Uncertainties in Predictions of Material Performance using Experimental Data that is Only Distantly Related to the System of Interest

Wayne E. King[1], Athanasios Arsenlis[1], and Charles Tong[2], and William L. Oberkampf[3]

[1] Physical and Life Sciences Directorate, [2] Computations Directorate
Lawrence Livermore National Laboratory
Livermore, CA, USA
{weking,arsenlis,tong10}@llnl.gov
[3] Consulting Engineer
Georgetown, TX, USA
wloconsulting@gmail.com

Abstract. There is a need for predictive material "aging" models in the nuclear energy industry, where applications include life extension of existing reactors, the development of high burnup fuels, and dry cask storage of used nuclear fuel. These problems require extrapolating from the validation domain, where there is available experimental data, to the application domain, where there is little or no experimental data. The need for predictive material aging models will drive the need for associated assessments of the uncertainties in the predictions. Methods to quantify uncertainties in model predictions, using experimental data that is only distantly related to the application domain, are discussed in this paper.

Keywords: uncertainty quantification, model form uncertainty, model uncertainty, nuclear energy, neutron damage, ion damage, irradiation effects scaling, extrapolation

1 Introduction

There is a growing need to make predictions of material performance in extreme environments and over very long periods of time where there is little or no experimental data. This is particularly the case in the prediction of the effects of "aging" on material performance where desired material lifetimes can exceed by a large margin what is practical for validation under normal application conditions. In large, complex engineering systems, the costs are often too high and/or the times too long to carry out desired validation experiments under actual operating conditions. In the case of aging, the required extrapolations can be orders of magnitude beyond the validation domain.

Further, model development can require reliance on accelerated experiments. Experiments can be accelerated by changing temperature to take advantage of

Arrhenius behavior or the rate of application of the experimental forcing function can be increased. Results from accelerated experiments, which are also outside of the application domain, require extrapolation, perhaps over orders of magnitude in rate, to the actual operating conditions, by way of a model.

The need for predictive models is particularly acute in the nuclear energy industry, where applications include (i) the desire for life extension of existing reactors to 80 years, (ii) the development of high burnup fuels, and (iii) the imperative for dry cask storage possibly to hundreds of years.

The nuclear materials community has adopted an experimental strategy involving experiments that are not conducted in commercial power plants. The nuclear industry and regulators require a "sound and defensible case for the relevance of these techniques to actual service conditions" [1]. This implies the ability (i) to make predictions across diverse irradiation energy spectra, irradiation rates, and irradiating particles including thermal neutrons, fast neutrons, and energetic ions, (ii) to scale the prediction to the relevant reactor conditions, and (iii) then to extrapolate to the application domain of interest.

Predictions are most useful in the presence of quantified uncertainties. In an engineered system, high uncertainty can lead to excessive conservatism and thus necessarily increased margins, which can adversely affect cost, schedule, and system performance. Thus, in addition to the predictive model, estimates of uncertainties in predictions are also required. Uncertainty quantification (UQ) provides a framework within which uncertainties for predictions can be estimated. The case that we are interested in here is uncertainties due to prediction in domains where there is little or no experimental data. This is generally referred to as model extrapolation.

Extrapolations using physics-based models differ from extrapolations using regression curve fits of the system responses quantities of interest. When making an assessment of uncertainties, there are several sources that must be considered, including model inputs and numerical solution approximation. When making an extrapolation using a model, the assumptions associated with the mathematical model itself result in a source of uncertainty usually referred to as model form uncertainty[1], and it must also be considered [2,3]. The inclusion of model form uncertainty represents a specialized field within UQ [3]. Approaches to dealing with uncertainties in model inputs are well established and can be implemented. The approach to uncertainties associated with model form is less well established, particularly for cases where models are assessed in a validation domain that does not fully overlap or overlap at all with the intended application domain.

In this paper, we discuss a predictive modeling problem in the nuclear energy area that requires large extrapolation. We also review possible approaches to the extrapolation problem. Throughout, we focus on the special case of predictions of models validated with experimental data that are only distantly related to the system of interest.

[1] Model form uncertainty is referred to as model uncertainty, model bias, or structural uncertainty.

2 Background

2.1 Accelerated experimentation for nuclear energy materials applications [4]

The nuclear industry needs models that predict the time dependence of microstructural and fission product evolution in structural materials and fuels. The most challenging extreme environment to study is that of high irradiation dose. Models developed to address this extreme are difficult to validate because of the inability to reach these doses using existing neutron-irradiation facilities in reasonable amounts of time and at modest costs. Furthermore, reactor facilities are problematic experimental venues for combining the various aspects of the extreme environments into a quantitative in situ study of material behavior.

Understanding radiation damage using ion irradiation is not a new idea. It has a long history of significant contributions spanning several decades. In fact, much of our understanding of material behavior under irradiation comes from well-controlled ion-irradiation experiments.

However, a key challenge is the scaling, or extension, of ion irradiation experiments and data to actual in-service conditions. Fig. 1 illustrates the particular case of scaling and extrapolation for the damage rate parameter. Scaling refers to use of models to bridge two unconnected validation domains. Extrapolation refers to the use of models to project into an application domain where there is no experimental data. The plot shows the range of damage expected for advanced reactors, GEN IV reactors, GEN III reactors, and light water reactors (LWRs). The plot further shows the damage levels that could be obtained in a 5-year irradiation experiment in a number of test reactors and the damage levels that could be obtained by ion irradiation in 5 days [5]. Consequently, a scientifically defensible argument for the applicability of models developed using accelerated experiments to neutron irradiation environments is critically needed. This should include rate scaling, effects of recoil energy spectra, and the ability to extrapolate to dose regimes not explored by neutrons. This irradiation-effects scaling is identified as a priority research direction in the Science for Energy Technology Workshop Report [1]. By definition, models developed for materials under neutron irradiation conditions at the extreme of high irradiation dose cannot be validated because little or no neutron irradiation data exists in that domain. Consequently, UQ will be particularly important in dose regimes that have not been explored using neutrons.

2.2 Other Relevant Science and Technology Application Areas

"Aging" of materials for nuclear energy applications is not the only area requiring scaling across and extrapolation outside of the validation domains. Fusion energy applications have a similar problem. In fusion machines, materials that can withstand very high levels of radiation damage are required. Without a fusion-relevant neutron source, the fusion materials community has adopted a research strategy similar to the fission community.

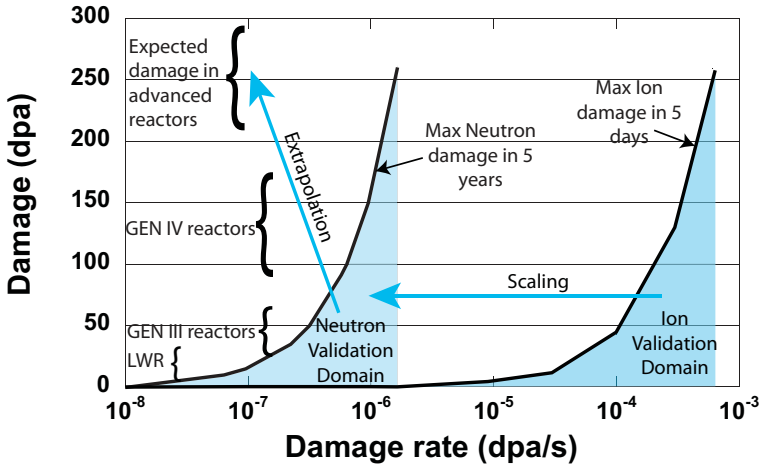


Fig. 1. Plot of radiation damage, measured in displacements per atom (dpa), as a function of the rate at which the damage is produced [5].

Uncertainty in the calculation of the depletion of nuclear fuels is composed of a number of individual components. Some of these include uncertainty in the cross section at a given neutron energy, uncertainty in core composition and other externally driven parameters such as power level and temperatures, and model approximations made to accommodate computer modeling capabilities.

These uncertainties will affect the prediction of the evolution of the fuel isotopic components in time. The isotopic component distribution at any given time represents the integration of the depletion conditions and uncertainties over all previous time. This effect can compound the effects of uncertainty or approximation. This compounding effect may also be limited by compensating effects. A common figure of merit for depletion is the energy extraction per mass of fuel, which is quantified as giga-watt-days per metric tonne of initial heavy metal (GWD/t). Present commercial fuel is depleted to about 60 GWD/t and the effects of uncertainties are well benchmarked for this range of operation. Advanced fuels may go beyond today's 60 GWd/t burnup. To operate outside the range of experience will require identifying the compounding effects for a given fuel type at a specific higher burnup so that they can be applied to the known uncertainties derived from inside the benchmarked region.

The effects of aging of materials are also of interest in both intermediate and long-term storage of nuclear waste. Analyses of the possible behavior of radioactive waste in a repository at Yucca Mountain (YM), Nevada, were conducted between 1982 and 2008. Early analyses (termed performance assessments or PAs) were for selecting the site and determining the feasibility of the disposal concept. Numerous parameter values for the numerical models were required and were mostly assigned by individual analysts and scientists for the early analyses. However, in 1987, Congress asked the YM Project (YMP) to evaluate the

viability of a repository at YM [6]. For this viability assessment (PA-VA), the YMP formed five panels to examine: (1) groundwater flow in the unsaturated zone, (2) groundwater flow and radionuclide transport in the saturated zone, (3) the near-field effects of heat on the region around the engineered barrier system, (4) waste form degradation, and (5) waste package (WP) degradation. These five panels assigned parameter values by aggregating disparate data available in the literature, prior to completion of project experiments, and estimated the uncertainty present as literature-based information was often for conditions and spatial and temporal scales that differed from those required for the PA-VA. The analysis underlying the license application for the Yucca Mountain repository in 2008 [7] considered a total of 392 uncertain analysis inputs (see Ref. [8], Table K3-3, for a complete listing of these inputs and additional sources of detailed information). The Waste Isolation Pilot Plant (WIPP), a repository for transuranic radioactive waste in southern New Mexico, also had similar needs [9,10,11].

Another closely related application is the Qualification Alternatives to the Sandia Pulsed Reactor (QASPR) project at Sandia National Laboratories [12,13]. In this case, pulsed ion beams are being used to understand radiation effects in semiconducting materials. The challenge is to extrapolate those results to the relevant pulsed neutron environment in the absence of a relevant neutron source.

There is a related field known as accelerated testing. Accelerated tests are used to obtain timely information on product-life or performance degradation over time [14]. Ref. [15] provides a comprehensive discussion of useful models and statistical methods for accelerated testing. Ideally, predictions from accelerated tests should be based on models of physics of failure. In practice, however, users of accelerated tests often use a combination of past experience and empirical fitting of data to statistical models. Although these procedures seem to have been adequate in the past, it is generally recognized that the path forward for large extrapolations must be based on physics-based modeling.

2.3 The Problem Recast in More General Terms

One of the goals of UQ is to provide a means to evaluate a model's predictive capability. Fig. 2(a) schematically illustrates the synergistic use of modeling, experiments, and UQ to make a prediction [3]. In this case, the system response quantity of interest is shown as a function of two system or environmental parameters. The application domain is highlighted in light brown, is the range of parameters #1 and #2 that are of interest in the application. The validation domain, highlighted in burgundy, is the range of parameters #1 and #2 where validation experimentation are carried out. The response surface is also shown with the application domain highlighted. In this example, the validation domain fully contains the application domain and predictions could be obtained through various types of interpolation over the validation domain. Likewise, uncertainties in predictions could be quantified by interpolation of uncertainties over the validation domain, or by direct calculation using the input uncertainties in the model.

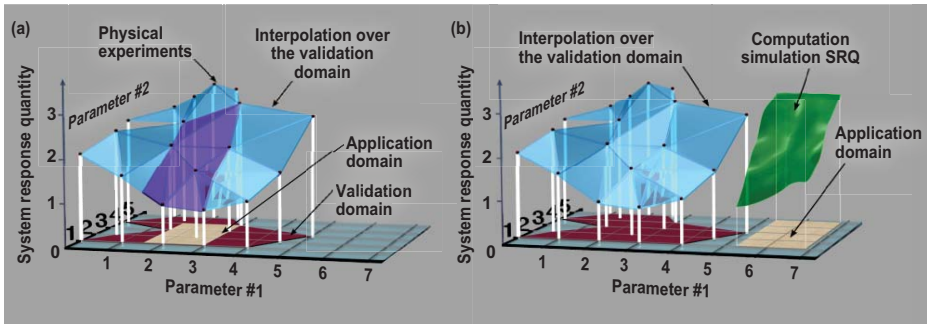


Fig. 2. Possible relations of the validation domain to the application domain. (a) Complete overlap and (b) no overlap [3].

One of the main reasons why we rely on modeling and simulation is to make predictions in domains where there is little or no experimental or observational data. Fig. 2(b) shows the situation where the application domain is no longer within the validation domain and there is no overlap of the application domain with the validation domain. While there is a clear path to quantify uncertainties in a prediction when the application domain is inside the validation domain, the approach to quantifying uncertainties for the case shown in Fig. 2(b) is less clear.

Although Fig. 2 is illustrative, it fails to capture the complexity inherent in today's multiphysics simulations. Another view of the extrapolation issue is given in Fig. 3, which illustrates the validation hierarchy for a complex engineering system model [16]. At each level of the hierarchy, the problem is broken down into smaller and smaller pieces until it is reduced to the unit problem (or unit mechanism) level. This is a more physics-based, or system-based, perspective than in Fig. 2. Each box at the unit problem level contains individual physics models that can be validated using targeted experiments that may not be in the same conditions as would apply to the next higher level in the hierarchy. At each higher level, the individual effects are brought together forming coupled physics and coupled subsystems and systems interactions.

In the case of small extrapolations, we would expect this physics-based approach to enable extrapolation of uncertainties, including model form uncertainty, to the application domain. However, for large extrapolations, it is possible that new physics could appear at higher levels, or that unexpected coupling could emerge at the higher levels. These possibilities are present in the nuclear materials aging problem, where (i) scaling is required across diverse irradiation energy spectra, irradiation rates, and irradiating particles and (ii) large extrapolation is required to the application domain. Extrapolation of uncertainty includes both (a) extrapolation to the application domains where there is little or no experimental data, Fig. 2, and (b) effects present in higher levels of system complexity in the validation hierarchy, Fig. 3, that are not anticipated in the modeling [17].

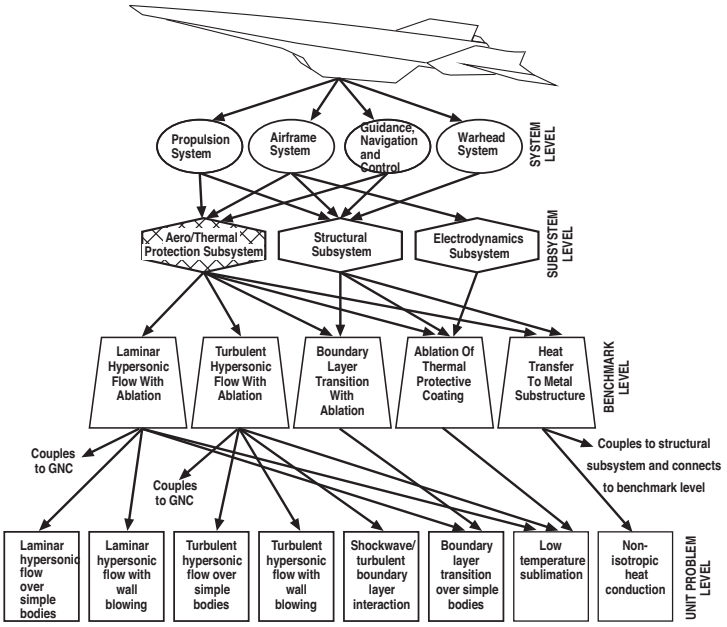


Fig. 3. A hierarchical validation structure for the hypersonic cruise missile (from [3]).

2.4 Prediction-Coupling of Accelerated Experiments with Physics Models [4]

It takes a long time to develop a new material or investigate the properties of materials that undergo low-dose-rate irradiation, such as the pressure vessel or core internals including the fuel and cladding. Experiments using neutron irradiation can take up to 7 years, including the irradiation time, the radioactive cool-down time, and the post-irradiation examination. Incorporating the effects of high temperature, stresses and a corrosive environment along with irradiation make the problem multidimensional and extremely complicated. Translating that into a program to satisfy a regulatory requirement for a new material or new fuel design can lead to a multi-decadal process. Such a timescale is unacceptable to efficient progress, and yet, it is the present-day norm.

One pathway to accelerate this process is to carry out accelerated experiments either inside the reactor core (in the case of the low-dose-rate regimes for the pressure vessel) or using external radiation sources such as ion beams (in the case of core internals that would see high neutron doses over their lifetimes). Using ion beams, one can investigate a large parameter space (in terms of external forcing functions) for irradiation effects on microstructure and macroscale properties. The phase space includes temperature, ion type, ion dose rate, ion energy, and total dose. It also includes the ability to apply in situ mechanical loading, chemical environments, and coolant fluids.

In some special cases, it is possible to create microstructures and material properties that are very similar to those that would be found in a particular nuclear reactor irradiation experiment. However, microstructures, mechanical properties, or other physical properties that deviate from neutron irradiations significantly, will also be observed. The challenge is to employ all of those observations to develop a science-based understanding of material degradation and performance, specifically, to establish a scientific basis for the key mechanisms of material performance.

The best approach to quantifying such scientific understanding is to build a model that captures all of the relevant physics, which is where modeling and simulation come into play. One seeks to understand the ion-beam forcing function and the material response to that forcing function. With that understanding, if a different boundary condition was applied (in terms of say temperature, ion type, or dose rate), it is reasonable to expect to be able to predict the material response. That is, one could have sufficient confidence that, with this robust model, interpolation and reproduction of an experimental result is possible.

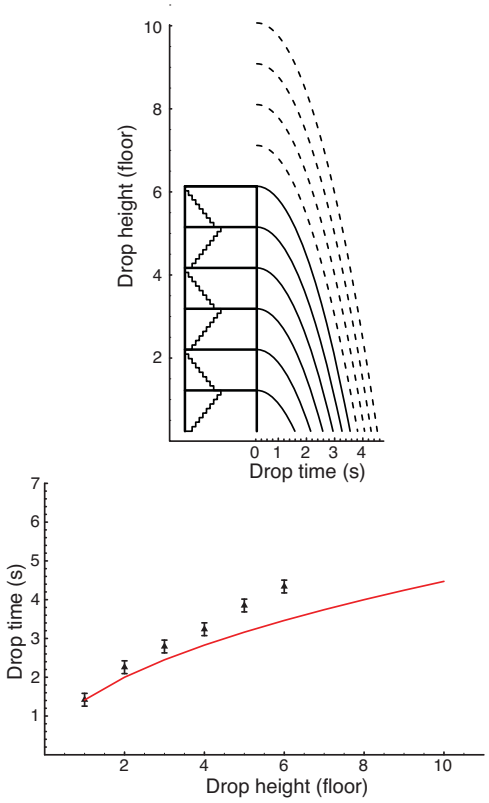
The question then becomes: How does one extrapolate, given a model, to high-dose neutron-irradiation environments? Compared with ion irradiations, neutron dpa rates are much, much lower ($\sim 10^2 - 10^3$ lower). In addition, physical mechanisms, such as transmutation and chemistry changes, occur simultaneously with the neutron bombardment and displacements in a material. With a robust model, the boundary conditions can be altered, while not perturbing any of the model internals, and an extrapolation can be made to project material performance to the neutron-irradiation environment. To make that extrapolation, researchers must also quantify the quality or accuracy of the extrapolation. This forward extrapolation and the qualification of the quality of the predicted extrapolation is where uncertainty quantification becomes important.

3 Approaches to Extrapolation of Uncertainty

3.1 Calibrate Model Parameters Over the Validation Domain and Ignore Model Form Uncertainty

This approach uses the physics model as it is (assuming no model form uncertainties) together with available experimental data for validation and calibration (given a number of model input parameters). The posterior distributions of the calibration parameters are then used in forward uncertainty propagation (through the computational model) to predict the extrapolated configuration and to estimate the corresponding uncertainties. To speed up calibration, response surface or surrogate modeling using Gaussian process, polynomial regression or polynomial chaos is often used. This approach has the inherent assumption that the computational model has captured all essential physics (except that there are uncertainties about some physics parameters that can be estimated using data) and the surrogate models are adequate for the extrapolated regime.

A simple example of this method is shown in Fig. 4, where a prediction and an associated uncertainty are required for the time required for an object to drop



Experimental data:
The time it takes an object to drop from each of 6 floors of a tower is recorded. There is an uncertainty in the measured drop times of about ±0.2 seconds. Predictions for times are desired for drops from floors 7 through 10, which do not exist yet.

Simulated drop times:
A (trivial) computational model is developed to predict the drop times as a function of height. The simulated drop times (red line) are systematically too low when compared with the experimental data (symbols). the error bars around the observed drop times show the observation uncertainty.

Fig. 4. A simple UQ example using experimental data and a computational model to predict drop times from new heights [18].

from yet unconstructed floors of a building [18]. Measurements are made of the time to drop the object from the first six floors and a model is constructed. Using the available experimental data from the drops, a Bayesian inference methodology is used to update/calibrate the uncertain input parameters in the model. Once the model is calibrated it is then used for prediction outside the calibration range. This is the most commonly used form of extrapolation: the model form uncertainty is assumed to be zero.

Continuing with the example in Fig. 4, when comparing the model against experimental data, systematic errors are observed which cannot be resolved satisfactorily by calibration. Model bias (or discrepancy) and its dependence on the drop height are evident. As shown in Fig. 5, an extrapolation would have led to a predicted time and uncertainty that, in fact, would not have predicted what would have been observed had an experiment been carried out. This illustrates that in this case, as in many others, the model form uncertainty can dominate the extrapolated uncertainties. The experiments revealed the presence of physics that was not accounted for in the original model.

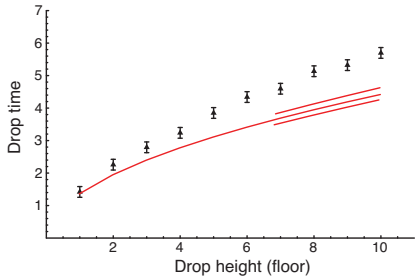


Fig. 5. The impact of ignoring model form uncertainty on the extrapolation and uncertainties [18].

3.2 Calibrate Parameters and Identify a Discrepancy Function to Characterize Model Inadequacy

This approach explicitly assumes a functional form for the discrepancy between the simulation and the actual physical process. The full method estimates from data simultaneously both the posteriors for the calibration parameters as well as the parameters in the discrepancy function. Some simplifications were suggested to reduce the complexity of the estimation process (see [19] for details). This approach assumes both the functional forms of the simulation model and the discrepancy stay the same in the extrapolated regime.

Kennedy and O'Hagan proposed a Bayesian approach that represents the model form uncertainty using a discrepancy function (in terms of some input parameters) to characterize model inadequacy [19]. This discrepancy function is created using experimental data as well as a selected regression or statistical emulator such as Gaussian process. Predictions are performed by incorporating the discrepancy function evaluated at the extrapolated points, in addition to the uncertainty due in posterior distributions of the calibrated parameters. This approach has the inherent assumption that the discrepancy function essentially captures the misrepresented physics.

In their approach, the system is modeled by

$$z = \zeta(\mathbf{x}) + e = \rho\eta(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x}) + e \tag{1}$$

where $\mathbf{x}$ is a vector of input parameters; $\boldsymbol{\theta}$ is a vector of calibration parameters; $\eta(\mathbf{x}, \boldsymbol{\theta})$ denotes the function for the simulation model; z is the observation; e is the observation error (independent normal distribution); $\zeta(\mathbf{x})$ is the true value of the process being modeled; ρ is an unknown regression parameter to be determined; and $\delta(\mathbf{x})$ is a function to describe model inadequacy, which is independent of $\eta(\mathbf{x}, \boldsymbol{\theta})$.

The full Bayesian calibration of this system is very complicated. Instead, Kennedy and O'Hagan proposed a multi-step approach:

– Build a Gaussian process model for $\eta(\mathbf{x}, \boldsymbol{\theta})$ based on sampling different values of $\mathbf{x}$ and $\boldsymbol{\theta}$ (that is, to estimate the "hyper-parameters" used to describe a Gaussian process model).

- Use data ($\{z\}$) to estimate the regression parameter ρ; the standard deviation of the observation error e; the hyper-parameters in the Gaussian process model of the model inadequacy function $\delta(\mathbf{x})$.
- Use the data ($\{z\}$) and the model ($z = \rho\eta(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x}) + e$) for calibration to get the posterior distribution of $\boldsymbol{\theta}$.
- Use the model ($z = \rho\eta(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x}) + e$) and the posteriors of $\boldsymbol{\theta}$ for prediction and uncertainty analysis.

Model inadequacy is defined as "the difference between the true mean value of the real world process and the code output at the true value of the inputs" [19]. There is debate across the field regarding the use of this definition of model inadequacy to describe the difference between a simulation and an experiment (see Sects. 2.2.3, 2.4, and Sects. 12.1-12.3 in [3].

This method is suitable if there is (a) sufficient experimental data from different input parameter configurations to characterize the discrepancy function, (b) reason to believe that the assumption concerning the discrepancy functional form will be valid in the extrapolation regime, and (c) that the discrepancy function is more significant than the observation error.

An alternative approach is where the candidate response surface methods for $\eta(\mathbf{x}, \boldsymbol{\theta})$ in Eq. (1) and the discrepancy functions are based on generalized polynomial chaos (or stochastic collocation) methods [20]. The advantages of polynomial chaos UQ methods are their efficiency and their utility for representing and propagating large uncertainties through complex models [21]. Both intrusive and non-intrusive applications of the polynomial chaos method are reviewed in [21].

This method is suitable if there is sufficient experimental evidence that the discrepancy function has polynomial form in the input parameters, and that this form will also be valid in the extrapolation regime. Moreover, another requirement is that the discrepancy function be more significant than the observation error.

3.3 Validation Metric Approach [2,3]

Oberkampf and Roy argue that model form uncertainty should be estimated as part of the process of model validation. They estimate the model form uncertainty in the validation domain using a validation metric which they define as "a mathematical operator that requires two inputs: the experimental measurements of the system response quantities of interest and the prediction of the system response quantities at the conditions used in the experimental measurements" [2]. Oberkampf and Roy in Sects. 13.2, 13.4, and 13.5 of [3] and in [22], describe two validation metrics: the confidence interval approach and the method of comparing cumulative distribution functions (CDFs) from the model and the experiment. In the confidence interval approach, they define the validation metric for model form uncertainty as the difference between the mean of model prediction and the estimated mean of the experimental data. In the CDF

method, the validation metric is defined as the area between the experimental and simulation CDFs.

Once a validation metric is estimated over the validation domain, the critical issue is how this error structure should be extrapolated to the application conditions of interest. One simple method for extrapolation is to construct a regression fit of the error structure over the validation domain using a low degree polynomial function [2,3,22]. The regression function is then evaluated at the application conditions, along with the statistical prediction interval at those conditions. The estimate of the model form uncertainty is increased by the prediction interval not only because of the imprecision of the regression function to fully represent the model form uncertainty, but also because of the random measurement uncertainty that is present in each experimental measurement. A level of statistical confidence is chosen for the prediction interval, say 90 or 95%, and the upper bound on the prediction interval is used as the estimate of the model form uncertainty at the application conditions of interest.

This estimated model form uncertainty is considered as an epistemic uncertainty, i.e., an uncertainty whose source is lack of knowledge as opposed to randomness, for the prediction of the system response quantities of interest at the application conditions. It has been found [2,3,22] that even if the model form uncertainty is relatively small over the validation domain, but the magnitude of the extrapolation is large in the multi-dimensional input space over which data are available, the estimate of model form uncertainty is typically quite large at the application conditions of interest. The model form uncertainty is clearly represented to the user of the simulation results, e.g., a designer or decision maker, as a probability-box, or p-box. The p-box is an interval-valued CDF, where the range of possible probabilities of the system response quantity reflects the epistemic uncertainty due to the model form.

3.4 Method of Alternate Plausible Models [3]

An approach for assessing uncertainty, both model form uncertainty and parametric uncertainty, is to compare predictions from alternative plausible models. This method is also referred to as the method of competing models. While simple in concept, it is not commonly used because of the time and expense of developing multiple models for a system. Examples of applications of this method include hurricane forecasting (Fig. 6) (e.g., see [23,24]), climate prediction, and long-term storage of nuclear waste.

The approach requires multiple models developed by independent groups. This approach does not actually provide an estimate of model form uncertainty; it only provides an indication of the similar or dissimilar nature of each model prediction. Because of the cost and time involved, this approach will likely be limited to application in matters of very high priority.

Since in many applications simulating the full physics is very consuming in terms of time and computational resources, a viable approach is to use simplified models for some of the physics components in the system (for example, see [25,26]). For example, in computational fluid dynamics, a popular simplification

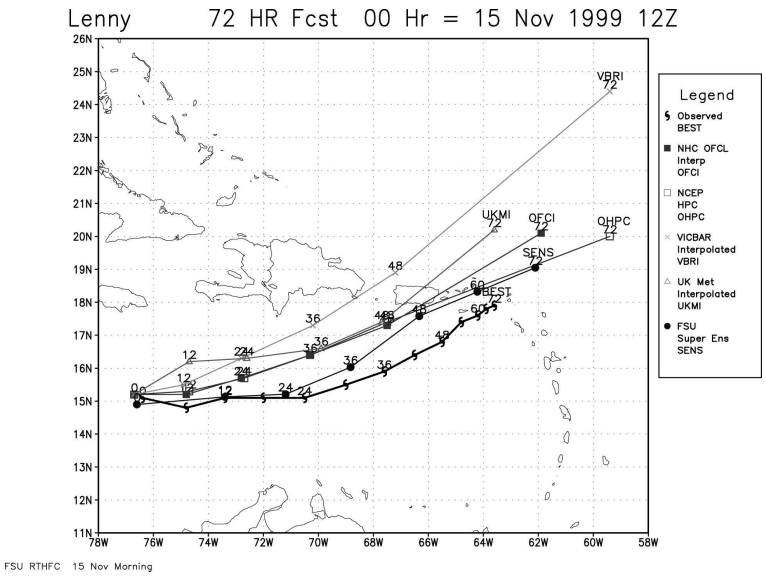


Fig. 6. Superensemble track forecast of Hurricane Lenny with predicted tracks of some member models and associated superensemble track shown [23].

is the use of the Reynolds-averaged Navier Stokes (RANS) model in place of the more complex large eddy simulation (LES) or even direct numerical simulation (DNS). These simplified models are often benchmarked or calibrated against the more complex counterparts. Outside the benchmark/calibration regime, the predictions from these two models of different physics fidelity can be compared. As a result, the model form uncertainty in the lower fidelity model could be estimated by comparison of the predictions with the higher fidelity model predictions at a limited number of conditions that are similar to the application conditions of interest. If it is concluded that the lower fidelity model accuracy is judged to be inadequate for the application of interest, then one may (a) increase the modeling fidelity of the lower fidelity model so as to attain the needed accuracy, or (b) characterize the model form uncertainty in some appropriate way so that the predictive uncertainty is recognized by the user of the simulation results.

4 Discussion

4.1 The Role of Model Form Uncertainty

For the case of extrapolations to application domains outside of the validation domain or to a higher level of system model, it is likely that model form uncertainty will dominate the extrapolated uncertainty. Extrapolating model form uncertainty is complex because it is extrapolating the error structure of a model, combined with the uncertainty in the experimental data, in a high dimensional space.

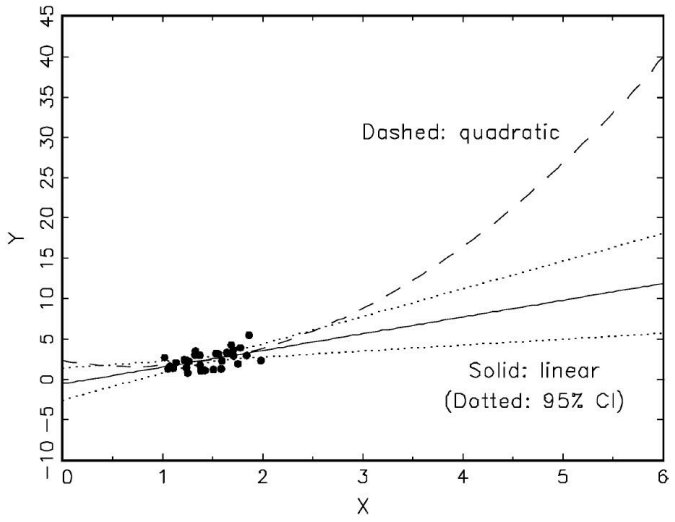


Fig. 7. Model dependence of the extrapolation with equivalent goodness of fit in the validation domain [27].

Uncertainty can be reduced, compared with for example, extrapolating a regression fit of the measured system response quantities, by taking advantage of the physics incorporated into the model [3]. Take for example Fig. 7, where there is no physical basis to the quadratic or linear models used to describe the data, and then extrapolate to large values of x. The fitted models over the range of the data are indistinguishable but in the range of extrapolation, the quadratic falls far outside of the confidence interval (CI) of the linear fit. A physics based model would help constrain the extrapolated uncertainty.

4.2 Use of UQ to Manage Extrapolation Uncertainties

The goal of a simulation is to produce a prediction along with an estimate of the effect of all of the relevant uncertainties on the system response quantities of interest. In addition to estimating the uncertainty of an extrapolation, UQ can serve to reduce that uncertainty through methodologies and mathematical methods for [28]:

- tuning (or calibrating) a simulation model to match with experimental results,
- establishing the integrity of (i.e., validate) a simulation model,
- assessing the region of validity of a simulation model,
- characterizing the output uncertainties of a simulation model,
- identifying the major sources of uncertainties of a model,
- providing information on which additional experiments are needed to improve the understanding of a model.

These methodologies can be used to understand how the model performs over the validation domain and improve it, if necessary. As in Fig. 5, these improvements serve to improve the quality of the extrapolation and reduce the uncertainty in the extrapolation. While there might be significant uncertainty on the magnitude of extrapolated uncertainties, the fact that the uncertainty has been reduced and by what fraction is useful in itself.

4.3 Missing Physics (Unknown-Unknowns)

In distant extrapolations, there is always the potential for missing physics to be present either as new unit mechanisms or as new coupling of mechanisms at higher levels in the validation hierarchy, see Fig. 3. Therefore, we discuss two elements aimed at reducing the unquantified uncertainty arising from missing physics for the case where there is no recognized disagreement between experiment and simulation in the validation domain. (If there is disagreement in the validation domain, as there is in Fig. 5, an unknown-unknown becomes a known-unknown).

The Role of Peer Review to Address Missing Physics. Because extrapolation is more of a physics endeavor than a statistics endeavor, scientific peer review plays a critical role in the extrapolation process. A nearly analogous issue was faced by Theofanous and co-workers in the application of the risk oriented accident analysis methodology (ROAAM) for low probability, high consequence hazards [29]. The basic premise is that once the selected sample of the community of experts in the problem area is convinced that the model reflects to the extent possible all of the relevant physics, the problem may be considered characterized. By this we mean that what is obtained is the best that can be done at the present time with the committed resources. One outcome may be that additional resources need to be committed to the problem or that additional resolution simply cannot be obtained, as there is no known path forward to gain such resolution. This peer review process must be traceable and scrutable [29].

The Importance of Data Assimilation to Mitigate Missing Physics. Unfortunately, as systems become more complex, a point is reached at which whole system models simply cannot be validated, in the sense of comparison with experimental results. The potential for new unit processes or coupling at higher levels of the validation hierarchy must be acknowledged when making an extrapolation [17]. This does not diminish the value of the extrapolated uncertainties but requires additional attention through data assimilation.

Take for example the case of hurricane forecasting (see, for example, [17,30]). These forecasts save millions of dollars by limiting evacuation areas. But in the early stages of the prediction of the track of a storm, models sometimes predict tracks that do not coincide with the actual track that the storm eventually follows. Hurricane forecasters effectively use data assimilation to constantly update their model predictions and uncertainties. As data is assimilated over time, the

uncertainties decrease because the period of prediction (i.e., extrapolation) becomes shorter as a hurricane nears the region of interest (i.e., the location where it will make landfall).

The same will be true for prediction of aging of nuclear reactor materials. As materials are irradiated in a real fission environment, an accompanying data assimilation effort must be in place, the model must be constantly updated, missing physics revealed, and as a result, uncertainties in predictions can be expected to decrease.

5 Conclusions

The problem posed by the nuclear industry, referred to as irradiation effects scaling [1], is an ideal example of a high impact uncertainty quantification problem requiring scaling across validation domains and extrapolations to application domains where there is little or no experimental data. The need for predictive material aging models will drive the need for associated uncertainties in the predictions.

The case of extrapolation is more of a physics endeavor than a statistics endeavor. The goal is to produce a prediction with scientifically defensible and acceptable uncertainty. Most extrapolation methods do not deal with missing physics, i.e., they only estimate (extrapolate) the error structure of known-unknowns. Therefore, the process involves using validating experimentation and more detailed physics-based models that capture the essential physics thus enabling the required scaling and extrapolation and reducing uncertainties.

The idea that uncertainty increases when extrapolating outside of the validation domain is clear. However, exactly how the uncertainty increases is not well understood, in addition to being model and situation dependent [17]. This is likely due to the fact that the uncertainty increase is very tightly coupled with the physics-basis of the model. Inaccuracies inherent to models that approximate the relevant physics, i.e., model form uncertainty, will likely dominate the uncertainties. Methods to quantify uncertainties in predictions of models, particularly model form uncertainty, are lacking and are topics requiring further fundamental research.

Acknowledgements. The authors are grateful to (i) Jon Helton of Sandia National Laboratories for his contributions to the document in general and to Sect. 2.2 in particular, (ii) Rob Rechard of Sandia National Laboratories for providing input on the Yucca Mountain Project, (iii) David Higdon of Los Alamos National Laboratory for permission to use Fig. 4 and Fig. 5, (iv) T. N. Krishnamurti of The Florida State University for permission to use Fig. 6, (vi) Gary King of Harvard University for permission to use Fig. 7 (vii) Richard Kochendarfer AREVA Federal Services LLC for a contribution to Sect. 2.2, (viii), Len Lorence of Sandia National Laboratories for providing input on the QASPR project, and (ix) Tony O'Hagen of Sheffield University, William Meeker of Iowa State University, J. Tinsley Oden of The University of Texas at Austin, Habib

Najm of Sandia National Laboratories, and Janet Allen of University of Oklahoma for extensive technical input. This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

1. Crabtree, G., Malozemoff, A.: Science for Energy Technology: Strengthening the Link between Basic Research and Industry. U.S. Department of Energy, http://science.energy.gov/~/media/bes/pdf/reports/files/setf_rpt.pdf (2010)
2. Roy, C.J., Oberkampf, W.L.: A Comprehensive Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing. Computer Methods in Applied Mechanics and Engineering 200, 2131–2144 (2011)
3. Oberkampf, W.L., Roy, C.J.: Verification and Validation in Scientific Computing. Cambridge University Press, New York (2010)
4. King, W., Allen, T., Arsenlis, T., Bench, G., Bulatov, V., Fluss, M., Klein, R., McMahon, D., Middleton, C., Morley, M., Turchi, P., Was, G.: Report on the Workshop on Accelerated Nuclear Energy Materials Development. Lawrence Livermore National Laboratory, LLNL-TR-436353, http://tinyurl.com/ANEMDworkshopreport (2010)
5. Fluss, M., King, W.E., Bench, G., Bulatov, V., De Caro, M.S.: Workshop on Science Applications of a Triple Beam Capability for Advanced Nuclear Energy Materials. Lawrence Livermore National Laboratory, LLNL-MI-413125, http://tinyurl.com/triplebeamworkshop (2009)
6. U.S. DOE: Viability Assessment of a Repository at Yucca Mountain. U.S. Department of Energy, Office of Civilian Radioactive Waste Management, DOE/RW-0508, http://www.osti.gov/bridge/product.biblio.jsp?osti_id=762971 (1998)
7. U.S. DOE: Yucca Mountain Repository Safety Analysis Report. U.S. Department of Energy, DOE/RW-0573, http://pbadupws.nrc.gov/docs/ML0907/ML090710096.pdf (2008)
8. Sandia National Laboratories: Total System Performance Assessment Model/Analysis for the License Application. U.S. Department of Energy Office of Civilian Radioactive Waste Management, MDL-WIS-PA-000005 Rev 00, AD 01 (2008)
9. Helton, J.C., Martell, M.-A., Tierney, M.S.: Characterization of Subjective Uncertainty in the 1996 Performance Assessment for the Waste Isolation Pilot Plant. Reliability Engineering and System Safety 69, 191–204 (2000)
10. Rechard, R.P., Tierney, M.S.: Assignment of Probability Distributions for Parameters in the 1996 Performance Assessment for the Waste Isolation Pilot Plant. Part 1: Description of Process. Reliability Engineering and System Safety 88, 1–32 (2005)
11. Rechard, R.P., Tierney, M.S.: Assignment of Probability Distributions for Parameters in the 1996 Performance Assessment for the Waste Isolation Pilot Plant. Part 2. Application of Process. Reliability Engineering and System Safety 88, 33–80 (2005)
12. Myers, S.M., Wampler, W.R.: The Science of QASPR (Qualification Alternatives to the Sandia Pulsed Reactor) in Sandia's Physical, Chemical, and Nano Sciences Center (PCNSC). Sandia National Laboratories, SAND2008-7810P (2008)

13. Paulsen, R., Castro, J., Hoekstra, R., Myers, S., Wampler, W., Romero, V., Rutherford, B., Griffin, P., Giunta, A., King, D.: QASPR Silicon Device Prototype Exercise and Results Summary. Sandia National Laboratories, SAND2010-8043 (2010)
14. Meeker, W.Q., Escobar, L.A.: Pitfalls of accelerated testing. IEEE Trans. Reliab. 47, 114–118 (1998)
15. Nelson, W.: Accelerated Testing: Statistical Models, Test Plans and Data Analyses. Wiley, New York (1990)
16. Oberkampf, W.L., Trucano, T.G.: Verification and Validation in Computational Fluid Dynamics. Prog. Aerosp. Sci. 38, 209–272 (2002)
17. Cacuci, D.G., Ionescu-Bujor, M.: Sensitivity and Uncertainty Analysis, Data Assimilation, and Predictive Best-Estimate Model Calibration. In: Cacuci, D.G. (ed.) Handbook of Nuclear Engineering. Springer, New York, London (2010)
18. Higdon, D., Klein, R., Anderson, M., Mark Berliner, Covey, C., Ghattas, O., Graziani, C., Habib, S., Seager, M., Sefcik, J., Stark, P., Stewart, J.: Uncertainty Quantification and Error Analysis. In: Messina, P., Bishop, A. (eds.) Workshop: Scientific Grand Challenges in National Security: the Role of Computing at the Extreme Scale, pp. 121–142. Department of Energy, Hilton Hotel Washington DC North, Gaithersburg, MD (2009)
19. Kennedy, M.C., O'Hagan, A.: Bayesian Calibration of Computer Models. Journal of the Royal Statistical Society Series B-Statistical Methodology 63, 425–450 (2001)
20. Eldred, M.S., Burkardt, J.: Comparison of Non-Intrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Quantification. In: 47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition, pp. AIAA-2009-2976. American Institute of Aeronautics and Astronautics (2009)
21. Najm, H.N.: Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. Annual Review of Fluid Mechanics 41, 35–52 (2009)
22. Ferson, S., Oberkampf, W.L., Ginzburg, L.: Model Validation and Predictive Capability for the Thermal Challenge Problem. Computer Methods in Applied Mechanics and Engineering 197, 2408–2430 (2008)
23. Vijaya Kumar, T.S.V., Krishnamurti, T.N., Fiorino, M., Nagata, M.: Multimodel Superensemble Forecasting of Tropical Cyclones in the Pacific. Monthly Weather Review 131, 574–583 (2003)
24. Krishnamurti, T.N., Kishtawal, C.M., LaRow, T.E., Bachiochi, D.R., Zhang, Z., Williford, C.E., Gadgil, S., Surendran, S.: Improved Weather and Seasonal Climate Forecasts from Multimodel Superensemble. Science 285, 1548–1550 (1999)
25. Oden, J.T., Prudhomme, S.: Control of Modeling Error in Calibration and Validation Processes for Predictive Stochastic Models. Int. J. Numer. Methods Eng. 87, 262–272 (2011)
26. Qian, Z.G., Seepersad, C.C., Joseph, V.R., Allen, J.K., Wu, C.F.J.: Building Surrogate Models Based on Detailed and Approximate Simulations. Journal of Mechanical Design 128, 668–677 (2006)
27. King, G., Zeng, L.C.: The Dangers of Extreme Counterfactuals. Polit. Anal. 14, 131–159 (2006)
28. Tong, C.: PSUADE User's Manual (Version 1.2.0). Lawrence Livermore National Laboratory, LLNL-SM-407882 (2009)
29. Theofanous, T.G.: On the Proper Formulation of Safety Goals and Assessment of Safety Margins for Rare and High-consequence Hazards. Reliability Engineering & System Safety 54, 243–257 (1996)
30. Kalnay, E.: Atmospheric Modeling, Data Assimilation, and Predictability. Cambridge University Press, New York (2003)