Improved best estimate plus uncertainty methodology, including advanced validation concepts, to license evolving nuclear reactors

C. Unal, B. Williams, F. Hemez, S.H. Atamturktur, P. McClure

Many evolving nuclear energy technologies use advanced predictive multiscale, multiphysics modeling and simulation (M&S) capabilities to reduce the cost and schedule of design and licensing. Historically, the role of experiments has been as a primary tool for the design and understanding of nuclear system behavior, while M&S played the subordinate role of supporting experiments. In the new era of multiscale, multiphysics computational-based technology development, this role has been reversed. The experiments will still be needed, but they will be performed at different scales to calibrate and validate the models leading to predictive simulations for design and licensing. Minimizing the required number of validation experiments produces cost and time savings. The use of multiscale, multiphysics models introduces challenges in validating these predictive tools – traditional methodologies will have to be modified to address these challenges.

This paper gives the basic aspects of a methodology that can potentially be used to address these new challenges in the design and licensing of evolving nuclear technology. The main components of the proposed methodology are verification, validation, calibration, and uncertainty quantification – steps similar to the components of the traditional US Nuclear Regulatory Commission (NRC) licensing approach, with the exception of the calibration step. An enhanced calibration concept is introduced here, and is accomplished through data assimilation. The goal of this methodology is to enable best-estimate prediction of system behaviors in both normal and safety-related environments. This goal requires the additional steps of estimating the domain of validation, and quantification of uncertainties, allowing for the extension of results to areas of the validation domain that are not directly tested with experiments. These might include the extension of the M&S capabilities for application to full-scale systems. The new methodology suggests a formalism to quantify an adequate level of validation (predictive maturity) with respect to existing data, so that required new testing can be minimized, saving cost by demonstrating that further testing will not enhance the quality of the predictive tools.

The proposed methodology is at a conceptual level. Upon maturity, and if considered favorably by the stakeholders, it could serve as a new framework for the next generation of the best estimate plus uncertainty (BEPU) licensing methodology that the NRC has developed. In order to achieve maturity, the methodology must be communicated to scientific, design, and regulatory stakeholders for discussion and debate. This paper is the first step in establishing that communication.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

An increase in the use of nuclear energy in the nation's energy portfolio has recently been suggested because of concerns about global climate change, the security of the energy supply, and the depletion of fossil fuel reserves. New nuclear technology concepts such as small modular and advanced nuclear reactors, alternative fuel cycle options, used fuel treatment processes, and waste forms optimizing economics, safety, proliferation, and environmental impacts of the nuclear reactor technology are under consideration in many countries. These new reactor technology concepts require enhanced approaches in design and licensing methodologies in order to be cost effective and timely.

The development of new reactor concepts, including new fuel and core designs and qualification and licensing for large-scale deployment, is a lengthy and expensive process, typically spanning two decades from conception to licensing. For example, fuel fabrication is still viewed by parts of the community as

* Corresponding author at: Los Alamos National Laboratory, Computer Computational & Statistical Sciences Division, Mail Stop B297, Los Alamos, NM 87545, United States. Tel.: +1 505 665 2539; fax: +1 505 665 0120.
E-mail address: cu@lanl.gov (C. Unal).
more of an art than a science, requiring many process iterations through a trial and error approach. In recent years, scientific solutions to some of the difficulties are playing a more prominent role in this empirically dominated process (GNEP, 2007; MacLean and Pasamehmetoglu, 2008). Nuclear fuel and reactor behavior is based on the interaction of multiple complex phenomena, and recent evolutionary approaches are being applied more on a phenomenon-by-phenomenon basis that targets localized problems (Watanabe et al., 2008; Yamada et al., 2000; Ramirez et al., 2006; Vega et al., 2008), as opposed to a systematic approach based on a fundamental understanding of all interacting parameters and processes, and requiring the use of multiscale, multiphysics modeling and simulation (M&S).

Evolving nuclear energy programs expect to use enhanced M&S capabilities, utilizing multiscale, multiphysics modeling approaches, in order to reduce both cost and time from the design through the licensing phases. There has been increased interest in the development of the multiscale, multiphysics approach in the last decade because of the need for predictive tools for complex interacting processes as a means of eliminating the limited use of empirically based model development. Complex interacting processes cannot be predicted by analyzing each individual component in isolation. In most cases, the mathematical models of complex processes and their boundary conditions are nonlinear. As a result, the solutions of these mathematical models often require high-performance computing capabilities and resources.

As an example, the Nuclear Energy Advanced Modeling and Simulation (NEAMS) program introduced an approach that is predicated upon transferring the recent advances in computational sciences and computer technologies to the development of new programs (Unal and Pasamehmetoglu, 2008; Carmack et al., 2010). The NEAMS Integrated Performance and Safety Computational (IPSC) tool development projects will couple computational science with recent advances in the fundamental understanding of physical phenomena, starting from ab initio and progressing to continuum modeling, using selective and reduced testing to leapfrog many fuel and reactor-development activities. The full benefits of this approach will likely take some time to be realized. However, it is important that the developmental activities for modeling and simulation are tightly coupled with experimental efforts in order to maximize feedback effects and accelerate both the experimental and analytical elements of the program towards a common objective. The close integration of M&S and experiments also subjects science-based predictions to rigorous verification and validation (V&V) and enables the explicit quantification of uncertainties embedded within these predictions.

The use of multiscale, multiphysics predictive tools brings another challenging dimension to the licensing methodology. Even though it may be possible to verify and validate the models and simulations, the credibility, objectivity, accuracy, and robustness of predictions that support licensing may or may not be acceptable when the validation is done with limited data. We expect that the required new data, physics models, and simulation tools will be developed simultaneously with the new concepts.

The licensing process that has historically been used by the US Nuclear Regulatory Commission (NRC) for fuels and reactors qualification is based on the use of a large body of existing experimental work to qualify and license a new fuel or reactor (see Appendix A for brief review of applicable NRC requirements and guidance to fuel and reactor licensing). If M&S with more directed experimentation is to be considered as an alternative approach for licensing, a framework must be developed early in the process. The use of M&S with limited experiments as a basis for demonstrating that a design can meet NRC requirements is not new, and has precedence in the NRC. This is the best estimate plus uncertainty (BEPU) approach, so-called because the goal is to compare the model value (best estimate) plus any uncertainty with a figure of merit (such as cladding temperature). This process, when performed for design-basis accidents, is often referred to as “margins assessment,” because the goal is to compare the best estimate answer plus its uncertainty to a predetermined figure of merit in order to establish the “margin” in the engineered solution.

The challenges for extending the BEPU method for fuel qualification for an advanced reactor fuel are reviewed by McClure et al. (2009) and driven by schedule, the need for data, the sufficiency of the data, the need for consistent terminology, identification of important phenomena, the process of validation (with a focus on the multiscale modeling), and the need to produce and extend BEPU uncertainty methodology. For reactor issues other than fuel, the challenges are similar and are generally characterized by a lack of tools with adequate V&V, a lack of necessary data to perform the V&V, and no basis for extrapolating or forecasting the model into areas where the data does not exist. In this paper we offer a proposed set of methods (derived from Nelson et al., 2010) that extends the current BEPU methodology to address most, if not all, of these new challenges. This methodology integrates the classical code V&V with uncertainty quantification and licensing, and addresses issues associated with a sparse data environment (the nuclear experiments are expensive and take time, and limited facilities exist for irradiation testing); the data, models, and simulation tools are developed concurrently. The new methodology suggests a formalism (predictive maturity) to quantify an adequate level of validation with respect to existing data so that the number of required new tests can be minimized by showing that further testing would not enhance the quality of predictive tools, thus saving cost.

The main components of the BEPU methodology are verification, validation, calibration, and uncertainty quantification. Calibration and validation are often sequential operations that are necessary for constructing physics models and quantifying the degree to which they represent physical reality. Calibration is accomplished through data assimilation. The goal is to enable best-estimate prediction of system behaviors with quantified uncertainty in both normal and safety-related environments. This goal requires the additional steps of estimating the domain of validation (i.e., the domain in which the model is executed to yield predictions) and quantifying uncertainties that allow for the extension of results to areas of the validation domain that are not directly tested with experiments – these might include extension of the M&S capabilities for application to full-scale systems. The enhanced calibration method offers an advantage over classical calibration methods because the uncertainties in calibration parameters are quantified as a part of the process. Another advantage is that the multiple data sets (separate effect tests [SET] and integral effect tests [IET]) and multiple models can be calibrated simultaneously.

Finally, we present an implementation of the methodology developed for the NEAMS program for the proposed new nuclear system licensing process. The methodologies presented in this paper should be considered to be potential approaches for enhancing the traditional and BEPU methodologies – a similar methodology has been used in the certification of defense applications. By adopting a proposed framework that can be universally applied to multiple issues in reactor design, the path forward for reactor design and licensing can be simplified and made more cost effective.

1.1. Starting a discussion

We present this methodology to start a discussion and scientific debate in the design, modeling, and regulatory community in order to identify the gaps and offer solutions for accelerat-
Fig. 1. The overview of the VU methodology providing VU-assessed capability for licensing calculations.

1.2. Organization of this paper

We expect our strategy to reduce the qualification and licensing time, identify other potential qualification/licensing issues that were not anticipated, and develop an early understanding of issues and their resolution by the design and regulatory organizations. This paper does not contain an integrated example of a solid application of this methodology to a technology program – rather, it presents a conceptual implementation process for multiscale, multiphysics fuels development. Section 2 discusses the methodology and how it is applied to licensing calculations. Section 3 presents state-of-the-art validation methodologies to address new challenges. Section 4 gives a conceptual example of how one applies the proposed methodology to a multiscale, multiphysics fuels development program.

2. The verification, validation, and uncertainty quantification methodology

The verification, validation, and uncertainty quantification (VU) methodology is a quantitative approach. It is used to determine how well a product (capability) meets the specifications (requirements) of the system it represents through the use of mathematical and numerical M&S. The overarching goal of the VU methodology is to provide a credible M&S capability by integrating information from multiple sources, such as those provided by historical datasets, physical tests, numerical simulations, and expert judgment. These sources are used to: (1) thoroughly verify the implementation of the models and the numerical solutions obtained by in these models, (2) validate the models through quantification of their predictive capability and identification of their origins of uncertainty, and (3) quantify the effect of uncertainty on predictions and the decisions that must be supported by these predictions. These three activities are paramount to the characterization of a predictive simulation capability.

2.1. Step 1: scope of project

Fig. 1 presents an overview of the VU methodology, integrating verification, validation, and uncertainty quantification to provide predictive capability that can directly support licensing calculations. The first step is to establish the scope of a new project. The new VU framework can be applied to the multiple areas of design-basis analysis, whether the issues involve new coolants, fuels, or reactor materials such as cladding. By adopting a new framework that can
be universally applied to multiple issues in reactor design, the path forward for reactor design and licensing can be simplified and made more cost effective. This approach applies to the development of a new fuel type for a particular reactor, or the development of new reactor concepts such as small modular reactors with an existing or new fuel.

2.2. Step 2: phenomenon identification and ranking tables

The phenomenon identification and ranking tables (PIRT) process was developed by the NRC and has evolved from its initial development and applications to its description as a generalized process (Wilson and Boyack, 1998; Wilson et al., 1990). A PIRT can be used to support several important decision-making processes as well as to identify programmatic needs when developing requirements for experimental and modeling work. The PIRT process in this framework is used to specify following multiple objectives:

- M&S requirements.
- Experimental requirements for validation of new models and simulations.
- Description of design calculations.
- Development of detailed, specific phenomenological issues for design and beyond design-basis accidents for the proposed project.

A good example of a PIRT that is applied to high burn-up fuel can be found in Boyack et al. (2001).

2.3. Step 3: modeling and simulation and new experimental requirements

We assume that M&S needs are determined using PIRT methodology for design concepts under consideration. The models and associated computer codes are then developed to meet the design and modeling requirements. The models can be simple algebraic equations, complicated mechanistic models, or may involve computer simulations from another scale or capability. A good example of modeling requirements for a new fuel development project is given in Unal and Pasamehmetoglu (2008).

2.4. Step 4: verification

Once the computer code has been built with the appropriate constitutive models (also called mechanistic models or physics models), the first process is verification. Verification is concerned with the degree to which the underlying mathematical and computational models are correctly implemented and solved. Verification is separated into two main categories, code verification and solution verification, with each category addressing different needs. Code verification encompasses the broad area of software quality engineering (SQE), which defines the processes and practices used for software design, implementation, testing, version release, documentation, training, and more. A software project will often adhere to (or have imposed upon it) a specific SQE framework by which it is measured and assessed.

Understanding that code verification is more than just SQE is critical. Included in the definition of verification is the process of quantifying the mathematical correctness of the numerical algorithms and their implementations. Standard SQE does not necessarily address the issue of mathematical correctness, but proper code verification does address it through techniques such as order-of-accuracy testing. To do this, it is necessary to have test problems with known exact or highly accurate solutions. The method of manufactured solutions (MMS) is a very useful technique for obtaining exact solutions for the many situations where analytical solutions do not exist, allowing for more thorough testing of the code.

Like code verification, solution verification is concerned with mathematical correctness. However, solution verification is carried out during the application of the code to a real analysis. In these cases, exact or extremely accurate solutions are almost never attainable. One of the main activities in proper solution verification is the quantification and/or control of the numerical discretization error to ensure that the numerical error is sufficiently small. To define “sufficiently,” clear knowledge of the intended use of the code is required. (Note that “intended use” could include simulations as diverse as an analysis for a licensing application or a validation study prior to code release.) Various methodologies exist for estimating and controlling discretization error, including various flavors of a posteriori error estimators and indicators, as well as adaptive mesh refinement. A detailed description of the verification process and state-of-the-art capabilities available to us are not given in this paper (see Nelson et al., 2010).

2.5. Step 5: validation

The next step in the process is code validation. Validation employs testing to determine if predictions are acceptable, taking into consideration the uncertainties in both the data and predictions. Double arrows in Fig. 1 indicate a potential flow of information in both directions to enhance or improve the M&S capability. In particular, arrows associated with the flow of information from and to verification and validation represent the testing process associated with each. Note that this is a change from the traditional NRC methodology where the data is assumed to already exist. This methodology allows for the validation effort to guide the development of new experimental data until sufficiently high robustness and sufficiently low bias are achieved in the validation. The foundation of the validation process is built from uncertainty quantification. M&S capabilities (codes), constructed to predict system designs and potential safety scenarios, are assessed using this foundation. However, the system design and safety and licensing calculations are considered after the M&S capability is validated to an acceptable level, as shown by the process in the right side of Fig. 1. The validation process produces a VU-assessed code that can be used in licensing calculations. The validation process quantifies key modeling parameters and their uncertainties, which need to be propagated in full-scale licensing calculations. The validation can consider already existing separate effects and integral effects data, as well as newly defined small-scale experiments supporting multiscale, multiphysics approaches.

Another difference from the traditional approaches used in reactor safety is the introduction of calibration. A Bayesian calibration methodology is used for selected groups of data as well as multiple data sets. A sequential calibration strategy is used for newly added data to enable decisions on the sufficiency of the new data for achieving acceptable levels of bias and robustness. A concept to measure the predictability of a model, a predictive maturity index (PMI), is introduced in the validation process. A PMI establishes the sufficiency of new data to decide whether the model solution is acceptable. In the past, expert judgment determined the completeness of validation— the proposed approach represents a potentially significant improvement. The new method can also give guidance to experimentalists for selecting new sets of experiments to improve predictive capability. The PMI concept, when used in a sequential calibration scheme, also answers the question of stability in the predictions—that is, when further testing does not add value to the current physics understanding. Details of these new concepts are discussed in Section 3. We discuss the expected results of this step in Section 2.
2.6. Step 6: analysis during validation

Through the use of a quantitative index for predictive maturity, the validation process ensures that predictions are robust and sufficiently matured for licensing calculations. Within each layer, confidence levels and their contributing uncertainties are determined and provide a measure of the code’s predictive capability. The numerical issues associated with nodalization are part of the validation process. Converged nodalization schemes for integrated tests are target schemes for the full-size-model development. The baseline full-size-plant model is prepared during the validation. The significant sensitivities, including parameter-to-parameter interactions (see Appendix B for details of the global sensitivity analysis methodology), are performed on modeling fidelity and nodalization. The new validation concepts are implemented for validating the extension of previous models using new data. The proposed forecasting model is validated in this step using available and potential new data sets.

The other key feature of this methodology that differs from traditional BEPU methodology is the use of feedback loops if desired performance is not obtained. Either new model development or new experiments may be required to fully meet the validation criteria. Feedback loops exist in the current evaluation methodology; however, the current code scaling, applicability, and uncertainty (CSAU, see Boyack et al. (1990)) methodology relies heavily on the assumption that the physics of the phenomenon of interest is known in advance. This may not be the case when experiments and modeling are being developed simultaneously. When the predictive ability of M&S tools, for example, decreases or does not converge to a desired level upon the addition of new data points to the calibration data set, a new PIRT (or PIRT-like process) may be needed because the deteriorating performance of tools may be due to unknown or inadequate physics issues. At this stage, multiple iterations on model assessments using the processes of PIRT and/or expert judgment may also be needed. This fact is formally recognized in the process. Although not expected, the new validation process may require design changes. Therefore, the design, safety, model, and code developers have to work together in parallel on implementing this approach. This feature may be responsible for saving time and money in design by producing VU-assessed designs and tools.

It is also worth mentioning here that the PIRT process we envision in this step should not be limited to processes described by Wilson et al. (1990), and can include new methodologies such as suggested by Luo et al. (2010) who proposed the use of the “Quantified PIRT (QPIRT)” process. QPIRT is based on the characteristic ratio of dimensionless numbers, which represent the physical processes as proposed in the two-tier scaling methodology that is used in the nuclear industry. QPIRT is designed to support PIRT with quantified judgments. QPIRT offers improved approaches for identifying important physics in integrated complex simulations. QPIRT, however, is applicable when the physical processes can be represented with well-defined mathematical equations and dimensionless numbers. While it is applicable to most of our areas of interest, it may not be an effective method when the knowledge of driving physical mechanisms is incomplete, such as in the area of irradiation of fuels that occurs at the atomistic scale, where models are primarily empirical representations of potential functions describing interactions between atoms. When the physics are known and can be expressed with scaling rules, QPIRT can offer not only more relevant sensitivity studies that identify key physics and associated uncertainty parameters, but also the technical basis for justification of extrapolations and scaling. Verifying the existence of similar key dimensionless numbers in the extrapolated regimes with regard to certain physical processes could be used as the basis for extrapolated predictions in scenarios for which there are no data.

2.7. Step 7: VU-assessed code

The robust VU-assessed code is produced at the end of the validation process. This code is used to simulate the full system behavior. The full system model is developed after the lessons learned are developed from the sensitivity studies done in the validation analysis step. The full plant model can be tuned with the plant data after the numerical model becomes robust. After some fine-tuning of this model, it is ready to be used for safety calculations.

2.8. Step 8: full system model and predictions with uncertainties

The manufacturing design uncertainties, scaling related uncertainties, and modeling uncertainties are combined in the licensing calculations using an actual system model. The licensing calculations are done for design-basis or beyond-design-basis accident scenarios, as well as under normal full system operating conditions for design optimizations. The safety-related scenarios are identified after a PIRT is developed for the design. An extended group, including designers, regulators, and the VU experts, is assembled to perform the PIRT. PIRT is also used to develop M&S requirements as depicted in Fig. 1.

The uncertainty in the full-size plant model results from modeling uncertainties, design uncertainties induced by the fabrication processes, and scaling and forecasting uncertainties. The key modeling parameters and their calibrated uncertainties are obtained from the validation analysis step. The fabrication tolerances can be defined from manufacturing tolerances. The simultaneous propagation of modeling and manufacturing uncertainties is a desired goal, but is very expensive because multiple full-size plant models are needed. If resources are available, both uncertainties should be propagated in the safety calculations. If this is not possible, then only modeling uncertainties should be propagated. If the full-size system will operate in an extrapolated regime, then a forecasting model should be used in this step. Additional uncertainties due to scaling should be incorporated as additional biases, as has been done in the BEPU methodology. Probabilistic calibration and manufacturing uncertainty considerations are new features of this methodology.

At the end of the process (a full system simulation), predictions and their uncertainties determine the safety margin for a defined metric or metrics (e.g., the peak rod clad temperature in a reactor is below the melting point of the clad material so that the rod stays intact and its fuel material remains inside).

BEPU methodology allows design changes and model reevaluation with or without new experiments as a possibility if margins are insufficient for regulatory criteria. This implies that the project may go through the whole process in order to obtain a desired margin acceptable to the regulatory body while recording all changes made. This process is believed to add more credibility to the methodology.

2.9. Step 9: scaling and forecasting biases

The quantified uncertainties in the validation phase are associated with the modeling and modeling fidelities (alternative models). The uncertainties in design and fabrication of the system are not considered in the validation step. These and additional uncertainties (associated with the scaling between testing conditions and the actual system) are addressed in the licensing calculations. Uncertainties associated with scaling issues occur in several aspects as validation data are collected. They can include geometric scaling and the use of simulant fluids and materials (for example, the use of electrically heated rods instead of nuclear rods). The use of models in extrapolation regimes is another difficulty.
Each aspect of these scaling issues requires a different approach. The validation methodology offers the option of accounting for the uncertainty when models are used in extrapolation regimes (forecasting uncertainty). This, again, is a feature that deviates from the traditional BEPU approach. In Fig. 1, an arrow is connected from the validation box to the scaling and forecasting box to demonstrate this ability of the methodology.

2.10. Step 10: margins

If the margin is not sufficient relative to regulatory guidelines, the largest uncertainties are identified – these provide a basis for further experimentation and/or model development, aimed at reducing particular uncertainties and improving margin. Thus the overall process has loops requiring multiple iterations in the process, shown by the gray arrows in Fig. 1. If this entire process is documented properly, a licensing agency then has sufficient information to make an informed decision about the system’s operation and safety. Since this process is proposed to be used simultaneously by design and safety organizations through a transparent process with the involvement of the licensing agency, the licensing agency may require less review time than is currently needed.

2.11. Step 11: design and safety analysis report

The documentation of the processes in each step and the final design and safety calculations are assembled for review by the licensing authority.

2.12. Detailed discussion

The above discussion summarizes the overview of the methodology. The remaining subsections discuss details of the new validation and uncertainty quantification concepts introduced in this paper.

3. Enhanced validation, calibration, and predictive maturity, including uncertainties

3.1. Statistical model for calibration with uncertainty quantification

Uncertainties are grouped into two types\(^1\): aleatory and epistemic. Aleatory uncertainties are random and irreducible. An example of aleatory uncertainty is the roll of dice. We often represent the variability of a process or experiment as aleatory because of our inability to precisely control inherent variability from sources such as materials fabrication and environmental conditions. Manufacturing uncertainties in design are random in nature and are thus aleatory.

Epistemic uncertainties occur due to lack of knowledge. An inadequate representation of a particular physics phenomenon, or missing physics, is an example of epistemic uncertainty. Multiple model inference is one approach to quantifying this sort of epistemic uncertainty. Within the calibration process, imprecise model parameters (also some times called calibration parameters, denoted as \(\theta\) below) are treated as aleatory uncertainties and adjusted based on available data.

To provide more clarity to this discussion, the following relationship is given, representing the “real world” as it relates to the modeling and simulation:

\[
y(x) = \xi(x) + \epsilon(x),
\]

where \(x\) is the vector of controllable parameters (design conditions, e.g., pressure), \(\xi(x)\) is the true response of the physical system under conditions \(x\) (reality), \(\epsilon(x)\) is the observation error, and \(y(x)\) is the observed system response (data).

Often the inferred observation error is associated only with the instrumentation. However, it is desirable to assess both components of observation error, the instrument error and experiment repeatability (irreducible statistical variability).

Next we represent the real world as:

\[
\xi(x) = \eta(x, \theta) + \delta(x),
\]

where \(\theta\) is the vector of unknown best values of the calibration parameters in the model, \(\eta(x, \theta)\) is the computer model, and \(\delta(x)\) is the discrepancy (model inadequacy) term – that is, the difference between reality and the best computer model.

An intrinsic component of the VU process is model calibration. Model calibration attempts, through the comparison of simulation predictions and experimental measurements, to gain an understanding of the imprecise model calibration parameters, \(\theta\), as well as inadequate representation or missing physics in the model, which are reflected through \(\delta(x)\).

Traditionally calibration determines the single best value of \(\theta\) under the assumption that \(\delta(x)\) is zero, for example:

\[
\theta = \arg \min_{\theta} \sum_{i=1}^{N} (y(x_i) - \eta(x_i, \theta^*))^2
\]

using data \(y(x_1), \ldots, y(x_N)\). Thus the model is forced to represent the data by adjusting the calibration parameters.

More recent calibration methodologies are based on estimating a best \(\theta\) while simultaneously inferring \(\delta(x)\):

\[
\theta = \arg \min_{\theta, \delta} \sum_{i=1}^{N} \left\{ \lambda_1 \sum_{i=1}^{N} (y(x_i) - \eta(x_i, \theta^*) - \delta(x_i))^2 + \lambda_2 \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma^{ij} \delta(x_i) \delta(x_j) \right\}
\]

where the \(\sigma^{ij}\) is derived from the \(\delta\) covariance structure.

Traditional calibration approaches ignore the uncertainty in \(\theta\) that is induced by the dual uncertainties associated with observation error and discrepancy bias, and thus uncertainties in predictions of interest are also neglected. This situation is remedied by using modern statistical methods to calibrate \(\theta\) with formal quantification of uncertainty in the calibration (see Appendix B for a discussion of a Bayesian approach to calibration and prediction with uncertainty quantification).

When fixed values of the model input, \(x\) and \(\theta\), are selected and fed into the model, an output \(\eta\) is calculated. Forward propagation of uncertainty refers to passing samples of these pairs of input vectors, taken from appropriate probability distributions, through \(\eta\) to obtain predictions of desired metrics.

Inverse uncertainty propagation (often called the adjoint method), in contrast, directly attributes the sources of uncertainty in the output to variation in the calibration parameters. In this case, one would obtain the quantity of interest \(\xi\) with uncertainty, and directly solve for \(\theta\). The existence of multivalued solutions with nonlinear models presents challenges to the use of adjoint methods.

Thus forward propagation of uncertainty results in a family of model runs, which are repeated for a sample of calibration parameters in order to observe the variability in the model outcomes.

\(^1\) It is not unusual for heated debates to occur about the classification of an uncertainty as one type or the other. Our view is that what it is called, aleatory or epistemic, is less important than its identification and how it will be handled within a particular study.
Fig. 2a provides results showing uncertainty in $\eta$ as a function of design space $x$ for uncalibrated uncertainty in $\theta$. In this case, the initial (prior) distribution of $\theta$ is uniform, as shown by the small insert in the upper left portion of Fig. 2a. Data with its uncertainty bars are shown with the black symbols. While the trend of the predictions is somewhat like the data, one can see the spread in predictions is much larger than the data uncertainty. A Bayesian inference approach, such as that introduced by Higdon et al. (2004), is used to update this initial uncertainty in $\theta$ by accounting for information in the available data to obtain a calibrated $\theta$ distribution, called the “posterior” distribution. The small insert in the upper left portion of Fig. 2b shows the calibrated $\theta$ distribution (note the blue distribution) and the resulting predictions (also in blue) in the figure. One should note that the spread in the predictions is much less than that associated with the prior distribution (shown in green). Is this calibration good enough? That question is addressed as part of the validation process on the left side of Fig. 1 and may be investigated by consideration of predictive maturity metrics in conjunction with “acceptance criteria” defined by the customer.

As long as a large enough sample of calibrated $\theta$ values is generated, the resulting sample of calibrated predictions is representative of the true distribution of model predictions, given currently available experimental data. The number of necessary samples depends on model complexity and on the adopted sampling approach.

Not shown in this example is the discrepancy term $\delta(x)$. Effectively, discrepancy is used to compensate for modeling inaccuracies that might be due to an inadequate functional form or missing physics. A desired characteristic of $\delta(x)$ is that $\delta(x) \approx 0$. When this is true, the model is a reasonable approximation to the data. But when that characteristic is not true, more work is required to revise or develop a new model. Wang et al. (2009) quantify the concept of achieving near-zero discrepancy into criteria for validating complex computer codes.

For calibration to be meaningful, both types of uncertainty, aleatory and epistemic, must be investigated, and the quantification of uncertainty both in simulation predictions and experimental measurements must be an integral part of the calibration process. Thus, among the activities of VU, calibration is particularly demanding on resources, as it requires a large number of repeated computational solutions and experimental measurements.

Epistemic uncertainty that results from reliance on particular physical or model forms can introduce systematic bias in predictions of interest, particularly if they represent extrapolations. Alternative physical models or specifications may be simultaneously considered for desired predictive inferences. Each included model form must be validated and its predictive maturity quantified. Assessments required to support licensing would combine calibrated inferences across the models, which are appropriately weighted to reflect fidelity to existing experimental data and parsimony.

In current practice, the notion of completion for these model calibration activities is established according to factors that are often driven by nonscientific constraints. Such factors include, for instance, budgetary and time constraints placed on experimental facilities and the testing that is done. For proper management of resources, however, an objective, quantitative, and scientifically dependable definition of completion is needed.

3.2. Data

Data integrity must be ensured before its use. The data can then be used as part of the validation or calibration processes. The effort to validate data can place significant demands on resources (time and money); however, a comparison to “corrupt or inaccurate” data and the associated problems of the (1) costs of trying to correct a non-existent M&S problem arising from a “poor comparison of prediction to bad data,” and (2) credibility loss that results with a sponsor or licensing agency due to finding the data problem later rather than sooner, justify the cost of assuring data integrity.

Data validation and quality assurance requirements will not be discussed further here, but their importance must be realized (see Nelson et al., 2010).

3.3. Coupling and calibration

Two further significant complications enter the calibration process. One complication results from the coupling that may or may not exist between physics phenomena. This coupling is typically characterized as weak or strong and will be discussed in greater detail in Section 3.3.1. The second complication results from the fact that the degree of coupling between phenomena can vary from one application (scenario) to another.

3.3.1. Calibration of weakly coupled physics phenomena

Two physics models, $\eta_1(x, \theta_1)$ and $\eta_2(x, \theta_2)$, are said to be weakly coupled when the physics phenomenon each one represents is
Fig. 3. Bayesian calibration process for many uncoupled physics phenomena to determine $\theta_i$ for each model using its specific separate effect test.

independent of the other – that is, the two are not correlated. Calibration for weakly coupled physics models is a straightforward process since separate effect tests (SET) can be done for each phenomenon within the desired design space $x$, and the corresponding $\theta$ and $\delta(x)$ can be determined using the Bayesian inference approach.

Extension of this process for many uncoupled physics phenomena is shown in Fig. 3. Basically, each model can be processed separately to provide a calibrated posterior $\theta_i$ distribution, where the $i$ subscript represents the $i$-th model. This process of building a complex model from simpler uncoupled (or weakly coupled) parts was first suggested by Simon (1962) and is often called “hierarchical decomposition.” In complexity theory by Auyang (1998), this process of building a complex system from its fundamental parts is often called “reductionism,” where the behavior of the system is realized from the sum of its independent parts.

Predictions of applications (scenarios) against IETs for an M&S capability are generated by forward propagating the posterior $\theta_i$ distribution through the relevant models. These predictions support validation of these models through formal comparisons to the IET data. This process from calibration through validation for uncoupled physics models is depicted in Fig. 4. It is useful to revisit Fig. 1 to emphasize progress through various hierarchies of data within this calibration and validation process.

3.3.2. Calibration of strongly coupled physics phenomena

The calibration of strongly coupled physics phenomena is generally much more complicated. Two strongly coupled models might be written as, $\eta_1(x,\theta_1,\theta_2)$ and $\eta_2(x,\theta_2,\theta_1)$, with the physics phenomena of each implicitly related to (dependent on) the other. This representation is required when the phenomena are present in an application (scenario) where their behavior is interrelated. However, the models should also be able to represent situations (applications or scenarios) where only one phenomenon is present, or their coupling is weak. It takes a great deal of expertise to differentiate between these cases if weak versus strong coupling is not suggested in advance – that is, one might hope that the PIRT phase of the process indicated in Fig. 1 would provide assistance, but that is not always the case.

Proper representation of the coupled physics phenomena is dependent on experiments that can measure some characteristic of each phenomenon to allow proper separation of the two (or more) models. If, for example, only measurements for overall global system behaviors are made, the most likely compensating errors among the two (or more) models cannot be eliminated.

Fig. 5 is a flow representation of this calibration process for a strongly coupled physics system. Note that the complete hierarchies of data in Fig. 5 are used within this calibration process where the posterior $\theta$ distributions are produced using all available data. Within this particular example, models 1 and 2 are coupled in IET 1 while models 1, 2, and 3 are coupled in IET 2.

3.3.3. Concept of stabilization of predictions

The two requirements for predictive maturity are (1) consistency with experiments, and (2) robustness to added new experiments. The concept of stabilization states that the addition of each new experiment provides fractional new information for model calibration and thus a fractional reduction in discrepancy bias. Therefore, the discrepancy term gradually converges to a stable value (either to a negligible value within experimental uncertainty, or to a constant non-zero value of acceptable magnitude).

Fig. 6 represents the desired behavior of discrepancy terms. Here, the first discrepancy estimate is obtained by calibrating the model with respect to a single experiment, the next is obtained by...
The notional behavior of the discrepancy term representing the model inadequacy as a function of the number of available experiments. Figure 6.

Calibrating the model to two experiments, and so forth. Once the discrepancy is stabilized, corresponding to the fifth experiment of Figure 6, allocating resources for additional experiments provides only a marginal reduction in discrepancy bias. Therefore, the maturity of a calibrated model must be judged on the basis of discrepancy bias stabilization rather than its goodness-of-fit.

3.3.4. Predictive maturity index – a quantitative metric

Several questions can arise for any validation methodology. These include the following:

- “When is the model solution good enough?”
- How many physical experiments must be conducted to reach the desired predictive maturity?
- What set of physical experiments maximize the improvement in predictive fidelity?

In the past, when no other method was evident to answer the first question, expert judgment was used to make this determination. For the approach used here, the idea of characterizing the “maturity” of a model result is introduced. Maturity can be defined as either a qualitative or a quantitative value. For example, early qualitative measures of maturity are suggested by Oberkampf et al. (2007). The value of a qualitative measure of maturity lies in the opinions the experts bring to bear on the question. Even as more quantitative measures are developed, discussion with the expert base will always remain of great value to the overall result.

One aspect a quantitative maturity measure provides is a measure of the “change” or “improvement” in a result – very close in principle to determining that a maturity is reaching an asymptote. This criterion is measured and used as a means of determining when (1) further experimentation is needed, or (2) the maturity of the model has been reached and no further reduction in uncertainty can be achieved, or (3) the figure of merit has been achieved.

The PMI (Hemez et al., 2010) is an objective, quantifiable metric, capable of evaluating progress in predictive capability as calibration activities take place.

The quantitative metric for PMI is continuing to evolve, but a basic understanding can be gained from its initial form denoted by:

\[ \text{PMI} = F(\gamma_c; N_K; \delta_5), \]  

where \( F \) is a functional form defined to satisfy desired characteristics, \( \gamma_c \) is the coverage ratio e.g. convex hull of data divided by convex hull of validation domain, \( N_K \) is the number of calibration parameters in the model, and \( \delta_5 \) is the goodness of fit related to the discrepancy \( \delta \) of Section 3.1. Appendix D discusses the PMI formulation and its application to a practical case of creep modeling of clad materials. A more detailed discussion of these quantities can be found in Hemez et al. (2010).

For the development of a PMI metric we adapt a batch sequential experiment design strategy (Williams et al., 2010), an approach that aims to define the optimum settings at which the physical experiments should be conducted for model calibration. Batch sequential design targets a desired level of predictive maturity with a minimal number of physical experiments.

An example of a recent Los Alamos National Laboratory (LANL) application of predictive maturity is shown in Figure 7; it is based on the results of Hemez et al. (2009). Figure 7 shows the influence of the sequential addition of data to the calibration of a material strength model (Preston et al., 2003, Preston–Tonks–Wallace (PTW)). One can observe that after the first seven tests are considered, little is gained by adding additional tests.

The same concept is applied to calibration of the Visco Plastic Self Consistent (VPSC) code that is a multipurpose, homogenization-based polycrystal plasticity computer code, based on the knowledge of the plastic deformation mechanisms at the single-crystal level and the crystallographic texture of the polycrystalline aggregate (Atamturktur et al., 2010) (see Appendix D for this example).

Regarding the second question (How many physical experiments must be conducted to reach the desired predictive maturity?), its answer would be particularly useful for simulating the performance of new generations of nuclear fuels, reactors, sep-
aration plants integrated with their safeguards, and waste forms where needs (in terms of experimentation to support model validation and calibration) could rapidly overwhelm the available resources. Furthermore, when the model is mature enough in the domain of interest, additional experiments may provide insignificant improvement, which cannot justify allocating resources for further experimentation.

While not discussed here, this approach can also address the third question (What set of physical experiments maximize the improvement in predictive fidelity?). This topic involves consideration of both the concepts of statistical experiment design and expert input regarding critical scientific issues of relevance to experimental campaigns.

3.3.5. Validation including predictive maturity

Very recently, several attempts were made to develop criteria for measuring the level of maturity for an M&S capability. Oberkampf and Trucano (2007) summarize the metrics suggested by the National Aeronautics and Space Administration (NASA) and others, used to objectively track maturity, leading to the credibility of an M&S effort. They propose four levels of maturity in considering the consequences of the analysis.

Although these efforts are very valuable and should be considered in the design of a new nuclear reactor certification process, they require the expert interpretation of results. Likewise, it is difficult to quantify the effect that different levels of maturity may have on risk and decision-making. These issues drive the need for another approach to develop a quantitative metric of predictive maturity in the M&S effort. This approach complements Oberkampf’s maturity model. Such a model is expected to increase the credibility of safety predictions and quantify the sufficiency of simulation predictions. Establishing a level of predictive maturity deemed acceptable will also enable us to justify a limited number of tests. The characterization of predictability using the maturity model becomes extremely important when the fuel performance tools are validated with limited data, and used in untested regions of design space.

Fig. 8 shows an expansion of the basic validation process represented by Fig. 1, but includes the use of predictive maturity. It shows the flow process we want to apply as part of the validation stage before we assess the full-scale testing to complete the final stage of the validation process. The objective is to move from point A (build the M&S product) to point E (validated product with quantified uncertainties) making the best decision possible (best investment in additional experiments or improved physics models) to provide a scientifically defendable program that produces the best capability at minimal investment.

The four boxes on the left side of Fig. 8 characterize the M&S assessment process with uncertainty quantification (UQ) but without predictive maturity. As noted, adding the predictive maturity step to this process produces significant value – it determines the value of additional work (experimental or model development), thus justifying and defending the required levels of this work. In this regard, Fig. 8 indicates that there is a possibility that sufficient maturity cannot be demonstrated and additional experiments may be required. In this case, the process returns to the design and experimental box to either change the design or obtain new experiments that address the validation issues in the physics models or integral simulations (point B–C–D), and then the methodology is applied again until sufficient maturity is attained. There is also a possibility that new data will not provide further improvement and an improved or new model will be required (path B–A).

3.3.6. A forecasting metric for predictive modeling

The discrepancy bias is defined in Section 3.1 as the difference between reality and the calibrated model as a function of the design settings. At the settings where experimental information is available, the discrepancy bias can be obtained with high fidelity. However, simulation models are rarely used to make predictions at settings where experiments are readily available. Therefore, to ensure the success of a licensing process dominated by forecasting predictions, the focus must go beyond the discrepancy bias and consider forecasting errors. In this section we will use the term forecasting error (introduced by Atamturktur et al., 2010) to represent the difference between reality and the bias-corrected calibrated model solutions at settings where experiments are unavailable. The difference between discrepancy bias and forecasting error is illustrated in Fig. 9. We incorporate a forecasting error component in
analyses designed to isolate inference about errors. To errors that are hopefully captured by the proposed forecasting model at the tested settings yields the discrepancy term. The disagreement between reality and calibrated model predictions at untested settings yields the forecasting error (see Fig. 9).

If the true form of the discrepancy function is known, model biases can be properly corrected. In practical applications, the true form of discrepancy function will inevitably be unknown and the estimated discrepancy bias will have inherent inaccuracies. This is of particular concern when extrapolation is required, ultimately contributing to forecasting error.

Here, $x'$ represents the settings at which a forecast is required, and reality $y(x')$ represents the truth observed at the settings of $x'$. The code calculation at settings $(x',\theta)$ is given by $\eta(x',\theta)$, while the discrepancy $\delta(x')$ is a stochastic process supplying the quantification of systematic bias error. The calibrated and bias-corrected statistical process model, therefore, yields our best possible estimate of reality, henceforth referred to as the best estimate model. At the tested settings of $x^t$ this best estimate model is expected to successfully infer reality. However, at settings other than $x^t$, this best estimate model may not maintain the same level of fidelity, especially for extrapolations, possibly leading to errors that are hopefully captured by the proposed forecasting error.

The forecasting error will also be treated as a stochastic process, with constraints on its behavior provided by experts or analyses designed to isolate inference about $E(x)$. For example, a conservative approach would entail specifying an upper bound for forecasting error applied to extrapolations. Note that alternative specifications, such as multiplicative forecasting errors, are also possible. As more experimental measurements become available, uncertainties in model calibration and discrepancy bias will be defined more accurately. As a result, we expect that forecasting error will also be reduced. This anticipated reduction is similar to the behavior expected from the discrepancy bias. Thus, the notion of forecasting error applied to extrapolations is similar to the representation of discrepancy bias as illustrated in Fig. 6.

For the fuels effort, example output quantities might include the number and depth of lower-length scale models may vary but typically will be limited to two levels. This approach must ensure that all uncertainty and bias in the lower-length scale data and models are taken into account. Therefore, when uncertainty is propagated in the engineering-level model, the uncertainties from the lower-length scale models must be propagated upwards as well.

Fig. 10 represents the multi-level validation framework at a conceptual level; it is possible for the engineering scale to use direct experimental data from the atomistic and meso scales. The details of this process will be decided after the atomistically informed meso-scale framework is defined.

The primary uncertainty quantification method we envision at this point is forward propagation. Our simulations will be much more complex than the current tools used for analysis and will involve multiphysics codes and data. The number of potential model parameters involved is very high. The methodology to reduce the number of significant parameters is a key part of our overall approach.

Figs. 11–13 show the processes we will follow. The initial model parameters involved is very high. The methodology to reduce the number of significant parameters is a key part of our overall approach. Figs. 11–13 show the processes we will follow to quantify the uncertainty in system performance output metrics. For the fuels effort, example output quantities might include temperature, strain, and stresses.

Physics models are developed using a classical hierarchical decomposition approach (Simon, 1962) from SETs. A hierarchical Bayesian model (HBM) is used to characterize the initial model parameter uncertainty induced by SETs. The models developed from SETs are applicable to a certain state space. We often extrapolate these models to more extreme conditions at which data are lacking. In order to justify these extrapolations we assess IET data where the state spaces are more typical of conditions we seek.

An atomistic scale simulation can accept measured model parameters and assumed functional forms as well as adjustable model parameters. The uncertainties in these inputs are constrained by SETs (these could be lattice measurements, etc.) and the resulting parameter distribution becomes a prior to atomistic scale simulations, as shown in Fig. 11 (prior drawn in red). The atomistic scale simulations are then used to further calibrate these parameters against integral data obtained for single crystals.

We start our process by determining all uncertain factors going into the atomistic scale simulations. Inputs are initially constrained by SET data, if available. If necessary, atomistic scale simulations are run for the purpose of developing a statistical surrogate for model
output, and an HBM is used to further calibrate the uncertain inputs to IET data. This results in a posterior distribution for the uncertain inputs arising from the atomistic level uncertainty quantification, as shown in Fig. 11 (drawn in green).

If extrapolation of the SET conditions to the IET conditions were not required, and the various physics models employed in the IET were truly uncoupled, then we would expect the prior and posterior distributions to be very similar, if not the same. Note also that some of the atomistic model parameters derive from ab initio models that may not need to be calibrated. If the atomistic model is successfully calibrated with these ab initio models, the associated parameters can be carried as is to the next stage in the process. However, if this calibration fails, alternatives to the assumed ab initio models may be necessary.

The outputs from this first stage are calibrated key atomistic modeling parameters with their uncertainties which now serve as prior inputs to the meso-scale UQ analysis. A similar inference process will then be applied with uncertain atomistic and meso-scale simulation inputs, and meso-scale SET and IET data. The only major difference from the atomistic level inference is that the meso-scale simulations accept atomistic level modeling parameters and their uncertainties, as well as uncertain meso-scale parameters and experimental SET data performed at the meso-scale. These experimental studies are currently being defined. The meso-scale will also have access to some IETs, such as grain growth experiments. These data will be used to further calibrate the adjustable meso-scale inputs. Fig. 12 illustrates the steps involved in this intermediate stage of UQ analysis.

Fig. 11. Uncertainty quantification methodology for atomistic scale simulations.

Fig. 12. Uncertainty quantification methodology for meso-scale simulations.
Fig. 13 shows the UQ process for the final stage of inference involving engineering-scale code calculations. This analysis accepts calculated meso-scale parameters, but also measured or calculated parameters, from atomistic scale simulations. As before, the second-stage calibration serves as a prior constraint on uncertain atomistic and meso-scale parameters, and an HBM is then applied to calibrate uncertain engineering-scale parameters to SET data. When these sources of uncertainty are all integrated into the engineering-scale calculations, complex compensating errors can occur — although they will be mitigated by adequate data constraints from the first two stages of this inference process. Engineering-scale IET data will be obtained from the Advanced Fuel Cycle (AFC) testing series. AFC testing uses prototypical pellet dimensions with shorter rods. We plan to use AFC data to further constrain input uncertainties in order to generate a final set of modeling parameters with quantified uncertainties that can be used in plant calculations without further calibration.

Data reduction models may be needed when processing output from calculations at one scale for input into calculations at another scale, potentially introducing an additional source of uncertainty that must be quantified. The calibrated parameter uncertainties resulting from this three-stage UQ process are combined with relevant manufacturing input uncertainties and propagated through performance calculations in order to obtain the predictive inferences used in fuel qualification. These assessments also generally require thresholds with rigorously quantified uncertainties for comparison with performance results.

5. Summary and conclusions

In this paper, we lay out the basic aspects of a methodology that can potentially be used to address new challenges in the design and licensing of evolving nuclear technology programs. The main components of the proposed methodology are verification, validation, calibration, and uncertainty quantification. An enhanced calibration concept is introduced and is accomplished through data assimilation. The goal is to enable best-estimate prediction of system behaviors with quantified uncertainties in both normal and safety-related environments. This goal requires the additional steps of estimating the domain of validation and quantification of uncertainties that allow for the extension of results to areas of the validation domain that are not directly tested with experiments — these might include extension of the M&S capabilities for application to full-scale systems. The new methodology suggests a formalism (predictive maturity) to quantify an adequate level of validation with respect to existing data so that required testing can be minimized to save cost by showing that further testing would not enhance the quality of the predictive tools.

The proposed methodology is at a conceptual level. Upon maturity, and if considered favorably by the stakeholders, it could provide a new framework for the next generation of BEPU licensing methodology developed by the NRC. In order to arrive at that level of maturity it is necessary to communicate the methodology to scientific, design, and regulatory stakeholders for discussion and debate. This paper is the first step in establishing this communication and we hope that a comprehensive assessment of gaps and difficulties in the proposed methodology will be conducted as stakeholders are actively engaged.

Acknowledgements

This work was sponsored by the US Department of Energy, Nuclear Energy Division, Nuclear Energy Advanced Modeling and Simulation Office (NEAMS), Fuels and Verification and Validation Program Elements. Authors are grateful to Dr. Keith Bradley, NEAMS National Director, Dr. Jim Peltz, NEAMS Program Manager, and Dr. Robert Versluis, NEAMS Program Manager, Mr. Alex Larzelere, NEAMS Office Director, and Dr. Ralph Nelson, NEAMS VU
To understand the NRC requirements concerning reactor fuel it must be understood that the reactor fuel is not separate from the reactor system. Said another way, the fuel and the reactor are licensed together and cannot be evaluated without input from each other. The reactor system must be analyzed for potential adverse operating and accident conditions. These conditions will then set the limits for the fuel system to be evaluated.

The NRC has published a white paper that outlines the basic requirements from the Code of Federal Regulations that apply to the licensing of reactor fuel (Clifford, 2005). The main requirements are derived from the General Design Criteria (GDC) given in 10CFR50 Appendix A. Appendix A of the paper by McClure et al. (2009) lists the appropriate GDCs.

From the NRC white paper the requirements provide “assurance – that (a) the fuel system is not damaged as a result of normal operation and Accident Operational Occurrences, (b) fuel system damage is never so severe as to prevent control rod insertion when it is required, (c) the number of fuel rod failures is not underestimated for postulated accidents, and (d) coolability is always maintained.” This statement summarizes the goals of fuel licensing, which is to analyze the reactor system and reactor fuel and provide the NRC assurance that the fuel will remain undamaged during normal operations and have predictable or bounded behavior during accident conditions.

The analysis of reactor fuel is typically done with an analytic model that has been validated from an extensive data set. Again quoting the NRC whitepaper: “The process for all new fuel licensing in the United States requires three basic steps; (1) approval of the analytic model, (2) approval of the analytic method, and (3) verifying that the licensee has shown that the analytic model and method are applicable to its new fuel design and all conditions and limitations are met. It is possible to license a new fuel design without research reactor data. The NRC believes that irradiated fuel data is required, however, to show how the analytic model and method are applicable to the new fuel design and how all conditions and limitations are met.”

The use of an analytic model (modeling and simulation) as means of demonstrating that a design can meet NRC requirements is not new and has precedence in the NRC. The NRC developed a methodology called CSAU (Code Scaling and Analysis of Uncertainty) with the goal of demonstrating compliance with fuel cladding temperature criteria. The NRC has provided guidance on the use of modeling for demonstrating compliance, and that guidance appears in two Regulatory Guides: (1) Reg-Guide 1.157 (NRC, Reg. Guide 1.157,1989) and (2) Reg-Guide 1.203 (NRC, Reg. Guide 1.203, 2005). Reg-Guide 1.157 is guidance focused on a specific type of LWR accident while Reg-Guide 1.203 is a more-generalized approach for using modeling and simulation.


As a means of demonstrating the use of Regulatory Guide (Reg-Guide) 1.1.57, the NRC sponsored a study by the national laboratories to test the use of “best estimate tools.” This study, “Quantifying Reactor Safety Margins, Application of Code Scaling, Applicability, and Uncertainty Evaluation Methodology to a Large-Break, Loss-of-Coolant Accident” (Boyack et al., 1994), used the NRC sponsored thermal-hydraulics code TRAC (Liles et al., 1986) as the “best estimate” tool for examining cladding temperatures during a Loss-of-Coolant Accident.

NUREG/CR-5249 summarizes the CSAU methodology very succinctly: “The CSAU evaluation methodology emphasizes a practical engineering approach that can be used to quantify code uncertainty.” The CSAU procedure can be most easily conceptualized as consisting of 14 primary steps, which can be grouped into three key elements (with the steps summarized below each element):

1. Requirements and capabilities – in which scenario modeling requirements are identified and compared against code capabilities to determine the code’s applicability to the particular scenario and to identify potential limitations. Basic steps are as follows:
   1. Specify scenario.
   2. Select nuclear power plant.
   3. Identify and rank phenomena (the PIRT, phenomena identification and ranking table process).
   4. Select frozen code.
   5. Provide complete documentation.
   6. Determine code capability.
2. Assessment and ranging of parameters – in which code capabilities to calculate processes important to the scenario are assessed against experimental data to determine code accuracy and scale-up capability and to specify ranges of parameter variations needed for sensitivity studies. Basic steps are as follows:
   1. Establish assessment matrix.
   2. Assess parameters – compare calculations to separate effects test and compare code to integral effects tests; this may change the nodalization of the model.
   3. Determine code and experiment accuracy.
   4. Determine the effects of scale.
3. Sensitivity and uncertainty analysis – in which the effects of individual contributors to total uncertainty are obtained and the propagation of uncertainty through the transient is properly determined.
   1. Determine the effect of reactor input parameters and state.
   2. Perform sensitivity calculations.
   3. Combine biases and uncertainties.
   4. Use total uncertainty to calculate specific scenarios for a specific plant.

Reg-Guide 1.157 was seen as a major step in the use of modeling and simulation as a means of meeting NRC requirements without resorting to excessive conservatism. Several reactor vendors have used a variation of the methodology to meet the ECCS requirements.

The importance of Reg-Guide 1.157 in the licensing of an advanced reactor fuel can be summarized as follows:

1. The method (CSAU) provides a precedent for future use of BEPU.
2. The method has been successfully applied.
3. The NRC has experience with the method and the concept of using modeling and simulation to fulfill licensing requirements.
4. The method developed vital techniques that will be essential for developing a framework for an advanced reactor fuel.
1. Determine requirements for the evaluation model.
2. Develop an assessment base consistent with the determined requirements.
3. Develop the evaluation model.
4. Assess the adequacy of the evaluation model.
5. Follow an appropriate quality assurance protocol during the evaluation model development and assessment process.
6. Provide comprehensive, accurate, up-to-date documentation.

Reg-Guide 1.203 is an excellent starting point for developing a framework for using BEPU in the licensing of an advanced reactor fuel. When paired with the techniques from CSAU (Reg. Guide 1.157) a beginning for a framework for using modeling and simulation for an advanced reactor fuel is in place. However, there are some challenges to using this methodology. Some of these issues are discussed by McClure et al. (2009). Reg-Guide 1.203 has six basic principles that are to be used when developing an evaluation model.

1. Determine requirements for the evaluation model.
2. Develop an assessment base consistent with the determined requirements.
3. Develop the evaluation model.
4. Assess the adequacy of the evaluation model.
5. Follow an appropriate quality assurance protocol during the evaluation model development and assessment process.
6. Provide comprehensive, accurate, up-to-date documentation.

Reg-Guide 1.203 is an excellent starting point for developing a framework for using BEPU in the licensing of an advanced reactor fuel. When paired with the techniques from CSAU (Reg. Guide 1.157) a beginning for a framework for using modeling and simulation for an advanced reactor fuel is in place. However, there are some challenges to using this methodology. Some of these issues are discussed by McClure et al. (2009).

US NRC Reg-Guide 1.203 (Evaluation Model Development and Assessment Process, EMDAP) is a generic methodology for performing BEPU for generalized transient and accident analysis. And although it in general contains more steps than CSAU, it also lacks the specificity that comes from a methodology that was developed for a specific accident application (evaluation of Large Break Loss Of Coolant Accident (LOCA) for Emergency Core Cooling System (ECCS)). We have in general kept many of the specifics from CSAU (such as PIRT).

Aspects of both EMDAP and CSAU still rely on engineering judgment to satisfy requirements in the methodology (example: when is the data sufficient?). These issues are not necessarily deficiencies. Our goal in this paper was to introduce more robust and defensible techniques that improve upon the overall approach. It will make it more “complicated”; however, experience in licensing techniques also indicate that a method that has less reliance on engineering judgment, but is based in sound science and mathematics, makes the technique more appealing to potential users.

**Appendix B. Global sensitivity analysis**

We summarize an approach to variance-based global sensitivity analysis commonly used with complex computational models (Saltelli et al., 2000) and apply this method to the simulation model used to illustrate statistical calibration methodology in Appendix C. Let \( \eta(z) \) denote code output calculated at input \( z \). Here \( z \) contains all inputs being varied, including design and calibration parameters. We assume that uncertainty in \( z \) is described by the probability density function (PDF) \( g(z) \). Let \( I \) denote an index set of elements in \( z \), and \(-I\) the complement index set of elements in \( z \) not in \( I \). The conditional PDF \( g(z|z_{-I}) \) describes uncertainty in \( z_I \) given particular values for \( z_{-I} \).

It has been shown (Sobol’, 1993) that there is a unique decomposition:

\[
\eta(z_1, \ldots, z_d) = \eta_0 + \sum_{k=1}^{d} \eta_k(z_k) + \sum_{1<k<l} \eta_{kl}(z_k, z_l) + \ldots + \eta_{12...d}(z_1, \ldots, z_d) \tag{B1}
\]

in which

1. \( \eta_0 = \int \eta(z)g(z)dz = E[\eta(z)] \) is the expected value of the code output with respect to the assumed uncertainty in the inputs \( z \),
2. the remaining components have zero expected value, \( E[\eta_k(z_k)] = \int \eta_k(z_k)g(z_k)dz_k = 0 \), and
3. any two distinct components are orthogonal, \( E[\eta_k(z_k)\eta_I(z_I)] = \int \eta_k(z_k)\eta_I(z_I)g(z)dz = 0 \) for \( I \neq J \).

Explicit expressions for each component are available. For example, the main effect for code input \( k \) is given by:

\[
\eta_k(z_k) = E[\eta(z)|z_k] - \eta_0 = \int \eta(z|z_k)g(z_k)dz_k - \eta_0 \tag{B2}
\]

The two-factor interaction effect for code inputs \( k \) and \( l \) is given by

\[
\eta_{kl}(z_k, z_l) = E[\eta(z)|z_k, z_l] - \eta_k(z_k) - \eta_l(z_l) - \eta_0, \tag{B3}
\]

where \( E[\eta(z)|z_k, z_l] = \int \eta(z)g(z_k, z_l)dz_kdz_l \) in a similar fashion, higher order effects \( \eta_{IJ}(z_I) \) for \( l \) having more than two elements are constructed by computing the joint effect:

\[
E[\eta(z)|z_{-I}] = \int \eta(z)g(z_{-I})dz_{-I} \tag{B4}
\]

and then subtracting all the lower order effects formed from elements of \( I \).

With these definitions in hand, we turn our attention to the subject of primary interest, variance-based global sensitivity analysis. The total variance of code output \( \eta(z) \) with respect to uncertainty in \( z \) is defined in the usual way:

\[
V = \int \eta(z)^2g(z)dz - \eta_0^2 \tag{B5}
\]

while the variance of individual components \( \eta_I(z_I) \) in the Sobol’ decomposition are computed as follows:

\[
V_I = \int \eta_I(z_I)^2g(z_I)dz_I \tag{B6}
\]

The sensitivity index corresponding to the effect \( \eta_I(z_I) \) is defined to be the variance ratio \( S_I = V_I/V \). If the inputs are assumed to be independently distributed, then the total output variance decomposes into a sum of the component variances:

\[
V = \sum_{k=1}^{d} V_k + \sum_{1<k<l<d} V_{kl} + \ldots + V_{12...d}, \tag{B7}
\]

referred to as the analysis of variance (ANOVA) decomposition. The total effect of input \( k \) is defined as follows,

\[
V_{ik} = V_k + \sum_{i \neq k} V_{il} + \ldots + V_{12...d} \tag{B8}
\]

that is, the sum of input \( k \) main effect variance \( V_k \) and all interaction effect variances involving input \( k \). Calculation of total effect variances is simplified when the ANOVA decomposition holds.

Monte Carlo sampling can often be used to approximate the integrals involved in computing effect functions or variance components (Saltelli, 2002). This is particularly true when the inputs are independently distributed or expressions for the required conditional distributions are available analytically. Standard orthogonal array (Williams et al., 2006; Morris et al., 2008) designs can also be used to estimate these quantities based on performing an ANOVA on the resulting computations. For complex computational models, run sizes are constrained by time. Direct sampling methods will thus not be available and runs must be used to construct a fast surrogate for the complex code, from which effect functions and variance components can be estimated (Oakley and O’Hagan, 2004).

We demonstrate ANOVA-based global sensitivity analysis on the example in Appendix C. Two parameters were varied, drop height...
(x) and drag coefficient (\(\theta\)), each independently and uniformly on the unit interval. The model computes drop time as a function of these two parameters. The main effect indices \(S_x\) and \(S_\theta\) are 71.0% and 23.6%, and the total effect indices \(S_{x\theta}\) and \(S_{\theta x}\) are 76.4% and 29.0%. Therefore, 5.4% of the total variation in drop time is due to an interaction effect between \(x\) and \(\theta\). Fig. B1 presents the main effect functions of drop time. Note the nonlinearity in these functions, which would not be accounted for by direct model computations on a two-level design. For more complex models, accurate estimation of the desired effects requires balancing the potentially large size of the designs required for direct computation against the use of a surrogate model, which introduces uncertainty but generally requires far fewer runs.

Local sensitivity methods are more commonly applied to complex codes. They are based on assessing the impact on the output of parameter perturbations from a baseline value \(z^0\) through a Taylor approximation:

\[
\eta(z^0 + \Delta z) = \eta(z^0) + \sum_{i=1}^{d} \frac{\partial \eta(z^0)}{\partial z_i} \Delta z_i + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2 \eta(z^0)}{\partial z_i \partial z_j} \Delta z_i \Delta z_j + \ldots
\]

(B9)

The partial derivatives \(\frac{\partial \eta(z^0)}{\partial z_i}\) are called first-order local sensitivities, \(\frac{\partial^2 \eta(z^0)}{\partial z_i \partial z_j}\) are called second-order local sensitivities, and so on. Local sensitivities are normalized to be independent of the units in which the inputs or outputs are reported, allowing local input effects to be compared. Specifically, the quantity:

\[
\left. \frac{z^0_i}{\eta(z^0)} \frac{\partial \eta(z)}{\partial z_i} \right|_{z^0}
\]

represents the relative change in code output at \(z^0\) for a 1% change in the \(i\)th output \(z_i\). Several methods are available for estimating or calculating the necessary partial derivatives, including finite-difference approximations, direct, and adjoint methods.

Assuming the inputs are uncorrelated, each with variance \(\sigma_i^2\), to first order the variance of the code output is given by:

\[
V = \sum_{i=1}^{d} \left[ \frac{\partial \eta(z)}{\partial z_i} \right]_0^2 \sigma_i^2
\]

(B11)

and therefore another measure of first-order local sensitivity of the \(i\)th input based on an approximate variance decomposition is given by:

\[
\left[ \frac{\partial \eta(z)}{\partial z_i} \right]_0^2 \frac{\sigma_i^2}{\sigma^2}
\]

(B12)

This calculation can, in principle, be extended to higher orders. In the drop time example above, we take the baseline value of the parameters to be the midpoint of the input region (0.5, 0.5). Assuming uniform distributions for \(x\) and \(\theta\), the approximate variance decomposition leads to first-order local sensitivities of 73.9% and 26.1% for \(x\) and \(\theta\).

As we are primarily interested in response variation in some volume of input parameter space, defined generally by either prior uncertainty in the inputs or probabilistic calibration of these inputs to experimental data, we emphasize the global sensitivity methods that decompose response variation with respect to input uncertainties. Local sensitivity methods will often not be sufficiently accurate due to output nonlinearities in the volume of input space of interest, although they often provide a useful order of magnitude approximation for assessing relative input importance, as seen in the drop time example above.

### Appendix C. Calibration and prediction

This appendix provides motivation for a statistical approach to calibrate a physics simulation model to experimental data, and to predict a future experiment given information from calculations and data already in hand (Williams and Higdon, 2007). These notions will be made more specific as this discussion proceeds. The statistical framework for the calibration and prediction methodology outlined below is Bayesian (Higdon et al., 2004; Kennedy and O’Hagan, 2001; Williams et al., 2006). To begin, assume that experimental data \(y\), as a function of system inputs \(x\), follows a basic measurement error model:

\[
y(x) = \zeta(x) + \epsilon(x)
\]

(C1)

Here, \(y(x)\) represents the data actually observed, \(\zeta(x)\) represents the true, unknown response of the physical system, and \(\epsilon(x)\) represents measurement error. The system inputs \(x\) are known to the experimentalist with certainty. They represent such quantities as physical constants (e.g., ambient temperature, nominal density) or variables specific to the geometry of the physical system under study, such as spatial coordinates or time. The concepts involved in what follows will be illustrated for the simple physical setting shown in Fig. C1.

The observed datum \(y(x)\) is a measurement of the time required for an object to impact the ground after being dropped at rest from height \(x\). In this setting there is one system input (height), but in general multiple system inputs are allowed. The system response \(\zeta(x)\) represents the true, underlying physical relationship between height \(x\) and impact time. This function is generally not observable, due to measurement error and other components of uncertainty inherent in collecting data, represented broadly here by the error term \(\epsilon(x)\). The function \(\zeta(x)\) is an object of statistical inference.

For complicated physical systems, actual experimental data may only exist for a very small number of \(x\) values. Replicated data may not be available. In this situation, inference about complex functions \(\zeta(x)\) is difficult or impossible if it based on the observed data alone. For this reason, intricate simulation models that incorporate advanced physics are employed with the goal of effectively modeling \(\zeta(x)\). These simulation models produce output as a function of the system inputs \(x\), but they often include additional parameters \(t\) that are commonly used in practice to tune the simulation model to experimental data. These additional parameters are sometimes referred to as knobs. They may represent actual physical concepts in the mathematical description underlying the simu-
The time taken for an object to drop from each of six floors of a tower (solid lines) is recorded. Predictions of unobserved times are desired for drops from higher floors (dashed lines).

Fig. C1. The time taken for an object to drop from each of six floors of a tower (solid lines) is recorded. Predictions of unobserved times are desired for drops from higher floors (dashed lines).

In many applications, the simulation model does not contain all of the relevant physics to explain the physical system being represented. This can be due to lack of knowledge about applicable physics for certain components of the system or inability to implement advanced physics due to computational complexity. A corollary to this situation is a simulation model that contains all of the (known) relevant physics but for which simplifications are required (such as collapsing over dimensions or choice of mesh) for computational tractability. Many applications involve inadequate physics and computational simplification. In this reality, a model for the physical system can be specified as follows:

\[ \zeta(x) = \eta (x, \theta) + \delta(x) \]  \hspace{1cm} (C5)

Here, \( \delta(x) \) is defined as the difference \( \zeta(x) - \eta (x, \theta) \) between the true physical system response \( \zeta(x) \) and the simulation model evaluated at the best value for the parameters \( \theta \), \( \eta (x, \theta) \). This term is often referred to as discrepancy or model inadequacy. Non-zero values for \( \delta(x) \) indicate that for system inputs \( x \), the simulation model evaluated at optimal settings for the parameters \( \theta \) is an incomplete representation of the physical system. For example, suppose the simulation model solves the differential equation \( \ddot{s}(\tau) = -1 \), subject to the initial conditions \( s(0) = x \) and \( s'(0) = 0 \). The impact time \( s(x,0) \) is the root of the equation \( s'(\tau) = 0 \). This model is appropriate for releasing objects in a vacuum but inadequate under any circumstances where air resistance affects the velocity of the falling object. If the latter applies, \( \delta(x) \) will be an increasing function of \( x \) as the effect of drag on impact time will increase with drop height.

Fig. C2 illustrates the concept of model inadequacy for the situation where the simulator produces output under the assumption of a vacuum (\( \theta = 0 \)) while data has been generated under the assumption of a positive drag coefficient (\( \theta = 0.5 \)) with a small amount of measurement error. In the notation of the previous development, the physical system response is \( \zeta(x) = \eta(x,0.5) \) and the discrepancy function is \( \delta(x) = \eta(x,0.5) - \eta(x,0) \), that is, the difference between the system response and the simulation model \( \eta(x,0) \). In this example, the simulation model is so simple that there are no calibration parameters. All statistical modeling of these data is concentrated on the discrepancy function.

It is important to note that statistical modeling does a reasonable job of predicting the system response within the range of available data, but does quite poorly when extrapolating to drop heights beyond the upper range of the data, as indicated by the rapidly increasing spread in the quantiles of the predicted system response. This highlights the difficulty of using extrapolations with an inadequate simulator. This behavior is a consequence of the fact that statistical modeling of discrepancy in this example does not incorporate any knowledge of physics to assist predictive capability in data-poor regimes. Statistical discrepancy modeling can be used to identify the possible presence of inadequate physics in simulation models under the following circumstances:

- Non-zero discrepancy predictions in the range of the data.
- Poor extrapolation of system response predictions.

Of course, absence of these indicators does not imply that the simulation model contains completely adequate physics.

In many applications, there will be no single value of the parameter \( \theta \) that best matches the data. Many values of \( \theta \) will give similar results. Multiple experimental datasets will generally be associated with different best estimates of \( \theta \). Changing experimental conditions could influence the best \( \theta \), such as the degree of laminar or turbulent flow on the best drag coefficient. Different values of \( \theta \)
may be appropriate across changing sets of conditions in dynamic simulations; for example, best material strength parameters varying as a function of strain rate and temperature. For these reasons, the statistical approach to inference about $\theta$ assumes that initial uncertainty about the value of $\theta$ is represented as a probability distribution. This prior assessment of $\theta$ is updated using experimental data and simulation runs to yield a refined assessment of uncertainty in $\theta$. This process is referred to as calibration of the simulation model to experimental data. It is distinct from tuning in that the end product is a probability distribution for $\theta$ while tuning produces a single best estimate for $\theta$. The calibration approach to gaining information about $\theta$ recognizes that typically there will be residual uncertainty in knowledge about $\theta$ even after updating this knowledge with data and calculations. This remaining uncertainty should be accounted for in any follow-on analysis (such as prediction) involving these parameters.

Summarizing, given initial (prior) information about the parameters, the calibration process results in a final (posterior) distribution for these parameters. Therefore, a unique best value of the parameters does not result from probabilistic calibration and would only be obtained in the limit of a perfect physics model and “infinite” data. However, the highest posterior density value is associated with at least one parameter setting, which could be regarded as an estimate of the best parameter value (as could the posterior mean or median). There will often be a single mode arising from any given analysis, as long as the physics models are well-behaved (e.g., sensitive to and monotonic in each parameter being varied) and the prior distribution is well-constrained (e.g., SETs provide useful information). Of course the mode is subject to change as the physics model(s) or available data changes. The priors given to the methodology may influence the calibrated posteriors. We pay attention to this issue and generally use information given to us by model developers (expert judgment). The methodology may also consider correlation structure between parameters through a specified or random covariance matrix. Therefore correlated parameters are not a limitation for the methodology.

A common inferential problem at this stage is to predict the value of experimental data at system inputs $x$, which are generally composed of input conditions at which no experiment has previously been conducted. That is, the goal is to provide a best estimate with uncertainty bounds for the value of data that would be generated by a future experiment. Several prediction problems are of interest: prediction of physical system response $\zeta(x)$ using the calibrated simulator, where the refined uncertainty in the knob parameters from the calibration analysis is essentially propagated through the simulator to predict $\zeta(x)$; prediction of discrepancy $\delta(x)$, where differences between data and simulation model are examined; prediction of $\zeta(x)$ using the discrepancy-adjusted calibrated simulator; and prediction of experimental data $y(x)$ based on the entire model, which incorporates uncertainties from measurement error in addition to the above sources (calibrated simulation and discrepancy). Predictions of discrepancy can be used as a diagnostic in quantitative model validation (Wang et al., 2009).

Fig. C3 shows results obtained by enhancing the physics of the drop time simulation model for vacuum environments of Fig. C2 to incorporate air resistance. The data set shown in Fig. C2 is used to calibrate this improved simulator. Initial uncertainty in the drag coefficient $\theta$ was assumed to be uniform on the unit interval [0,1]. The experimental data constrains this uncertainty substantially so that all probability in the calibrated $\theta$ is tightly concentrated around 0.5, the value used in generating the data for this example. It is also evident that the reduction in uncertainty of $\theta$ due to calibration leads to much more informative (less variable) predictions of the system response. The discrepancy is probabilistically zero in this analysis, so that the calibrated simulator is an effective predictor of the system response. It is evident that extrapolation of the calibrated simulator beyond the upper range of drop heights in the observed data leads to well-behaved predictions of the system response. This is to be expected when the simulation model incorporates all or most of the relevant physics.
Fig. C3. The left panel compares the initial uncertainty distribution of the drag coefficient $\eta$ (gray; 0.05, 0.1 (0.1) 0.9, 0.95 quantiles indicated) with the calibrated distribution (black; 0.05, 0.5 and 0.95 quantiles indicated). The right panel shows simulation model output for $\eta$ set to the indicated quantiles of its initial distribution (gray) and its calibrated distribution (black).

We conclude this appendix by noting that the calibration and prediction methodology described above has been extended to functional and multivariate computer model (and experimental data) outputs (Higdon et al., 2008). Furthermore, this framework has been applied to settings in which multiple data sources are used to simultaneously calibrate (subsets of) the parameters $\eta$ (Habib et al., 2007).

Appendix D. Predictive maturity of VPSC plasticity model through batch sequential calibration

This appendix illustrates the application of the predictive maturity concept to a practical, non-trivial problem: creep of core reactor clad and duct components subjected to in-service conditions of irradiation, stress, and thermal cycling. Section D.1 overviews the derivation of the PMI. Section D.2 introduces the physics model of interest and defines the input and output parameters. Section D.3 presents the use of a batch sequential design approach to reach desired predictive maturity with the least possible number of physical experiments.

D.1. Predictive maturity index

The PMI metric quantifies the degree of predictive capability of a simulation model by integrating three attributes of the model development, experimentation, and calibration processes: discrepancy, coverage, and complexity. The manner in which discrepancy is treated in the context of predictive maturity is fundamentally different than the commonly implemented goodness-of-fit metrics. Discrepancy represents the disability of a simulation model to

Fig. D1. Coverage and domain of applicability of experiments.
reproduce reality due to incomplete or inaccurate physics. Discrepancy is determined, not for a single experimental measurement, but for the entire domain of applicability. If the discrepancy of a simulation model increases, the predictive capability of the model is compromised and thus predictive maturity decreases. The second attribute, coverage, refers to the extent to which experiments explore the domain of applicability. One approach for quantifying coverage of the domain is based on the convex hull—that is, the smallest convex domain within which all physical experiments fit. Here, coverage can be calculated as the ratio of the convex hull of the physical experiments to the domain of applicability. If the coverage of the domain of applicability is insufficient, the predictive capability of the model is limited, and thus the predictive maturity decreases. Complexity, the third attribute, defines the sophistication of the physics of the model. A natural outcome of increasing sophistication of a model is the increased number of knobs that are the imprecise model parameters considered as candidates for calibration (fine-tuning). A model with an infinite number of knobs can predict virtually any physics measurements with practically no predictive maturity. Therefore, as the complexity of the model increases, everything else being equal, the predictive maturity decreases. These three attributes are combined to obtain a quantitative metric for the predictive maturity. The basic formulation of the PMI index is proposed in Hemez et al. (2010):

\[
PMI(\eta_C, N_K, \delta_S) = \eta_C \left( \frac{N_K}{N_E} \right)^{\gamma_1} (1 - \delta_S)^2 e^{(1 - \eta^2)\gamma_3 - \delta_S^2} \tag{D1}
\]

where \(\gamma_1\), \(\gamma_2\), and \(\gamma_3\) are weight coefficients controlling the effect of the three attributes of coverage, complexity, and discrepancy (quantified by \(\eta_C\), \(N_K\), and \(\delta_S\), respectively) and provide a user-defined adjustment over the calculation of PMI metrics.

D.2. Visco plastic self consistent material model

This section introduces the input and output parameters of the VPSC plasticity model specifically for predicting the creep strain rate in Face Centre Cubic (FCC) steel (Lebensohn et al., 2010). VPSC is a multipurpose numerical model, based on linearized homogenization of polycrystal plasticity. Numerous research groups worldwide presently use VPSC as a predictive tool for parameter identification, interpretation of experimental results, and multiscale calculations. VPSC is a meso-scale code for modeling creep of core reactor clad and duct components subjected to service conditions of irradiation, stress, and thermal cycling. Because enhancing performance of core reactor cladding and duct alloy components is fundamental for increasing fuel burn-up, the selected fuel-deployment problem is not only relevant to fuel licensing but also very timely for meeting the objectives when responding to the field of nuclear energy’s needs.

The VPSC code delivers the rate of the creep strain for given values of climb dislocation orientation and deviatoric stress input, which constitute the control parameters. Climb dislocation orientation loosely defines the importance of climb phenomena in the crystallographic thermal creep, and varies between 0 and 90° (0 and 1.57 rad). For an angle of 0°, climb dislocation remains inactive; however, for an angle of 90°, climb dislocation is fully activated. Deviatoric stress input is the stress input to the specimen to induce the creep strain. Herein, the lower and upper bounds for the climb dislocation orientation angle, [010, 0.6 rad], and for the deviatoric stress input, [900, 1100 MPa], are selected.

The three uncertain parameters of the VPSC code (knobs) are the rate sensitivity of glide, rate sensitivity of climb and the ratio of threshold stress for glide and climb. The maximum and minimum limits for these three imprecise parameters are determined according to expert judgment (Table D1).

<table>
<thead>
<tr>
<th>Table D1</th>
<th>Comparison of prior and posterior knowledge.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration parameters</td>
<td>Min value</td>
</tr>
<tr>
<td>Rate sensitivity exponent for glide</td>
<td>2</td>
</tr>
<tr>
<td>The ratio of threshold stress for glide and for climb</td>
<td>8000</td>
</tr>
<tr>
<td>Rate sensitivity exponent for climb</td>
<td>2</td>
</tr>
</tbody>
</table>

D.3. Batch sequential design (BSD)

Depending upon the given physical experiments, BSD selects the optimum settings where the next set of physical experiments must be conducted according to the given criteria. This appendix presents the preliminary results obtained through the batch sequential calibration techniques with the Expected Improvement for Global Fit (EIGF) criterion (Williams et al., 2010).

The batch sequential algorithm starts with an initial set of physical experiments. Herein, the control parameter settings for this initial set (i.e., the distribution of the experiments within the domain) are specified by a Latin hypercube design. The initial set of experiments is obtained by executing the “truth” function at three control parameter settings for the hypothetical true values of the calibration parameters (Table D1). The BSD augmentation is repeated six times with two simulated experiments in each batch. With the addition of six batches and 12 additional simulated experiments, batch sequential calibration reaches a total of 15 experiments.

A representative plot showing the location of the physical experiments in the domain of applicability is given in Fig. D1. Fig. D1 also presents the increase in the coverage of the domain of applicability. New batches of experiments are selected through BSD. Blue squares indicate the location of all 15 experiments, which corresponds to the final stage (Batch 6). In the initial stage (Batch 0), the coverage is only 5%. After the EIGF criterion selects a total of six batches, the coverage increases up to 80% as indicated by the solid red line in Fig. D1. The procedure is repeated five times. The normalized discrepancy remains below 1% for all repeats.

While constructing the PMI charts, \(\gamma_1\), \(\gamma_2\), and \(\gamma_3\) values are taken as 0.5, 0.25, and 2, respectively, to provide equal weight to all three attributes (Lebensohn et al., 2010). The PMI values are obtained when the reference number of knobs \(N_K\) and number of knobs \(N_S\) are taken as 3. In Fig. D2, PMI stabilizes as early as the second batch and converges to a value of approximately 90%. In a practical sense, if the EIGF criterions is used to select the experi-
mental settings, seven physical experiments are sufficient to reach a VPSC model with approximately 90% predictive maturity.

References