments, an empirically trained approximation of model form error (referred to herein as *discrepancy bias*) can be obtained and the model uncertainties can be reduced. As the available experimental data increases, this empirical discrepancy bias is expected to converge to a consistent and systematic level, and the model uncertainties are expected to continually reduce with diminishing returns. However, the convergence of the discrepancy bias and the reduction in model uncertainty both exhibit *path dependency;* i.e. the rates at which the discrepancy converges and the uncertainties reduce are dependent upon both the control settings of the physical experiments (i.e. settings in the domain of interest the experiments are

conducted) as well as the sequence in which those experiments are conducted. Therefore, the settings and the sequence of the experiments are vitally important for experiment-based validation.

Section 4.2 overviews techniques developed to exploit this path dependency, broadly known as *optimal experimental design*. One particular methodology (described in further detail in Section 4.3), the Batch Sequential Design (BSD), is particularly useful for the optimal design of validation experiments. BSD, proposed by Loeppky et al. (2010) for design of computer experiments and extended by Williams et al. (2011) and Atamturktur et al. (2012) for physical experiments, allows for experiments to be selected in groups, or batches, rather than single experiment optimizations available in the literature. The objective of the BSD methodology is to select the optimal settings for experiments to be conducted within each batch such that the discrepancy (treated with an empirically trained Gaussian Process Emulator (GPM)) is trained as effectively as possible with the fewest number of experiments.

The selection of optimal experiments through BSD can be completed using various forms of selection criteria. In Atamturktur et al. (2012), for instance, eight different BSD selection criteria, including entropy-based and distance-based criteria are implemented and compared against each other. Atamturktur et al. (2012) compared the selected criteria using a novel metric, the predictive maturity index (Hemez et al., 2010) (reviewed in Section 4.4), that incorporates not only the model discrepancy, but also how well the selected experiments cover the domain of applicability (i.e. coverage). In the

established literature, BSD operates solely on the discrepancy bias and neglects the coverage for the given problem.

This chapter describes improvements to the BSD algorithm, in particular to a BSD selection criteria, *Coverage Augmented Expected Improvement for Predictive Stability* (C-EIPS), such that the effect of both coverage and discrepancy are considered, resulting in the most efficient improvement in the predictive maturity of the model. As such, C-EIPS both *explores* the entirety of the domain of applicability by including the effect of coverage and *exploits* areas within the domain that exhibit high variability in the empirically trained model form error by also operating on the discrepancy bias. This new selection criterion, developed in Section 4.5 of this chapter, is demonstrated on a multi-variate problem, the Visco Plastic Self-Consistent (VPSC) code for modeling the plasticity of metals (reviewed in Section 4.6). Section 4.7 demonstrates, through the application of the VPSC code, that the functional form of the newly developed BSD criterion, C-EIPS, results in an efficient use of experimental resources. The performance of C-EIPS is compared against the two most effective criteria reported in Atamturktur et al. (2012) for a series of cases with varying levels of predefined model form error. Finally, conclusions are drawn in Section 4.8.

4.2. Review of Optimal Experimental Design

This section focuses on research endeavors that specifically use experimental design as a tool to improve the predictive maturity of a specific numerical model. One such class of experimental design stems from the replacement of computationally burdensome high-fidelity numerical models with fast-running statistical emulators as has

been widely used in the scientific computing community (Crombecq et al., 2009; Jin et al., 2005; Allen et al., 2003; Crary 2002). While training these emulators, it is advantageous to determine the optimal selection of settings for new computer experiments such that information gain and predictive accuracy are maximized. This intelligent selection approach is the main purpose of optimal experimental design. According to Dawid and Sebastiani (1999), the optimal experimental design falls into two categories: *criterion-based design* and *decision-based design*.

In the *criterion-based design*, the goal is to fulfill a particular optimality criterion to benefit the most from limited experiments, which is typically achieved through an optimization procedure. Some popular optimization criteria in this class of optimal experimental design include minimizing the average variance of parameter estimates (A-optimality), maximizing information (D-optimality), and minimizing the maximum variance in predictions (G-Optimality) (Dawid and Sebastiani, 1999; Thompson et al., 2010).

Based upon the principles of Bayesian decision theory, *decision-based design* is generally known as *Bayesian experimental design* in the literature. Lindley (1972) is credited for providing a unified theory in Bayesian experimental design. This approach allows for the incorporation of prior knowledge and uncertainties of imprecisely known quantities and uses a utility function considering the purpose and costs of the experiments; the optimal design seeks to maximize the expected utility function. Specific applications in the literature utilize different utility functions based upon the intended application. Examples include information gain based utility functions (Huan and

Marzouk, 2011a, 2011b; Terejanu et al., 2012; Sebastiani and Wynn, 1996) and utility functions based on the minimization of uncertainty (Davis and Prieditis, 1999).

Sequential design entails developing and applying optimal experimental design strategies to enhance the selection process of future experiments (Chernoff, 1959; Blot and Metter, 1973; Williams et al., 2000; Dror and Steinberg, 2008; Crombecq et al., 2009). Unlike a one-step experimental approach that requires predetermined settings at which experiments are conducted, sequential design allows updating and improving future experiments with the aid of the data already collected. Jiang and Mahadevan (2006) present one such study relevant to the model validation community. Their approach specifically seeks an optimum design of validation experiments through the implementation of a Bayesian cross-entropy methodology and a simulated annealing algorithm. These designs have adaptive ability on the basis of the characteristics of application and experimental space so as to reduce the number of experiments necessary. However, as many experiments are usually conducted during a single experimental campaign, it is advantageous to select multiple experiments at once (i.e., in batches). BSD can accomplish this by selecting a batch of experiments with a user-defined size.

BSD is appealing because of its efficiency and cost-effectiveness. Moreover, many of the optimization criteria mentioned earlier can be utilized in BSD. Applications of different BSD approaches include solving nonlinear problems using an asymptotic variance of the maximum likelihood-estimator as the design criterion (Muller and Poscher, 1989) and the use of Bayesian D-optimality criterion (to achieve computational efficiency) coupled with K-mean clustering algorithm to augment initial designs in

solving generalized linear models (Dror and Steinberg, 2008). Also, in the Bayesian context, two batch sequential strategies have been recently proposed with different motivations: Loeppky et al. (2010) used distance-based criteria to select new experiments to accelerate the maturity of an emulator, while to achieve predictive maturity of the calibrated model, Williams et al. (2011) used several selection criteria including integrated and maximum mean square error, maximum entropy, and two expected improvement criteria. While the work of Loeppky et al. (2010) focuses more on computer experiments, Williams et al. (2011) extend their efforts to the optimal design of physical experiments. An application and evaluation of the BSD criteria presented in Williams et al. (2011) is provided in Atamturktur et al. (2012). A further discussion of these three studies is provided in the next section.

4.3. Batch Sequential Design

Loeppky et al. (2010) employs batch sequential methods strictly for the training of the fast-running surrogate models (also known as emulators, meta-models or response surface models) to replace the computationally expensive physics codes. Specifically, the focus is on improving the training of GPM with a given number of available computer runs, which only partially describes the response surface of the given source code. As such, the motivation behind using BSD in Loeppky et al. (2010) is to select the optimal settings for computer experiments such that the residual errors between the actual model predictions and the GPM are minimized. In doing so, four BSD selection criteria including integrated mean square error, entropy, weighted distance, and distance criteria

are implemented. The distance criteria are observed to perform well for low dimensional problems.

Williams et al. (2011) extends the concepts applied in Loeppky et al. (2010) to select physical experiments and thus, to improve the training of a discrepancy GPM. Through this approach, the physical experiments are conducted in batches to be used in the calibration of the numerical models. A GPM is useful as a reliable statistical estimation only if it is trained properly with sufficient data. As seen in Figure 4.1, the variance of a GPM increases in regions away from the experimental settings. It is also intuitive that the trained GPM for discrepancy will be a better representation of the "true" model form error if more experiments become available. Figure 4.1 (left) demonstrates that with only two experiments, the mean GPM for discrepancy is a straight line and the variance is quite high. With three experiments however, the mean GPM better represents the true residual difference between model predictions and experimental observations, i.e. model form error, and the variance is significantly reduced simply because the maximum distance between two experiments decreased. This dependency on the number of experiments (and the distance between experiments) demonstrates the importance of selecting the settings for optimal experiments.

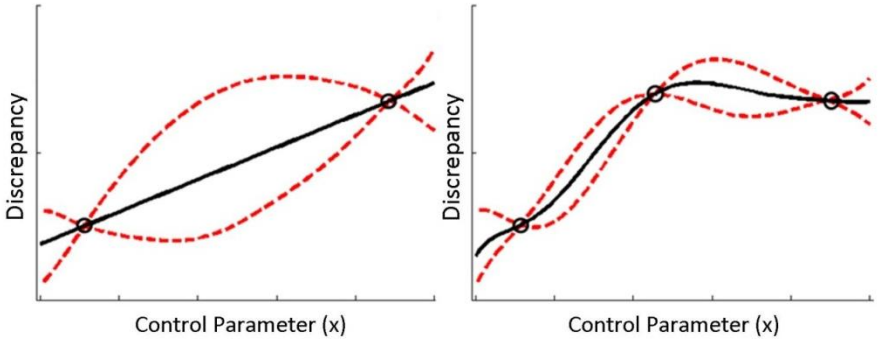

Figure 4.1: Training of Discrepancy GPM with Experiments

Given an initial set of experimental data and a numerical model, the BSD algorithm selects the optimum set of additional experiments based upon a given optimization criterion. Once the proposed new set of experiments become available, the model calibration process is repeated. Through model calibration, the uncertainties in model parameters are reduced, the estimated discrepancy bias is refined, and the BSD algorithm is executed to determine the next batch of optimal experimental settings (see Figure 4.2). The number of experiments for inclusion in the batch is naturally problem specific, and can be as low as one experiment per batch.

A modified Fedorov exchange algorithm is used to select the optimal experimental settings (design points) (Fedorov, 1972; Cook and Nachtsheim, 1980). In this algorithm, the starting values for the batch design points are generated using quasi-random designs such as Sobol Sequence, Scrambled Net Sobol Sequence, etc. or a uniform distribution. The criterion is then stringently optimized with respect to each new design point (while fixing the other design points in the batch) in turn, until negligible improvement in the criterion value is observed. The initial design points in the batch are

then replaced (exchanged) by the optimized points. The optimal design points selected for this batch are then augmented to the initial design (or previous batch). Each of the original design points are considered for exchange with a candidate point. Thus, the algorithm involves a search over all possible pairs of candidate and design points.

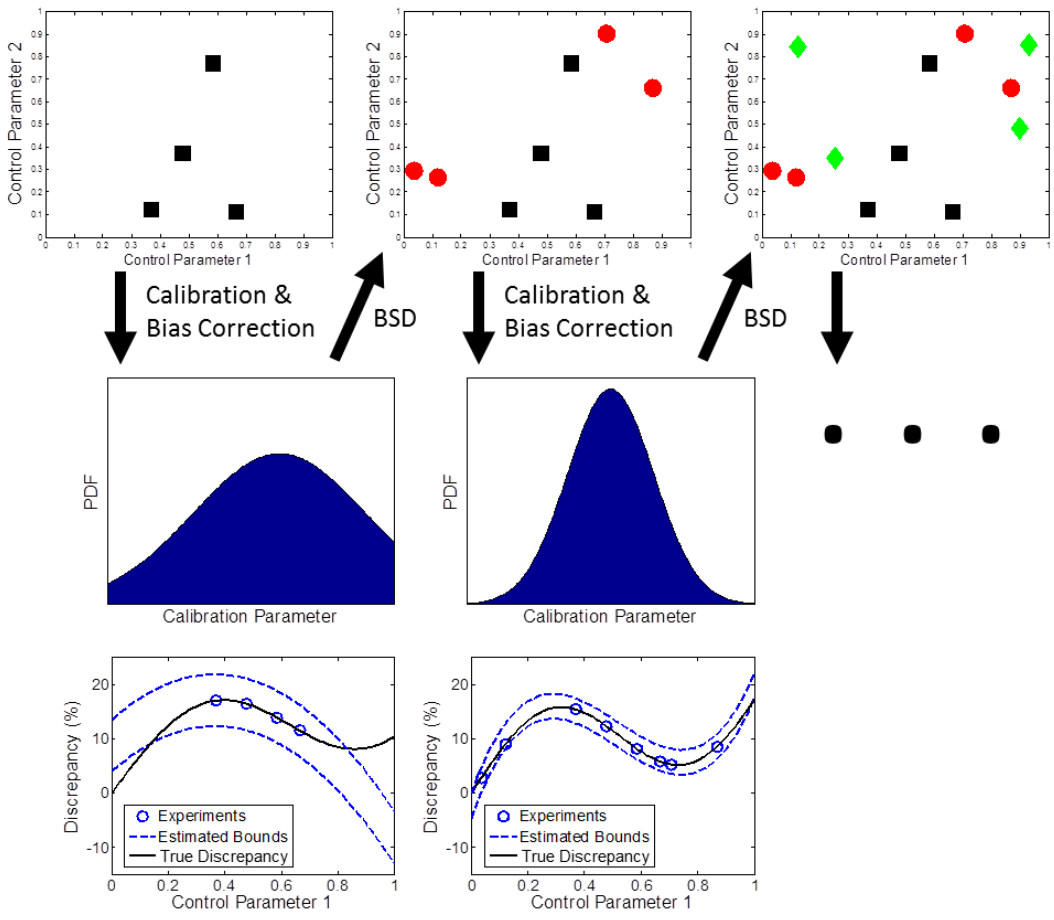

Figure 4.2: Selection of Optimal Experiments

Atamturktur et al. (2012) utilizes the BSD approach in Williams et al. (2011) to augment experiments for the VPSC code applied to creep strain rate in FCC steel. Several BSD criteria are investigated and assessed by the Predictive Maturity Index (PMI) (as

described in Section 4.4). The criteria employed by Atamturktur et al. (2012) to select optimal experimental settings include:

- Expected improvement for predictive stability criterion (EIPS): evaluates the Kullback-Leibler distance between the current and proposed future distributions of the discrepancy, by selecting design points that maximize the minimum information gain.
- Mean-square error-based criterion (MSE): selects design points based upon minimizing functions defining the posterior discrepancy variance.
- Maximum entropy criterion (ENT): selects future designs points by maximizing the determinant of the correlation matrix defined by relating the predicted discrepancy at the new design settings and the currently available data.
- Distance-based criterion: utilizes the measures such as Euclidean (EDIST) and Mahalonobis (MDIST) distances to minimize the maximum correlation between the predicted discrepancy values and the existing design.

Figure 4.3 compares the PMI metric for BSD selected and user-selected (Latin-hypercube (LHS) design in this figure) batches of experiments, where the more efficient convergence of the BSD-selected batches over sample user-selected batches is evident. In Atamturktur et al. (2012), the EIPS and EDIST criteria are found most effective for improving PMI.

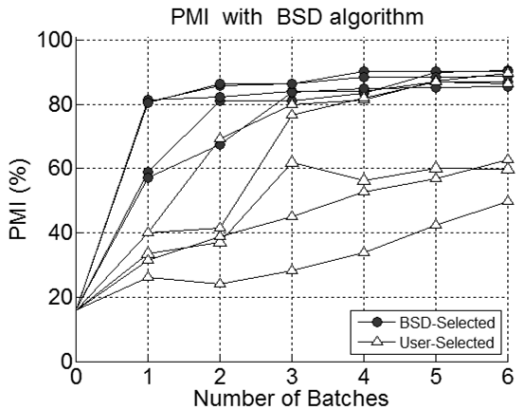

Figure 4.3: BSD vs. User Selected PMI Stabilization (reprinted with permission from Atamturktur et al., 2012)

4.4. Predictive Maturity Index (PMI)

The goal of a predictive capability metric entails assessing the ability of a model to make reliable predictions throughout a set domain of applicability, within which the models are intended to be used in a predictive manner. Recently many institutions have led efforts to develop metrics for predictive capability of complex numerical models (e.g. Harmon et al, 2005; Zang, 2008; NASA, 2007; Oberkampf et al., 2007; Sornette et al., 2006, 2007; Hemez et al., 2010). For a more detailed review of the literature on predictive capability metrics, see Hemez et al. (2010).

As documented in Hemez et al. (2010), very few investigations venture from a heavy reliance on qualitative expert judgment, or with the exception of Sornette et al. (2006) and Hemez et al. (2010), model accuracy alone. Furthermore, the existing methods used to measure predictive capability attempt to quantify the quality of the validation activities applied on a simulation model, and not necessarily the quality of the solutions of the simulation model. Therefore, these aforementioned approaches operate

on the unwarranted assumption that increased thoroughness in validation activities directly improves predictive capability of the model solutions. Moreover, these approaches are subjective in nature and could lead to different conclusions by different experts.

To supply a consistent evaluation approach, Hemez et al. (2010) proposed a novel concept based upon the stabilization of the systematic discrepancy bias between numerical predictions and physical measurements to assess the maturity of simulation models. In contrast to other earlier efforts, the Predictive Maturity Index (PMI) approach of Hemez et al. (2010) is quantitative, objective and repeatable. The PMI metric is based on several additional attributes that move beyond model accuracy: (i) a measure of discrepancy between physical observations and numerical predictions, (ii) a measure of complexity of the model, and (iii) a measure of the extent to which physical experiments used for calibration cover the domain of applicability. The original PMI metric of Hemez et al. (2010) has recently been modified by Stull et al. (2011) to include a refined definition of coverage, an additional model robustness term, and a modified functional implementation. These modifications are made to improve asymptotic properties of the PMI metric. Herein, PMI is implemented in its most recent form. The individual components and functional form of PMI are described briefly below. For a more rigorous description of each component see Hemez et al. (2010) and Stull et al. (2011).

4.4.1. Coverage of the Domain of Applicability:

Coverage of the domain of applicability refers to how well the experiments explore the domain of applicability. In Hemez et al. (2010) coverage is defined as the

ratio of the multi-dimensional volume encompassed by the convex hull surrounding the physical experiments to the multi-dimensional volume that defines the domain of applicability. However, in Stull et al. (2011), coverage is measured by taking the ratio of the sum of weighted experimental design points to the multi-dimensional volume that defines the domain of applicability. The weights for each experimental design point with respect to each control parameter are expressed as a percentage of the design domain that an individual experiment covers. While providing flexibility, these values are case and project specific and should be thoughtfully determined based upon expert judgment. In Stull et al. (2011) coverage is formulated as:

$$\eta_c = \frac{\sum_{i=1}^{M} \text{Volume}(\Omega_{E,i})}{\text{Volume}(\Omega_V)} \tag{4.1}$$

where $\Omega_{E,i}$ represents the convex hull around the i^{th} experiment, Ω_V represents the convex hull of the entire domain of applicability, and M is the total number of experiments. Figure 4.4 demonstrates how this definition of coverage provides a better representation of the intent of the coverage parameter: to assess how well the physical experiments explore the domain of applicability.

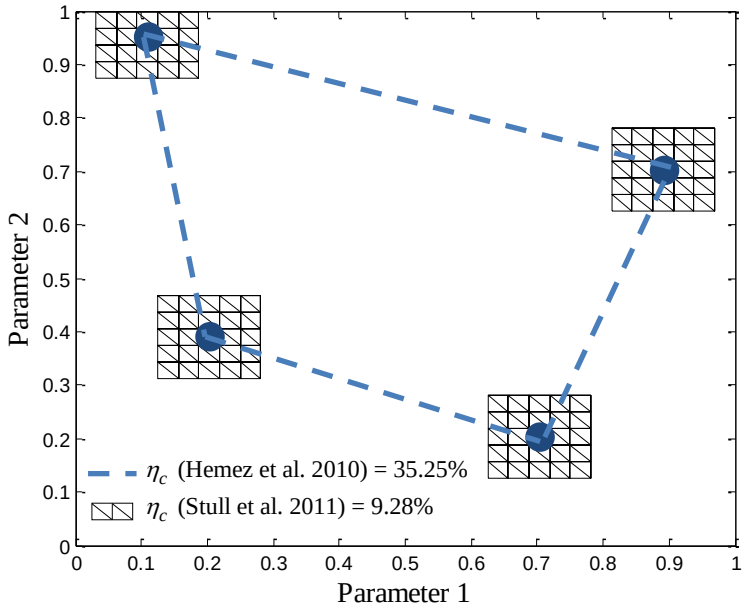

Figure 4.4: Coverage definition in Hemez et al. 2010 and Stull et al. 2011

4.4.2. Complexity of a Model or Numerical Simulation:

In Hemez et al. (2010) and Stull et al. (2011), the *complexity* attribute is determined by the number of knobs of the model. Knobs refer to the number of calibration parameters or uncertain parameters that cannot be controlled during testing and may be "calibrated" to experimental data. This choice is guided by the principle that, in general, more sophisticated models possess larger numbers of calibration parameters. However, a model with a larger number of calibration parameters is prone to over-fitting which can result in inaccurately estimated parameter values and unrealistically high confidence in the model. The number of calibration knobs is denoted by the symbol N_K and is normalized by a reference number of knobs N_R.

4.4.3. Level of Accuracy of Model Predictions (Discrepancy):

The ability of a model to accurately reproduce the phenomena of interest can conveniently be defined through the discrepancy of the model. This discrepancy represents the empirically trained error between model predictions and experimental data that cannot be rectified through model calibration. Typically, discrepancy, δ_S, is normalized as a percentage of the response feature of interest, y_{sim}, as shown in Equation (4.2).

$$\delta_S = \frac{\|\delta(P)\|}{y_{sim}} \tag{4.2}$$

where $\delta(P)$ is the functional representation of discrepancy throughout the domain of applicability, P, and ||.|| is an application specific norm.

4.4.4. Robustness

The *robustness*, α, for a given model is defined as the maximum amount of uncertainty allowed in the calibration parameters such that the maximum discrepancy realized (worst-case performance) from such uncertainty does not exceed a set discrepancy limit. As defined in Stull et al. (2011), the robustness term integrated into PMI operates on an info-gap uncertainty model of the calibration parameters (Ben-Haim, 2006), and the robustness for a single experiment is defined as:

$$\hat{\alpha}^k = \max_{\alpha}\left(\max_{X^k \in U^k(\alpha)} R(X^k) \le R_{\max} \right) \tag{4.3}$$

where $R(X^k)$ represents the difference between the model prediction and experiment *k*, R_{max} is the set maximum allowable discrepancy, and $U^k(\alpha)$ is the set of calibration

parameters as defined by the info-gap model. The term for robustness used in PMI is given in Equation (4.4).

$$\alpha_S = \left\| \alpha(P; R_{\max}) \right\| \tag{4.4}$$

4.4.5. Predictive Maturity Index (PMI) Metric:

The PMI metric defined in Stull et al. (2011) depends upon coverage, η_C, number of knobs, N_K, discrepancy, δ_S, and robustness, α_S. Its values are, without loss of generality, bounded in the interval $0 \leq \text{PMI} \leq 1$ for intuitive interpretation. PMI = 0 means that the model has no predictive maturity. PMI = 1 implies, on the other hand, perfect predictive maturity over the entire domain of applicability. Clearly, these two cases are asymptotes that cannot be reached with a finite number of physical experiments. Stull et al. (2011) proposes the following PMI metric that verifies the above-stated properties:

$$PMI\left(\eta_C; N_K; \delta_S; \alpha_S\right) = \prod_{i=1}^{5} \Psi_i \tag{4.5}$$

Table 4.1: PMI Term Definitions (Stull et al., 2011)

Term	Definition
Ψ_1	$\begin{cases} \tanh(\gamma_1 \times \eta_c) & \eta_c < 1 \\ 1 & \eta_c \geq 1 \end{cases}$
Ψ_2	$\tanh\left(\left(\frac{N_R}{N_K}\right)^{\gamma_2}\right)$
Ψ_3	$\left(1-\delta_S\right)^{\gamma 3}$
Ψ_4	$\left[1-\tanh\left(\frac{\gamma_4}{\alpha_S}\right)\right]$
Ψ_5	$e^{-[e^{-\eta_C} \times \delta_S \times e^{-\alpha_S}]}$

where γ_1, γ_2, γ_3 and γ_4 are strictly positive, user-defined coefficients to weight the effects of the four attributes of maturity and Ψ_i terms are defined in Table 4.1. In Hemez et al. (2010) and Stull et al. (2011), several mathematical and asymptotic properties of predictive maturity are proposed that constrain the definition of the metric. In general, those limits are stated as: (i) as $\delta_S \to 0, \eta_C \to 1, \alpha_S \to \infty$, PMI approaches 1, and (ii) as $\delta_S \to 1, \eta_C \to 0, N_K \to \infty, \alpha_S \to 0$ PMI approaches 0.

4.5. Coverage Augmented Expected Improvement for Predictive Stability (C-EIPS)

In Williams et al. (2011) and Atamturktur et al. (2012), optimization criteria solely operate on the empirically trained discrepancy model with the intention of reducing the variance of the GPM trained to represent the discrepancy bias as illustrated in Figure 4.1. Herein, a new criterion, Coverage Augmented Expected Improvement for

Predictive Stability (C-EIPS) is proposed which considers not only the reduction in the variance of the discrepancy but also the coverage of the domain of applicability.

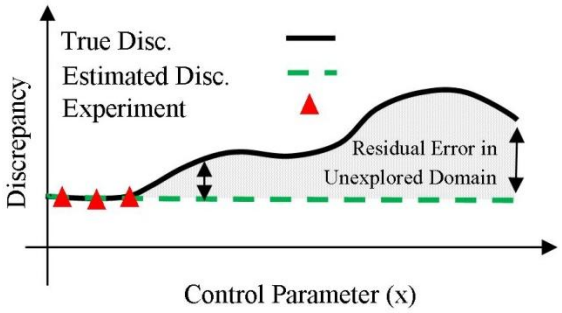

Figure 4.5: Potential Error in Discrepancy Estimation

In BSD, the importance of discrepancy as a measure of the model's accuracy, discussed at length in Williams et al. (2011) and Atamturktur et al. (2012), is clear. The importance of considering coverage, however, is demonstrated in Figure 4.5. Specifically, if all experiments are selected in one region of the domain of applicability, the potential inaccuracies in different regions of the domain may remain hidden. As such, estimates of discrepancy (both interpolative and extrapolative) may have significant residual error as shown by the dashed line for extrapolative prediction in Figure 4.5. By selecting experiments that better *explore* the domain of applicability, the discrepancy prediction can be better defined throughout the domain.

Maximizing PMI requires maximizing the coverage realized when adding a potential new design point. Since the total coverage can be computed directly for any new potential set of candidate design points, the coverage component in C-EIPS remains identical to its use in PMI as defined by Equation (4.1). On the other hand, in order to maximize PMI, the discrepancy of the model calibrated with an experimental campaign

that includes the potential batch of experiments must be minimized. However, the discrepancy estimate over the domain of applicability cannot be directly computed *a priori* to the availability of this new batch of experiments. Therefore, approaches similar to those used in Williams et al. (2011) and Atamturktur et al. (2012) that seek to minimize the variance of the discrepancy GPM must be utilized. By searching for potential design settings with the highest variance in the discrepancy GPM, it is possible to accomplish the greatest reduction in the variance of the overall model discrepancy. Therefore, the following criterion is considered for C-EIPS:

$$C-EIPS\left(\eta_C;\bar{\delta}\right)=\eta_c{}^{\lambda_1}x\bar{\delta}^{\lambda_2} \tag{4.6}$$

where η_c is the coverage as defined in Section 4.4, and $\bar{\delta}$ represents a term utilized to capture discrepancy in terms of potential reduction in the variance of the discrepancy GPM. Herein, for $\bar{\delta}$, a normalized EIPS criterion (described in Williams et al. (2011)) is used. In Equation (4.6) λ_1 and λ_2 represent user defined weighting coefficients such that more or less emphasis can be placed on either component depending on the application of interest.

As defined in the derivation of PMI, this value must be normalized such that the value of $\bar{\delta}$ is scaled between 0 and 1. To determine the normalization factor, prior to determining the first batch of experiments, the EIPS criterion values are computed for a given set of initial experimental data and randomly selected initial candidate point settings. This process is repeated a user-selected number of times, and the maximum EIPS criterion value is kept over all repeats. As the uncertainty in the discrepancy will

reduce as more experiments are added to the analysis, this maximum selected EIPS criterion value will serve as the normalization term for all $\bar{\delta}$ computations as shown in Equation (4.7):

$$\bar{\delta} = \frac{EIPS}{EIPS_{norm}} \tag{4.7}$$

where *EIPS* represents the computed EIPS criterion value for the given batch and candidate point and $EIPS_{norm}$ represents the maximum initially selected EIPS criterion reference value.

4.6. VPSC Application

In this section, the performance of the proposed C-EIPS criterion is assessed against two additional criteria, EDIST and EIPS, using the PMI metric. Herein, the C-EIPS criterion is applied to the climb and glide (C&G) version of the VPSC model derived in Lebensohn et al. (2010) for simulated 5182 Aluminum alloy experimental data measuring both the stress/strain response and textural evolution under varying strain rates and temperatures. A thorough description of the VPSC code is available in Lebensohn et al. (1993 and 2010).

The VPSC code for the application of interest has 12 uncertain parameters, which are calibrated against actual physical experiments on the 5182 Aluminum alloy in Hegenderfer et al. (2012). The posterior mean values for each of the calibration parameters, documented in Hegenderfer et al. (2012), are considered the "true" parameter values herein and used to generate the synthetic experimental data. On the other hand, the (intentionally) *imprecise and inexact* computer model that needs to be calibrated is

executed with 'uncertain' parameter values. The true parameter values used to generate the synthetic experiments along with the ranges defined in the calibration are listed in Table 4.2, along with the domain of applicability defined by the control parameters, strain rate and temperature. The response features of interest are the stress at the maximum measured strain of 0.6 and the textural intensities at the 001 and 101 poles.

A Bayesian inference calibration process is utilized in this study to simultaneously calibrate uncertain model parameters and train the discrepancy bias. First, the following relations describe the relationship between model predictions and observations:

$$y_{obs}(x) = \zeta(x) + \varepsilon(x) \tag{4.8}$$

$$\zeta(x) = y_{sim}(x, \theta) + \psi(x) \tag{4.9}$$

where x *and* θ represent control and the true values of the calibration parameters respectively, and experimental observations, y_{obs}, are a sum of the truth, $\zeta(x)$, and experimental error, $\varepsilon(x)$. Truth is further defined in Equation 4.9 as the sum of model predictions, $y_{sim}(x, \theta)$, and model form error, $\psi(x)$.

$$\hat{y}(x) = y_{sim}(x, \bar{\theta}) + \delta(x) + \varepsilon(x) \tag{4.10}$$

It follows in Equation (4.10), that the best estimate of truth, $\hat{y}(x)$, is the sum of calibrated model predictions, $y_{sim}(x, \bar{\theta})$, the estimated discrepancy bias between truth and predictions, $\delta(x)$, and experimental error, $\varepsilon(x)$ (Kennedy and O'Hagan, 2001). Herein, $y_{sim}(x, \bar{\theta})$ represents a trained GPM to serve as fast-running surrogate for the

VPSC predictions from which posterior distributions for the calibration parameters, $\bar{\theta}$, are inferred using Bayesian calibration. Similarly, the estimated discrepancy bias, $\delta(x)$, is simultaneously trained as a GPM to represent the true model incompleteness, $\psi(x)$ (see Higdon et al. (2008) for formulation details). Herein, posterior distributions are explored by a Markov chain Monte Carlo (MCMC) algorithm.

Two GPM emulators, one for model predictions and one for discrepancy, are trained using a 300-run LHS design. Uniform prior distributions for the calibration parameters are assumed according to the bounds given in Table 2 and 10,000 MCMC iterations are used to estimate the posterior distribution of the calibration parameters. For this study, model predictions are made using 500 linearly spaced samples of the posterior distributions of the calibration parameters.

A calibration process that exploits the BSD selected experiments should successfully retrieve the "true" parameter values (used to generate the synthetic experiments) with minimal number of experiments. The performance of the C-EIPS is assessed by comparing its final PMI values against those obtained using the two best performing criteria from Atamturktur et al. (2012), EIPS and EDIST. As this application is multivariate, the functional forms of the EIPS and EDIST criteria are utilized. The performance of the C-EIPS criterion is then checked under both absence and presence of an artificially imposed model form error.

Table 4.2: Calibration Parameters for C&G Model

	Parameter	Posterior Mean	Min	Max
Control Parameters	Temperature (C°)	n/a	200	550
	Strain Rate (s^{-1})	n/a	0.001	1.00
Calibration Parameters	n_g	3.7683	2.5000	4.5000
	a_g	3.2375	2.5000	4.5000
	b_g	3094.6447	2939.9124	3249.3769
	c_g	-0.0083	-0.0086	-0.0081
	d_g	278.6182	264.6873	292.5491
	n_c	-0.0040	-0.0042	-0.0039
	a_c	26529.0665	25202.6132	27855.5198
	b_c	-0.0119	-0.0122	-0.0115
	c_c	1600.4681	1520.4447	1680.4915
	d_c	-0.0079	-0.0081	-0.0076

4.7. Results and Discussion

The BSD algorithm is run for the three aforementioned criteria: C-EIPS, EIPS, and EDIST for six different specifications of model form error (listed in Table 4.3). For the C-EIPS criterion the λ values are set to 1 and 2 respectively, and the weight applied to each experiment for the coverage computation is set to 25% for each axis in the domain of applicability. For each criterion and in each case, two different sets of three initial experiments are used to initiate BSD (Table 4.4). The BSD algorithm is executed for each case until a total of ten batches with two experiments in each batch are selected.

The standard deviation of experimental error is set to 5% of the mean experimental observation for each feature. The artificially imposed model form error from Case 4 is shown as a representative example in Figure 4.6. The imposed model form error is assumed to be in a Gaussian form in both the strain rate ($\dot{\varepsilon}$) and temperature (T) axis. Table 4.3 shows the maximum amplitude, *A*, and standard deviation, σ, as a percentage of the feature value for each case.

For each case and after each batch, the model output and discrepancy bias predictions are obtained. Predictions are obtained at 81 *hold-out* experiments which uniformly cover the domain of applicability (forming a nine-by-nine grid). Since the hold-out experiments represent the truth, this simulated study permits a direct comparison of both the model predictions to truth and the predicted discrepancy bias to the true model form error. Similar to the forecasting error term introduced in Atamturktur et al. (2011), this study seeks to determine not only the estimated discrepancy but also the residual differences between that estimate and the true model form error. This residual difference in addition to the residual differences between calibrated model predictions and truth is referred to herein as *absolute error* and is described below.

$$\varphi_y(x) = \left|(\zeta(x) - \psi(x)) - y_{sim}(x, \bar{\theta})\right| \tag{4.11}$$

$$\varphi_\delta(x) = \left|\psi(x) - \delta(x)\right| \tag{4.12}$$

$$\varphi(x) = \varphi_y(x) + \varphi_\delta(x) \tag{4.13}$$

The absolute error defined herein, is influenced by: (i) the residual error between the calibrated simulations, $y_{sim}(x, \bar{\theta})$, and the simulation predictions with the true values

of the calibration parameters, $y_{sim}(x,\theta) = (\zeta(x) - \psi(x))$, (Equation 4.11); and (ii) the residual error between the true model form error, $\psi(x)$, and the estimated discrepancy, $\delta(x)$ (Equation 4.12). Therefore, absolute error, $\varphi(x)$, is defined in Equation (4.13) as the sum of $\varphi_y(x)$ and $\varphi_\delta(x)$. This calculation of absolute error allows for a true assessment of both the calibration of model parameters in the model predictions and the training of the discrepancy bias, while avoiding potential compensating effects realized when the direct sum of the discrepancy and predictions are compared to truth. Although this analysis may not be possible in real-world applications, it is necessary in this controlled study to properly assess the effectiveness of the proposed BSD criterion. Insights from the results of this simulation-based study will allow an effective and meaningful comparison of the proposed BSD criterion, C-EIPS, with the existing established criteria.

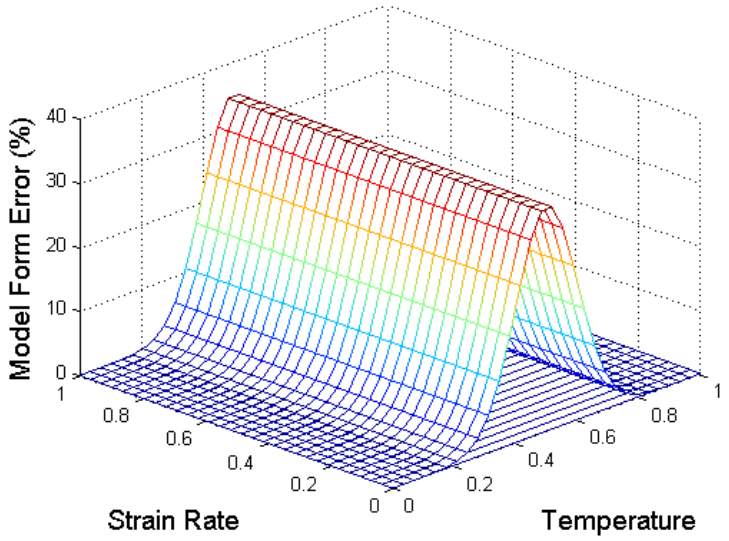

Figure 4.6: Model form error for stress calculations as a function of strain rate and temperature (Case 4).

Table 4.3: BSD Cases

Case	Feature											
	Stress				Text001				Text101			
	T Axis		$\dot{\varepsilon}$ Axis		T Axis		$\dot{\varepsilon}$ Axis		T Axis		$\dot{\varepsilon}$ Axis	
	A (%)	σ (%)	A (%)	σ (%)	A (%)	σ (%)	A (%)	σ (%)	A (%)	σ (%)	A (%)	σ (%)
0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	35	10	0	0	30	10	30	10
2	65	20	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	40	15
4	35	10	0	0	0	0	0	0	0	0	0	0
5	-40	10	-40	18	65	20	0	0	0	0	0	0

Table 4.4: Initial Experimental Settings

	Experiment	Temperature (˚C)	Strain-Rate (s^{-1})
Set 1	1	277.7	0.216
	2	238.9	0.002
	3	355.4	1.000
Set 2	1	316.7	0.022
	2	511.1	0.001
	3	472.2	0.216

For each of the three predicted features (stress at the maximum measured strain of 0.6 and the textural intensities at the 001 and 101 poles), normalized absolute error is computed for the mean model predictions and discrepancy biases. These error

calculations are then averaged over all three features to obtain a single value to represent the discrepancy for the PMI metric.

The PMI metric is computed according to Equation (4.5) for each criterion after each batch using γ values of 3, 0, 0.5, and 0, corresponding to coverage, complexity, discrepancy and robustness attributes, respectively (values are chosen within the recommended ranges provided in Stull et al. (2011)). Note that these choices of γ values put a higher weight on the coverage term than the discrepancy term, which requires high coverage in early batches for PMI to be maximized (these results are compared to alternative choices of γ values later in this discussion).

The coverage is computed in each case using the procedure described in Section 4.4. The resulting PMI values are plotted against the number of batches for each criterion in Figure 4.7. For brevity, only the results from Set 1 of the initial experimental settings are shown; however, note that the general trends and results shown are consistent with the results from Set 2 of the initial experimental settings. With the exception of Case 2, Figure 4.7 shows that in each case, the C-EIPS and EDIST criteria outperform the EIPS criterion. In Case 2, the C-EIPS and EIPS criteria perform very similarly and both outperform the EDIST criterion. Figure 4.7 also shows that the C-EIPS criterion outperforms EDIST through the first two batches in all cases and outperforms EDIST through the first five batches in a majority of the cases. Furthermore, for Set 1, over the first five batches of Cases 0-5 the PMI value for C-EIPS is on average 13.4% higher than EIPS and 4.8% higher than EDIST. For the remaining batches (6-10) the C-EIPS criterion outperforms the EIPS criterion by an average of 11.2% and is marginally

outperformed by EDIST by an average of 1.0%. Similarly for Set 2, the PMI value for C-EIPS is on average 7.3% higher than EIPS and 9.4% higher than EDIST over the first five batches. For the remaining batches (6-10) the C-EIPS criterion outperforms the EIPS criterion by an average of 7.1% and is marginally outperformed by EDIST by an average of 1.6%.

To investigate the behavior of the C-EIPS criterion for varying γ values, the previous analysis is completed for the case in which the γ values of 1 and 2 are chosen for the coverage and discrepancy terms, respectively. Unlike the first set of γ values, this case clearly puts more weight on the discrepancy term in the PMI calculation. However, the resulting trends are similar to the first analysis. For Set 1, over the first five batches, C-EIPS outperforms EDIST and EIPS by an average PMI of 3.3% and 9.2% respectively. In the remaining batches, C-EIPS outperforms EIPS by 14.1% while it is outperformed by EDIST by 3.2%. Set 2 shows similar results. While the trends among the performance of the criteria are similar for both sets of γ values, one noticeable difference is the larger residual difference between EDIST and C-EIPS in the later batches. This difference surfaces from the increased coverage required to maximize the hyperbolic tangent function (see Table 4.1) when a smaller γ value is used, especially when coverage is high; therefore, the criterion exhibiting higher coverage in the later batches, EDIST, will have a better performance than other criteria when the γ value associated with coverage is low.

These results indicate that when the selection of future experimental setting is most crucial (i.e. when discrepancy is high, coverage is low, and henceforth PMI is low) the C-EIPS criterion is the most efficient criterion for improving PMI. Furthermore, the

EDIST criterion only marginally outperforms the C-EIPS criterion when PMI is above 86% and at least four batches (11 experiments in this case) are collected (this trend is magnified when lower γ_1 values are utilized), which in many applications may correspond to an already acceptable level of discrepancy and coverage. Nonetheless, if very high PMI and coverage values are desired, the EDIST criterion should be considered for the latter batches.

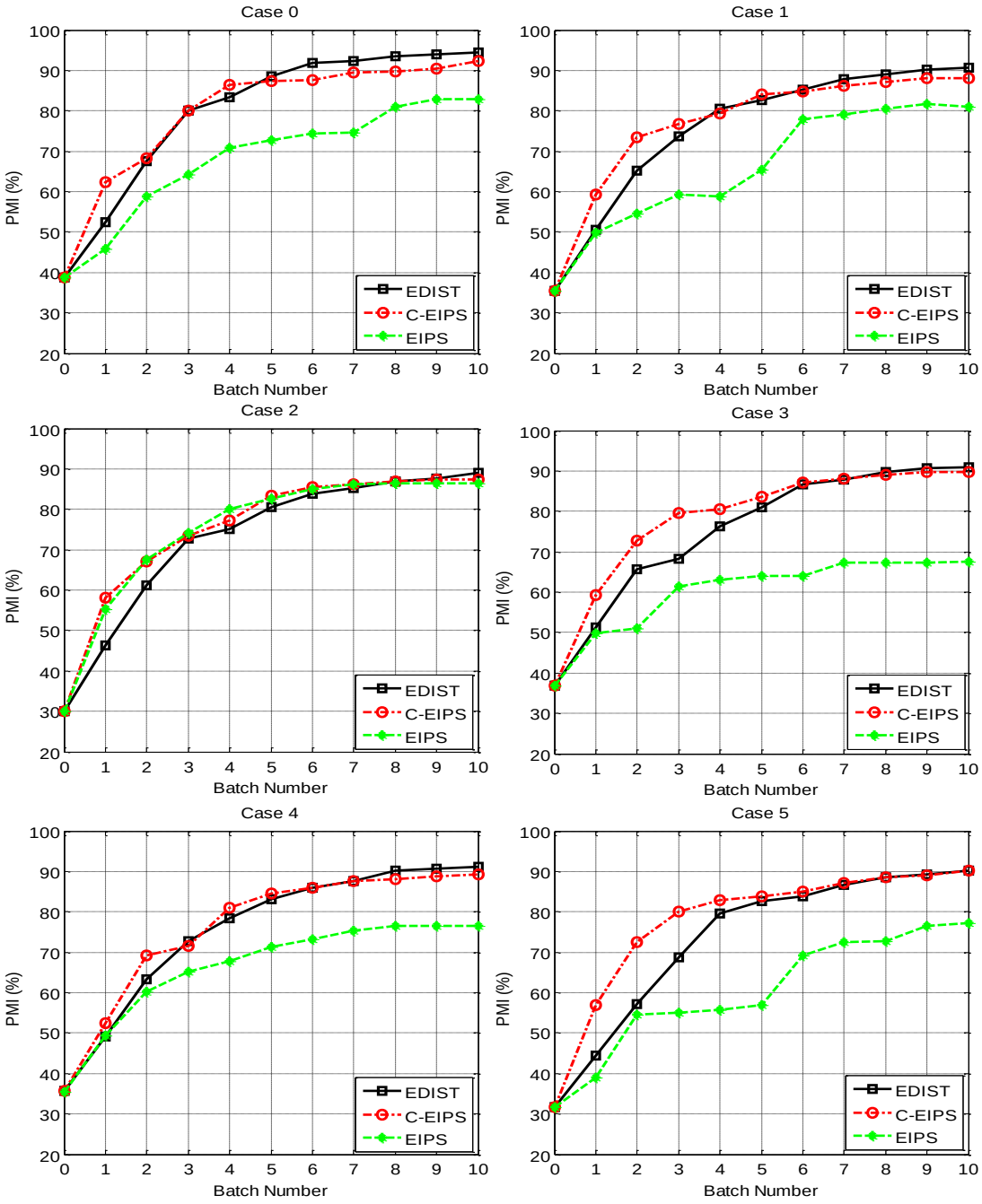

Figure 4.7: PMI vs. Batch Number for six different discrepancy cases for initial experimental set 1.

In Figure 4.8, the coverage at each batch is plotted against the computed absolute error for Set 1 of the initial experimental settings (similar trends are evident in Set 2). The

optimal criterion would result in the highest coverage and lowest absolute error with the least possible number of experiments. Figure 4.8 shows that with the exception of Cases 0 and 1, the C-EIPS results in the lowest final absolute error; the EDIST criterion results in the highest final coverage value, however. Each case shows a consistent reduction in absolute error and increase in coverage for each criterion (with the exception of Case 3). Note that for a majority of the cases, as confirmed in Figure 4.7, the C-EIPS criterion provides lower absolute error and higher coverage than the other criteria when coverage is relatively low and discrepancy is relatively high; however, in the later stages of the validation campaign, as discrepancy becomes small and coverage large, the EDIST criterion provides higher coverage than the other criteria, and in some cases, lower absolute error.

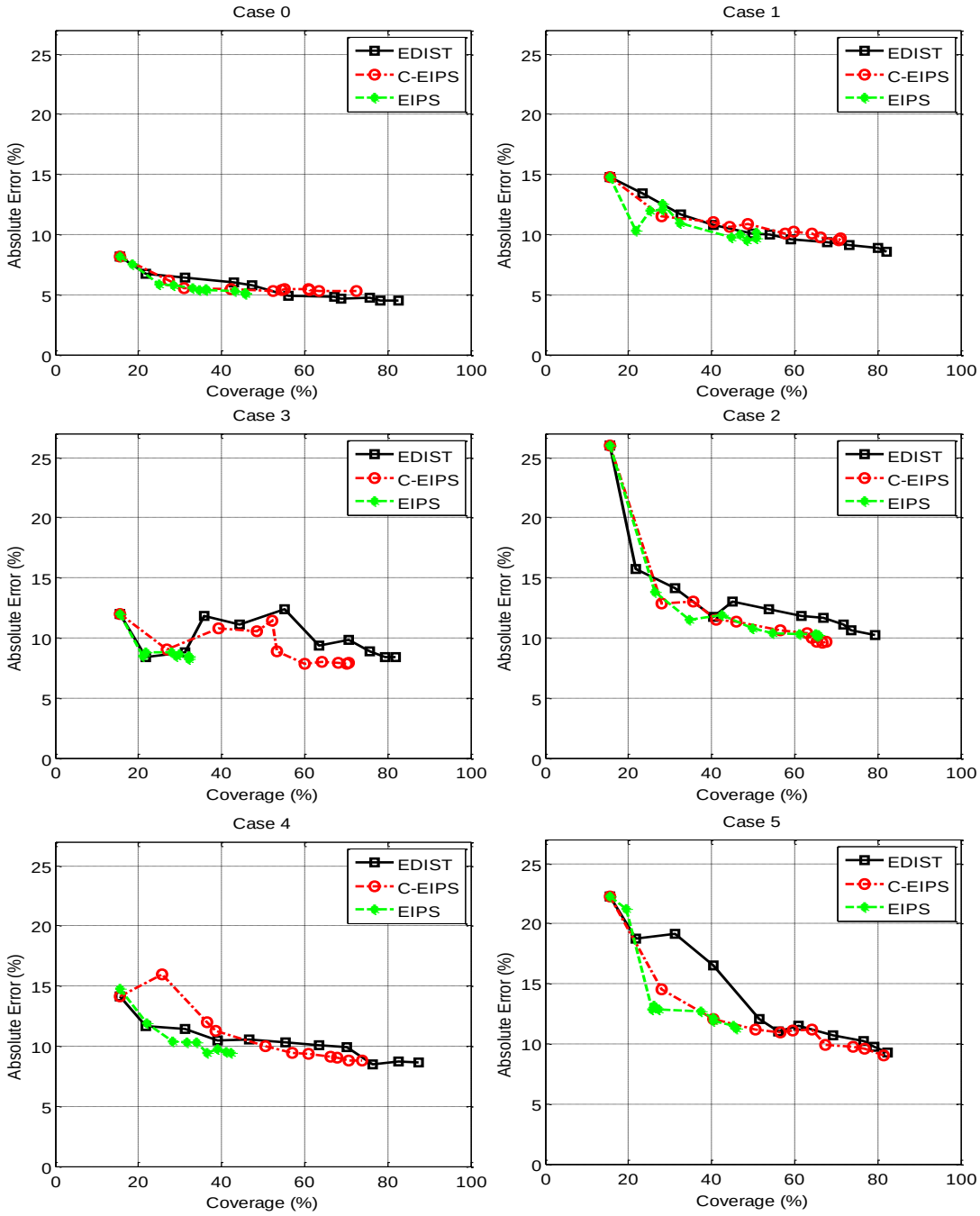

Figure 4.8: Coverage vs. Absolute Error for initial experimental set 1

4.8. Conclusions

In this chapter, a new optimal experiment selection criterion, C-EIPS, to be used in the BSD algorithm, is developed, the purpose of which is to efficiently improve the PMI of the numerical model. The C-EIPS is a metric that is comprised of both a coverage and discrepancy component. The coverage is directly computed for each potential design candidate and the discrepancy term is a normalized EIPS criterion, an effective criterion in minimizing the variance of the trained discrepancy GPM. The effectiveness of the C-EIPS is compared against the EIPS and EDIST criteria through a simulated case study using the climb and glide version of the VPSC code. The results of six different cases with varying degrees of imposed model form error for two different sets of initial experimental settings show that the C-EIPS is more effective than other criteria in the first half of the batches. This indicates that when optimal selection of experimental settings is most crucial for a model validation campaign (when PMI is low), the C-EIPS criterion is the most effective criterion. In the latter batches, as the PMI increases (approximately 86% or greater) the EDIST criterion performs equally well or marginally better than C-EIPS. In future studies, to improve the performance of the C-EIPS criterion in later batches, an adaptive weighting methodology can be considered. Additionally, further application of the proposed criterion to additional numerical models is necessary to provide additional confidence in the BSD methodology and criterion.

4.9. References

Allen, T.T., Lu, L., & Schmitz, J. (2003), An experimental design criterion for minimizing meta-model prediction errors applied to die casting process design, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52, 103-17.

Atamturktur, S., Williams, B.J., Hamutcuoglu, M.O., & Unal, C. (2012), Batch sequential calibration for improved predictive maturity in physics-based modeling, *International Journal for Numerical Methods in Engineering*, (under review).

Atamturktur, S., Hemez, F., Williams, B., Tomé, C., & Unal, C. (2011), A forecasting metric for predictive modeling, *Computers and Structure*, 89, 2377-87.

Ben-Haim, Y. (2006), Info-gap decision theory: decisions under severe uncertainty (2nd Edition ed.), Oxford, United Kingdom: Academic Press.

Blot, W.J. & Meeter, D.A. (1973), Sequential experimental design procedures, *Journal of the American Statistical Association*, 68(343), 586-93.

Chernoff, H. (1959), Sequential design of experiments, *The Annals of Mathematical Statistics*, 30(3), 755-70.

Cook, R.D. & Nachtsheim, C.J. (1980), A comparison of algorithms for constructing exact D-optimal designs, *Technometrics*, 22(3), 315-24.

Crary, S.B. (2002), Design of computer experiments for metamodel generation, *Analog Integrated Circuits and Signal Processing*, 32, 7-16.

Crombecq, K., De Tommasi, L., Gorissen, D., & Dhaene, T. (2009), A novel sequential design strategy for global surrogate modeling, Proceedings of the 2009 Winter Simulation Conference (WSC), 731-42.

Davis, R. & Prieditis, A. (1999), Designing optimal sequential experiments for a Bayesian classifier, *Pattern Analysis and Machine Intelligence*, IEEE Transactions, 21(3), 193-201.

Dawid, A.P. & Sebastiani, P. (1999), Coherent dispersion criteria for optimal experimental design, *The Annals of Statistics*, 27(1), 65–81.

Dror, H.A. & Steinberg, D.M. (2008), Sequential experimental designs for generalized linear models, *Journal of the American Statistical Association*, 103, 288-98.

Fedorov, V.V. (1972), Theory of optimal design, *New York: Academic*.

Harmon, S.Y. & Youngblood, S.M. (2005), A proposed model for simulation validation process maturity, *Journal of Defense Modeling and Simulation*, 2 (4), 179-90.

Hegenderfer, J., Atamturktur, S., Williams, B., Lebensohn, R., Yonten, K., & Unal, C. (2012), Framework for experiment-based validation of numerical models of complex systems: application to VPSC code for metals, *International Journal for Numerical Methods in Engineering*, (submitted for review).

Hemez, F., Atamturktur, S., & Unal, C. (2010), Defining predictive maturity for validated numerical simulations, *Computers and Structures Journal*, 88, 497-505.

Higdon D., Gattiker J., Williams B., & Rightley M. (2008), Computer model calibration using high-dimensional output, *Journal of the American Statistical Association*, 103(482), 570-83.

Huan, X. & Marzouk, Y. (2011a), Optimal Bayesian experimental design for combustion kinetics, 49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, 4-7 January 2011, Orlando, Florida.

Huan, X. & Marzouk, Y. (2011b), Simulation-based optimal Bayesian experimental design for nonlinear systems, *Journal of Computational Physics*, (submitted for review).

Jiang, X. & Mahadevan, S. (2006), Bayesian cross-entropy methodology for optimal design of validation experiments, *Meas. Sci. Technol*. 17, 1895-908.

Jin, R., Chen, W., & Sudjianto, A. (2005), An efficient algorithm for constructing optimal design of computer experiments, *Journal of Statistical Planning and Inference*, 134, 268-87.

Kennedy M. & O'Hagan A., 2001, Bayesian calibration of computer models (with discussion). *Journal of the Royal Statistical Society Series B,* 68:425–64.

Lebensohn, R.A., Hartley, C.S., Tomé, C.N., & Castelnau, O. (2010), Modeling the mechanical response of polycrystals deforming by climb and glide, *Philosophical Magazine,* 90 (5), 567-83.

Lebensohn, R.A. & Tomé, C.N. (1993), A self-consistent anisotropic approach for the simulation of plastic deformation and texture development of polycrystals: application to zirconium alloys. *Acta metal. Mater*, 41 (9), 2611-23.

Lindley, D.V. (1972), Bayesian Statistic: A Review. SIAM, Philadelphia, USA.

Loeppky, J.L., Moore, L.M., & Williams, B.J. (2010), Batch sequential designs for computer experiments, *Journal of statistical Planning and Inference*, 140, 1452-64.

Muller, W.G. & Potscher, B.M. (1989), Batch sequential design for a nonlinear estimation problem, *Forschungsbericht,* Research Memorandum No. 259.

NASA (2007), Standard for Models and Simulation. Technical Standard NASA-STD-7009, National Aeronautics and Space Administration, Washington, DC; November 2007.

Oberkampf, W.L., Pilch, M., & Trucano, T.G. (2007), Predictive capability maturity model for computational modeling and simulation, Sandia National Laboratories Technical Report SAND-2007-5948.

Sebastiani, P. & Wynn, H.P. (2000), Maximum entropy sampling and optimal Bayesian experimental design, *Journal of the Royal Statistical Society,* Series *B* (Statistical Methodology), 62(1), 145-57.

Sornette, D., Davis, A.B., Ide, K., & Kamm, J.M. (2007), Theory and example of a new approach to constructive model validation, NATO RTA AVT-147 symposium on computational uncertainty; December 3-6, 2007; Athens, Greece. (Also, Los Alamos National Laboratory Technical Report LA-UR-07-7013.)

Sornette, D., Davis, A.B., Kamm, J.R., & Ide, K. (2006), 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, September 9-14; Lake Tahoe, California; Berlin: Springer.

Stull, C.J., Hemez, F., Williams, B.J., Unal, C., & Rogers, M.L. (2011), An improved description of predictive maturity for verification and validation activities, Los Alamos National Laboratories Technical Report, LA-UR-11-05659.

Terejanu, G., Upadhyay, R.R., & Miki, K. (2012), Bayesian experimental design for the active nitridation of graphite by atomic nitrogen, *Experimental Thermal and Fluid Science*, 36, 178-93.

Thompson, D.E., McAuley, K.B., & McLellan, P.J., (2010), Design of optimal sequential experiments to improve model predictions from a polyethylene molecular weight distribution model, *Macromolecular Reaction Engineering*, 4(1), 73-85.

Williams, B.J., Loeppky, J.K., Moore, L.M., & Macklem, M.S. (2011), Batch sequential design to achieve predictive maturity with calibrated computer models, *Reliability Engineering and System Safety*, 96, 1208-19.

Williams, B.J., Santner, T.J., & Notz, W.I. (2000), Sequential design of computer experiments to minimize integrated response functions, *Statistica Sinica*, 10, 1133-52.

Zang, T.A. (2008), Perspectives on uncertainties (and margins) in NASA engineering decisions, 10th AIAA Non-Deterministic Approaches Conference, April 7-10, 2008, Schaumburg, Illinois.

CHAPTER 5: PRIORITIZATION OF CODE DEVELOPMENT EFFORTS IN PARTITIONED ANALYSIS

5.1. Introduction

Numerical simulation has become a viable tool for investigating complex physical systems and processes that are encountered in many civil engineering disciplines, especially when solely experiment-based, empirical studies are infeasible. While numerical models offer versatility in simulating various operational conditions or design scenarios, they can only provide an approximation of reality, and thus, there is a need to check the validity of model solutions against experimental observations. Numerical models therefore come with the burden of quantifying the uncertainties and biases in model predictions through tasks that fall under the broad concept of *model validation*.

To guide the *model validation efforts*, many frameworks, such as those presented in Unal et al. (2011), Jung (2011), Bayarri et al. (2007), and Jiang and Mahadevan (2007), and this dissertation have been developed to rigorously and quantitatively assess model biases and uncertainties (Figure 5.1). Each of these frameworks demonstrates that model validation efforts can be improved through either additional experimentation or further code development. If additional experimentation is necessary, the efficient and optimal selection of experimental settings can be achieved through algorithms such as those presented in Jiang and Mahadevan (2007) and Williams et al. (2011). However, if code development is selected as the next step in model validation, the particular elemental components of the numerical model requiring more sophisticated modeling of physics or engineering principles remains unknown. Naturally, unlimited time and

resources would render the prioritization of such code development efforts irrelevant. However, given the inevitable limitations in resources, major model deficiencies must be pinpointed and code development efforts must be focused to achieve the greatest reduction in model uncertainty and bias. The scientific problem of code prioritization has been studied to optimize software fault detection (Korel, 2009) and parallel processing of finite element computer codes (Zeyao and Lianxiang, 2004); however, quantitative approaches for code prioritization of complex numerical models are lacking in the published literature.

Such need becomes especially amplified for complex, heterogonous systems that are driven by the interaction of functionally distinct but strongly coupled constituents commonly tackled by partitioned analysis procedures. The focus herein is therefore on models that are comprised of multiple, isolated, components (or substructures), herein referred to as *constituents*. These constituents interact to form a collective system, herein referred to as a *coupled model*. Such focus is relevant since many numerical models in civil and infrastructure engineering are indeed coupled models that are an amalgam of multiple constituents or systems of constituents; see for instance, the published literature on soil-structure interaction (Provenzano, 2003; Qian and Zhang, 1993), fluid-structure interaction (Kutay and Aydilek, 2009; Caracoglia et al., 2009), human-structure interaction (Macdonald, 2009; Wang et al., 2011), and the broad field of sub-structuring, which essentially focuses on structure-structure interaction. While aiming to improve the predictive accuracy of such coupled simulation models, one obvious question arises: to

achieve the greatest reduction in model uncertainty and bias, which constituent must be given the highest priority for further code development?

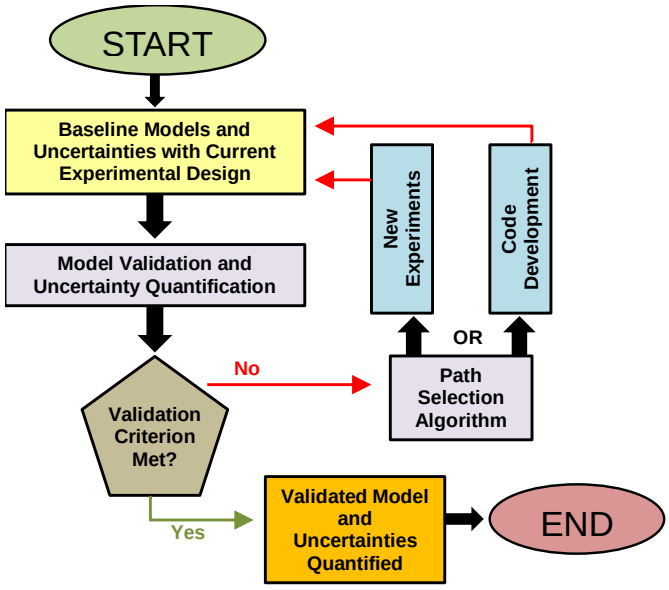

Figure 5.1: Generic simulation model validation framework.

This chapter presents a code prioritization approach for coupled numerical models through the use and extension of a well-established ranking system, in which code development efforts are guided to effectively improve the predictive capability of the coupled model. A novel, quantitative code prioritization index is proposed and demonstrated using a proof-of-concept example, in which the structural system consists of a two-story steel frame built in the laboratory with semi-rigid connections. An initial, simplified finite element (FE) model is developed for the frame system using beam elements, where the beam-to-column connections can be modeled as either fixed or pinned connections, neither of which are representative of reality. One approach commonly utilized in the literature entails adding fictitious *'knobs'* (empirical or arbitrary

parameters used in place of detailed physics modeling) to represent these inherently semi-rigid connections. This initial FE model can then be improved by refining the definition of these knobs, such as through the development of high-fidelity three-dimensional, nonlinear FE models (constituents). Three possible constituent connection models are identified to be coupled to the initial simplified frame FE model. These three constituents are ranked using the proposed Code Prioritization Index (CPI) that combines knowledge level, importance, and error contribution. The frame model is then improved incrementally by coupling the three high fidelity connection models in the sequence selected by CPI, and the resulting improvement in predictive accuracy of the coupled model is quantified.

The chapter is organized as follows: Section 5.2 overviews the pertinent literature on coupling algorithms and ranking procedures that are used as the foundation for the code prioritization index described in Section 5.3. Section 5.4 introduces the case study application, in which the predictions of the initially inexact FE model of the steel frame are compared against experimentally obtained static and dynamic characteristics. In Section 5.5, the code prioritization index is deployed to rank model constituents, which are developed, verified and validated in Section 5.6. The coupling procedure used to integrate the constituent models to the frame model is described in Section 5.7. Section 5.8 presents a comparison of the initial FE model as well as its improved variants against experimental data. Finally in Section 5.9, concluding remarks are made, limitations of the proposed approach are summarized and the future direction is overviewed.

5.2. Background

5.2.1. Phenomenon Identification and Ranking Table

Originally proposed as part of the U.S. Nuclear Reactor Commission's (NRC) Code Scaling, Applicability and Uncertainty (CSAU) evaluation procedure (Boyack et al., 1989), the Phenomenon Identification and Ranking Table (PIRT) is currently being used by the U.S. NRC at the start of new programs to rank constituent phenomena from the perspective of resource allocation (ARRIA, 2003; Olivier and Nowlen, 2008; Tregoning et al., 2009). The purpose of PIRT is to effectively gather expert opinion about the importance and knowledge level of a set of phenomena (Diamond, 2006). The PIRT process, typically completed by a committee of multi-disciplinary experts, provides a systematic, structured, and hierarchical methodology to rank phenomena of interest for resource allocation problems (Boyack, 2009). The two major scoring components of PIRT, (i) the importance of a phenomena and (ii) the level of current knowledge, constitute the particular phenomena's sensitivity or impact on the evaluation metric and the current lack of knowledge about the phenomena (and/or parameters associated with that phenomena), respectively. Although historically PIRT has mainly been used in the nuclear reactor and safety fields, the general and high-level approach to the PIRT process makes it highly adaptable to many different areas of science and engineering.

5.2.2. Partitioned Analysis Procedures

In partitioned analysis procedures, constituent models are viewed as discrete entities with data transferred at the interface between the individual constituent codes through coupling algorithms (Rugonyi and Bathe, 2001). This type of analysis most often

results in an iterative procedure involving prediction, substitution, and organization techniques (Felippa et al, 2001; Larson et al., 2005; Larson, 2009). The advantage of the partitioned approach stems from the ability to exploit independent modeling strategies developed in different domains (Leiva et al., 2010) in addition to time and space discretization that is most appropriate for each constituent (Kassiotis et al., 2011; Joosten et al., 2009). This flexibility obtains solutions for highly complex coupled problems while making efficient use of well-established codes and expert knowledge in various fields (Ibrahimbegović et al., 2004). Furthermore, the ability to solve complex problems through integrated parallelization on multiple sets of processors can make the partitioned approach particularly efficient (Park and Felippa, 1983).

Coupling is most commonly classified according to the nature of the mutually dependent parameters. In weak coupling (also known as loose or explicit coupling), the state of the constituents does not mutually interact with each other's inputs or outputs (Matthies and Steindorf, 2002a; Wang et al., 2004). In partitioned analysis, if the state of the constituents are affected by model outputs, the coupling interface is referred to as strong coupling (also known as tight or implicit coupling) (Matthies and Steindorf, 2003; Zhang and Hisada, 2004). In strong coupling, the repeated execution of constituent models to achieve self-consistent state solutions becomes necessary (Ramanath, 2011). Input parameters in strong coupling problems are defined as either dependent or independent. The dependent input parameters are functions of the output of another constituent; therefore, the coupling algorithm must evaluate and substitute these parameters as input for the appropriate constituent. Finding the correct values for these

dependent variables is the main question to be solved in strong coupling problems. Various strong coupling algorithms are proposed in the literature (Figure 2a-d): the Block-Jacobi method (Matthies et al., 2006; Fernandez and Moubachir, 2005), the Block Gauss-Siedel method (Joosten et al., 2009; Matthies et al., 2006), gradient-based Newton-like methods (Heil, 2004; Matthies and Steindorf, 2002b; Matthies and Steindorf, 2003; Fernandez and Moubachir, 2005), and optimization-based methods (Farajpour and Atamturktur, 2012).

Note that in partitioned analysis, quantification, propagation, and mitigation of uncertainties in input parameters play an important role for the complete validation of coupled models (Avramova and Ivanov, 2010), which is currently an active research area. The focus of this chapter however, is exclusively on improved *sophistication* of constituent models to reduce the systematic bias in coupled models predictions; therefore, uncertain input parameters are treated as deterministic best estimates.

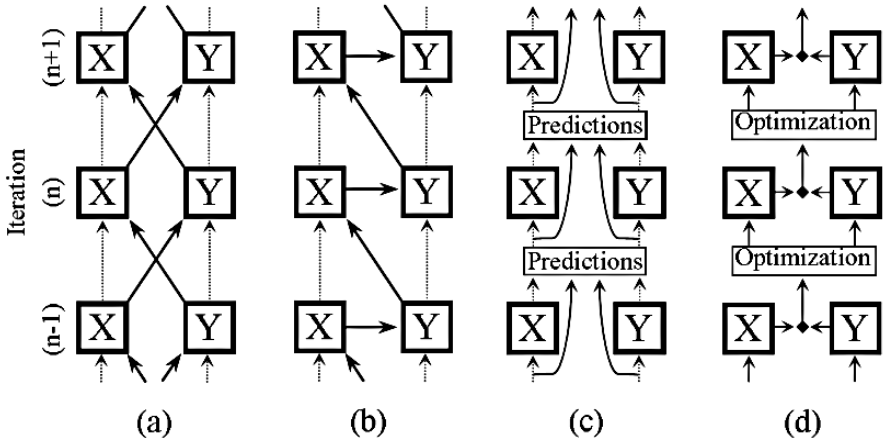

Figure 5.2: Coupling algorithms: (a) Block-Jacobi, (b) Block Gauss-Seidel, (c) Block Newton, and (d) Optimization-based Coupling.

5.3. Code Prioritization Index

The PIRT, overviewed in Section 5.2.1, supplies an inherently qualitative and subjective evaluation of a system that is driven by the interaction of coupled constituent phenomena. As such, this section details the transformation of PIRT into a quantitative approach through the Code Prioritization Index (CPI). Following an approach similar to Hemez et al. (2010), CPI is formulated considering the three factors that determine the importance a constituent has on the overall predictive accuracy of a coupled simulation: (i) the sensitivity of the coupled system behavior to the constituent output, (ii) the estimated biases in the constituent output and (iii) the estimated uncertainties in the constituent. CPI, therefore, takes the following form:

$$CPI_i = \|SA_i\| \times \|EA_i\| \times \|UA_i\| \tag{5.1}$$

where SA_i, EA_i, UA_i respectively represent the importance (*Sensitivity Analysis*), estimated error (*Error Analysis*), and current level of knowledge (*Uncertainty Analysis*) for a given constituent *i*, in an appropriately chosen norm, $\|\bullet\|$.

Sensitivity Analysis: The SA term is defined as the measure of the variability in the model output due to unit variability in a given constituent. This term can be quantified using various established screening, local or global sensitivity analysis techniques; however, for complex problems that are nonlinear in nature and that involve a large number of input parameters of varying uncertainty, global sensitivity analysis techniques should be preferred (Cukier et al. 1973). The global sensitivity analysis inherently takes the range and shape of the probability distributions of input parameters into account. Moreover, in global sensitivity analysis techniques, such as the analysis of variance

(ANOVA) (Moaveni et al., (2009); Frey and Patil, 2002), the sensitivity estimates for each parameter are computed in the presence of uncertainty of all other factors of interest; thus taking the possible correlations and dependencies between input parameters into consideration.

Error Analysis: The EA term is a representative estimate of a constituent's error contribution to the total error of the system in comparison to the experimental evidence. This term can be quantified using various metrics including model form error estimates for separate effect experiments as defined in Higdon et al. (2008), Oberkampf and Barone (2006), Rebba et al. (2006) and Atamturktur et al. (2011).

In the present study, for a given constituent, EA is defined as a measure of the constituent 'correction' necessary such that error in the coupled model predictions is minimized (given a set of uncertainty attributed to other constituents) as shown in Equations (5.2)-(5.3):

$$P_j = \{P : \delta_j(P, \alpha) < \delta_j^t\} \tag{5.2}$$

$$EA = \frac{1}{n}\sum_{j=1}^{n}\left(\min\left|P_0 - P_j\right|\right)\Big/P_0 \tag{5.3}$$

where n is the number of experiments; P_0 represents the nominal value for the *fictitious knob*; P_j is the selected parameter value for the constituent at the j^{th} experiment; $\delta_j(P, \alpha)$ is the discrepancy of the constituent at the j^{th} experiment defined as a function of collective parameters of the constituents, and; α is the uncertainty assigned to all other

uncertain parameters. δ_j^t is the discrepancy threshold value at the j^{th} experiment for the constituent of interest, which is derived in Equations (5.4)-(5.7).

An absolute minimum discrepancy exists if Equations (5.4)-(5.5) given below are both satisfied.

$$\frac{\delta_j^k - \delta_j^{k-r}}{P(\delta_j^k) - P(\delta_j^{k-r})} < -\gamma \tag{5.4}$$

$$\frac{\delta_j^{k+r} - \delta_j^k}{P(\delta_j^{k+r}) - P(\delta_j^k)} < +\gamma \tag{5.5}$$

where γ is the limiting slope as indicated in Figure 5.3 and *r* represents the number of runs over which the slope is computed (see Figure 5.3); *k* is the index number of computer runs corresponding to the minimum discrepancy over all simulation runs, and δ_j^k is the minimum computed discrepancy across all simulation runs. In this case, the threshold value for discrepancy is selected according to Equation (5.6); otherwise, a converged minimum discrepancy exists and the threshold value for discrepancy is calculated according to Equation (5.7);

$$\delta_j^t = \delta_j^k \tag{5.6}$$

$$\delta_j^t = \beta\left(\delta_j^S - \delta_j^k\right) + \delta_j^k \tag{5.7}$$

where β represents a percentage of the total allowable reduction discrepancy to limit excessive changes in the nominal value of the knob (see Figure 5.3) and δ_j^S represents the maximum computed discrepancy across all simulation runs.

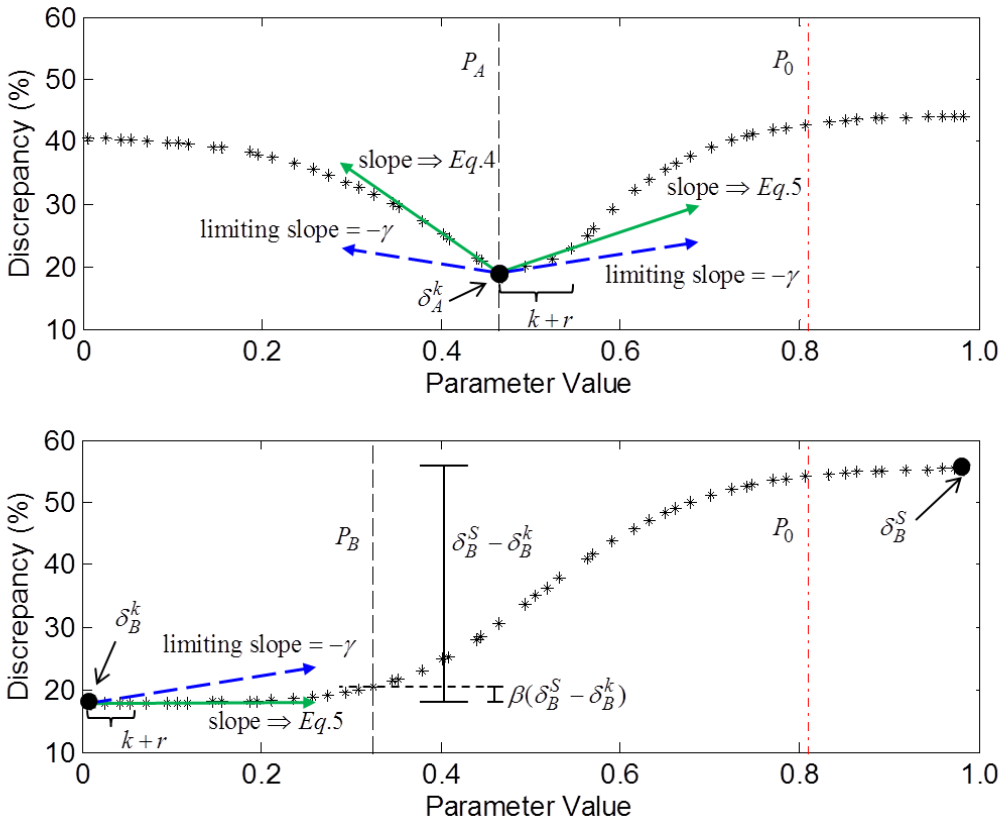

Figure 5.3: Computation of P_j : Absolute Minimum Discrepancy: Equation (5.6) (Top) and Converged Minimum Discrepancy: Equation (5.7) (Bottom).

The EA term in Equation (5.3) seeks to determine the necessary deviation from the initial parameter value over all available experiments such that the relative difference between model predictions and experimental results (referred to herein as discrepancy) is minimized. To achieve this minimization, the numerical model parameters of interest are sampled within a plausible range defined for parameter of interest, P, and a set percentage variability, α, for all other parameters, thus accounting for potential cross-correlations between parameters through the α variability term. Computer runs can be

generated using various sampling designs that thoroughly explore possible values for P in an efficient manner. The EA value realized from Equation (5.3) is represented as a percentage of the initially assumed parameter value, P_0. Therefore, higher EA values indicate a greater estimated error for the parameter of interest.

The values of *r*, γ and β (from Equations (5.4)-(5.7)) are assigned according to the specific application of interest. If discrepancy throughout the investigated domain is constant (the calculated slope is below the limiting value of γ) or if the discrepancy behavior is divergent for any specific experiment *j*, then a meaningful and confident evaluation of P_j is difficult. Thus, if the given experiment is not sensitive to parameter P, a selection of P_j would be unfounded; therefore, the contribution from such experiments is excluded from the analysis.

Note that if model calibration of constituents is completed prior to the utilization of the CPI metric, the calibrated parameter values for the parameters would be used as the initial values for the EA analysis. When such calibration activities are not available a priori, the proposed method gives the best possible guidance by directing the focus of the decision makers and code developers to the constituent that is responsible for the inaccuracies of the coupled model prediction. The inaccuracies in the selected constituent may originate from either the incompleteness of the modeled physics or engineering principles or the imprecision in the model parameters.

Uncertainty Analysis: The UA term can be quantified for a particular model constituent according to experimental test results or qualitative expert opinion. Herein, however, the current level of knowledge, i.e. uncertainty analysis, for each constituent

will be treated as a binary number, where zero represents a phenomenon for which the knowledge level is mature and where one represents a phenomenon that is not yet well known. In effect, as recommended in the PIRT construction process, the UA term restricts code prioritization efforts only to poorly or partially understood phenomena.

Therefore, the CPI term is maximized when SA, EA, and UA are all high. Herein, a type of norm for each of the CPI terms is chosen such that each term is scaled to a maximum of one over all possible constituents; therefore, CPI values are, without loss of generality, bounded between zero and one, and higher CPI terms are considered as higher priority.

5.4. Overview of the Case Study Application

5.4.1. Case Study Structure: Laboratory Specimen

The prototype structure is a two story single bay steel frame shown in Figure 5.4, in which all members of the frame are made of mild steel. The connections are secured with two bolts on each side, and all bolts are torqued to 9.0 Newton meters. The geometric dimensions of the test frame are given in Table 5.1.

Table 5.1: Geometric properties of the laboratory frame structure.

Steel Frame Member Geometric Properties			
Member	Length (cm)	Cross-Section Type	Cross-Section Dimensions (cm)
Columns	63.5	Angle	5.08 x 5.08 x 0.32
Beam	124.46	Flat	5.08 x 0.32
Base Connecting Tabs	5.08	Angle	5.08 x 5.08 x 0.32
Column Base Plates	1.27	Flat	15.24 x 15.24
Frame Base	1.27	Flat	121.92 x 243.84
Bolt Properties			
Bolt Type		Bolt Diameter (cm)	
Grade 5		0.66 cm	

Figure 5.4: Frame structure built in the laboratory (Left) and frame connections (Right): Top connection (Top), Middle connection (Middle), Base connection (Bottom).

5.4.2. Initial Numerical Model

The baseline finite element (FE) model for the frame (henceforth referred to as Model #1) is created in ANSYS v.13.0 using BEAM188 elements as shown in Figure 5.5. These are two-noded six-degree-of-freedom linear elements that consider the cross-section and orientation of the member (ANSYS, 2011). An isotropic linear elastic constitutive model is used with Young's modulus specified as 200 GPa (29,000 ksi) and Poisson's ratio as 0.33, which are typical values for mild structural steel. Similar to the approach used by Zapico et al. (2008), geometric offsets are considered when modeling the beam-to-column connections (also known as modeling to centroid). The two-bolt connections of the frame are expected to exhibit semi-rigid behavior with unknown stiffness characteristics.

In the literature, such semi-rigid connections are often approximated as either fixed or pinned depending upon the number of bolts, the bolt pattern, and expert judgment (Lee and Moon, 2002; Galvão et al., 2010). Therefore, an initial numerical model (*Model #1*) is built with similar approximations commonly implemented in practice. For Model #1, to determine a suitable connection type, either fixed or pinned connection, model predictions are compared against experimentally obtained static and dynamic response. In the fixed model, all connections are assumed to be fixed. However, for stability, in the pinned model the rotational degrees of freedom at the base connections as well as the top and middle level torsional degrees of freedom are fixed, while all remaining rotational degrees of freedom at the top and middle levels are pinned. After the selection of a suitable connection type, Model #1 will ultimately be used as a

reference while determining the improvement gained in predictive accuracy by refining the way in which beam-to-column connections are represented in the FE model.

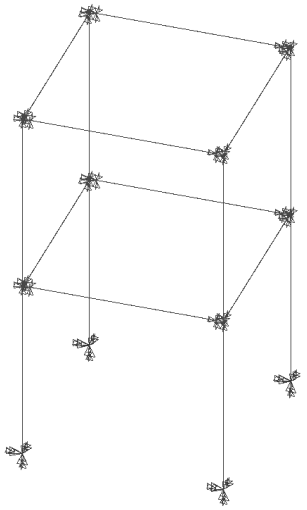

Figure 5.5: Baseline FE model (Model #1) for the frame structure.

5.4.3. Experimental Campaign

Static Testing: Static testing is conducted using a cable and pulley system designed to apply a point load in the horizontal direction as shown in Figure 5.6. Federal C8IS Dial Deflection indicators are positioned at the locations shown in Figure 5.6 to measure horizontal deflection. Six separate static tests are conducted using point loads with increased amplitude and applied at a different location on the structure as shown in Figure 5.6, while displacements are measured at four points (L_1-L_4) along a single column for Tests 1-3 and two points (L_5-L_6) on the beams for Tests 4-6. The displacement results for the static tests, shown in Figure 5.7, are compared to both the pinned and fixed connection FE models in Table 5.2. The static tests are completed at

force levels where the material has not yielded, i.e. it is still in the elastic regime as is demonstrated in the linear response shown in Figure 5.7.

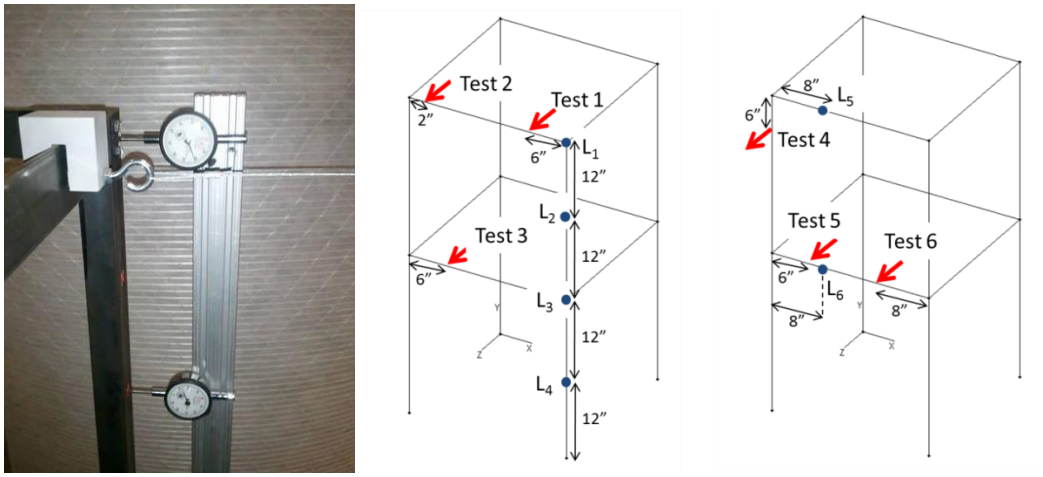

Figure 5.6: Static experiment set-up: Cable and pulley system (Left), Measurement locations (Right).

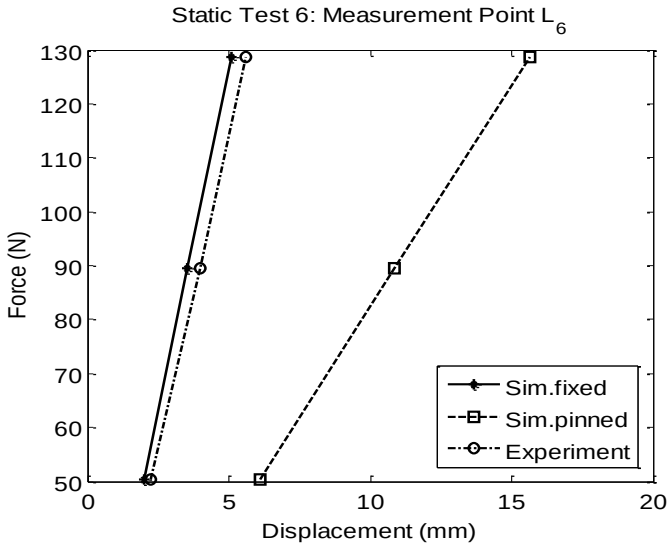

Figure 5.7: Static test comparison between two alternative FE models (pinned connections and fixed connections) and experimental data.

Modal Testing: Model 4507B Brüel and Kjær (B&K) unidirectional accelerometers are distributed to 88 locations, 6 inches apart throughout all beams and columns (see Figure 5.4), and a Model 8207 B&K modal impact hammer is used to excite the structure. Each hammer strike is repeated five times to reduce degrading effects of noise through averaging. Using Reflex software from B&K, natural frequencies (see Table 5.3) and mode shapes (see Figure 5.8) are extracted from the experimental data. The rational fraction polynomial parameter estimation technique (Richardson and Formenti, 1982) is implemented to generate stability diagrams for mode selection. The first three global modes of the structure are identified as shown in Figure 5.8 and listed in Table 5.3. As a means of verifying the orthagonality of the experimentally collected modes, the *modal assurance criterion* (MAC) (Allemang, 2003) is calculated (Table 4). The low off-diagonal terms in Table 5.4 demonstrate the linear independence of each of the experimentally determined modes.

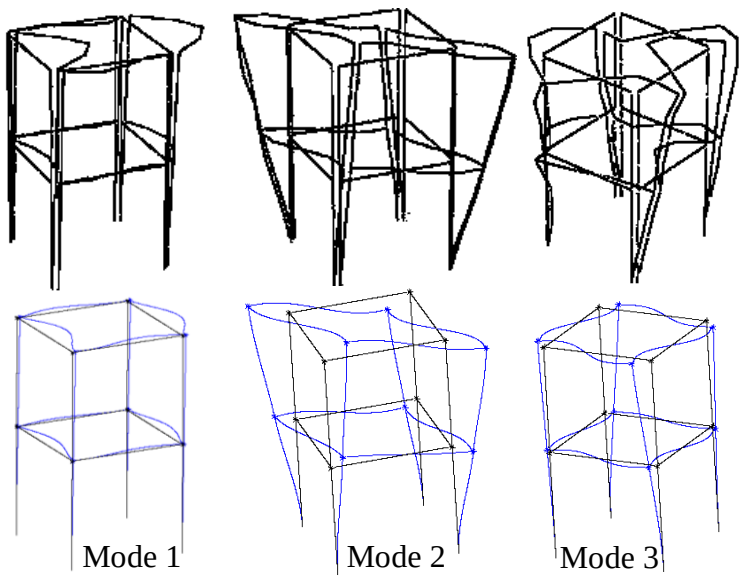

Figure 5.8: Mode shapes: Experimental testing (Top), and FE model predictions (Bottom).

5.4.4. Test-Analysis Correlation

Tables 5.2 and 5.3 show that the baseline model with fixed connections underestimates the deformations and overestimates the natural frequencies, while the baseline model with pinned connections overestimates the deformations and underestimates the natural frequencies. As seen, neither the fixed nor the pinned connections are suitable to represent the connections of the frame of interest. However, the FE model with fixed connections more accurately predicts the static tests with a 46.1% lower average error than the FE model with pinned connections. Also, the FE model with fixed connections predicts the values of the natural frequencies in the correct order of modes, while pin connections yield an incorrect mode sequence.

Although the FE model with fixed connections yields a closer agreement to experiments and has been selected as the connection type to be used in Model #1, it still has an approximately 20% average error for natural frequencies and 45% average error over all static tests, necessitating further improvement of the model sophistication.

Table 5.2: Comparison of baseline models and experimental data for static tests.

Test	*1*	*2*	*3*	*4*	*5*	*6*	*Mean*
Fixed FE Average Error (%)	44.4	84.8	55.6	46.8	24.3	18.9	45.8
Pinned FE Average Error (%)	32.8	196.5	76.7	18.8	100.5	126.3	91.9

Table 5.3: Natural frequencies from experimental testing and baseline models.

Mode	*Description*	*Test: Frequency (Hz)*	*Fixed FE Frequency (Hz)*	*Error (%)*	*Pinned FE Frequency (Hz)*	*Error (%)*
1	Sway	23.09	27.97	21.13	16.30	29.41
2	Bending	29.58	33.21	12.27	14.01	52.64
3	Torsion	37.04	47.56	28.40	25.96	29.91

Table 5.4: MAC of experimentally collected modes.

Mode	*1*	*2*	*3*
1	1	0.003	0.002
2	0.003	1	0.001
3	0.002	0.001	1

5.5. Identification and Prioritization of Constituents

The structure of interest is comprised of two main constituents: the steel members and the connections. Though the material behavior of mild steel is widely studied, the bolted semi-rigid steel connections between beam and column are highly uncertain and are of great interest in the design and analysis communities. Lee and Moon (2002) and Eurocode 3 (1993) provide provisions for approximate analysis of semi-rigid steel connections; however, many assumptions are made while establishing these provisions, such as small deformation in the connection, negligible deformation in the beam and column compared to the deformation in the connection, and negligible slip deformation.

Several investigators acknowledged and demonstrated that traditional assumptions of connections being either fixed or pinned can lead to significant errors in the analysis of static (Lee and Moon, 2002) and dynamic response (da Silva et al., 2008;

Galvão et al., 2010; Türker et al., 2009), design and sizing of members (Lee and Moon, 2002) as well as design against progressive collapse (Liu et al., 2010). The findings obtained in Section 5.4 of this study also support these earlier studies that neither the fixed nor the pinned connections are acceptable to represent the two-bolt connection used in the steel frame studied herein. Therefore, while demonstrating the application of the proposed CPI metric, the present study demonstrates mitigation of potential errors originating from the incomplete and inexact modeling of the semi-rigid steel connections through a rigorous modeling approach, where connections are parameterized and treated as constituents.

Similar to da Silva et al. (2008), three linear, rotational springs are added at each connection of the initial FE model (Model #1), which are used to capture rotations of the connection in all three directions. The springs are modeled using linear, COMBIN14 (spring-damper) elements (ANSYS, 2011). As the spring constants tend to zero, the connection behaves as a pin, and as the spring constants tend to infinite the connection becomes fixed. The stiffness of the bolted connections utilized in the frame structure however, lie somewhere between these two extremes with unknown spring constants. Also note that though this linear representation of semi-rigid springs supplies an improvement beyond a fixed or pinned connection, it fails to incorporate nonlinear effects.

The FE model can be further improved by developing three-dimensional, nonlinear FE models of the connections that account for friction between the members and pretension in the bolts (as it will be discussed in Section 5.6). These high-fidelity,

nonlinear FE models of connections must be developed separately for top, middle and base connection; however, the limitations on project resources may inhibit model development of all three connections. Thus, to efficiently utilize the available resources, one must prioritize among these three possible constituent models to achieve the highest improvement in predictive capability. Such prioritization is achieved herein using the linear FE models of the constituent along with the CPI metric introduced in Section 5.3. The CPI requires a sensitivity analysis, and error estimate analysis, which are discussed in the following sections.

5.5.1. Sensitivity Analysis

Herein, a global effect sensitivity analysis, ANOVA, is utilized. The selection of the plausible ranges, within which spring constants may vary, is critical in ANOVA; therefore, plausible ranges are determined by varying individual spring constants from free to an almost fixed condition one at a time, while the remaining spring constants are left constant at their nominal values. See for instance Figure 5.9, where the resulting natural frequencies are plotted as the strong axis bending stiffness constants for top level beams is varied. As one spring constant is varied, the natural frequencies vary between an upper bound (corresponding to a fixed connection) and a lower bound (corresponding to a pinned connection) in an asymptotic manner. Appropriate ranges for each spring constant can then be selected to ensure a semi-rigid behavior. As demonstrated in Figure 5.9, for the strong axis bending stiffness constant at the top level, the range is chosen to be $1.1\text{x}10^{1}$ –$1.1\text{x}10^{6}$ newton meters as indicated with dashed lines.

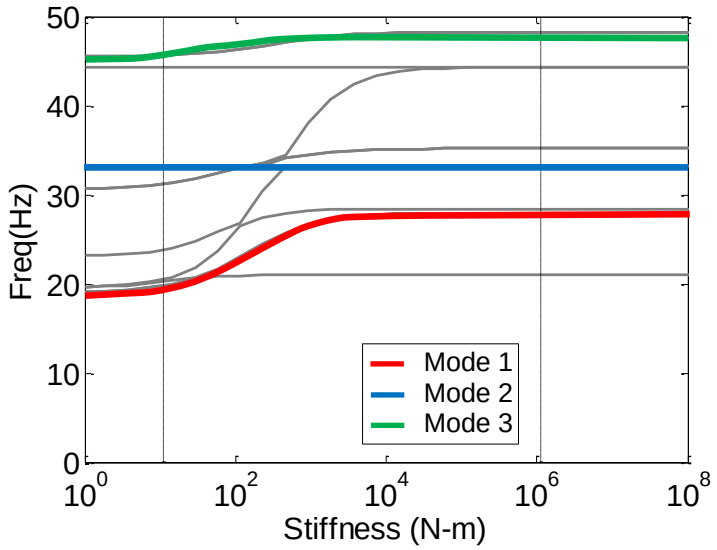

Figure 5.9: Sensitivity of natural frequencies to strong axis bending stiffness constant at top the level.

After selecting the ranges for each stiffness constant, the sensitivities of natural frequencies and lateral displacements to the spring constants are analyzed through ANOVA. As seen in Figure 5.9, the analytically computed modes can swap their order when the stiffness constant is varied. For proper implementation of ANOVA on natural frequencies, a *mode-pairing algorithm* is developed, which operates on the spatial properties of each mode to properly reorder swapped modes according to a reference mode sequence. Table 5.5 lists the relative contribution of the stiffness constants to the overall variability in the predicted features, where the normalized sensitivities are summed for each connection level. Results indicate that the most sensitive constituents to model output are the middle, base, and top connections, respectively. The torsional stiffness constants are insensitive to the predicted features for all connection levels and thus, are not included in further discussion and are held at nominal (i.e. fixed) values.

Table 5.5: Sensitivity analysis with normalized R^2 (%) values for each parameter.

			Natural Frequency			*Max Horizontal Displacement*				
		Parameter Number and Description of Stiffness	*Mode 1*	*Mode 2*	*Mode 3*	$\Delta_{Beam\ Top}$	$\Delta_{Beam\ Mid}$	$\Delta_{Coln.}$	Σ	*Total*
Base	1	Rotational (strong axis)	47.7	36.9	65.3	18.1	7.3	37.2	212.3	212.3
	2	Torsional	0.00	0.00	0.00	0.00	0.00	0.00	0.01	
Middle	3	Rotational (strong axis)	33.2	51.8	23.1	17.1	5.9	35.2	166.3	251.6
	4	Rotational (weak axis)	2.2	0.02	0.17	0.08	82.8	0.00	85.3	
	5	Torsional	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Top	6	Rotational (strong axis)	12.0	11.1	11.2	13.4	4.0	27.6	79.2	136.1
	7	Rotational (weak axis)	4.9	0.30	0.32	51.3	0.02	0.00	56.9	
	8	Torsional	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

5.5.2. Error Analysis

The parameter of interest, *P* (see Table 5.5 for parameter numbers and corresponding constituents), is allowed to vary from $1.1x10^1$ -$1.1x10^{10}$ newton meters; α is chosen to be equal to 10%, γ is set to 0.5, *n* is 4, and β is chosen to be equal to 20%. One hundred scenarios are generated through Latin-Hypercube design (LHS). Figure 5.10 plots discrepancy as a function of spring constant for Parameter 1 for Static Tests 1 and 3, where the discrepancy is computed as a percentage of the experimental value for each experiment. The plot shows the initial parameter value and algorithm chosen parameter values for each experiment respectively.

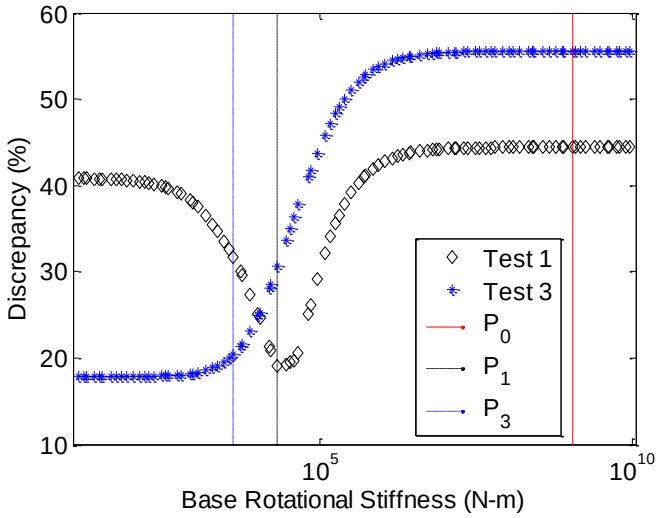

Figure 5.10: Discrepancy vs. Parameter 1 and selected parameter values for Static Tests 1 and 3.

In Figure 5.10, P_0 represents the nominal value for the knob, which in this case is the stiffness constant of the base connection; and P_1 and P_3 represent the selected values, P_j, that minimize the discrepancy of Model #1 against Static Tests 1 and 3, respectively, according to Equations (5.4)-(5.7). P_1 shows a case where an absolute minimum discrepancy exists. Here, the selected parameter value, P_1, is the parameter value at the computed minimum discrepancy. P_3 shows a case, where the discrepancy converges to a minimum. Here, the selected parameter value, P_3, is selected with a β value of 20%, i.e. at the point where the discrepancy is 20% of the total possible reduction in discrepancy.

Table 5.6 lists the EA term computed as a percentage of the nominal value, P_0, according to Equations (5.2)-(5.3). For this application the discrepancy tends to vary on a logarithmic scale for parameter values. Therefore, the percentages computed in Equation

(5.3) are modified to represent log based percentages. In Table 5.5, the EA terms are averaged for each connection. Results indicate that the largest error is associated with the top, middle, and base connections, respectively.

Table 5.6: Error analysis of the three constituent models.

Constituent	*Description of Stiffness*	*EA(%)*	*Average Constituent EA*
Base	Rotational (strong axis)	49.8	49.8
Middle	Rotational (strong axis)	51.2	59.1
	Rotational (weak axis)	66.9	
Top	Rotational (strong axis)	54.0	61.9
	Rotational (weak axis)	69.8	

5.5.3. CPI Calculation

The CPI values of the three connections models, calculated according to Equation (5.1), are listed in Table 5.7. Since the knowledge level for each connection stiffness is low (i.e. constituents are highly uncertain), the UA terms for each connection are set to one. The SA and EA terms are normalized such that the highest value in each respective category is one. This normalization bounds all three terms in the CPI calculation between zero and one.

Table 5.7: CPI for the three constituent models.

Constituent	$\|\|SA\|\|$	$\|\|EA\|\|$	$\|\|UA\|\|$	*CPI*
Base Connection	0.84	0.80	1.00	0.68
Middle Connection	1.00	0.95	1.00	0.95
Top Connection	0.54	1.00	1.00	0.54

Table 5.7 prioritizes the three constituents in the following order of importance: middle, base, and top suggesting that the available resources should be devoted to obtain an improved representation of the middle connection first and the base connection next. Accordingly, three FE models with gradually improved representation of connections are developed in three successive phases: first, the initial frame FE model (Model #1) is coupled with only the middle connection model resulting in Model #2. Second, the base connection model is added to Model #2 to obtain Model #3. Finally, Model #4 is constructed by adding the top connection model to Model #3.

5.6. Development, Verification, and Validation of Constituents

5.6.1. Constituent Model Development

The connection models are developed as three-dimensional non-linear FE models in ANSYS v. 13.0. As suggested by Kim et al. (2007), Bursi and Jaspart (1997a, 1997b, and 1998), and Selamet and Garlock (2010), both the members of the connections and the bolts are modeled using three-dimensional solid finite elements (as opposed to simplified bolt models such as the coupled bolt model, spider bolt model, and no-bolt models as discussed in Kim et al. (2007)). SOLID187, a ten-noded tetrahedral element with quadratic displacement interpolation is used to model steel members and bolts. Effects of friction, represented using contact and target elements available in ANSYS v.13.0, CONTA174 and TARGET170, are modeled at all interfaces including: beam to column, bolt to column/beam, and bolt hole to bolt shank. The coefficient of friction is assumed for a Class A surface and set to 0.35 (AISC, 2008). The mesh size for all elements is selected through a mesh refinement study to assure that the numerical discretization is

sufficiently fine to properly capture the deformation and stress concentration in critical regions. Figure 5.11 presents the FE models for all three connections.

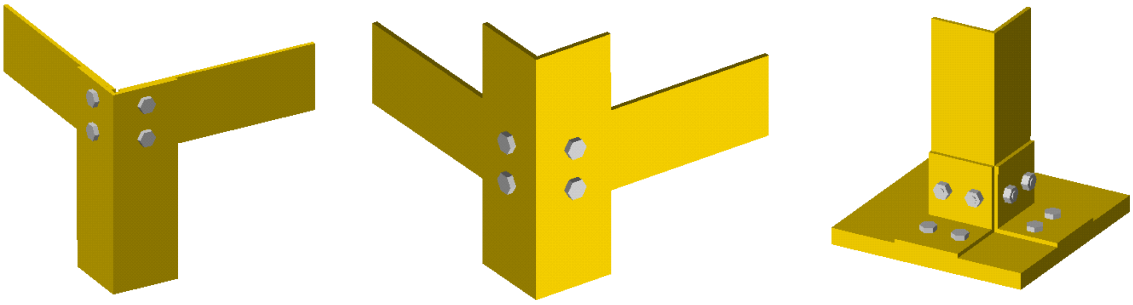

Figure 5.11: FE connection models: Top (Left), Middle (Middle), and Base (Right).

As suggested by Citipitioglu et al. (2002), Kim et al. (2007) and Pirmoz et al. (2008, 2010), pretensioning of the bolts is included in the model. However, as opposed to using temperature or initial displacement methods commonly found in the literature, pretension elements available in ANSYS v.13.0, PRETS179, that directly apply a pretensioning force to the bolt volumes are implemented. This pretensioning force is determined by converting the torque applied on the bolts using Equation (5.8) (Fastenal, 2005).

$$F_{PT} = \frac{T}{Kd} \tag{5.8}$$

where F_{PT} is the pretensioning force, T is the applied torque, K is the dimensionless nut factor, and d is the nominal bolt diameter. The recommended K for zinc-plated fasteners is 0.17-0.22 (Fastenal, 2005).

The beam and column material model is defined as a multilinear isotropic hardening model using the von Mises yield criterion and the typical stress-strain

relationship shown in Figure 5.12 (Hibbeler, 2008). The bolts are modeled similarly except a bilinear model is utilized (Fastenal, 2005) (Figure 5.12).

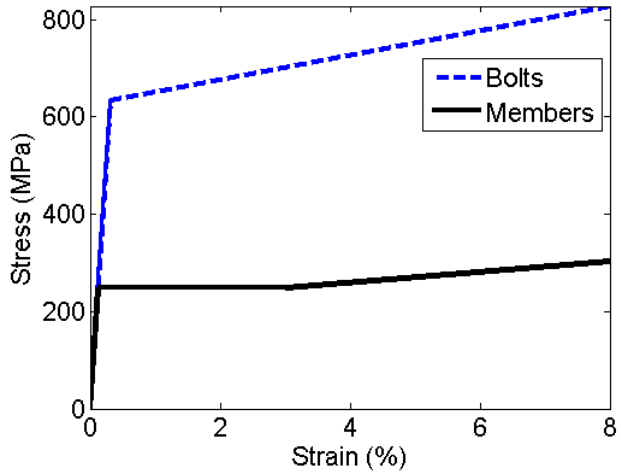

Figure 5.12: Material models for bolts and steel frame members.

5.6.2. Constituent Model Test-Analysis Correlations

To ensure that the constituent FE models exhibit sufficient fidelity, the model predictions are checked against experimental measurements. Since the nonlinear features of the contact elements and pretensioning effects are not considered in modal analysis of the FE models (ANSYS only considers initial contact state), the clamping effects of the pretensioning are modeled by gluing a conical section formed through the beam and column between the bolt heads and nuts as described in Kim et al. (2007). Modal analysis is also conducted with scaled laboratory models of the connections supported with fixed boundaries (Figure 5.13), results of which are summarized in Figures 5.14-5.15 and Table 5.8.

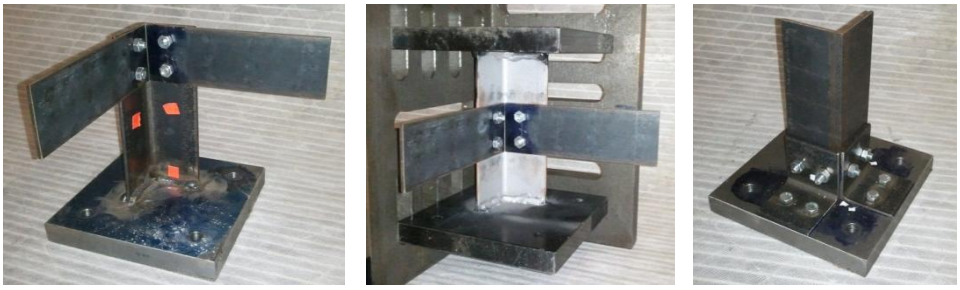

Figure 5.13: Connection test specimens: Top connection (Left), Middle connection (Middle) and Base connection (Right).

The errors between the constituent FE model predictions and the experimental measurements of the first three natural frequencies range from a 10.9% average error for the middle connection to an average error of less than 7.8% for the top connection (Table 5.8). To validate the accuracy of the mode shapes predicted by the FE model, the MAC (used in Section 5.4.3 to check the orthagonality of the experimental modes) is calculated for each connection. In this case, the experimental mode shape vectors are calculated against the FE predicted mode shape vectors. The results shown in Tables 5.9-5.11demonstrate that with an average diagonal term value of 0.67, 0.80, and 0.88 for the base, middle, and top connection modes respectively, the FE model predicted mode shape vectors are highly correlated to the experimental mode shape vectors and are an accurate representation of the experimentally collected modes. The accuracy shown in the dynamic testing provides sufficient evidence for the proper modeling of mass and stiffness distribution of the connections. Therefore, with an average error of less than 9% for the natural frequencies and an average diagonal MAC value of 0.77 across all modes of all models, the connection models are determined to be useful for making improved estimates of connection stiffness.

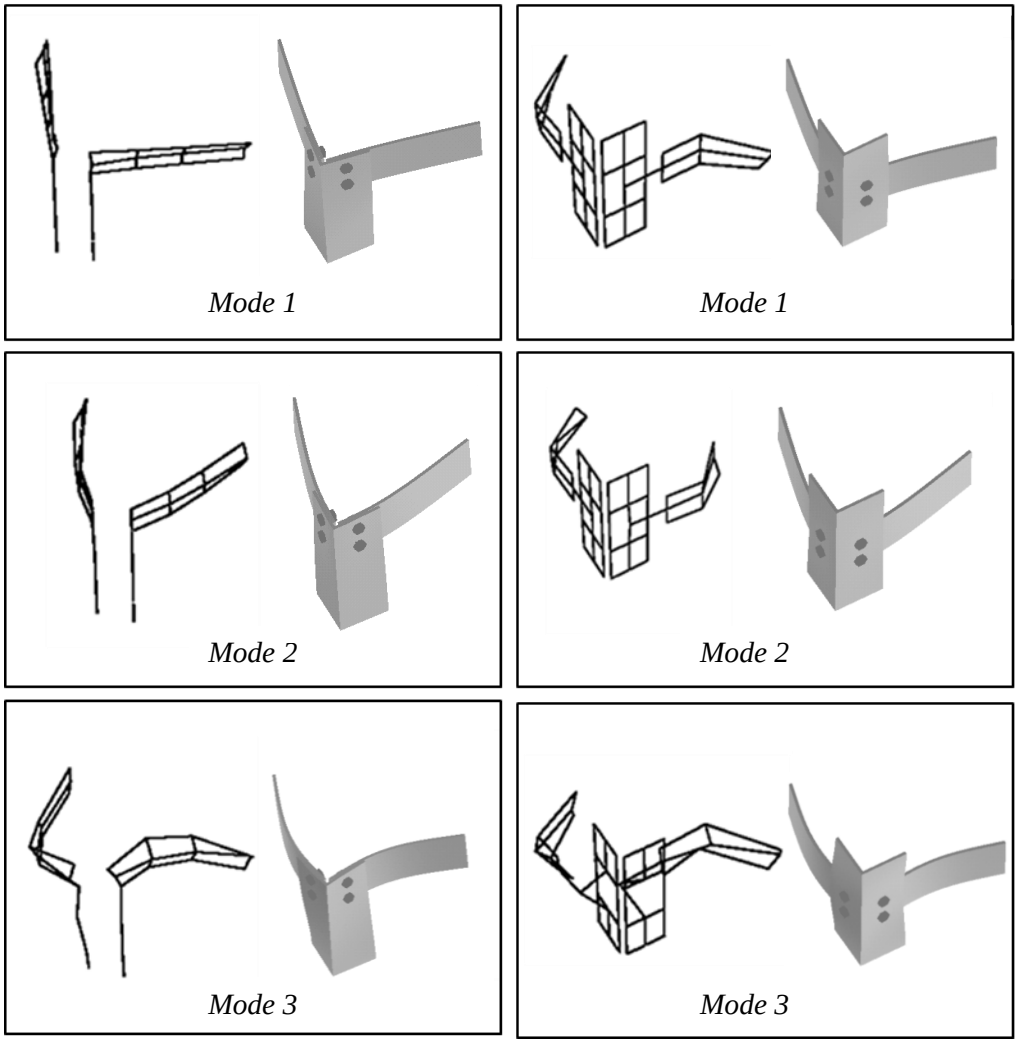

Figure 5.14: Mode Shapes (Experimental on left, FE on right): Top connection (a), Middle connection (b).

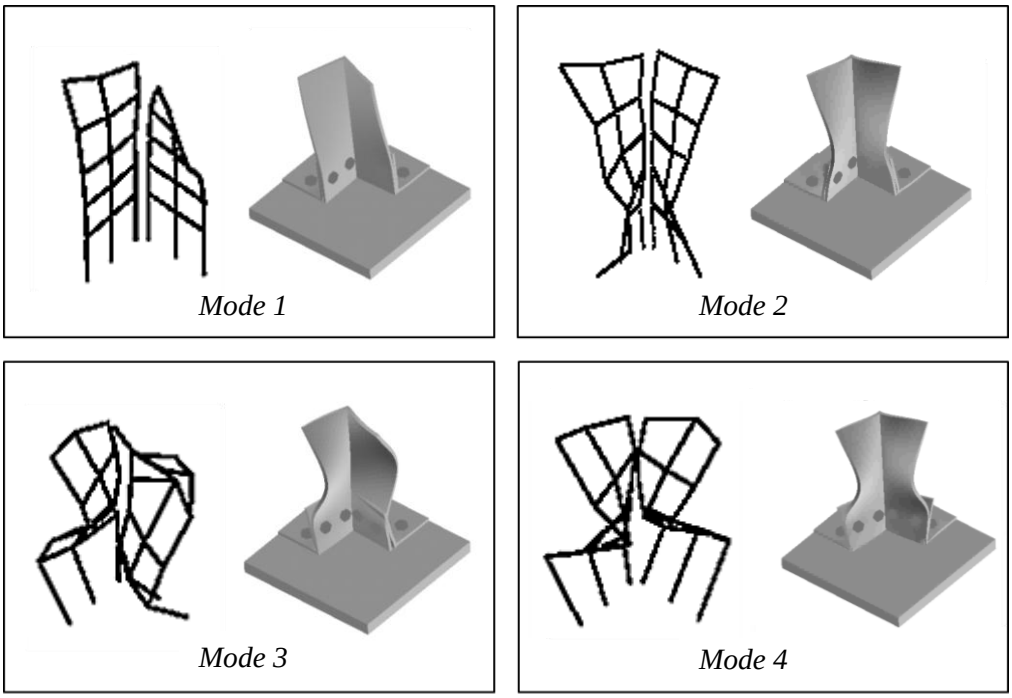

Figure 5.15: Mode Shapes (Experimental on left, FE on right): Base connection.

Table 5.8: Connection test-analysis correlation for connection models.

	Mode	*$Freq_{exp}$ (Hz)*	*$Freq_{sim}$ (Hz)*	*% Error*
Base	1	473.4	425.6	10.1
	2	2043.6	2003.0	2.0
	3	2636.1	2295.4	12.9
	4	3005.4	2744.7	8.7
Middle	1	142.9	126.6	11.4
	2	157.0	135.1	13.9
	3	685.6	635.6	7.3
Top	1	79.7	75.1	5.8
	2	143.3	124.68	13.0
	3	459.0	438.4	4.5

Table 5.9: MAC analysis for Base connection.

		FE Model			
	Mode	*1*	*2*	*3*	*4*
Experiment	*1*	0.576	0.000	0.008	0.000
	2	0.008	0.783	0.010	0.019
	3	0.017	0.037	0.576	0.000
	4	0.000	0.054	0.000	0.764

Table 5.10: MAC analysis for Middle connection.

		FE Model		
	Mode	*1*	*2*	*3*
Experiment	*1*	0.870	0.004	0.030
	2	0.000	0.662	0.011
	3	0.005	0.000	0.880

Table 5.11: MAC analysis for Top connection.

		FE Model		
	Mode	*1*	*2*	*3*
Experiment	*1*	0.985	0.000	0.007
	2	0.045	0.908	0.002
	3	0.029	0.004	0.750

5.7. Coupling of Constituent Models

The coupling of the constituent connection models with the frame model is completed in two phases: the first step is an initial investment in computing the moment-rotation curves of the connections, and second step is coupling the trained moment-rotation curves with the frame model. The goal of the first phase is to determine the appropriate stiffness of the connections over a range of possible loading conditions; thus, in this phase, the deformations of columns in the frame must be taken into account.

Determining the connection stiffness solely from the connection FE model with assumed boundary conditions would be invalid (note that this approach in literature, would be considered as weak (or loose) coupling). Therefore, a strong (or tight) coupling approach is taken and implemented using a Block Gauss-Siedel algorithm (Figure 5.2).

As shown in Figure 5.16, the moment-rotation curves are developed by first inducing a global loading condition on the frame FE model at the midpoint of the beam for top and middle constituents and 1/8 point from the bottom of the column for the base constituent. The force, F_g, is applied in the direction that critically affects the stiffness of each specific connection. The rotational stiffness constants are held at nominal (i.e. fixed) values, P_0. Next, displacements and rotations at the points where the connection models are cut-off, BC_g, and the moment at the connection, M_g, are calculated by the frame FE model and passed to the coupling interface. The displacements and rotations at the cut-off points of the beams and columns are entered as inputs for the connection FE model, which in turn is executed to calculate the corresponding rotation at the connection, θ_g. Similar to approaches by Bursi and Jaspart (1997a; 1997b) and Pirmoz et al. (2008), the rotation of the connection is computed by taking the relative difference between the rotation of the column and the rotation of the beam as expressed in Equation (5.9).

$$\theta_g = \left|\theta_{beam} - \theta_{column}\right| \tag{5.9}$$

The calculated rotation is then passed to the interface. The stiffness constant, P_g^h, at iteration *h* for load F_g is computed from the ratio of the computed moment to rotation, and then compared to the stiffness constant computed in the previous iteration, *h-1*. The

algorithm is terminated if the change in the computed stiffness constant between the last two iterations is less than a set threshold value (set to be 0.005 herein). Otherwise, the new stiffness constant is passed back to the frame model and the process is iterated until convergence is achieved.

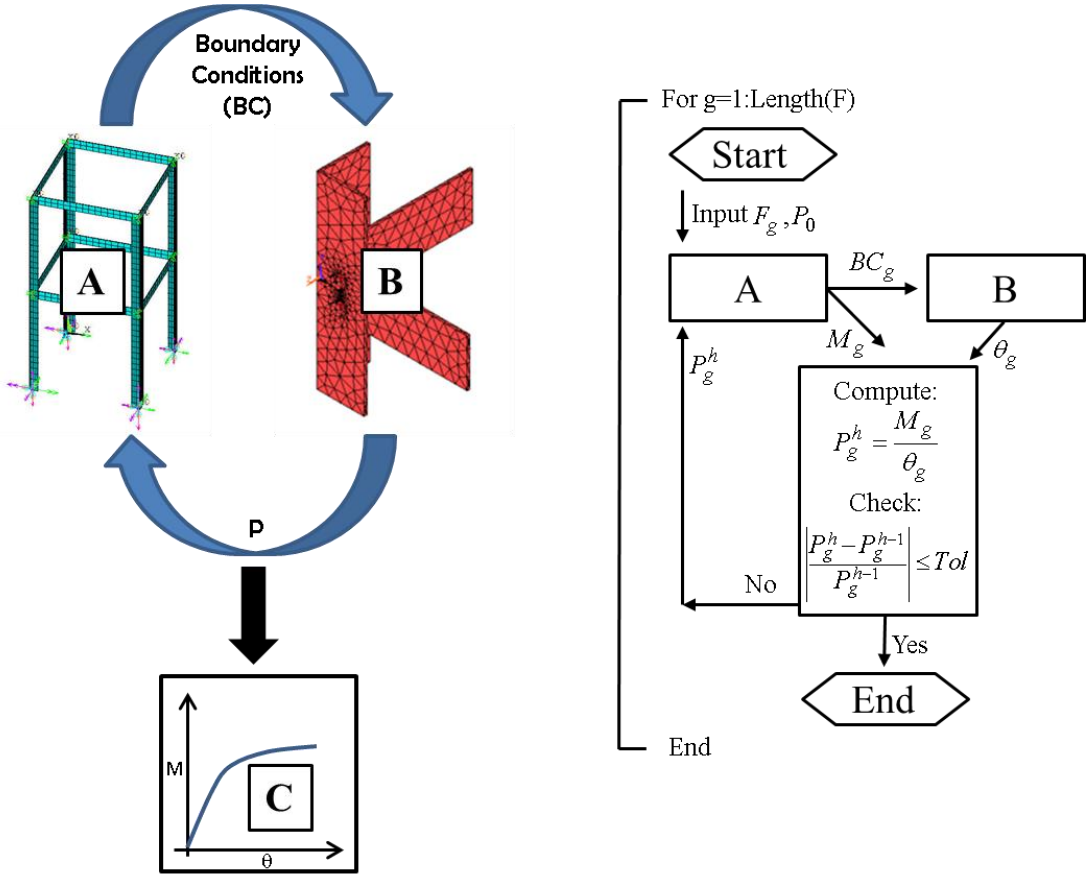

Figure 5.16: Coupled process: (a) schematic representation, and (b) numerical algorithm.

The above-described process is repeated for increasing loads, F_g, to develop the moment-rotation curves, which are shown in Figure 5.17 for all rotational springs of interest. Note that for rotations less than 0.00017 radians, the moment/rotation data points for both the Top and Middle connection vertical bending (Top-R_{xz} and Middle-R_{xz}) are

inconsistent with the linear portion of the respective moment-rotation curves potentially due to the numerical round-off errors. Thus, these data points are removed from the curves for the analysis that follows.

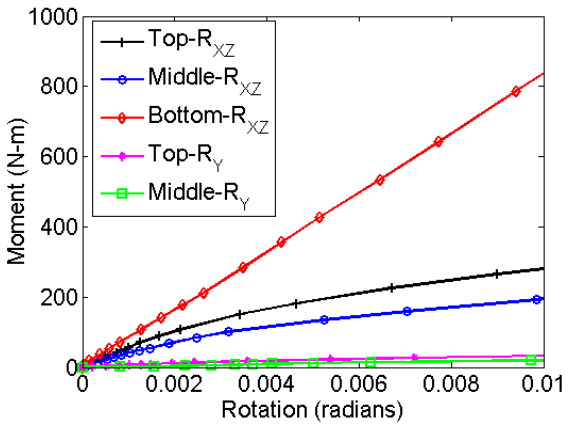

Figure 5.17: Moment-rotation curves.

In the second phase of the coupling process, the moment-rotation curves are directly incorporated into the frame FE model as nonlinear moment-rotation relationships using the nonlinear spring elements COMBIN39 available in ANSYS v.13.0.

The coupled solution is obtained for three variants of the frame FE model with increasing sophistication (Model #2-Model #4). The predictions of these models along with the initial FE model (Model #1) are compared against the static and modal experimental data as discussed in the next section.

5.8. Results and Validation

All four variants of FE models are compared directly to modal and static experimental data. The disagreement between model predictions and experiments (referred to herein as discrepancy) is determined for each test as a percentage of the

experimental measurements (Tables 5.12 and 5.13). Figure 5.18 compares the force-displacement plots of Static Test 1 for Model #4 and Model #1 to experimental data.

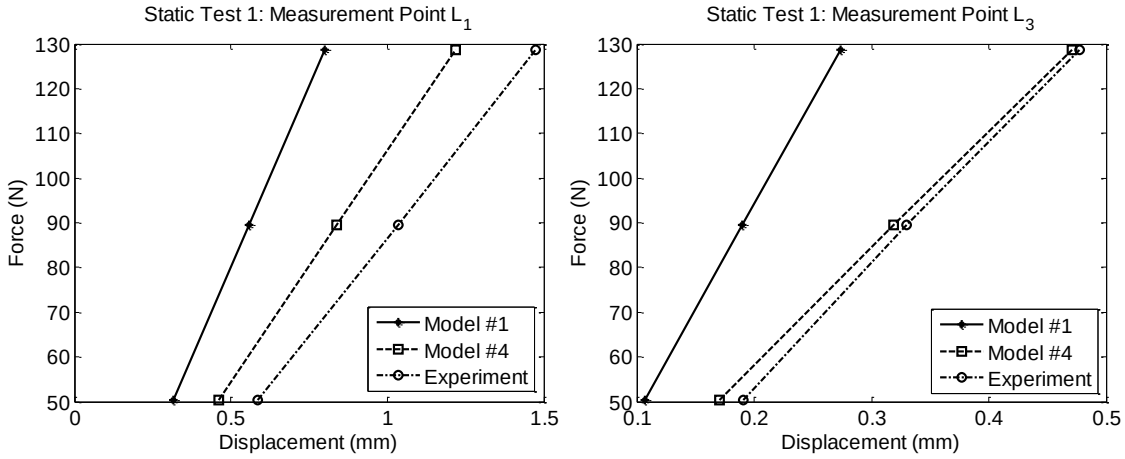

Figure 5.18: Static test comparison: Model #1 vs. Model #4.

The computed discrepancy for each of the four models is compared for modal and static tests as shown in Tables 5.12 and 5.13. As the physics sophistication of the model increases from Model #1 to Model #4, the discrepancy between model predictions and experimental results decrease for nearly all predicted features, with the exception of Mode 2, which is underestimated by Model #4 approximately the same amount it is overestimated by Model #1.

Table 5.12: Comparison of natural frequencies (Model #1-Model #4).

Mode	*Exp. (Hz)*	*Model #1 (Hz)*	*Disc. (%)*	*Model #2 (Hz)*	*Disc. (%)*	*Model #3 (Hz)*	*Disc. (%)*	*Model #4 (Hz)*	*Disc. (%)*
1	23.09	27.97	21.13	26.50	14.78	26.26	13.73	25.30	9.56
2	29.58	33.21	12.27	28.55	3.48	28.38	4.06	27.11	8.35
3	37.04	47.56	28.40	42.70	15.29	41.65	12.45	38.47	3.86

Likewise, for the static case, the discrepancy between model predictions and experiments in general decreases as the model sophistication is increased. For instance, the discrepancy is significantly reduced by approximately 75% as the model sophistication is increased from Model #1 to Model #4 for Test 1. On average over all six static tests, Model #4 yields an agreement with a 29.59% discrepancy.

Table 5.13: Comparison of static displacements (Model #1-Model #4).

Test	*Model #1: Mean Disc.(%)*	*Model #2: Mean Disc.(%)*	*Model #3: Mean Disc.(%)*	*Model #4: Mean Disc.(%)*
1	44.42	33.12	18.91	11.08
2	84.83	89.00	79.68	73.73
3	55.60	52.57	48.74	46.29
4	46.75	37.35	25.37	18.05
5	24.26	19.39	17.42	16.43
6	18.93	13.98	12.75	11.96

Recall that in Section 5.3, the three connections models are ranked using the CPI metric to determine the order to prioritize development of the connection FE models. In Tables 5.12 and 5.13, the gain in predictive accuracy on average is 6.41% from Model #1 to Model #2, 5.11% from Model #2 to Model #3, and 3.74% from Model #3 to Model #4 (Figure 5.19). These results indicating that the most information is gained by adding the middle connection model, then the bottom connection, and finally the top connection, are consistent with the code prioritization findings from CPI in Section 5.5.3.

To further validate the effectiveness of the CPI, all possible sequences of model development are investigated. Table 5.14 shows the average reduction in discrepancy for each incremental model development over all possible cases for the order, in which

constituents (B–bottom connection, M–middle connection, and T–top connection) are improved through the coupling of a nonlinear FE connection model. The CPI selected path (highlighted in Table 5.14) is the only path that yields a monotonic increase in model accuracy; and thus, this path most efficiently improves the FE model. Therefore, in this application, the model accuracy is shown to be improved as physics sophistication is increased, and a prioritization effort (CPI) is demonstrated to be effective in prioritizing constituents of a coupled model.

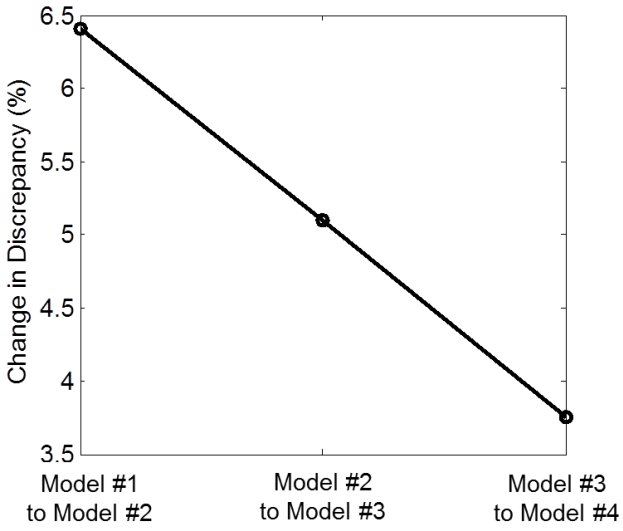

Figure 5.19: Relative change in discrepancy with respect to physics sophistication.

Table 5.14: Comparison of discrepancy reduction for varying model increment cases.

	Average Reduction in Mean Discrepancy (%)		
Case Constituent Order	*Model #1 to Model #2*	*Model #2 to Model #3*	*Model #3 to Model #4*
M-B-T	6.41	5.11	3.74
M-T-B	6.41	3.08	5.77
B-M-T	3.47	8.04	3.74
B-T-M	3.47	3.27	8.51
T-M-B	3.05	6.44	5.77
T-B-M	3.05	3.70	8.51

5.9. Concluding Remarks

Complex physical phenomena are amenable to decomposition in various forms depending upon the specific analysis or design objectives. Such decomposition into constituents can be steered by considerations regarding the physical, functional, or computational nature of the problem. While the inherent inaccuracy of the numerical models used for simulations of complex coupled systems is acknowledged, the biases and uncertainties in model predictions can be reduced through further improvement of the simulation models, and rigorous verification, validation and uncertainty quantification activities. There exists a need to prioritize these improvements such that resources are allocated to the model constituents that most effectively reduce model biases and uncertainties.

In this chapter, an index for code prioritization efforts of such complex numerical models, Code Prioritization Index (CPI) is developed for coupled numerical systems based upon fundamentals guiding the well-estabilished PIRT ranking procedure. The CPI goes beyond the qualitative ranking system instituted in PIRT and supplies a quantitative assessment of the knowledge (uncertainty) and importance level (sensitivity) of a

constituent phenomenon. Moreover, CPI incorporates an error analysis term (initial error). Through the combination and normalization of the three ranking categories (uncertainty, sensitivity, and initial error), the CPI represents a quantitative and defendable metric for ranking code prioritization efforts such that the largest reduction in model error can be achieved using minimal resources. The fundamental concept presented herein and its associated metrics requires an ability to parameterize the constituent models. Therefore, the study presented herein is most amenable for partitioned analysis procedures.

The CPI metric is demonstrated on a two-story steel frame with semi-rigid bolted connections. Predictions from an initial FE model (Model #1) of the steel frame indicate a high disagreement with experimental results. Deficiencies in the model are identified in the fixed beam-to-column connection assumptions. The stiffness of the three possible connection configurations (top, middle, and bottom) are treated as model constituents, and through the use of linear rotational springs implemented in the numerical model, these constituents are prioritized using the CPI metric. Next, high-fidelity, three-dimensional, nonlinear FE models of the connections are built for coupling with Model #1 in the order of their ranking as identified by CPI. Ultimately, through this coupling, Model #2 through Model #4 are obtained.

A comparison of modal and static test data to model predictions shows that as the physics of the model improve (from Model #1 to Model #4) the discrepancy reduces for nearly all test cases. Furthermore, the relative improvement from one sophistication to the next decreases as the model progresses from Model #1 to Model #4 indicting that the

prioritization effort properly identified the order in which the constituent models should be improved.

While this chapter shows the importance and usefulness of the proposed CPI metric, further work in the implementation and quantification of the uncertainty term is being considered. Additionally, the CPI metric proposed in this present chapter is just one of many metrics that can be developed to incorporate all three terms, and investigations into the format, possible interactions among terms, and additional case studies is necessary. Also, as each component in CPI represents a specific facet of the model that may be more or less important to a specific project or research endeavor, user-defined weighting terms can be considered for each of the three components respectively. However, future research is necessary to develop guidelines regarding the possible values (or ranges of values) for such weighting factors.

Additionally, while the connections used in the case study herein are not typical of traditional, large steel structures; they serve to demonstrate not only the validity of the CPI metric but also the potential for using the coupling methodology presented herein on full-scale semi-rigid connections. Although the experimental campaign in the proof of concept application in this study is demonstrated in the linear/elastic regime, the proposed metric makes no assumptions that limit its application to such a regime. The proposed metric is versatile and generally applicable for any stable, non-chaotic responses (linear, nonlinear, elastic, inelastic, etc.). For newly developed semi-rigid connection types, the moment-rotation curves predicted through the coupled finite

element model could prove to be vital in providing a better understanding of semi-rigid connections and a reducing the need for extensive laboratory testing.

5.10. References

Advanced Reactor Research Infrastructure Assessment (ARRIA). (2003), Rev. 3, pp. 122-123.

AISC (American Institute of Steel Construction). (2008), Steel Construction Manual, 13th Edition, Chicago, 16.1-349.

Allemang, R.J. (2003), The modal assurance criterion (MAC): Twenty years of use and abuse, *S V Sound and Vibration*, **37**(8), 14-23.

ANSYS User's Manual, Release 13.0. (2011), ANSYS, Inc.

Atamturktur, S., Hemez, F., Williams, B., Tomé, C., & Unal, C. (2011), A forecasting metric for predictive modeling, *Computers and Structures*, 89(23-24), 2377-87.

Avramova, M.N. & Ivanov, K.N. (2010), Verification, validation, and uncertainty quantification in multi-physics modeling for nuclear reactor design and safety analysis, *Progress in Nuclear Energy*, 52, 601-14.

Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C.H., & Tu, J. (2007), A framework for validation of computer models, *Technometrics*, 49(2), 138-54.

Boyack, B.E. (2009), PIRT Workshop, Los Alamos National Laboratory presentation.

Boyack, B.E., Duffey, R.B., & Griffith, P., et al. (1989), Quantifying reactor safety margins: application of code scaling, applicability, and uncertainty methodology to a large-break, loss-of-coolant accident, US NRC Reprt, NUREG/CR-5249.

Bursi, O.S. & Jaspart J.P. (1997a), Benchmarks for finite element modeling of bolted steel connection, *Construct. Steel Res.*, 43(1-3), 17-42.

Bursi, O.S. & Jaspart J.P. (1997b), Calibration of a finite element model for isolated bolted end-plate steel connections, *Construct. Steel Res.*, 44(3), 225-62.

Bursi, O.S. & Jaspart J.P. (1998), Basic issues in the finite element simulation of extended end plate connections, *Computers and Structures*, 69, 361-82.

Caracoglia, L., Noè, S., & Sepe, V. (2009), Nonlinear computer model for the simulation of lock-in vibration on long-span bridges, *Computer-Aided Civil and Infrastructure Engineering*, 24, 130-44.

Citipitioglu, A.M., Haj-Ali, R.M., & White D.W. (2002), Refined 3D finite element modeling of partially-restrained connections including slip, *Journal of Constructional Steel Research*, 58, 995-1013.

Cukier, R.L., Levine, H.B., & Schuler, K.E., (1978), Nonlinear sensitivity analysis of multiparameter model systems, *Journal of Computational Physics*, **26**, 1-42.

da Silva, J.G.S., de Lima, L.R.O., da Vellasco, P.C.G., de Andrade, S.A.L., & de Castro, R.A. (2008), Nonlinear dynamic analysis of steel portal frames with semi-rigid connections, *Engineering Structures*, 30, 2566-79.

Diamond, D.J. (2006), Experience using Phenomena Identification and Ranking Technique (PIRT) for nuclear analysis, BNL-76750-2006-CP.

European Committee for Standardisation (1993), Eurocode 3 ENV-1993-1-1. Revised Annex J. Design of Steel Structures. Doc. CEN/TC250/SC3-N419E.

Farajpour, I. & Atamturktur, S., (in press 2012), Optimization-based strong coupling of multiple single-solver models, *ASCE Computing in Civil Engineering*.

Fastenal (2005), Technical Reference Guide, s7028, rev. 9. http://www.fastenal.com/content/documents/FastenalTechnicalReferenceGuide.pdf <. Accessed on May 14, 2012>.

Felippa, C.A., Park, K.C., & Farhat, C. (2001), Partitioned analysis of coupled mechanical systems, *Computer Methods in Applied Mechanics and Engineering*, 190, 3247-70.

Fernandez, M.A. & Moubachir, M. (2005), A Newton method using exact Jacobians for solving fluid–structure coupling, *Computers and Structures*, 83, 127-142.

Frey, H.C. & Patil, S.R. (2002), Identification and review of sensitivity analysis methods, *Risk Analysis*, 22(3), 553-78.

Galvão, A.S., Silva, A.R.D., Silveira, R.A.M., & Gonçalves, P.B. (2010), Nonlinear dynamic behavior and instability of slender frames with semi-rigid connections, *International Journal of Mechanical Sciences*, 52, 1547-62.

Heil, M. (2004), An efficient solver for the fully coupled solution of large-displacement fluid–structure interaction problems, *Computer Methods in Applied Mechanics and Engineering*, 193, 1– 23.

Hemez, F.M., Atamturktur, S.H., & Unal, C. (2010), Defining predictive maturity for validated numerical simulations, *Computers and Structures*, 88, 497-505.

Higdon D., Gattiker J., Williams B., & Rightley M. (2008), Computer model calibration using high-dimensional output, *Journal of the American Statistical Association*, 103(482), 570-83.

Hibbeler, R. C. (2008), Mechanics of Materials, 7th edition, Pearson Prentice Hall, Upper Saddle River, New Jersey, pp. 90.

Ibrahimbegović, A., Knopf-Lenoir, C., Kučerová, A., & Villon, P. (2004), Optimal design and optimal control of elastic structures undergoing finite rotations and deformations, *International Journal for Numerical Methods in Engineering*, 61, 2428- 60.

Jiang, X. & Mahadevan, S. (2007), Bayesian risk-based decision method for model validation under uncertainty, *Reliability Engineering and System Safety*, 92, 707-18.

Joosten, M.M., Dettmer, W.G., & Peric, D. (2009), Analysis of the block gauss–seidel solution procedure for a strongly coupled model problem with reference to fluid–structure interaction, *International Journal for Numerical Methods in Engineering*, 78, 757- 78.

Jung, B. (2011), A hierarchical framework for statistical model validation of engineered systems, Dissertation: University of Maryland.

Kassiotis, C., Ibrahimbegović, A., Niekamp, R., & Matthies, H. (2011), Partitioned solution to nonlinear fluid-structure interaction problems. Part I: Implicit coupling algorithms and stability proof, *Computational Mechanics*, 47, 305- 23.

Kim, J., Yoon, J.-C., & Kang B.-S. (2007), Finite element analysis and modeling of structure with bolted joints, *Applied Mathematical Modelling*, 31, 895-911.

Korel, B. (2009), Experimental comparison of code-based and model-based test prioritization, IEEE International Conference on Software Testing Verification and Validation Workshops, Denver, CO, United States; April 1-4, 77-84.

Kutay, M.E. & Aydilek, A.H. (2009), Pore pressure and viscous shear stress distribution due to water flow within asphalt pore structure, *Computer-Aided Civil and Infrastructure Engineering*, 24, 212-24.

Larson, J.W. (2009), Ten organizing principles for coupling in multiphysics and multiscale models. *ANZIAM Journal*, 48, C1090- 111.

Larson, J., Jacob, R. & Ong, E. (2005), The model coupling toolkit: a new fortran90 toolkit for building multi-physics parallel coupled models, *International Journal of High Performance Computer Applications*, 19(3), 277- 92.

Lee, S.S. & Moon, T.-S. (2002), Moment-rotation model of semi-rigid connections with angles, *Engineering Structures*, 24, 227-37.

Leiva, J.S., Blanco, P.J., & Buscaglia, G.C. (2010), Iterative strong coupling of dimensionally heterogeneous models, *International Journal for Numerical Methods in Engineering*, 81, 1558-80.

Liu, Y., Xu, L., & Grierson, D. E. (2010). Influence of semi-rigid connections and local joint damage on progressive collapse of steel frameworks, *Computer-Aided Civil and Infrastructure Engineering*, 25, 184-204.

Macdonald, J.H.G. (2009), Lateral excitation of bridges by balancing pedestrians, *Proceedings of the Royal Society A - Mathematical Physical and Engineering Sciences*, 465, 1055-1073.

Matthies, H.G., Niekamp, R., & Steindorf, J. (2006), Algorithms for strong coupling procedures, *Computer Methods in Applied Mechanics and Engineering*, 195, 2028–49.

Matthies, H.G. & Steindorf, J. (2002), Fully coupled fluid–structure interaction using weak coupling, *Proceedings in Applied Mathematics and Mechanics*, 1(1), 37-8.

Matthies, H.G. & Steindorf, J. (2002), Partitioned but strongly coupled iteration schemes for nonlinear fluid–structure interaction, *Computers and Structures*, 80, 1991–9.

Matthies, H.G. & Steindorf, J. (2003), Partitioned strong coupling algorithms for fluid–structure interaction, *Computers and Structures*, 81, 805-12.

Moaveni, B., Conte, J.P., and Hemez, F.M. (2009), Uncertainty and sensitivity analysis of damage identification results obtained using finite element model updating, *Computer-Aided Civil and Infrastructure Engineering*, 24(5), 320-34.

Oberkampf, W.L. & Barone, M.F. (2006), Measures of agreement between computation and experiment: Validation metrics, *Journal of Computational Physics*, 217, 5-36.

Olivier, T.J. & Nowlen, S.P. (2008), A Phenomena Identification and Ranking Table (PIRT) exercise for nuclear power plant fire modeling applications, NUREG/CR-6978 SAND2008-3997P.

Park, K. & Felippa, C.A. (1983), Partitioned analysis of coupled systems, *Computational Methods for Transient Analysis*, 1, 157-219.

Pirmoz, A., Ahmadi, M.M., Valadi, E., Farajkhah, V., & Balanoji, S.R. (2010), Performance of the PR connections under combined Axial-tension and moment loading, *Electronic Journal of Structural Engineering*, 10, 66-73.

Pirmoz, A., Daryan, A.S., Mazaheri, A., & Darbandi, h. E. (2008), Behavior of bolted angle connections subjected to combined shear force and moment, *Journal of Constructional Steel Research*, 64, 436-46.

Provenzano, P. (2003), A fuzzy-neural network method for modeling uncertainties in soil-structure interaction problems, *Computer-Aided Civil and Infrastructure Engineering*, 18, 391-411.

Qian, L. & Zhang, H. (1993), An improved non-reflecting boundary method for soil-structure interaction analysis, *Computer-Aided Civil and Infrastructure Engineering*, 8(4), 291-7.

Ramanath, G. (2011), Collaborative Research: Understanding mechanical and thermal properties and their coupling at nanomolecularly modified metal-ceramic, NSF Award Abstract #1100933.

Rebba, R., Mahadevan, S., & Huang, S. (2006), Validation and error estimation of computational models, *Reliability Engineering & System Safety*, 91, 1390-7.

Richardson M.H. & Formenti, D.L., (1982), "Parameter Estimation From Frequency Response Measurements Using Rational Fraction Polynomials," Proceedings of the 1st International Modal Analysis Conference, Orlando, FL.

Rugonyi, S. & Bathe, K.J. (2001), On finite element analysis of fluid flows fully coupled with structural interactions, *Computer Modeling in Engineering and Sciences*, 2, 195-212.

Selamet, S. & Garlock, M. (2010), Guidelines for modeling three dimensional structural connection models using finite element methods, *International Symposium: Steel Structures: Culture & Sustainability* 2010, Istanbul, Turkey, Paper No. 14.

Tregoning, R.T., Apps, J.A., Chen, W., Delegard, C.H., Litman, R., & MacDonald, D.D. (2009), Phenomena Identification and Ranking Table evaluation of chemical effects associated with generic safety issue 191, NUREG-1918.

Türker, T., Kartal, M.E., Bayraktar, A., & Muvafik, M. (2009), Assessment of semi-rigid connections in steel structures by model testing, *Journal of Constructional Steel Research*, 65, 1538-47.

Unal, C., Williams, B., Hemez, F., Atamturktur, S.H., & McClure, P. (2011), Improved best estimate plus uncertainty methodology, including advanced validation concepts, to license evolving nuclear reactors. *Nuclear Engineering and Design*, 241, 1813–33.

Wang, D., Gao, S.Q., Kasperski, M., Liu, H.P., & Jin, L. (2011), Simulation of the dynamic characteristics of the coupled system structure, *Applied Mechanics and Materials*, 71-78, 1507-1510.

Wang, J.G., Nogami, T., Dasari, G.R., & Lin, P.Z. (2004), A weak coupling algorithm for seabed–wave interaction analysis, *Computer Methods in Applied Mechanics and Engineering*, 193, 3935-56.

Williams, B.J., Loeppky, J.K., Moore, L.M., & Macklem, M.S. (2011), Batch sequential design to achieve predictive maturity with calibrated computer models. *Reliability Engineering and System Safety*, 96, 1208-19.

Zapico, J.L., Gonzalez-Buegla, A., Gonzalez, M.P., & Alonso, R. (2008), Finite element model updating of a small steel frame using neural networks, *Smart Materials and Structures*, 17(045016), 1-11.

Zeyao, M. & Lianxiang, F. (2004), Parallel Flux Sweep Algorithm for Neutron Transport on Unstructured Grid, *The Journal of Supercomputing*, 30, 5-17.

Zhang, Q. & Hisada, T. (2004), Studies of the strong coupling and weak coupling methods in FSI analysis, *International Journal for Numerical Methods in Engineering*, 60, 2013- 29.

CHAPTER 6: CONCLUSIONS

6.1. Summary of Research Program:

Advancements in science and technology in the past decades have led scientists and engineers to develop and use numerical models to predict for increasingly complex physical systems. However, for these models to be useful in design and analysis, the uncertainties in their predictions must be identified, quantified, and reduced as much as possible. This process is known as *experiment-based model validation*. During the model experiment-based validation process, one must assess not only if the predictions are within an acceptable accuracy and have a sufficient precision, but also if the experiments being used to validate the model are sufficient both in terms of quality and quantity. An appropriately developed model validation framework can address these needs through a quantitative and systematic approach. Many model validation frameworks are available in literature, but they seldom provide decision makers with the tools to not only determine the suitability of a model for an intended use, but to also guide decision makers in allocating resources in three distinct areas: (i) if the model is not considered sufficiently accurate or precise for the given application, should additional experiments be conducted or should the model itself be improved though implementation of additional mathematical and/or physical principals of the model; (ii) if more experiments are to be conducted, at what settings in the domain of applicability should experiments be conducted to efficiently improve the predictive capability of the model; and (iii) if model physics sophistication is necessary, which model constituent(s) should be improved such that the greatest increase in model accuracy is realized at a least possible cost. The

framework developed in this dissertation provides methodologies for each of the above questions as an integral part of an experiment-based model validation framework.

For a given numerical model and a current set of experimental data, the framework developed herein (see Figure 1.1) utilizes a Bayesian inference based back calculation technique that exploits a Markov chain Monte-Carlo (MCMC) random walk to infer probability distributions of unknown model parameters (model uncertainty) and an independent estimate of the residual difference between model predictions and experimental data, i.e., the discrepancy bias. Once the model uncertainty and accuracy are quantified, a validation metric is employed to assess whether or not the model is suitable for use in its intended domain of applicability. If the validation metric is satisfied, then the model validation process ends; however, if the validation metric is not satisfied the framework implements a path selection algorithm. This path selection algorithm analyzes the convergence rate of the discrepancy bias along two axes: the number of experiments and the number of model physics sophistications. The path selection algorithm chooses the path, which yields the highest convergence rate indicating that either additional experiments should be conducted or the physics sophistication of the model should be improved. Of course, this algorithm can be adjusted to account for more or less weight being put on either axis, as is demonstrated for the case where experimentation is considered less costly than code development and hence the algorithm continues to select the experimentation path until no further gains in either precision or accuracy are realized.

After the choice in path is made, either additional experiments are to be collect or the model is to be improved. However, instead of arbitrarily selecting experimental settings or code constituents, the framework herein provides methodologies for selecting either the optimal experimental settings through the use of Batch Sequential Design (BSD) or the most effective model constituents for code development of coupled numerical models through the use of the Code Prioritization Index (CPI).

The BSD methodology is proven in literature to be an effective tool for selecting batches (as small as one or as large as needed) of physical experimental settings such that a provided optimization criterion is met. Recall, that at this stage in the validation framework, it is necessary to improve the empirical estimation of the model's discrepancy bias and reduce model uncertainty throughout the domain of applicability; therefore, in this study, a new optimization criterion, C-EIPS, is developed such that the goal of the BSD is to select experiments that will not only reduce the uncertainty in the discrepancy bias but also adequately cover the domain of applicability. Experiments are then conducted at the BSD selected experimental settings and the validation process repeats.

If, however, the physics sophistication path is selected, this framework implements the CPI metric for ranking the model constituents of a coupled numerical model according to the potential reduction in the model form error throughout the domain of applicability. This metric is based upon three principal quantities: sensitivity of model predictions to variability in the constituent, the estimated initial error of the constituent, and the uncertainty of the constituent. Each component listed above can be numerically

calculated using metrics available in literature or those provided in this dissertation. The computed CPI values for each constituent are ranked such that model constituents receiving a higher CPI are considered higher priority for improvement. After the selected constituents are improved and a new. more sophisticated numerical model becomes available, the main loop in validation framework repeats.

The framework described above is demonstrated within this dissertation using a variety of non-trivial, real-life case studies. These studies include the stress, strain, and texture response of metals (that are used as cladding materials in the nuclear reactors) exposed to high temperatures and strain rates as well as static deflection and vibration responses of a steel frame. The types of models utilized in these studies include continuum-based phenomenological models and mesoscopic models to consider microstructural behaviors. While the proposed validation framework is demonstrated on a few specific case studies in this dissertation, its generality allows its application across all science and engineering fields.

6.2. Major Findings of the Presented Research:

The previously summarized research campaign has resulted in the following findings:

Findings from the development and application of the general framework, its algorithm and associated metrics (Chapters 2 and 3):

- As hypothesized, the case studies presented in this dissertation show that more sophisticated models may have higher uncertainty and require additional experiments to realize convergence of the systematic bias and entropy-based

information gain metric. This finding is realized during the model calibration phase of the framework presented herein, and can be utilized to help determine the sufficient number of experiments.

- In the case study applications, it is shown that as additional experiments are introduced into the calibration process, both the information gain and systematic bias of all response features of interest of the more sophisticated model converge to a consistent systematic level while the convergence trend is shown to not occur for some of the response features of the less sophisticated model.
- With both case study applications and conceptual examples herein, the systematic bias is shown to reduce as the physics sophistication of the model is improved. However, it is important to note that this may only become evident once sufficient experimental data becomes available such that the discrepancy bias has converged.
- The discrepancy based convergence algorithm (computed using the C_M value) calculated along the prior experiments or prior code development paths is demonstrated to be an effective tool for selecting the next step in the model validation process (experimentation or code development) as demonstrated in a conceptual mathematical model and the PTW model.
- As hypothesized, for a suitable[16] numerical model parameter uncertainty is mitigated through additional experiments. This mitigation is shown through the

[16] Model's with extremely high model form error (i.e. poor predictors of specific phenomena) do not necessarily demonstrate this trend.

increase and eventual convergence of the information gain metric as well as a narrowing of the posterior distributions of the calibration parameters. It is also shown that for a numerical model that is very poor at predicting one or more of the features of interest parameter uncertainty may not be mitigated.

- As hypothesized, model form error is reduced through increased physics/code development as demonstrated in each case study presented herein.
- An approach to treat imprecise input parameters that are dependent upon the control parameters is developed. These imprecise input parameters are functionally trained such that the trained-function coefficients are independent of the control parameters that define the domain of applicability. These coefficients are then treated as the calibration parameters of the numerical model and are successfully calibrated.
- The validated, more sophisticated models are shown to have significantly more accurate predictions when compared to their predecessors.
- The proposed framework is demonstrated on both the VPSC and PTW models to be an effective decision making tool for use in the validation of numerical models. Moreover, the developed framework is shown to provide reliable metrics for choosing whether additional experiments and/or code development are necessary when a reduction in model form error and/or model uncertainty is needed.

Findings from the development of a new algorithm for prioritization of experiments and its application to the VPSC model (Chapter 4):

- BSD can be utilized as an effective tool for selecting the experimental settings at which new experiments should be conducted. As demonstrated in the case study, BSD selects experimental settings that efficiently increase the PMI of the numerical model.
- The C-EIPS criterion, developed herein for use with BSD and based upon both the discrepancy bias and the coverage of the domain of applicability, is shown to effectively increase predictive maturity, measured through PMI, and outperform its predecessor EIPS in a simulated case study using the VPSC code.
- A negative correlation is shown between absolute error and coverage until either full coverage or a zero absolute error is realized in which case the slope of the trend line becomes zero. This correlation provides further evidence that as more experimental data becomes available (resulting in increased coverage) the discrepancy bias of a suitable model reduces and converges.

Findings from the development of a new algorithm and metric for code prioritization and application to the steel frame finite element model (Chapter 5):

- The herein developed error analysis, EA, term for investigating the initially estimated error of a constituent of a coupled numerical model, is shown through the steel frame case study to be an accurate predictor of the degree/magnitude of the constituents' initial error.
- The quantitative CPI metric is capable of effectively ranking the most important model constituents for consideration in model validation campaigns as demonstrated on the semi-rigid steel frame case study. When compared to all

other possible combinations of the order in which model constituents could be improved, the CPI chosen order provided not only the most efficient reduction in model form error, but was also the only order that resulted in a monotonic decrease in model form error.

- To alleviate the widely-known 'mode swapping' phenomena, a mode reordering algorithm is developed to properly order mode shapes for comparison in sensitivity analysis. This algorithm uses a weighted metric combining many mode shape comparison tools (general shape, coordinates of global max/min, coordinates of local max/min) and modes are shown to be successfully reordered for proper sensitivity analysis.
- A strongly coupled numerical model, coupling a frame model with nonlinear rotational springs at connections and a nonlinear connection finite element model is used to infer the moment-rotation curves of each connection in the prototype steel frame. Predictions made using the computed moment-rotation curves provide significantly improved correlation with both vibration and static test results over those that are obtained with initially estimated connection stiffness values.

6.3. Limitations, Remaining Issues, and Recommendations for Future Work:

Limitation/Assumptions:

While the model validation framework developed herein is designed to be a generally applicable framework for use in simulation-based science and engineering fields, limitations and assumptions are present. The following assumptions are made in

the development of framework: (i) conducting further experiments is not prohibited by either resources or statute (e.g. insufficient funding for additional experiments or the prohibited testing of nuclear devices), (ii) there is known additional first-principal knowledge of missing physical laws that could be used to improve the numerical model (e.g. quantum physics models are lacking accuracy however, no additional first-principal knowledge currently exists), (iii) the numerical model has undergone the necessary verification such that numerical errors/uncertainties are sufficiently small, and (iv) at least two experiments and physics sophistications of the code are available such that the path selection algorithm can be properly implemented. A numerical model violating one or more of these assumptions would require a modification to the framework in its current form.

Some notable limitations faced during case study applications should be mentioned. First, both the BSD criterion and the model calibration procedure utilized herein rely on training Gaussian Process Models as surrogate models for the discrepancy bias and numerical model respectively. The underlying assumption with GPM's is that the input/output relationship being trained is smoothly varying and non-chaotic. Therefore, if the physical process being modeled produces chaotic or discontinuous output within the domain of applicability then the GPM surrogate model needs to be replaced with a model that can better handle such circumstances. Second, a best estimate of uniform prior distributions for calibration parameters are chosen throughout this study; as the Bayesian inference methodology is dependent on these prior distributions, it becomes paramount to assume the most informed prior distributions possible for the

given numerical model and parameters. Finally, the CPI metric developed herein, assumes that the coupled numerical model being analyzed can be subdivided into constituents whose important parameters can be sampled independently of other constituents such that global output sensitivity to constituent input and initial error analysis can be conducted.

Suggestions for Future Work:

The framework developed herein with subsidiary components, BSD and CPI are novel, complete, and of immediate use to the science and engineering community; however, as with many research endeavors, many exciting and necessary areas of further research exist to further refine and solidify these tools.

The framework presented herein, can be extended to include a cost factor associated with experiments and/or code development respectively. This cost factor serves as a weight on the computed convergence rate metric such that the relative cost weighs the path selection. An extreme example of this type of weighting, where experiments are considered considerably less costly than code development, is evident in the VPSC case study in Chapter 3. Furthermore, as technology continues to advance, codes are being coupled across multiple physical states and scales; as such, the framework provide herein can be extended to account for multi-physics, multi-scale simulation models analyzing additional issues such as interface discrepancy, quantification of errors and uncertainties propagated through code constituents and across scales. The framework developed herein can also be improved by expanding to include

quantifiable model verification activities to create an integral, scientific, and quantifiable verification and validation framework.

The remaining recommendations for future work focuses on the CPI metric. While the CPI metric is shown to be effective in the case studies presented in this dissertation, this metric is newly developed and should be implemented in many case studies from various fields to identify possible areas for improvement. Some such possibilities include investigating the interaction of the three terms used in the CPI and incorporating weighting terms on each component such that the metric may be adapted for specific uses across many fields.

In closing, the framework and associated algorithms, metrics, and methodologies presented herein provide a step forward in the model validation and uncertainty quantification communities for use not only in the fields demonstrated in the case studies presented in this dissertation, but also other forefront topics in civil engineering such as fluid-structure interaction, soil-structure interaction, and advanced design and analysis software as well as research topics across many other science and engineering disciplines. The assumptions, limitations, and recommendations discussed herein provide a springboard for the continued study and advancement of experiment-based model validation strategies.