

Comparison of Several Model Validation Conceptions against a “Real Space” End-to-End Approach

2011-01-0238

Published
04/12/2011

Vicente Romero
Sandia National Laboratories

doi:10.4271/2011-01-0238

ABSTRACT

This paper¹ explores some of the important considerations in devising a practical and consistent framework and methodology for working with experiments and experimental data in connection with modeling and prediction. The paper outlines a pragmatic and versatile “real-space” approach within which experimental and modeling uncertainties (correlated and uncorrelated, systematic and random, aleatory and epistemic) are treated to mitigate risk in modeling and prediction. The elements of data conditioning, model conditioning, model validation, hierarchical modeling, and extrapolative prediction under uncertainty are examined. An appreciation can be gained for the constraints and difficulties at play in devising a viable end-to-end methodology. The considerations and options are many, and a large variety of viewpoints and precedents exist in the literature, as surveyed here. Rationale is given for the various choices taken in assembling the novel real-space end-to-end framework. The framework adopts some elements and constructs from the literature (sometimes adding needed refinement), rejects others (even some currently popular ones), and adds pivotal new elements and constructs. Crucially, the approach reflects a pragmatism and versatility derived from working many industrial-scale problems involving complex physics and constitutive models, steady-state and time-varying nonlinear behavior and boundary conditions, and various categories of uncertainty in experiments and models. The framework benefits from a broad exposure to integrated experimental and modeling activities in the areas of heat transfer, structural

mechanics, irradiated electronics, and combustion in fluids and solids.²

1. INTRODUCTION

Methodologies for modeling and prediction in the presence of uncertainty are being actively researched and formulated by the modeling and simulation (M&S) community. Comprehensive and detailed frameworks are still elusive. References [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73], [74], [75], [76], [77], [78], [79], [80], [81], [82] highlight various lines of thinking and progress in these areas. Refs. [43], [44], [45], [46], [47], [48], [49], [50], [51] and [71] develop key aspects of the author's particular paradigm of modeling and prediction summarized in this paper. The paradigm and procedures were developed within a larger context of integrated experimental-modeling-analysis programs with end-to-end scope (experiment design and analysis, to data conditioning, model conditioning³, model validation, hierarchical modeling, and extrapolative prediction under uncertainty).

The main subject of this paper is the validation element of the framework, although the other elements will be developed as necessary to derive and explain the model validation element.

¹This paper is a work of the United States Government and is not subject to copyright protection in the U.S.

²Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

³**Model conditioning** is a term signifying a superset of model initialization and corrective adjustment techniques in the presence of uncertainty. Model conditioning includes but extends beyond procedures otherwise variously known as parameter estimation, model calibration, model updating, etc.

The validation approach was arrived at by working backwards from an end objective of “best estimate with uncertainty” (BEWU) modeling and prediction. The approach reflects a pragmatism and versatility derived from working many industrial scale validation problems in the following application areas:

- device thermal response device thermal response and initiated failure [44], [47]
- device thermal-structural response and failure [18]
- foam thermal pyrolysis and vaporization [29], [34], [52]
- fire and object-response modeling [41], [50]
- radiation-damaged electronic device response [55].

Working from this broad basis, the framework's pragmatic and versatile “real-space” model validation (and conditioning) approach (summarized in sections 3.2 and 3.3) has evolved to address random and systematic correlated and uncorrelated uncertainties of aleatory, epistemic, and combined type that commonly arise in modeling and experiments, including:

- experimental variability in multiple repeat experiments
- associated epistemic uncertainty from limited numbers of repeat experiments
- measurement uncertainties in experimental inputs and outputs
- uncertainties that arise in data processing and inference from raw experimental and simulation outputs
- model parameter and discrete model-form uncertainties
- numerical solution uncertainty from model discretization effects.

Among the many model validation paradigms, frameworks, and methodologies surveyed in this paper, the real-space formulation has been found by the author to be uniquely capable (adaptable, robust, and practical) to handle the diverse set of challenging validation problems and attributes cited above. The challenges emanate from:

- analysis and processing of the experimental data and associated interpretation of results;
- model accuracy characterization-deciding on the formulation or metric for characterizing discrepancy between model and experiment results;
- model adequacy characterization-deciding the threshold or criterion for model adequacy (acceptable agreement with reality);
- dealing with extrapolation of validation information/results/products-deciding how to project results and outcomes from

validation settings to other prediction settings beyond the model validation activity. (This item is not part of model validation, but is an essential consideration in the formulation of a useful validation approach.)

These will be central items of attention in the paper. Before getting into specific methodology in Section 3, Section 2 considers some fundamental issues and conceptions regarding models and model validation. As will be surveyed, many model validation frameworks treat the above items (or some of them) as separable, and then devise sub-frameworks to address them separately. However, the present framework considers the items to be strongly interdependent. It is reasoned in the following that only by considering them together can a comprehensive and workable end-to-end framework be formulated.

This interconnectivity bears not only on the technical procedures of model validation, but also on the interpretability and usability of the information and products from the model validation activity. This is of paramount importance. For a validation project to be recognized to add value and justify the resources allocated to it, and to be relevant beyond just qualitative insights, there must be a plan for how the information gained will be packaged and interpreted to provide a useful quantitative characterization of model accuracy and adequacy in a larger context of downstream model use for predictions beyond the model validation setting. These considerations pertain similarly to model conditioning and are addressed by the framework similarly.

2. MODEL VALIDATION-WHAT IS IT? WHAT DOES IT ENTAIL AND IMPLY?

2.1. TERMINOLOGY, AMBIGUITY, AND VARIOUS INTERPRETATIONS OF ‘MODEL’, ‘MODEL VALIDATION’, AND ‘VALIDATED MODEL’

There seems to be fairly uniform agreement in the validation community on what model validation implies at a vague conceptual level, i.e., at the level of the following brief contemporary definition: *Model Validation is the process of determining the degree to which a computer model is an accurate representation of the real world from the perspective of an intended use of the model.* This statement and close variants of it are the accepted standard definition in the recognizing bodies [1], [2], [3], [4], [5], [6] (there may be others).

This statement is a concise expression of some fundamental notional aspects of model validation. However, despite broad

agreement on this notional definition, at a detailed interpretational and implementational level there is room within the definition for considerable debate among validation methodology developers and practitioners concerning the specific procedures, steps, and products of model validation. As a result, phrases like “the model has been validated” can mean very different things to different people even within the model validation community.

Various conceptions of model validation are discussed below. A large variety of paradigms produce profound differences in the validation outlook, approach, procedures, and end products—including validation criteria, results interpretation and usage, and ultimately how the validation activity supports the objective of BEWU extrapolative prediction beyond the validation setting. For example, the definition statement above is originally from the DoD document [1], arising largely from a tradition of operations research and systems simulation. The statement was adopted consecutively in [2], [3], [4], [5], [6] essentially without change. However, the nature of the models and decision scopes pertinent in [1] are largely different from those in [2], [3], [4], [6], and perhaps [5]. The models addressed in [1] vary from discrete-event and agent type models, e.g. for supply chain logistics, battlefield simulations, and theatre defense simulations, to computational physics type models for simulation of fluid mechanics, heat transfer, and structural mechanics phenomena, etc. The type of models addressed in [2], [3], [4], [6], and to large measure in [5] are computational physics type models. Interpretation of the “standard” definition statement above is also substantially different in the DoD validation community versus the others cited. According to personal discussions with several validation practitioners in DoD, validation includes not only the process of assessing model performance vs. empirical data (if suitable data exists), but also the subjective process of adjudicating whether it *appears* that the model will perform acceptably for specified *intended uses* outside the conditions where the model has been calibrated or tested against data. In contrast, the author perceives that [2], [3], [4], [5], [6] do not consider validation to involve a process for judging whether the model will extrapolate acceptably; this is seen as a largely subjective and situation-dependent endeavor that is a different element of the M&S process (e.g., BEWU extrapolative prediction, model accreditation, etc.).

Other ambiguities in the standard definition are discussed in the following. Other validation definitions from the literature are considered. These are leveraged to offer a new definition intended to clear up ambiguities in relation to validation of computational physics models (according to the author's paradigm). Other terminology refinements are introduced to capture what is meant in this paper by ‘model’, ‘model validation’, and ‘validated model’.

2.2. DEFINITIONS OF ‘MODEL’ IN A VALIDATION CONTEXT

It is important to identify what model or set of models is being validated in a validation activity. Often, validation activities are ambiguous in this respect, which can lead to confused interpretation and improper usage of the validation results, not to mention improper accounting steps (regarding uncertainties) in the validation procedure.

Consider a 1-D heat conduction heated-rod experiment. If a finite-element model is built to model the experiment, how does one target or differentiate whether the 1-D heat diffusion equation (partial differential equation, PDE) alone is being validated or the validation applies instead to the larger model of the particular physical instantiation of the PDE? This would include the PDE and a representation of the particular physical conditions in the experiment, like rod diameter, length, material properties, boundary conditions, etc. Furthermore, if, for example, temperature-dependent material properties are present, then other models/equations/relationships besides the PDE would be involved. The ‘model’ to be validated could also include the discretization scheme and solution algorithm present in the computational model of the experiment. What, specifically, is to be validated?

It is necessary to specify exactly which model or set of models is the focus of a validation exercise. This significantly affects how one plans and performs the validation activity and interprets the results. The present framework's stance for specific identification of the ‘model’ to be validated is explained below as the “*traveling model*” ([48]).

In model validation, the solution of the mathematical ‘model’ (to be defined below) is compared to experimental data to assess the agreement between experiment and prediction. Then are ‘model result validation’ or ‘solution validation’ more accurate terms for what is being done in validation? The answer is ‘no’. Fundamentally, validation is intended to test the predictiveness of a model in tests relevant to its eventual use. The solution results (assuming that solutions can be readily obtained) are the outward tangible evidence of the model's predictive ability. The solution results are the accessible means by which the model's predictivity is assessed, but the predictivity or lack thereof comes from the underlying model itself (its *basis* of physics equations, geometry description, parameter values/ranges, boundary conditions, discretization scheme, etc.). It is really the predictivity of this underlying basis that is the object of the validation assessment. This basis is considered to comprise the ‘model’ in the validation activity and the term ‘model validation’ is accordingly more appropriate than ‘solution validation’ or ‘model result validation’.

2.2.1. “Experiment Model” and “Traveling Model”

Here we get more precise about the definition of ‘model’ by recognizing two different scopes to ‘the physical problem’ in model validation, conditioning, and extrapolation.

The useful notion of a *traveling model basis* or simply *traveling model* is employed to delineate those physical-representation models in the model validation or conditioning activity that will be used in downstream predictions. This is a subset of the larger set of models employed in the validation activity, referred to as the *experiment model* or *e-model*. This larger set of models usually cannot be avoided. No matter what the model of downstream or traveling interest is-whether a set of PDEs, a material-behavior constitutive model, a model of a hardware device, a process model, etc.-the experiment will usually have elements that need to be modeled which are *auxiliary* to the elements in the traveling model of interest.

For example, a physical system of traveling interest could be an electromechanical device, for which a model is being validated. The physical device is part of a larger hardware system. A finite-element (FE) model of the device is to be validated under certain loading in the validation experiments. The model is then to be used in a hierarchical system-level model to make predictions of device response under *other* loads that the device may experience within the assembled system subjected to loading at the system level. Here, the hardware device in the validation experiment is the traveling physical system, the FE model is the traveling model, and the applied loads in the validation experiment are not part of the traveling system or model. Here the traveling model includes the PDE equation sets and the descriptions of the device geometry and material properties.

In contrast to the definite image that a traveling hardware device and FE model evoke, the traveling physical system and model can have an amorphous image. This occurs when the physical system of traveling interest is a generic phenomenological behavior such as material property behavior, turbulence, etc. Here the traveling model is a phenomenological equation set with associated parameter values/ranges, but without a specific traveling geometrical-object form.

Amorphous phenomenological models can only be validated through some embodiment or instantiation of a particular physical problem. For example, to develop or validate a constitutive equation set for elastic-plastic behavior of a certain material, the experiments must involve a specific specimen of material, shape/geometry, initial state, and applied loads in the experiment. A corresponding e-model is built of the experimental instantiation. Any e-model aspects such as applied loading and specimen shape/geometry that

are not part of a physical setting for which the constitutive model is later used are outside the traveling model of interest.

Although only a certain portion of the experiment model travels to application/prediction settings away from the validation experiment, the uncertainty that issues with the traveling model is often accumulated from various uncertainties in the experiment, including those from the non-traveling elements. It is also important to recognize that uncertainty in the various aspects of the experiment are necessarily treated differently if the uncertainties are affiliated with traveling aspects versus non-traveling aspects ([50], [51]).

The traveling model can contain “*model-intrinsic uncertainties*” that are inherently affiliated with the traveling model, such as an uncertainty range on parameter values in a turbulence or material model. These uncertainties are defined prior to the current experiment (they are not determined by or in the experiment) and come to the experiment model as *a priori* uncertainties in model form and/or parameter values. See [18], [29], [41], [47], [50], [55] for illustration of how the framework handles model-intrinsic uncertainties in model validation and conditioning.

The model that travels from a validation or conditioning activity could even be the representation of the loads or boundary conditions in an experiment. One may want to develop or validate a model of an experimental facility's applied loading or excitation, e.g. the spatial/temporal radiation intensity profile in a testing chamber. The model would be the subject of continuing interest for downstream use in predicting device response to simulated radiation shots in the test chamber.

The following is a poignant example where the traveling model concept would have been helpful in an eventful exchange ([63], [64], [65]) in the groundwater modeling literature regarding the boundaries of the ‘model’ being validated. Konikow & Bredehoeft [63] give an example where a calibrated groundwater flow model under 10 years of precipitation conditions was used to predict the next 10 years of ground-water flow, assuming that precipitation in the calibration period would be representative of the future. The actual precipitation (stream recharge) happened to be substantially different in the following 10 years, contributing to disagreement with the actual ground-water flow measured in that period. De Marsily et al. [64] point out that Konikow & Bedeloft are “... mixing apples and oranges...In no way has any groundwater flow model claimed that it can predict climate variability [precipitation and stream recharge] in the future...the ten years following the model calibration are more humid than those used for the calibration, why should this in any way invalidate the groundwater flow model?”

In the vernacular of the present paper, Konikow & Bredehoeft include the model inputs of stream recharge rates as being in their traveling model; the recharge rates are included as part of the model from the 10-year calibration period that travels to the new prediction setting and is used to predict behavior in the next 10 years. De Marsily et al. argue that recharge rates should not be part of the model on trial; that recharge should not be assumed to remain consistent over the next 10 years. Instead, a variety of recharge scenarios should be run in the predictions to reflect the uncertainty (historic climatic variability) of future recharge in the next 10 years. Whether it was reasonable for Konikow & Bredehoeft to include the recent 10-yr. history of stream recharge rates in their traveling model is for the groundwater community to discuss and decide. Such a discussion would be focused and facilitated by the concept and terminology of the ‘traveling model’. In general, a crisply defined traveling model should be a routine aspect of the thinking in model conditioning, validation, and prediction.

It is the traveling model that is the subject model of a validation or conditioning activity. The traveling model carries an implicit connotation that the physical attributes and phenomena it represents are proposed to carry-over consistently to the new prediction settings, such that the traveling model applies there as well. Hence, Konikow & Bredehoeft were implicitly proposing that the stream recharge rates measured in the previous 10 years would remain stable over the next 10 years. This didn't turn out to be the case, although the geology was probably stable over the time period and hence that part of the model would carry over well to the new prediction setting.

2.2.2. Strongly and Weakly Defined Models

A related contribution of interest was put forward by Leijnse & Hassanizadeh [65]. They observed that “there is no unanimity within the groundwater modeling community on what constitutes a model.” They further observed that much of the difference of opinion between De Marsily et al. and Konikow & Bredehoeft was that ‘model’ and ‘validation/validated’ were defined differently in the two view points. The present paper recognizes their different definitions of ‘model’ as a difference in their scopes for the traveling model, i.e., where the boundaries of the traveling model were drawn.

Leijnse & Hassanizadeh go on to introduce the concept of ‘weakly defined’ models like Darcy's law and the Navier-Stokes equations. Such governing equations are considered weakly defined in that they do not involve set parameter values or boundary conditions, geometry, etc. (An *amorphous* model in the author's terminology.) When an instantiation of the equations of a weakly defined model occurs for a particular application, specific parameter values, boundary conditions, geometry, etc. become defined and the collection

comprises a ‘strongly defined’ model in the Leijnse & Hassanizadeh terminology. They observe that models can exist anywhere in the spectrum between weak and strong extremes, with most working models involving both strongly and weakly defined aspects.

They argue that the rift between the view of Konikow & Bredehoeft that groundwater models cannot be validated, and the opposite view of De Marsily et al., is explainable by the fact that the two camps were talking about validation of models at very different places on the weak/strong spectrum. Leijnse & Hassanizadeh contend that it is easier for De Marsily et al. to validate their more weakly defined model (using measured precipitation for input to the model) than for Konikow & Bredehoeft to validate their more strongly defined model (involving stream recharge rates modeled as those from the previous 10 year period). Leijnse & Hassanizadeh generalize that it is easier to validate weakly defined models than strongly defined ones.

A different view is held in the current paper. The difficulty that Konikow & Bredehoeft experienced was not necessarily that they defined their future recharge rates strongly, it was that the proposed recharge rates were inaccurate. If the strongly defined recharge rates were accurate, the model's prediction accuracy would have been much better. Conversely, it is not necessarily true that a more weakly defined model would have been easier to validate. Consider if the model was more weakly defined by modeling the future recharge rates as uncertain, say with a range $\pm x\%$ about the strongly defined rate $R(\text{time})$. Then at one end of the uncertainty range, say $(1 + x\%)R(\text{time})$, the model output results would be closer to the actual measured 10-yr. data. But at the opposite extreme of the uncertainty range, $(1 - x\%)R(\text{time})$, the model predictions would presumably be further from the data. How does one deal with this new complication presented by the weaker model? A detailed discussion is given in sections 3.2 and 3.3 in connection with [Figure 1](#). Anyhow, it is not necessarily easier to validate weaker models than stronger ones.

More broadly, the following views are supported in this paper:

1. Completely strong models with no uncertainty in the parameter values and/or model form cannot be validated to capture reality.
2. Weaker models which explicitly contain a suitable degree of uncertainty in their parameter values and/or model form are validatable in the sense described in section 3.3 (can be substantiated to capture reality).
3. Validating models defined in the weak limit (generic governing equation sets like Darcy's law) is vastly more difficult than a determination for situations 1 and 2. This is because validation of a generic equation set proposed to govern over a large range of physical instantiations requires

non-rejection over an immense number of strong or semi-strong instantiations of the weak model under diverse parameter values, geometries, boundary conditions, etc. In fact, the testing never stops with regard to laws and theories.⁴ Darcy's law is said to hold, period (not “to hold within $\pm 10\%$ ” as might be the relaxed standard of acceptance for a more strongly defined application model).

For reasons associated with item 3, it is not realistic to speak of *code* validation because a particular code comprises all possible instantiations of the featured equation sets. Rather, we speak of instantiation-specific *model* validation.

In the spirit of what Leijnse & Hassanizadeh may intend, the present author, as a long-time modeler and non-purist, finds value in loose “*anecdotal* validation” of weak models by way of a broad empirical basis of loosely quantitative affirming experiences with the model. References [42] and [6, Appendix C-9] recognize consensus community acceptance as a legitimate mechanism of model validation. Still, all seek a more rigorous approach to model validation for instantiation-specific models not at the weak limit.

The assessment in this paper is that the rift between De Marsily et al. and Konikow & Bredehoeft viewpoints is less an issue of where their respective models lie on the weak-strong spectrum, and more aligned with what De Marsily et al. cite: A) a difference in their strictness of definitions of Validation; and B) their already-discussed difference on viable scope of the traveling model-whether the past 10 yr. record of recharge rates should have been proposed to represent the future 10 years. Regarding their definitions of validation, in the vernacular of the present paper Konikow & Bredehoeft use a strict internal-correctness criterion for model validation, while De Marsily et al. use a more relaxed definition that concerns model effectiveness-whether the model, though unavoidably erred to some degree, is good enough to be useful for particular analysis and forecasting needs. These two viewpoints will be discussed in Section 2.3-Part I.

The current paper largely takes the De Marsily et al. view, a widely held view in the validation community. However section 3.3 takes a different view than expressed in the last paragraph of De Marsily et al., “...we do not validate our models...we try to show that they are not invalidated by the data.” There is a nuanced problem with this view as discussed at the end of Section 2.3-Part I.

2.2.3. Cases where Discretization Dependency is part of the Traveling Model

Traditionally in the computational mechanics model validation community (e.g. in [2], [4], [6]), discretization dependency is not considered to be part of the model basis subject to validation. It is envisioned that discretization dependency of solution results can be reduced to negligible levels in the model validation setting and in subsequent model-use settings. Therefore, discretization considerations can be “converged out” of the validation problem; the traveling model basis consists of the governing equations which are solved essentially exactly, so the equations alone embody the essence of the model.

However, when prediction results depend not only on the governing equations but also on the discretization scheme and numerical solution procedures, then these unavoidably become part of the traveling model basis. Ref. [71] develops this idea further, into Class I and Class II discretization-dependent traveling models, according to whether the spatial discretization scheme is designed to be, or defaults to become, an aspect of the traveling model. The experience of the author is that discretization effects are not dismissable in most large computational physics models (see e.g. [71] for citation of many examples in heat transfer, fluid mechanics, and solid mechanics). In fact, mesh dependency is innate to some models like material failure models and turbulence spatial filtering schemes. As part of the traveling model basis, the discretization and solution procedures are tested for extrapolative consistency (along with the physics equations) in extrapolative tests of the model. The discretization and solution procedures in Class II traveling models are validated on par with the physics equations as first-order governors of the extrapolative success (or not) of the model. There appear to be no alternative paradigms in the computational mechanics verification and validation (V&V) literature for dealing with discretization dependency in model validation. Certainly there are no popular or broadly accepted paradigms at present.

This situation is clearly not as clean as the idealized one where discretization effects are converged out of the traveling model, making the physics equations alone the determiners of extrapolative performance. However, the realities demand that the computational mechanics community deal with the more difficult situation, so more research must be devoted to this topic. As pointed out in [42], other communities like the climate, groundwater flow, and ocean modeling communities

⁴Popper ([40]) argues that a theory cannot be proven true because at some point evidence can come along to overturn the theory. This has happened repeatedly in history. The best that can be done is to continually subject the theory to as many diverse tests as possible and if/while it survives the tests, not reject the theory. A related issue is that non-rejection of a hypothesis in individual tests does not imply trueness of the hypothesis because noisy and uncertain experimental data can leave much room for bias error to go undetected (see Section 3.3.1-Validation Setting I). However, the odds that the theory is valid would appear to increase as non-rejection evidence mounts over a large number of diverse tests; non-rejection in individual cases is not convincing, but taken together the assembly of evidence becomes convincing-though not absolute proof. It is conceivable that statistical arguments can be made that non-rejection in a large number of diverse cases statistically averages out the bias “parallax” in the individual tests such that a probability can be estimated that the law/theory/model is valid. However, the author is aware of no such methodology applied to positively substantiate (within estimated statistical uncertainty) weak models such as Darcy's law or the Navier-Stokes equations. Their broad acceptance is assumed to be based on less rigorous anecdotal evidence, although a lot of it.

have mesh-dependent models wherein the mesh dependence is not converged out or intended to be converged out. Those modeling communities might hold the key to successful paradigms for dealing with discretization dependency in computational mechanics model validation.

2.3. ACCURACY AND ADEQUACY ASPECTS OF MODEL VALIDATION

Part 1: ‘accuracy’ pertains to the model's mapping of inputs to output results, not to accuracy/correctness of the model's internal representations of reality

In the standard definition statement the following words “... *degree to which the model is an accurate representation of the real world...*” can engender a misconception of what model validation is capable of ascertaining. Without going too deeply into scientific philosophy and logic in the following discussion, it is necessary to be aware of the fundamental constraints on what model validation can actually establish.

An important aspect of the present framework is the recognition of the basic nature of models and modeling. A concise and popular expression that contrasts model validation from code verification is the following: model validation is concerned with “*solving the right governing equations*” whereas code verification is concerned with “*solving the governing equations right*.” (This phraseology was first used in [13], cited as a computational fluid dynamics adaptation of the software engineering phraseology in [83]: “*Validation is doing the right job and verification is doing the job right*.”) At one level the succinct statement concerning model validation is useful, but it requires further refinement when thinking at a deeper level about models, modeling, and model validation. Because models are mathematical abstractions of physical properties, processes, and behaviors of real systems, it is not proper or productive to approach model validation as an activity in assessing or judging how “right” or “correct” in a strict sense particular modeling equations and parameters are.

For example, a homogeneous material property treatment of mass density in a continuum description of a steel plate might be very effective in modeling certain thermal or mechanical behaviors. Nonetheless, this does not correspond to equation/parameter correctness; the steel material is in reality mostly vacant space with mass concentrated at numerous discrete locations (atoms). Although clearly not correct, the model representation is effective in this case. Would turbulence modelers claim that their equations model the rich phenomena in turbulence in a mechanistically correct or right manner? What about crack-propagation models, or thermal contact resistance models, or convection correlation models? Hence, when equations/models seem to work well in particular circumstances, ‘model effectiveness’ seems the

better way to think about and term things than ‘model correctness’ does. Zeigler ([72]) frames this as ‘predictively valid’ models versus ‘structurally valid’ models. In the latter the model structure agrees with the internal workings of the real system. Predictively valid models predict well even though they may not faithfully represent all the internal workings of the system.

Continuum mechanics partial differential equations (PDEs), as the foundation of the engineering models of interest in this paper, while very effective for certain modeling tasks and purposes, are not correct in a strict sense. These simplifications of reality are only superficially representative of the underlying systems and phenomenology. They do not and cannot represent the full rich physics in real systems.

The words ‘right’ and ‘correct’ also convey a notion of model uniqueness. The implication is that a uniquely right or correct mathematical model/basis exists that describes the underlying phenomena in some system for which a model is being validated. However, even if the model exactly matches the output data of the system it cannot logically be concluded that the present model is the only one capable of this. Other models that are perturbations of the said model, or even dramatically different from it, might also be able to produce exact matches to the experimental data. For example, in parameter estimation and model calibration one must deal with the usual circumstance that non-unique combinations of parameter values give the same output results of the model. There are even numerous non-unique model paradigms and structures that yield effectively the same solutions in regimes of overlap, e.g. Lagrangian vs. Eulerian formulations of transport phenomena, Newtonian Mechanics vs. Einstein Relativistic Mechanics, and Navier-Stokes vs. Euler vs. Potential Flow.

Therefore, even if a model exactly matches the output data of a validation experiment, it cannot be concluded that the model is uniquely the right one; that the equations/parameter values are the right ones; or that the model is correct in a strict sense.

For even our most revered and fundamental continuum-level behavioral models (PDEs) and simple geometries and boundary conditions, matching the behavior of real systems requires adjustable parameter values to make up for the limited causal representation in the models.⁵ For more complex models with many governing equations, materials/properties, complex geometry, multiple parts and pieces with boundary interactions like thermal contact resistance and frictional contact, and coupled nonlinear phenomenological interactions, it is expected that *a priori* parameter values input to the model cannot possibly make up for the various model-form deficiencies. This reflects back to point 1 in section 2.2.2. An important implication is that it seems

therefore inappropriate to propose validation as a determination of whether model results are different from experimental results. The expectation is that they will be different. The relevant issues are how much different and whether the difference is acceptable for an envisioned use purpose of the model (for some design, analysis, or decision-making purpose).

If a validation determines that the model output results are very close to experimental results (for now assuming exact solution of the equations and experiments with no error or uncertainty), the hope is that this is the result of negligible errors in all the aspects of the model. Unfortunately this cannot be strongly concluded. Numerous possibilities exist where relatively large errors occur in the various modeling aspects but the errors offset each other to a large degree. In fact, the probability of “right for the wrong reasons” would appear to be substantially greater than the probability of “right for the right reasons” in situations where complex models have not been assembled incrementally with careful experimental validation as each new significant element of the model is added.

However, such incremental validation cannot occur with more granularity than physical and experimental factors allow. This limits the granularity with which “right for the right reasons” can be established. It can be difficult or impossible to separate strongly comingled physics for individual validation of the contributing elements. For example it is not possible experimentally to build up to a fire (as the full system of interest) by incrementally adding combustion, then turbulence, then soot generation/agglomeration, then radiative participation, etc. Here one cannot claim that agreement of model and experimental results at the full system level is “right for the right reasons”. Nonetheless, model credibility in this respect can be alternatively built in the spirit of Footnote 4 by investigating a diverse set of conditions/scenarios in a validation matrix and showing that the model performs well (robustly) over the diverse tests.

Another type of granularity constraint in engineering projects is that time and resource constraints present practical restrictions which almost always limit hierarchical decomposition to something much coarser than what a detailed in-depth understanding of the system would require. As a practical matter it is often not necessary or advantageous to pursue understanding of system internal workings beyond a certain level of granularity. Beyond that level,

characterizing what's “under the hood” becomes counter-productive from cost and unnecessary-complexity standpoints.

In general, a validation activity at a given level of modeling can only be used to make assertions on the predictivity of the tested model at that level, but not to make statements on the accuracy or correctness of the individual elements of the model. Any statements on the accuracy of the underlying elements must come from probing and characterizing them individually.

The validation concerns brought up in this section are prevalent in the natural sciences modeling community and are cogently discussed in e.g. [63], [64], [65], [66], [67], [68], [69], [70]. It is not possible with natural systems (geological, biological, ecological, climatological, astrophysical, etc.) to pursue incremental system construction and characterization accompanied by incremental hierarchical model validation. Much of the thinking from those modeling realms is directly applicable to engineering validation where incremental model validation is not or cannot be pursued (e.g. modeling of as-built and in-place structures and systems).

The author agrees with the sentiment in [63], [64], [65], [66], [67], [68], [69], [70] that model uniqueness and internal causal correctness are not definitively establishable via validation or otherwise. The author also agrees with the majority sentiment that ‘absolute validation’ of internal causal correctness is not what model validation is intended to establish anyway (contrary to some views in that community, e.g. [63] and [67]). While it might be that absolute causal correctness of a model (if even a theoretical possibility) could ensure accurate extrapolative prediction beyond the validation conditions, this does not require that absolute validation of model causal correctness is the goal. A more practical and productive goal is to assess whether the model, even though having inexact causal description, predicts acceptably well in significant extrapolative tests relevant to specific use purposes of the model.

Hence the wording of the standard definition statement is somewhat troublesome in that the following interpretation in brackets [~] could easily be taken: “...*determining the degree to which a computer model is an accurate [internal causal] representation of the real world...*” The following early (1979) consensus definition of model validation seems less troublesome in this regard, although it has other weaknesses as will be discussed. From the SCS Technical

⁵Parameters like thermal conductivity, modulus of elasticity, and viscosity in the governing equations are state-variable-dependent (e.g., temperature-dependent conductivity). This signifies that the extent of the modeling scheme's predictive capability is limited. For example, Fourier's Law of heat diffusion does not explicitly include terms for the increased molecular vibration at elevated material temperatures, which normally increases the effective thermal conductivity. This lack of explicit representation in the model is compensated by inverse-calculating the conductivity values that produce the best match to experimental results at various temperatures. The outward appearance is a temperature-dependent material property (thermal conductivity) which, when coupled with Fourier's Law outwardly appears to be predictive over a range of temperatures. However, this is only true in a post-dictive sense, where a temperature dependence relationship for conductivity is developed over the temperature range to prevent empirical divergence of the model predictions otherwise. Even so, models that explicitly incorporate the essential physics mechanisms and principles would appear to have the best opportunity for predicting well in extrapolation, so would be preferable for extrapolative predictions purposes—although this is not necessarily true for interpolative prediction purposes, where physics-free statistical models may do better if sufficient data is available for these models.