AIAA

SAND2000-1656C

AIAA 2000 - 2549

Validation Methodology in Computational Fluid Dynamics (invited)

W. L. Oberkampf and T. G. Trucano
Sandia National Laboratories
Albuquerque, New Mexico

Fluids 2000
19-22 June 2000 / Denver, CO

Validation Methodology in Computational Fluid Dynamics

William L. Oberkampf* and Timothy G. Trucano**
Sandia National Laboratories
Albuquerque, New Mexico 87185-0828

Abstract

Verification and validation are the primary means to assess accuracy and reliability in computational simulations. This paper presents an extensive review of the literature in computational validation and develops a number of extensions to existing ideas. We discuss the early work in validation by the operations research, statistics, and CFD communities. The emphasis in our review is to bring together the diverse contributors to validation methodology and procedures. The disadvantages of standard practice of qualitative graphical validation are pointed out and the arguments for and the literature on validation quantification are presented. We discuss the attributes of a beneficial validation experiment hierarchy and then we give an example for a complex system; a hypersonic cruise missile. We present six recommended characteristics of how a validation experiment is designed, executed, and analyzed. Since one of the key features of a validation experiment is a careful experimental uncertainty estimation analysis, we discuss a statistical procedure that has been developed for improving the estimation of experimental uncertainty. One facet of code verification, the estimation of computational error and uncertainty, is discussed in some detail, but we do not address many other important issues in code verification. We argue for the separation of the concepts of error and uncertainty in computational simulations. Error estimation, primarily that due to numerical solution error, is discussed with regard to its importance in validation. In the same vein, we explain the need to move toward nondeterministic simulations in CFD validation, that is, the propagation of input quantity uncertainty in CFD simulations which yield probabilistic output quantities. We discuss the relatively new concept of validation quantification, also referred to as validation metrics. The inadequacy, in our view, of hypothesis testing in computational validation is discussed. We close the paper by presenting our ideas on validation metrics and we apply them to two conceptual examples.

*Distinguished Member Technical Staff, Associate Fellow

**Distinguished Member Technical Staff

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract No. DE-AC04-94AL85000.

1. Introduction

During the last three or four decades, computer simulations of physical processes have been used in scientific research and in the analysis and design of engineered systems. The systems of interest have been existing or proposed systems that operate at design conditions, off-design conditions, failure-mode conditions, or accident scenarios. The systems of interest have also been natural systems, for example, surface-water quality analyses and risk assessment of underground storage of toxic and nuclear wastes. These kinds of predictions are beneficial in the development of public policy, in the preparation of safety procedures, and in the determination of legal liability. Because of the impact that modeling and simulation predictions can have, the credibility of the computational results is of great concern to engineering designers, public officials, and those who are affected by the decisions that are based on these predictions.

For engineered systems, terminology such as "virtual prototyping" and "virtual testing" is now being used in engineering development to describe numerical simulation for the design, evaluation, and "testing" of new hardware and even entire systems. This new trend of modeling-and-simulation-based design is primarily driven by increased competition in many markets such as aircraft, automobiles, propulsion systems, and consumer products. The need to decrease the time and cost of bringing products to market is intense. This new trend is also driven by the high cost and time that are required to test laboratory or field components and complete systems. In addition, the safety aspects of the product or system also represent an important, sometimes dominant element of testing or validating numerical simulations. The potential legal and liability costs of hardware failures can be staggering to a company, the environment, or the public. The reliability, robustness, or safety of some of these computationally-based designs are high-consequence systems that cannot ever be tested. Examples are the catastrophic failure of a full-scale containment building for a nuclear power plant, fire spread or explosive damage to a high-rise office building, and a nuclear weapon exposed to a transportation crash and fire environment.

The critical issue is: How should confidence in modeling and simulation be critically assessed? Verification and validation of computational simulations are the primary methods for building and quantifying this confidence. Briefly, verification is the assessment of the accuracy of the solution to a computational model. Validation is the assessment of the accuracy of a computational simulation by comparison with experimental data. In verification, the relationship of the simulation to the real world is not an issue. In validation, the relationship between computation and the real world, i.e., experimental data, *is* the issue. Stated differently, verification is a mathematics issue; validation is primarily a physics issue.

At the national level, the Defense Modeling and Simulation Office (DMSO) of the Department of Defense has been the leader in the development of the fundamental concepts and terminology for verification and validation (V&V).[1,2] Recently, the Accelerated Strategic Computing Initiative (ASCI) of the Department of Energy (DOE) has also taken a strong interest in V&V. The ASCI program is focused on computational physics and computational mechanics, whereas the DMSO has traditionally emphasized high level systems engineering, such as ballistic missile defense systems. Of the work conducted by DMSO, it has recently been observed: "Given the critical importance of model validation..., it is surprising that the constituent parts are not provided in the (DoD) directive concerning...validation. A statistical perspective is almost entirely missing in these directives."[3] We believe this comment properly reflects the state of the art in V&V. That is, the state of the art has not developed to the point where one can clearly point out all of the actual methods, procedures, and process steps that must be undertaken for V&V. It is our view that the present method of qualitative "graphical validation," i.e., comparison of computation and experiment on a graph, is inadequate. This inadequacy especially affects complex engineered systems that heavily rely on computational simulation for understanding their predicted performance, reliability, and safety. We recognize, however, that the complexities of the quantification of V&V are substantial, from both a research perspective and a practical perspective. To indicate the degree of complexity, we suggest referring to quantitative V&V as "validation science."

It is fair to say that researchers in the field of computational fluid dynamics (CFD) have been pioneers in the development of methodology and procedures in computational validation. However, it is also fair to say that the development of CFD has proceeded along a path that is largely independent of experimental validation. There are diverse reasons why CFD has not perceived a strong need for code verification and validation, especially validation. A competitive and frequently adversarial relationship (at least in the U. S.) has often existed between computational modelers and experimentalists, which has led to a lack of cooperation between the two groups. We see computational simulation and experimental investigations as complementary and synergistic. Some will say, "Isn't that obvious?" We would answer, "It should be, but they have not always been viewed as complementary." The "line in the sand" was formally drawn in 1975 with the publication of the article "Computers versus Wind Tunnels."[4] We call attention to this article only to demonstrate, for those who claim it never existed, that a competitive and adversarial relationship has indeed existed in the past. This ambivalent relationship was of course not caused by the quoted article; the article simply brought to the foreground the competition and conflict. In retrospect, this situation is probably understandable because it is the classic case of a new technology rapidly growing and attracting a great deal of visibility and funding support that had been the domain of the older technology.

During the last few years the relationship between computation and experiment has improved significantly. There has been a growing awareness that competition does not best serve the interests of either computational modelers or experimentalists[5-15]. Even with this awareness, there are significant challenges in implementing a more cooperative working relationship between the two groups, and also in making progress toward a validation science. From the viewpoint of some experimentalists, one of the challenges is overcoming the perceived threat that CFD poses. Validation science requires a close and synergistic working relationship between computationalists and experimentalists, rather than competition. Another significant challenge is the required changes in most experimentalist's perspective toward validation experiments. We argue that validation experiments are indeed different from traditional experiments, i.e., they are designed and conducted for the purpose of code validation. For example, rigorous experimental uncertainty estimation is critically needed in validation experiments; it is not an optional element in experiments. Similarly, quantitative numerical error estimation by CFD analysts is a must. For complex engineering problems, this requires *a posteriori* error estimation; not just formal error analyses or *a priori* error estimation. An finally, we believe validation science will require the incorporation of nondeterministic simulations, i.e., multiple deterministic simulations that reflect uncertainty in model parameters, initial conditions, and boundary conditions that exist in the experiments that are used to validate the computational models.

This paper presents an extensive review of the literature in computational validation, as well as extensions to existing work. We trace the beginning of validation terminology and methodology development to the operations research community. We discuss the early work in validation by the CFD community and also the ambiguous meaning of the terms verification and validation from the perspective of other engineering disciplines and computer science. We briefly summarize portions of the first engineering standards document, the <u>AIAA Guide</u>, published on verification and validation.[16] The disadvantages of qualitative "graphical" validation are pointed out, and the literature on validation quantification is presented. In Section 3, we discuss the attributes of a beneficial validation hierarchy, and then we give an example for a complex system, namely, a hypersonic cruise missile. We present the six recommended characteristics of how a validation experiment is designed and executed. Since one of the key features of a validation experiment is a careful experimental uncertainty estimation analysis, we discuss a statistical procedure that has been developed for improving the estimation of experimental uncertainty. In Section 4, we discuss one facet of code verification, which is the estimation of computational error and uncertainty, but we do not address many other important issues in code verification. We argue for the separation of the concepts of error and uncertainty in computational simulations. Error estimation, primarily that due to numerical solution error, is discussed with regard to its importance in validation. In the same vein, we explain the need to move toward nondeterministic simulations in CFD validation, i.e., the propagation of input quantity uncertainty in CFD simulations which yield probabilistic system responses. In Section 5, we deal with the issue of the quantitative comparison of computation and experiment in validation. We present our views explaining why the traditional statistical technique of hypothesis testing is inadequate in computational validation. We then discuss the topic of validation metrics introduced by Coleman and Stern.[17] We close this topic by using two simple examples to present our ideas on validation metrics. We conclude the paper with recommendations for future work.

2. Review of the Literature

2.1 Methodology and Terminology Developments

The issue of validation of mathematical and computational models of nature touch on the very foundations of science. The reason for these deeply rooted issues in science is the question of how can formal constructs (models) be tested by physical observation. The renowned 20th century philosophers of science Popper[18,19] and Carnap[20] laid the foundation for the present day concepts of validation. The first technical discipline that began to struggle with the methodology and terminology of verification and validation was the operations research (OR) community.[21-56] In the OR activities, the complexity of the systems analyzed could be extraordinary, for example, industrial production models, industrial planning, marketing models, national and world economic models, and war fighting models. These complex models commonly involve a strong coupling of complex physical processes, human behavior, and computer controlled systems. For these complex systems and processes, fundamental conceptual issues immediately arise with regard to assessing credibility of the model and the resulting simulations. Indeed, the credibility of most of these models cannot be validated in any meaningful way.

The issue of credibility of the models is necessarily related to the meaning of the terms verification and validation. Verification and validation (V&V) are fundamentally tools for the assessment of the accuracy and the demonstration of correctness of a model. For much of the OR work, the assessment is so difficult, if not impossible, that V&V become more associated with the issue of credibility, i.e., the quality, capability, or power to elicit belief. In science and engineering, however, quantitative assessment of accuracy, at least for some physical cases, is mandatory. For certain situations, assessment may only be possible using physical models that are subscale, or the assessment may not have all of the physical processes active at one time. Regardless of the difficulties and constraints, methods must be devised for measuring the accuracy of the model for as many conditions as are appropriate for the uses of the model.

The meaning and the clarity of the terms verification and validation took a major step forward in 1994 with the definitions developed by the Defense Modeling and Simulation Office (DMSO) of the Department of Defense.[1] The DMSO definitions were formulated based on the foundational work of the OR community referenced above. In 1998, the AIAA Computational Fluid Dynamics Committee on Standards adopted the definitions of DMSO.[16] The definition of verification given by the <u>AIAA Guide for the Verification and Validation of Computational Fluid Dynamics Simulations</u> is:

> *Verification:* The process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model.

This definition slightly modifies the DMSO definition in order to make it clear that the solution to the model is included in verification. In computational physics and engineering, the numerical solution of the continuum partial differential equations is a dominant issue, whereas in the OR community it is a minor issue. The informal meaning of verification asks the question: "Did you solve the mathematical model correctly?" This statement has numerical algorithm implications, as well as software quality assurance implications.

The <u>AIAA Guide</u> definition of validation was taken verbatim from the DMSO definition:[1,2,16]

> *Validation:* The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.

The informal meaning of validation asks the question: "Did you solve the correct mathematical model?" Although this informal meaning is more intuitive than the formal definition, it also is somewhat misleading in the following sense. It implies that mathematical models are correct or incorrect; valid or invalid. For complex engineering systems this is an untenable view. This point was succinctly stated a number of years ago by George Box:[57] "All models are wrong, but some are useful."

In science and engineering, CFD was one of the first fields to seriously begin developing concepts for validation methodology and validation experiments[5-7, 9-11,13,58-96]. Much of this early work dealt with issues such as fundamental methodology, terminology, development of the concepts and procedures for validation experiments, confidence in predictions based on validated simulations, and methods of incorporating validation into the engineering design process. Essentially all of this early work dealt with CFD for aircraft and reentry vehicle aerodynamics, gas turbine engines, and turbopumps. In parallel with the aerospace activities and the OR work mentioned above, there were significant efforts in validation methodology in the field of surface and subsurface water quality modeling and safety assessment of underground radioactive waste repositories.[97-103] This water quality work is significant for two reasons. First, it addresses validation for complex processes in the physical sciences where validation of models is extremely difficult, if not impossible. The reason for the difficulty is that one of the key elements in the modeling is extremely limited knowledge of underground transport and material properties. For situations such as this, one must deal with *calibration* or *parameter estimation* in models, prior to considering validation. Second, because of this difficulty these fields have adopted statistical methods of validation assessment. As will be discussed shortly, we believe CFD must also begin adopting statistical methods of validation. Examining the literature from these diverse disciplines in operations research, earth sciences, and CFD clearly shows that each discipline developed concepts and procedures essentially independently.

A final comment should be made concerning the nonuniformity of the usage and meaning for the terms verification and validation. It is still common in CFD for people to misuse terms, for example, one refers to verification when one means validation. There is, however, a fundamentally different meaning of the terms verification and validation in other fields that must be noted. In 1984, the Institute of Electrical and Electronics Engineers (IEEE) defined verification as follows:[104, 105] "The process of evaluating the products of a software development phase to provide assurance that they meet the requirements defined for them by the previous phase." IEEE defined validation as:[104,105]: "The process of testing a computer program and evaluating the results to ensure compliance with specific requirements." Comparing these definitions with the DMSO/AIAA definitions given previously, it is immediately clear they mean something completely different. IEEE definitions are entirely referential, i.e., the value of the definition is related to the specification of "requirements defined for them by the previous phase" and "compliance with specific requirements." The substance of the meaning must be provided in the specification of additional information. Because those requirements are not stated in the definition, the definition does not contribute much to the intuitive understanding of verification and validation. These same IEEE definitions for verification and validation have been adopted by the software quality assurance and computer science communities,[106,107] the nuclear reactor safety community,[108,109] and the International Organization for Standardization (ISO)[110, 111].

The IEEE definitions for V&V are pointed out for two reasons. First, these definitions provide a distinctively different perspective toward the entire issue of verification and validation. This perspective asserts that because of the extreme variety of requirements for modeling and simulation, the requirements should be defined in a separate document for each application, not in the definition of validation. Second, the IEEE definitions are the more prevalent definitions used in engineering, and one must be aware of the potential confusion when the DMSO/AIAA definitions are used in mixed disciplines. The IEEE definitions are dominant because of the worldwide influence of this organization. As a result, we expect long-term ambiguity and confusion.

2.2 AIAA Guide

In 1992, the AIAA Computational Fluid Dynamic Committee on Standards began a project to formulate the basic terminology and methodology in the verification and validation of CFD simulations. After 6 years of discussion and debate, the project culminated in the publication of Guide for the Verification and Validation of Computational Fluid Dynamics Simulations.[16] The Guide defines a number of key terms, discusses fundamental concepts, and specifies general procedures for conducting verification and validation in CFD. AIAA Standards documents are segregated into three levels of the state of the art: guides, recommended practices, and standards. The V&V Guide is at the first level, reflecting the early stage of development of concepts and procedures in V&V. It is also the first standards document to be published by any engineering organization on the topic of V&V. The American Society of Mechanical Engineers is in the early stages of forming a new standards committee and developing a similar document in the field of solid mechanics.[112]

This section briefly reviews portions of the AIAA Guide that deal with fundamental V&V methodology. A few comments will be made concerning verification methodology in order to more clearly separate the topic from validation. Then validation terminology and methodology will be reviewed in more detail.

Verification is the process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model. The fundamental strategy of verification is the identification and quantification of error in the computational model and its solution. In verification activities, accuracy is generally measured in relation to highly accurate solutions of simplified model problems. Highly accurate solutions refer to either analytical solutions or highly accurate numerical solutions. Verification, thus, provides evidence (substantiation) that the conceptual (continuum mathematics) model is solved correctly by the discrete mathematics embodied in the computer code. The conceptual model does not require any relationship to the real world. As a result, verification is only a mathematics and computer science issue; *not* a physics issue. Figure 1 depicts the verification process of comparing the numerical solution with various types of highly accurate solutions.

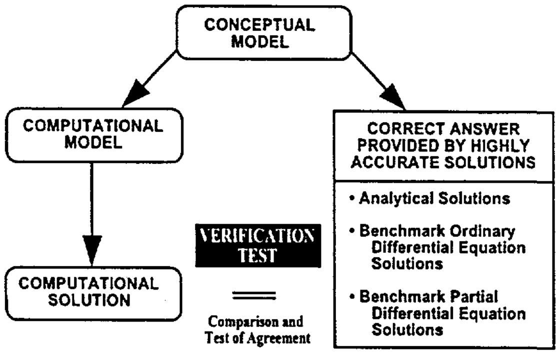

Figure 1: Verification Process[16]

Validation is the process of determining the degree to which a model is an adequate representation of the real world based on the intended uses of the model. The fundamental strategy of validation is to identify and quantify the error and uncertainty in the conceptual and computational models, quantify the numerical error in the computational solution, and also estimate the experimental uncertainty, and then make the comparison between computation and experiment. That is, accuracy is measured in relation to experimental data, our best measure of reality. This strategy *does not* assume that the experimental measurements are more accurate than the computational result. The strategy only asserts that experimental measurements are the only true reflections of reality. Validation requires that the estimation process for error and uncertainty must occur on both sides of the coin: mathematical physics and experiment. Figure 2 depicts the validation process of comparing the computational results of the modeling and simulation process with various types of experimental data.

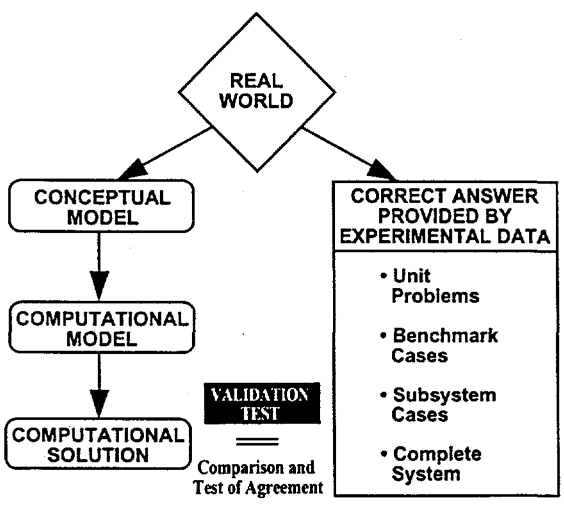

Figure 2: Validation Process[16]

Because of the infeasibility and impracticality of conducting true validation experiments on complex systems, the recommended method is to use a building-block approach.[13,16,67,81,85,113] This approach divides the complex engineering system of interest into three progressively simpler tiers: subsystem cases, benchmark cases, and unit problems. (Note that in AIAA Guide the building-block tiers are referred to as phases.) The strategy in the tiered approach is to assess how accurately the computational results compare with experimental data (with quantified uncertainty estimates) at multiple degrees of physics coupling and geometrical complexity (see Fig. 3). The approach is clearly constructive in that it (1) recognizes that there is a hierarchy of complexity in systems and simulations and (2) recognizes that the quantity and accuracy of information that is obtained from experiments varies radically over the range of tiers. It should also be noted that additional building-block tiers beyond the four that are discussed here could be defined, but additional tiers would not significantly alter the recommended methodology.

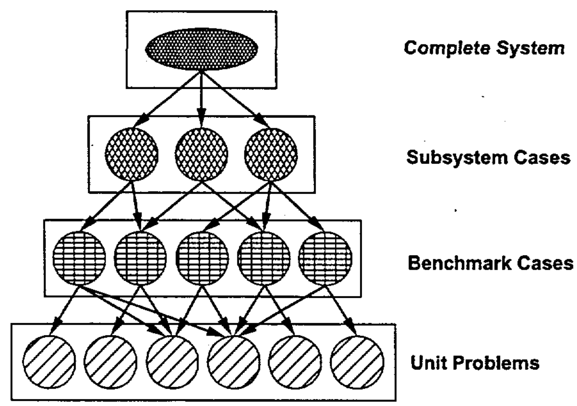

Figure 3: Validation Tiers[16]

The complete system consists of the actual engineering hardware for which a reliable computational tool is needed. Thus, by definition, all the geometric and physics effects occur simultaneously. For typical complex engineering systems, e.g., a gas turbine engine, multidisciplinary, coupled, physical phenomena occur together. Data are measured on the engineering hardware under realistic operating conditions. The quantity and quality of these measurements, however, are essentially always very limited. It is difficult, and sometimes impossible, to quantify most of the test conditions needed for computational modeling, e.g., various fluid flow rates, thermophysical properties of the multiple fluids, and coupled, time dependent, boundary conditions. Not only are many needed modeling parameters unmeasured, there is generally no experimental uncertainty analysis conducted.

Subsystem cases represent the first decomposition of the actual hardware into simplified systems or components. Each of the subsystems or components is composed of actual hardware from the complete system. Subsystem cases usually exhibit three or more types of physics that are coupled. Examples of types of physics are fluid dynamics, structural dynamics, solid dynamics, chemical reactions, and acoustics. The physical processes of the complete system are partially represented by the subsystem cases, but the degree of coupling between various physical phenomena in the subsystem cases is typically reduced. For example, there is normally reduced coupling between subsystems as compared to the complete system. The geometric features are restricted to the subsystem and its attachment, or simplified connection, to the complete system. Although the quality and quantity of the test data are usually significantly better for subsystem cases than for the complete system, there are still limited test data for subsystem cases. Some of the needed modeling data, initial conditions, and boundary conditions are measured, particularly the most important data.

Experimental data from complete systems and data from subsystem tests are always specific to existing hardware and are available mainly through large-scale test programs. Existing data from these tests have traditionally focused on issues such as the functionality, performance, safety, or reliability of systems or subsystems. For large-scale tests, there is often competition between alternative system, subsystem, or component designs. If the competition is due to outside organizations or suppliers of hardware, then the ability to obtain complete and unbiased validation information becomes even more difficult. Such tests generally provide only data that are related to engineering parameters of design interest and high-level system performance measures. The data that are obtained typically have large uncertainty bands, or no attempt has been made to estimate uncertainty. The test programs typically require expensive ground-test facilities, or the programs are conducted in uncontrolled, hostile, or unmeasured environments. Commonly, the test programs are conducted on a tight schedule and with a limited budget. Consequently, it is not possible to obtain the complete set of physical modeling parameters (e.g., thermochemical fluid and material properties), initial conditions, and boundary conditions that are required for quantitative validation assessment. Also, there are certain situations where it is not possible to conduct a complete system validation experiment. Such situations could involve public or environmental safety hazards, unattainable experimental-testing requirements, or international treaty restrictions.

Benchmark cases represent the next level of decomposition of the complete system. For these cases, special hardware is fabricated to represent the main features of each subsystem. By special hardware, we mean hardware that is specially fabricated with simplified properties, materials, or both. For example, benchmark hardware is normally not functional hardware nor is it fabricated with the same materials as actual subsystems or components. For benchmark cases, typically only two or three types of coupled flow physics are considered. For example, in fluid dynamics one could have turbulence, combustion, and two-phase flow, but one would eliminate any structural dynamics coupling that might exist at the subsystem level. The benchmark cases are geometrically simpler than those cases at the subsystem level. The only geometric features that are retained from the subsystem level are those that are critical to the types of physics that are considered at the benchmark level. Most of the experimental data that are obtained in benchmark cases have associated estimates of measurement uncertainties. Most of the needed modeling data, initial conditions, and boundary conditions are measured, but some of the less important experimental data have not been measured. The experimental data, both code input data and system response data, are usually documented with moderate detail. Examples of important experimental data that are documented include detailed inspection of all hardware, specific characterization of materials and fluids used in the experiment, and detailed measurement of environmental conditions that were produced by the experimental apparatus or testing equipment.

Unit problems represent the total decomposition of the complete system. At this level, high-precision, special-purpose hardware is fabricated and inspected, but this hardware does not resemble the hardware of the subsystem from which it originated. One element of complex physics is allowed to occur in each of the unit problems that are examined. The purpose of these problems is to isolate elements of complex physics so that critical evaluations of mathematical models or submodels can be evaluated. For example, unit problems could individually involve turbulence, laminar flow combustion, and laminar gas/liquid droplet flows. Unit problems are characterized by very simple geometries. The geometry features are commonly two-dimensional, but they can be very simple three-dimensional geometries with important geometric symmetry features. Highly instrumented, highly accurate experimental data are obtained from unit problems, and an extensive uncertainty analysis of the experimental data is prepared. If possible, experiments on unit problems are conducted at separate facilities to ensure that bias (systematic) errors in the experimental data are identified. For unit problems, all of the important code input data, and initial and boundary conditions are accurately measured. These types of experiments are commonly conducted in universities or in research laboratories.

Experimental data from benchmark cases and unit problems should be of the quantity and quality that are required in true validation experiments. If, however, significant parameters that are needed for the CFD simulation of benchmark cases and unit problem experiments were not measured, then the analyst must assume these quantities. For example, suppose for benchmark cases and unit problems that careful dimensional measurements and specific material properties of the hardware were not made. In this case, the computational analyst typically assumes reasonable or plausible values for the missing data, possibly from an engineering handbook. An alternative technique, although never done in CFD, is for the analyst to assume probability distributions with specified means and variances of the unmeasured parameters. Multiple computations are then performed using these assumptions, and likelihoods are computed for output quantities used to compare with experimental data. In existing or older published experimental data for benchmark cases and unit problems, it is common that a significant number of parameters are missing from the description of the experiment. Experiments in the past were typically conducted for the purpose of improving the physical understanding of specific phenomena or for determining parameters in models, rather than for the validation of CFD models. That is, these experiments were used inductively to construct mathematical models of physical phenomena, rather than deductively to evaluate the validity of models.

2.3 Validation Quantification

The standard validation procedure in CFD is to graphically compare computations with experimental data. If the computations "generally agree" with the experiment, the computations are declared "validated." Sometimes a favorable comparison motivates the code developer to declare the *entire* computer code "validated." There are two weakness and a fallacy with this procedure. The first weakness is that a comparison of computation and experiment on a graph is only somewhat better than a qualitative comparison. With a graphical comparison, one does not commonly see quantification of the numerical error or quantification of uncertainties due to little information of needed modeling parameters. Also, an estimate of experimental uncertainty is also not typically quoted, and in most cases it is not even available. A graphical comparison also gives little quantitative indication of how the agreement varies over the range of the independent variable, e.g., space or time, or the parameter of interest, e.g., Reynolds number or geometric parameter. The second weakness is that a graphical comparison does not quantify what is a "satisfactory" or "good" agreement of computation and experiment, i.e., graphical comparison is only qualitative. In Section 5.2, we discuss the distinction between a measure of agreement between computation and experiment, and whether the given measure is adequate for a particular application. Finally, the fallacy should be obvious in declaring an entire computer code "validated" with one, or even several, favorable comparisons. If this is not obvious, see the AIAA Guide for a discussion of this topic.[16] Unjustified validation declarations such as this are commonly driven by marketing and competitive pressures. Broad claims of validation based on little evidence should always be challenged. It is well known that the code developer and the code salesman require little convincing of the code's validity, whereas the code user, the code purchaser, and the decision maker relying on the code should require a great deal of quantifiable and reproducible evidence.

The critical issue then is how might comparisons of computation and experiment be better quantified. We suggest that validation quantification should be considered as the evaluation of a metric, or various metrics, for measuring the consistency of a given computational model with respect to experimental measurements. This metric quantifies errors and uncertainties in both the computational and experimental activities. Several researchers have pursued this topic since the late 1980s, however, only one of these researchers[17] is in the field of aerospace systems. We will now review in some detail those authors who have addressed validation quantification.

Beck,[97] in a long review article, dealt with four closely related topics: uncertainty about model structure, uncertainty in the estimated model parameter values, the propagation of prediction errors, and the design of experiments in order to reduce the critical uncertainties associated with a model. The first two topics are actually issues in validation quantification. His insight 13 years ago that validation can be thought of as test of both model structure (model form) and model parameters has escaped most researchers in the field. The application area he discusses is surface and ground water flow modeling. Because of the complex physical and chemical processes occurring, the large number of unknown parameters in the partial differential equations (PDEs), and the unknown state of most of the initial and boundary conditions, he approaches validation as a statistical estimation problem. His philosophical approach to validation, and toward modeling and simulation in general, is condensed in his comment: "If there were a longer term view to be taken, current research activities might be interpreted as a swing for the pendulum away from determinism toward indeterminism."

LeGore[100] dealt with validation quantification for problems in ground water transport of radionuclides and toxic materials. He discusses how Monte Carlo sampling methods can be used to compute the system response measures that are measured experimentally. For ground water flow problems there are stochastic parameters in PDEs which describe the hydraulic conductivity of a porous medium. A "realization" from the computational model for the purpose of comparing with the physical experiment is created by randomly selecting a value for the stochastic parameter using the predetermined, or assumed, probability distribution for the parameter. He also points out how three different stochastic system response measures from the model and the experiment could be compared.

Gass[46] discusses a method of model accreditation that would apply to any type of computer modeling. He defines accreditation as "an official determination that a model is acceptable for a specific purpose." Accreditation is a much more formal statement of model credibility than model validation. He recommends the Analytical Hierarchy Process (AHP) of Saaty[114] for accreditation. A numerical score is computed for a code that would be a weighted sum of scores in areas such as logical verification, documentation, code verification, face validation, independent review, data validation, comparison with laboratory data, and comparisons with system level data. If the code's total score were greater than some specified value, then the code would be accredited. Although this type of credibility measure is quite different than other authors discussed here, it does reflect a much broader view that would be appropriate to use when high consequence systems are modeled, e.g., nuclear reactor safety assessment.

Lee and Poolla[115,116] present a general theoretical approach for statistical model validation that they carefully define as: "Given experimental time-domain input-output data, certain *a priori* information about the true plant, and a hypothesized uncertainty model, determine the validation probability that the uncertainty model is consistent with both the prior information and the data record." They point out that by using a probabilistic point of view they are focusing on the *likelihood* that a given model is valid. They use a Bayesian method for updating prior probabilities when more data, typically more experimental data, become available. The also use certain statistical ideas from hypothesis testing, but they do no take the hypothesis testing approach. (More discussion of hypothesis testing is given in Section 5).

Draper[117] and Laskey[118] both discuss validation quantification, but in the context of the broader subject of model prediction uncertainty, i.e., given comparisons of the model with experimental data, what is the uncertainty of foretelling future events using the model.[16] Following Beck,[97] they develop statistical methods for segregating model structural (model form) validation and parametric validation (estimation). Draper's article is particularly valuable because it is both a review article and, in an addendum, it contains comments and criticisms from 27 other researchers in the field, plus Draper's response to each. Draper and Laskey take a statistical approach to validation quantification, and they both use a Bayesian approach to updating the model and parametric uncertainties when new experimental data become available. Commenting on the situation of incomplete experimental data to conclude whether a model is valid, Laskey comments "it may be necessary to acknowledge that in the presence of irreducible [epistemic] model uncertainty there may be no single 'right answer' and reasonable people may disagree."

Kleijnen[53] uses a case study approach toward validation of an acoustical model for the detection of mines on the sea floor. He points out the rapidly increasing interest in model validation, but he appropriately summarizes the state of the art as: "Unfortunately, this interest has not resulted in a standard theory on validation. Neither has it produced a standard 'box of tools' from which tools are taken in a natural order." In his case study he takes a statistical approach to validation quantification, but he takes a frequentist approach as opposed to a Bayesian approach. He recommends that before comparisons of theory and experiment are made, statistical analysts should work with the code user/analysts to design experiments that vigorously test the code. He believes that finding ways of "breaking the model" are best achieved by sensitivity analysis of computational model. Sensitivity analysis finds the local rate of change of any output quantity with respect to any input quantity. As a result, sensitivity analysis could then be used to find what input quantities cause the most uncertainty in important system response measures.

McKay[119] discusses model prediction uncertainty and sensitivity analyses, but these techniques are useful in parametric validation. He considers the case where certain input parameters are unknown, but they are represented by probability distributions. These parameters can be, for example, coefficients in the differential equations, initial conditions, or boundary conditions. He determines which of these parameters are the most important by comparison of the prediction distribution with conditional prediction probability distributions. He uses replicated Latin hypercube sampling, also referred to as stratified Monte Carlo sampling, for the propagation of the uncertainty through the model. His method does not depend on model linearity or monotonicity, which usually accompanies regression-based methods.

Coleman and Stern[17] are the first to deal with validation quantification in the field of aerodynamics/hydrodynamics CFD. They take a statistical approach that combines the random error from the experimental data and the simulation uncertainty. They define the simulation uncertainty as sum of the numerical solution error, simulation modeling uncertainty arising from using previous experimental data, and the modeling uncertainty due to modeling assumptions. As many authors in CFD, they refer to all numerical errors (e.g., spatial and time-step discretization, artificial dissipation, intra-step and global iterative non-convergence, computer round-off, etc.) as numerical uncertainty. They define a validation uncertainty metric as the sum of the uncertainty in the experimental data, the numerical solution error, and the simulation modeling uncertainty arising from using previous experimental data. They contend that this validation uncertainty metric sets the level at which validation can be achieved. Their criterion for validation is that the magnitude of the comparison error between experiment and computation must be less than their validation uncertainty metric. Using their approach, they give an example of modeling uncertainty due to multiple turbulence models in a Reynolds averaged Navier-Stokes code.

Hanson[120] uses a probabilistic framework for assessing uncertainties in simulation predictions that arise from model parameters derived from uncertain measurements. The focus is on the parametric uncertainties in physics models, how they propagate through the model, and what is the probabilistic result on various system response measures. These probabilistic modeling results are then compared to experimental data that are given by multiple replications of experiments. He stresses the importance of having a hierarchy of experiments, as discussed above, for aiding in identifying weaknesses in physics submodels and coupled physics that occurs in experiments of higher complexity. He uses Bayesian estimation for updating prior probability distributions as one moves from one experiment to the next. He also uses the Markov chain Monte Carlo technique for generating a sequence of random parameter vectors drawn from an arbitrary target probability density function (PDF). His continuous updating with hierarchical experiments allows one to determine weak spots in the submodels and then design and conduct new experiments for correcting the weaknesses.

Hills[121] develops two statistical approaches for validation and applies them to computations from CTH, a solid dynamics code emphasizing shock waves in diverse materials, and experimental data. Code predictions and measurements are compared for one system response measure: the shock wave speed through an aluminium bar after impact by an identical aluminium bar. The first approach takes the more traditional statistical approach by comparing many experimental measurements for different shock wave speeds and comparing them with a deterministic code prediction for similar conditions. The second approach propagates parametric input uncertainty through the code to develop a statistical model of the code output. This approach is more appropriate when it is easier, or cheaper, to statistically characterize the uncertainty in the model input parameters and propagate them through the code, than it is to conduct a large number of validation experiments.

In our review of the literature we have stressed two elements. First, the fundamental validation methodology that has been adopted by the AIAA Guide. Second, we have stressed the research that has begun to develop metrics for validation quantification. As can be seen, both validation methodology and quantification are in their early stages of development. In the remainder of this paper we will present our extensions to both validation methodology and quantification.

3. Validation Experiment Methodology

3.1 Construction of a Validation Hierarchy

As discussed in Section 2, the validation hierarchy is constructed starting from the complete system of interest. Stated differently, the validation hierarchy must be application driven, not code driven. As one constructs each lower tier the emphasis on the code increases, but the focus on the actual operating conditions of the complete system should not be lost. The construction of these hierarchical tiers and the identification of the types of experiments that should be conducted at each tier is a formidable task. There are many ways of constructing the tiers; no single construction that is best for all cases. We would draw the analogy of constructing validation hierarchies to the construction of control volumes in fluid dynamic analyses. Many varieties of control volumes can be drawn; some lead nowhere, and some are very useful for the task at hand. The construction should emphasize the modeling and simulation capability that is desired to be validated, whether it be CFD or other computational disciplines. Analogous tier structures can be developed for structural dynamics and electrodynamics, for example, when the engineering system of interest involves these disciplines.

A good hierarchical tier construction is one that accomplishes two tasks. First, it carefully disassembles the complete system into tiers in which each lower level tier has one less level of physical complexity. For complex engineered systems, this may require more than the three building-block tiers shown in Fig. 3. The types of physical complexity that could be uncoupled from one tier to the next are spatial dimensionality, temporal nature, geometrical complexity, and physical process coupling. The most important of these to decouple or segregate into separate effects experiments, from one tier to the next, is physical process coupling. This element commonly contains the highest nonlinearity of the various contributors. It is important to recognize the nonlinear nature of all of the contributors in the construction of the tiers because the philosophy of the tier construction rests heavily on linear system thinking. That is, confidence in the computational capability for the complete system can be built from assessment of computational capability of each of its parts. The complete systems of interest clearly do not have to be linear, but the philosophy of the hierarchical validation approach loses some of its utility for strong nonlinear coupling from one tier to the next.

Second, the individual experiments in a tier should be chosen so that they are practically attainable and able to produce validation quality data. In other words, the individual experiments should be physically achievable given the experimental test facilities, budget, and schedule, and they can produce quantitative experimental measurements of multiple system response measures that can test the code. As discussed in Section 2, the ability to conduct a true validation experiment at the complete system tier is extremely difficult, if not impossible, on complex systems. At the subsystem tier it is usually feasible to conduct validation experiments, but it is still quite difficult and expensive. At this tier, one usually chooses a single hardware subsystem or group of subsystems that are closely related in terms of physical processes or functionality. For complex subsystems, one might want to add a new tier below subsystems called components. As with subsystems, this would be actual operational hardware components. When one defines the individual experiments at the benchmark tier level, then special hardware, i.e., non-operational, non-functional hardware must be fabricated. This tier is probably the most difficult to construct because it represents the transition from a hardware focus in the two top tiers, to a physics-based focus in the bottom tiers of the hierarchy. At the bottom tier, unit problems, one should identify simple geometry experiments that have one element of physical process complexity. As with the subsystem tier, an additional tier may need to be added in order to attain only one element of physics at the bottom tier. Also, the experiment must be highly characterized so as to provide the needed data to the computational code, and it must be conducted so that experimental uncertainty can be estimated precisely. As discussed in Section 2, high quality validation experiments are practically attainable at the benchmark and unit problem tiers.

We will now discuss an example of how to construct a hierarchical tier structure for a complex system. In the only published example of validation hierarchy construction[85] the application was for impeller/diffuser interaction on rotating machinery. For the present example we consider a complex, multidisciplinary system: an air-launched, airbreathing, hypersonic cruise missile. Assume the missile has an autonomous guidance, navigation, and control (GNC) system and an onboard optical target seeker. Figure 4 shows the complete hierarchical validation structure that we will discuss. We refer to the missile as the complete system and the following as systems: propulsion, airframe, GNC, and the warhead. These systems would normally be expected in engineering design of such a vehicle; however, additional elements could be added or the named elements could be subdivided to emphasize systems of importance to the computational analyst. The launch aircraft is not included at the system level because its location in the hierarchy would be at the next higher level, i.e., above the cruise missile.

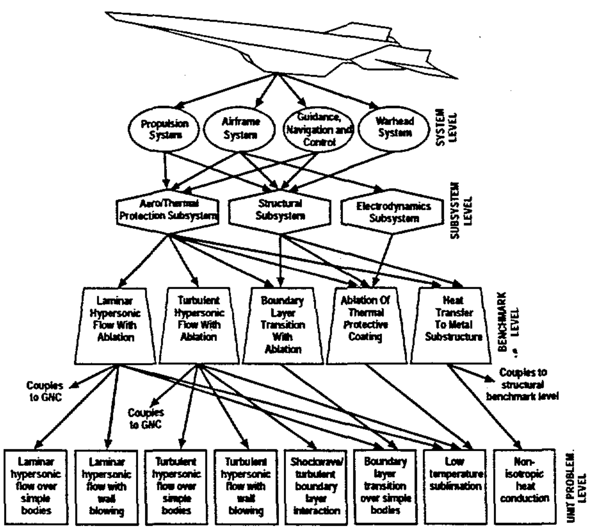

Figure 4: Validation Hierarchy for a Hypersonic Cruise Missile

At the subsystem tier we have identified the following elements: aero/thermal protection, structural, and electrodynamics. Electrodynamics deals with the computational simulation of radio frequency detectability of the cruise missile, e.g., radar cross-section at different viewing angles of the missile. Only these elements are identified at this tier because they are the primary engineering design features that deal with the airframe. Arrows drawn from the system tier elements to the subsystem tier elements indicate the primary elements that influence the lower tier. Recall at this tier that each element should be identified with functional hardware of the cruise missile. Notice, however, how one would begin to conduct validation experiments at this tier depending on the computational discipline of interest. For example, the aero/thermal subsystem would contain the actual thermal protective coating over the metal skin of the missile, the actual metallic skin of the vehicle, much of the substructure under the skin of the vehicle, all of the functional lifting and control surfaces, and the internal flow path for the propulsion system. However, the aero/thermal subsystem probably would not contain any other hardware inside the vehicle, unless some particular heat conduction path was critical. If one were interested in validation experiments for a structural dynamics code, then the structural system identified would be quite different from the aero/thermal subsystem. For example, it would contain essentially every piece of hardware from the missile because every part of the structure is mechanically coupled by every other part of the structure. Structural modes are influenced by all mechanically connected hardware, some to a lesser extent than others. Certain simplifications of the hardware, however, would be appropriate. For example, one could substitute mass-mockups for certain systems, such as the warhead and the completely functional propulsion system, with little loss in fidelity. The structural excitation modes of the propulsion system must still be considered in the structural subsystem.

At the benchmark tier the following elements are identified: laminar hypersonic flow with ablation, turbulent hypersonic flow with ablation, boundary layer transition with ablation, ablation of thermal protection coating, and heat transfer to the metal substructure. At this tier one fabricates specialized, non-functional hardware. For example, the laminar, turbulent, and boundary layer transition elements may not contain the actual ablative coating of the missile, but a simpler material could be used. One that would produce wall blowing and possibly gases or particles that may react within the boundary layer, but yet simpler than the typically complex gas and particle chemistry that results from actual ablative materials. The arrow from the structural subsystem to the boundary layer transition element is drawn to show that structural vibration modes of the surface can influence transition. The element for ablation of the thermal protection coating may use the actual material on the missile, but the validation experiment may be conducted, for example, at conditions that are attainable in arc-jet tunnels. An additional arrow is drawn from each of the elements for hypersonic flow with ablation that are marked "GNC". These arrows indicate a significant coupling of the flow field to the optical seeker in the GNC hierarchy (not shown here). The element for the heat transfer to the metal substructure shows an arrow that would point to elements in the structural subsystem hierarchical tree. This arrow indicates the coupling to the thermal induced stresses and the temperature dependent material properties into the structural simulation.

At the unit problem tier the following elements are identified: laminar hypersonic flow over simple bodies, laminar hypersonic flow with wall blowing, turbulent hypersonic flow over simple bodies, turbulent hypersonic flow with wall blowing, shock wave/turbulent boundary layer interaction, boundary layer transition over simple bodies, low temperature sublimation, and non-isotropic heat conduction. Many other elements could be identified at this tier, but these are representative of the types of validation experiments that would be conducted at the unit problem tier. The identification of elements at this tier are easier than the benchmark tier because these elements are more closely related to traditional experiments in fluid dynamics and heat transfer.

A clarification comment should be made concerning experiments at the lower levels of the hierarchy, particularly at the unit problem level. Some have referred to experiments, such as laminar hypersonic flow in a wind tunnel, as a "simulation" of the complete flight vehicle in the atmosphere. From a project engineer's point of view, this is appropriate. From a validation experiment point of view, this confuses the issue. That is, an experiment conducted at the benchmark or unit problem level *is* a physical realization of a process whose results are compared to a computation *simulation* of the actual physical experiment conducted. The relationship of the physical experiment to some higher level complex system is immaterial with regard to comparison of computation and experiment.

Even after a validation hierarchical structure such as illustrated in Fig. 4 has been constructed, another issue remains: identifying which validation experiments are the most important and, as a result, should be conducted first. To aid in prioritizing the validation experiments, we recommend a procedure that has been developed for nuclear power reactor safety assessment. In reactor safety, a procedure referred to as the Phenomena Identification Ranking Table (PIRT) has been developed for prioritizing which physical phenomena are the most important to analyze and understand.[122] This procedure focuses attention on the application of the code to the operating conditions of interest for the complete system. Although this procedure has not been used in the aerospace industry, we believe the PIRT procedure can be used in conjunction with the present hierarchical structure to aid in prioritizing validation experiments.

To better explain how the validation hierarchy of the airframe system is related to the validation hierarchy of the propulsion, GNC, and the warhead systems, consider Fig. 5. The validation hierarchy of each of the four systems could be viewed as the primary facets of a four-sided pyramid. The airframe facet divides into three additional facets, each representing the three subsystems: aero/thermal protection, structural, and electrodynamics. The propulsion system could be divided into four additional facets that could represent the subsystems: compressor, combustor, turbine, and thermal signature. On the surface of this multifaceted pyramid one could more clearly and easily indicate the coupling from one facet to another. For example, the coupling of laminar and hypersonic flow with ablation to the optical seeker on the GNC facet of the pyramid would be shown by an arrow connecting these elements on different facets of the pyramid.

The validation pyramid stresses the system viewpoint, as opposed to a specific discipline viewpoint, in modeling-and-simulation-based design. Each facet of the pyramid can then be devoted to identifying validation experiments for each computational code responsible for part of the design of the system. As one traverses around

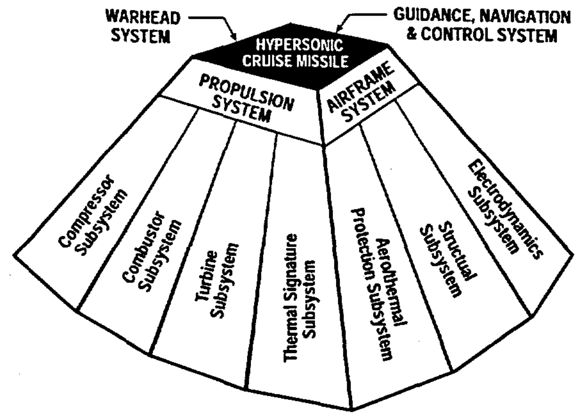

Figure 5: Validation Pyramid for a Hypersonic Cruise Missile

the top of the pyramid, the number of facets is equal to the number of systems that are identified. As one traverses around the bottom of the pyramid, the number of facets is equal to the total number of major computer codes used in the analysis of the system, i.e., the number of codes that require validation activities. For the hypersonic cruise missile example, if the code that simulates surface ablation is a separate code from the aerodynamics code, then an additional facet on the pyramid is added on the aero/thermal subsystem facet. We strongly believe this type of system-level thinking is necessary to increase the confidence in complex systems that are designed, manufactured, and deployed with reduced levels of testing.

Two final comments are in order concerning the construction of a validation hierarchy. First, the location of a particular validation experiment within the hierarchy must be determined relative to the complete system of interest, i.e., it must be appropriately related to all of the experiments above it, below it, and in the same tier. Stated differently, the same validation experiment can be at different tiers for validation hierarchies that are constructed for different complex systems of interest. For example, the same turbulent separated flow experiment could be at the unit problem tier in a complex system and at the benchmark tier in a simpler engineering system. Second, a validation hierarchy is constructed for a particular engineered system operating under a particular class of operating conditions, for example, normal operating conditions. A new hierarchical pyramid would be constructed if one were interested in computationally analyzing other classes of system operating conditions. For example, if one where interested in failure or crash scenarios of the system, then one would construct a different pyramid because different modeling and simulation codes would come into play. Another example is if the system would have to function under hostile conditions, e.g., under physical or electromagnetic attack, then a different pyramid would also be constructed.

3.2 Characteristics of Validation Experiments

Many researchers, analysts, and designers ask the question: "What is a validation experiment?" or "How is a validation experiment different from other experiments?" These are appropriate questions. We suggest that traditional experiments could generally be grouped into three categories. The first category comprises experiments that are conducted primarily for the purpose of improving the fundamental understanding of some physical process. Sometimes these are referred to as physical discovery experiments. Examples of these are experiments that measure fundamental turbulence characteristics, detailed reaction chemistry in combustion experiments, and experiments for flows in thermochemical nonequilibrium. The second category of experiments are those conducted primarily for constructing or improving mathematical models of fairly well understood flows. Examples are experiments to measure model parameters in finite rate reacting flows, shock wave-boundary layer interaction, and measurements to determine thermal emissivity of particles or surfaces. The third type of traditional experiment includes those that determine or improve the reliability and performance of components, subsystems, or complete systems. These are commonly called tests of engineered components or systems. Examples of these are tests of new combustor designs, compressors, turbopumps, gas turbine engines, and rocket engines. We would also include in this category traditional wind tunnel tests conducted for the purpose of measuring vehicle and control-surface forces and moments.

We argue that validation experiments are a new type of experiment. They are conducted for the primary purpose of determining the validity, or predictive accuracy, of a computational modeling and simulation capability. That is, an experiment that is designed, executed, and analyzed for the purpose of quantitatively determining the ability of a mathematical model and its embodiment in a computer code to simulate a well characterized physical process. In other words, in a validation experiment "the code is the customer." Only during the last 10 to 20 years has computational simulation matured to the point where it could even be considered as a customer. As modern technology increasingly moves toward engineering systems that are designed, and possibly even fielded, based on modeling and simulation, then modeling and simulation itself will increasingly become the customer of experiments.

As reviewed in Section 2, a number of researchers, particularly experimentalists, have been slowly developing the concepts of a validation experiment. A group of researchers at Sandia National Laboratories has been developing philosophical guidelines and procedures for designing and conducting a validation experiment. Although the following six guidelines and procedures were developed in a joint computational and experimental program conducted in a wind tunnel, they apply over the entire range of fluid dynamics:[9,14,15,65,80,123,124].

Guideline 1: A validation experiment should be jointly designed by experimentalists and code developers or code users working closely together throughout the program, from inception to documentation, with complete candor about the strengths and weaknesses of each approach. No withholding of limitations or deficiencies is permitted, and failure or success of any part of the effort must be shared by all. Without this level of cooperation, openness, and commitment in a team environment, the process is likely to fail or fall short of its potential.

Although this may sound easy to do, it is *extraordinarily* difficult to accomplish in practice. Some reasons for the difficulty have been discussed in Section 1. We give just two examples of why this is difficult to accomplish in reality between organizations and within an organization. First, suppose that the CFD team is in one organization and the wind tunnel facility is in a completely separate organization, e.g., the wind tunnel facility is contracted by the organization of the CFD team to conduct the experiments. The wind tunnel facility will be extremely reluctant to expose its weaknesses, limitations, or deficiencies, especially if the facility was, or will be, in competition with other facilities to win contracts. In our experience, we have learned that validation experiments require much greater depth of probing of the limitations of facility and the limitations of a CFD capability than any other experiment.

Second, suppose that the CFD team and the wind tunnel team are sister organizations in the same corporation. Here also, both teams will be very cautious to discuss weaknesses in their CFD or experimental facility capability. Because of the background and technical training of theoretical and experimental personnel, significant cultural hurdles must be overcome in dealing with one another. One of the most common detriments to achieving close teamwork and openness between computationalists and experimentalists is competition between the CFD organization and experimental organization for funding or recognition. If it is advantageous for one of the organizations to diminish the image or reputation of the other organization in the "computers versus wind tunnel" mentality, there can be little chance for a validation experiment. Management must make it clear, in word and deed, to both the CFD team and the experimental team that there is no success or failure of either side; there is only success or failure of the joint endeavor.

Guideline 2: A validation experiment should be designed to capture the essential physics of interest, including all relevant physical modeling data and initial and boundary conditions required by the code. By essential physics of interest we mean spatial dimensionality, temporal nature, geometrical complexity, and physical flow processes. For example, is one interested in conducting a 2-D experiment and computation, or is a full 3-D experiment and computation required? In this context, we note that no physical experiment can be truly planar 2-D; axisymmetric or spherical 2-D experiments are closely attainable. Experimentalists must understand the code assumptions so that the experiment can match, if possible, the code assumptions and requirements. If the parameters that are initially requested for the calculation cannot be satisfied in the proposed experimental facility, it may be feasible to alter the code inputs and still satisfy the primary validation requirements. Or, it may be necessary to look elsewhere for a facility. For example, can the CFD-requested boundary-layer state on a model be ensured? Is the type and quantity of instrumentation appropriate to provide the required data in sufficient quantity and at the required accuracy and spatial resolution? Conversely, CFD analysts must understand the limitations of the physical experiment, ensure that all the relevant physics are included, and define physically realizable boundary conditions. As noted above, the level of detailed understanding that is required can be achieved only if the validation experiment is planned and conducted as part of a team effort.

The most common reason that published experimental data can be of only limited use in the validation of CFD codes is that insufficient detail is given in the documentation of the experiment concerning physical modeling parameters, initial condition, and boundary conditions. By physical modeling parameters we mean quantities such as thermophysical quantities of fluids and solids, flow rates, surface roughness, and particle size distribution in two-phase flow. Published data in corporate reports or university reports occasionally contain sufficient characterization of the experiment, but rarely is sufficient detail contained in journal articles and conference papers. The following are a few examples of the level of detail that is needed for boundary conditions on a vehicle in a wind tunnel. Differences will exist between the nominal and the actual model dimensions (e.g., straightness and out-of-round), surface condition, location of instrumentation, and angular orientation in the wind tunnel. An important detail for aircraft configurations, wings, and deformable bodies is the measurement of the actual deformed geometry under the load, or an accurate calculation of the deformed structure. In long-duration hypersonic wind tunnels the deformation of the vehicle due to aerodynamic heating should also be estimated or measured, and then reported. Also for hypersonic tunnels, surface temperature can vary significantly with time and space over the surface of the body.

In wind tunnels, the highest priority for boundary conditions is probably calibration of the free-stream flow. This typically means spatially averaged quantities over the test section volume of the tunnel, such as free-stream Mach number, total pressure and static pressure, and total temperature and static temperature. For turbulent flow and transition simulations, the calibration should also include free-stream turbulence intensity, scale, and frequencies. Some facility managers may be reluctant to share such detailed flow-quality data with users (and competitors). However, for CFD validation experiments, these data are critical.

For supersonic wind tunnels, the spatial nonuniformity of the flow in the test section could be used to set the flow properties as location-dependent boundary conditions just upstream of the bow shock wave of the vehicle. Such a procedure, although conceptually feasible, has not yet been demonstrated, to the authors' knowledge. This approach might appear excessive at this stage of CFD code development for validation experiments in wind tunnels of high-flow quality. However, we believe this approach is necessary for validation experiments in high-enthalpy-flow facilities in which rapid expansions combine with finite-rate chemistry. In such facilities, the flow is typically highly nonuniform and poorly characterized, which makes it extremely difficult, if not impossible to accurately compare experimental data to code predictions.

For subsonic wind tunnels the question of boundary conditions becomes much more complex. For low-speed wind tunnels, even with low levels of blockage, one of the first issues that must be addressed by the CFD analyst is "Should I model the flow in the entire test section of the tunnel, or assume an infinite size tunnel?" This question could be restated as "For the quantity that will be measured in the tunnel and compared with the CFD computation, what is the change in this quantity if I make a finite versus infinite size tunnel assumption?" Although the authors have not seen any detailed analyses addressing this question, from our limited experience we believe that the sensitivity to tunnel blockage will be significant even at low blockage. For transonic flow tunnels, the large sensitivity of test section measurements to solid walls versus perforated walls is well known. For perforated-wall tunnels, the tunnel wall boundary conditions are very poorly characterized. We believe that a significant computational and experimental research program is needed to improve the mathematical modeling of perforated-wall tunnels. Unless this is done, the ability to conduct high quality validation experiments in perforated-wall tunnels will greatly suffer. Solid wall tunnels provide a well characterized boundary condition for the CFD analyst, even if wall interference and wave reflections occur in the tunnel. This point emphasizes one of the key differences mentioned above concerning the difference between a test and a validation experiment: the code is the customer, not the project group which is only interested in zero-wall interference.

Guideline 3: A validation experiment should strive to emphasize the inherent synergism between computational and experimental approaches. By a "synergism," we mean an activity that is conducted by one approach, whether CFD or experiment, but which generates improvements in the capability, understanding, or accuracy of the other approach. And in this exchange, both computational and experimental approaches benefit. Some who are discovering the benefits of validation experiments claim that this is the primary value of validation experiments. Discovering the strong positive reinforcement of computation and experiment working closely together can be surprising, but validation experiments contribute much more than this. We give two examples of how this synergism can be exemplified.

First, the strength of one approach can be used to offset a weakness of the other approach. Consider the example of perfect-gas, laminar flow in a supersonic wind tunnel. Assume that a wind-tunnel model is designed so that it can be easily re-configured from simple to complex geometries. For the simple geometry at a low angle of attack one should be able to compute solutions with very high confidence, exclusive of the separated flow in the base region. This may require independent CFD codes and analysts, but it is certainly possible with present CFD technology. With very high accuracy solutions one can then compare them with wind tunnel measurements to detect a wide variety of shortcomings and weaknesses in the facility, as well as errors in the instrumentation and data recording system. In our experience we have demonstrated that this can occur. The high accuracy solution can also be used for *in situ* calibration of the free-stream flow in the test section. When the complex vehicle geometry is tested one may have strongly 3-D flows, separated flows, and shock wave-boundary layer separation. The experimental measurements would then be more accurate than the CFD simulation. The complex geometry case would then be viewed as a validation experiment to test the code.

Second, one can use CFD simulations in the planning stages of a validation experiment to dramatically improve the design, instrumentation, and execution of the experiment. For example, one can compute shock wave locations and their impingement on a surface, separated flow and reattachment locations, high heat flux regions, and vortical flows near a surface. This allows the experimentalist to improve the design of the experiment and especially the type and location of instrumentation. One can take this strategy a step further by optimizing the design of the experiment so as to most directly stress the code, i.e., design the experiment to "break the code." This can be done by optimizing the physical modeling parameters, such as Reynolds number and Mach number, modifying the boundary conditions, such as geometry and wall surface conditions, and also changing the initial conditions on an initial-value problem. Sometimes the code developers cannot get too excited about this strategy.

<u>Guideline 4: Although the experimental design should be developed cooperatively, complete independence must be maintained in obtaining both the computational and experimental results.</u> The reason for this recommendation, of course, is that it is so common that CFD codes are calibrated to the experimental measurements that many people do not recognize when they are calibrating versus validating.[16] As example of this, CFD analysts have been known to say: "Why should I do anymore grid refinement when the code agrees with the experimental data?" The ability of the CFD analysts to calibrate results is infinitely greater than the experimentalist.

It is difficult to accomplish the close cooperation of the CFD analysts and the experimentalists and yet keep the independence of each result. However, it can be done by careful attention to procedural details. We give the following example. We have stressed in preceding examples the close working relationship needed in the design and execution of the validation experiment. However, when the experimental measurements are reduced and analyzed, the CFD team should not be given the results initially. They should be given the complete details of the physical modeling parameters and the initial and boundary conditions of the experiment, *exactly* as it was conducted. That is, everything that is needed for the CFD analysts to compute solutions must be provided; but no more. The CFD analysts must quantify the errors and uncertainties in the CFD solution and then present the results for comparison with experiment. (Computational error and uncertainty quantification will be discussed in the Section 4.) When that is completed, the validation test between computation and experiment is made.

After analyzing and discussing the comparison, there can be various outcomes. If the error and uncertainty bounds on the computational side are very narrow, there is a large amount of data to estimate the mean values of the measurements, and the agreement is uniformly good, then one can crisply conclude whether the computation is validated or invalidated. (Technically one can not validate the code in a general sense: one only can claim "not invalidated.") In our experience, this rarely happens. The more typical outcome is that there will be agreement on some measured quantities and there will be disagreement on other quantities. Or, one side or the other will say: "I need to go back and check a few things." After further checks, sometimes the agreement will improve; sometimes it won't. This discussion and iterative process is very beneficial to both sides of the team. We have found that the discussions will always lead to a deeper understanding of both the computations and the experiment.

As a final procedural comment, we recommend that management *not* be involved in the initial comparisons and discussions. The discussions should just involve CFD analysts and experimental staff. Nothing will poison the teamwork more quickly than one side of the team telling management: "Look how far off they were."

<u>Guideline 5: A hierarchy of experimental measurements of increasing computational difficulty and specificity should be made, for example, from globally integrated quantities to local measurements.</u> As one moves from global to locally measured quantities, the challenge to the CFD analysts and the experimentalists increases significantly. In wind-tunnel experimentation for a flight vehicle, the following hierarchy is suggested:

- Total body forces and moments
- Control surface forces and moments
- Surface pressure distributions
- Surface heat flux, shear stress, or both
- Flow field distributions of pressure, temperature, and velocity components
- Flow field distributions of Reynolds stresses

The ordering of difficulty in the above hierarchy indicates that each lower level, i.e., higher level of complexity, is either the spatial or temporal derivative of the level above it, or it is a subset of the level above it. In other words, the integration or selection of a larger set is a powerful mathematical smoothing process. Thus, there is a relationship between this hierarchy and the levels of credibility in validation activities. That is, in the "process of determining the degree to which a model is an accurate representation of the real-world," one must specify what physical quantities, or system response measures, have been validated. For example, validation of total body normal force *does not* imply that surface heat flux to the model has been validated to the same degree of accuracy. There are two separate reasons for this. First, the physical modeling fidelity that is required to predict these two quantities is remarkably different. For example, total body normal force on many geometries can be computed fairly accurately with inviscid flow, whereas the difficulty of predicting surface heat flux in a turbulent boundary layer is well known. Second, the computational grid size that is required to predict these two quantities is strikingly different. For example, consider a steady, compressible, laminar, attached flow over a delta wing at a low angle of attack. To achieve the same level of computational accuracy for total body normal force as compared to surface heat flux, one would need approximately a factor of 100 to 1,000 times the number of grid points for the heat flux computation as compared to the normal force computation.

The predictive difficulty for a code that is illustrated by the above hierarchy is also very similar to the difficulty in experimentally measuring each of these quantities. The experimental uncertainty increases as one proceeds down this list, probably at a factor of two for each level of the hierarchy. With the recent development of experimental techniques such as particle-imaging-velocimetry (PIV) and planar-laser induced-florescence (PLIF), we believe there is a significant increase in quantity of flow field velocity measurements attainable. The quantity of data, for example, over a large volume surrounding a body, permits much more stringent tests for CFD code validation than simply measurements at individual locations.

The last recommendation concerning the hierarchy of measurements is that in a validation experiment one should, if possible, make measurements at multiple levels of the hierarchy. That is, do not design the experiment with the philosophy that only one detailed level of measurements will be made. The more complicated the flow field or the physical processes taking place in the flow field, the more important this recommendation is. For example, on a complex turbulent, reacting flow, do not just measure surface heat flux over a portion of the body. Also measure flow field temperatures and surface pressures over a different portion of the body. Flow field visualization and surface flow visualization can also provide valuable additional pieces of information. With sophisticated post-processing capability, the CFD solution can be used to simulate the experimental flow field visualization and surface flow visualization. For example, the computed flow field solution can be used to compute a Schlieren photograph that can then be compared with the experimental photograph.

<u>Guideline 6: The experimental design should be constructed to analyze and estimate the components of random (precision) and bias (systematic) experimental errors.</u> The standard technique for estimating experimental uncertainty in wind-tunnel data has been developed over a number of years by members of the AGARD Fluid Dynamics Panel.[77] The standard procedure is well documented in a recent AIAA standards document[125] and also in the text of Coleman and Stern.[126] We believe it is the minimum level of effort required for uncertainty estimation in validation experiments. The standard technique propagates components of random and bias uncertainty through the entire data flow process. The technique estimates these components and their interactions at each level of the data flow process, from the sensor level to the experimental result level. As with all experimental uncertainty estimation techniques, the ability of estimating random uncertainty is much better than estimating bias uncertainty.

During the last 15 years, we have developed an experimental uncertainty estimation procedure that takes a very different approach from the standard wind-tunnel procedure. Instead of propagating individual uncertainty components through the data flow process we have taken an approach referred to in the <u>AIAA Standard</u> as an "end-to-end" approach. Our approach compares multiple experimental measurements for the same experimental quantity and then statistically computes the uncertainty. The traditional approach could be viewed as an *a priori* approach, whereas ours is an *a posteriori* approach. Just as in comparing *a priori* and *a posteriori* error estimation in CFD, we believe our procedure provides not only a better estimate of random and bias errors, but it is also able to quantify important contributions due to components that cannot be estimated in the traditional approach. We will discuss this new procedure in detail in the next subsection.

As a final comment on experimental uncertainty estimation, we recommend that the same validation experiment be conducted, if possible, in different facilities. For example, in a wind tunnel experiment the same physical model should be used and the experiment conducted at the same freestream conditions. Satisfactory agreement of results from different facilities lends significant confidence that there are no inadequately understood facility-related bias errors in the data, e.g., condensation effects, wave focusing, and excessive flow angularity. This procedure, especially for simple model geometries, would also serve to uncover inaccuracies and inconsistencies in the flow calibration data for each facility that is used. The same personnel should oversee the execution of the experiment at each site, and these personnel should also have access to all operational and performance data in the facility. When the same model is used in different facilities surprising results are always discovered, usually to the dismay of the facility owner.

3.3 Experimental Uncertainty Estimation

A nontraditional approach has been developed by the first author and his colleagues for estimating random and bias experimental errors.[9,14,15,58,123] Although the approach is new to wind tunnel testing, it is based upon accepted statistical uncertainty estimation techniques.[127,128] Our approach uses statistical methods to compute both the random and correlated bias uncertainties in the final experimental result. Our method uses symmetry arguments for the flow in the test section and symmetry of the model, and then conducts carefully selected comparison runs. The method uses the fact that a freestream flowfield that is uniform has an infinite number of planes of symmetry. Also, the method uses the fact that a perfectly constructed wind tunnel model commonly has one plane of symmetry and some have a larger number of planes of symmetry. For example, a typical aircraft has one plane of symmetry, and a four-finned missile has four planes of symmetry. Using these symmetry arguments and then comparing certain experimental measurements in the real wind tunnel and on the real model, one can statistically compute these correlated bias error contributions to the total uncertainty.

Even though our procedure improves the estimate of the total random uncertainty compared to the traditional approach, we believe the most important contribution is that it improves the estimate of correlated bias errors. The damaging effect of bias errors was sharply emphasized in the classic paper by Youden.[129] He pointed out that systematic errors commonly exceed the estimate of random errors, yet the magnitude of systematic errors is normally unknown. As an example, he lists 15 experimental measurements of the Astronomical Unit over the period from 1895-1961. Each experimentalist estimates total uncertainty in his or her measurement. In every case, the next measurement made of the AU was *outside* the experimental uncertainty of the predecessor. Youden states "If we have accurate knowledge of the magnitude of the systematic errors in the components of the equipment much of the discrepancy among results from different investigators would be accounted for." Youden believes that the most effective method for estimating systematic errors is to conduct experiments by multiple investigators, with different equipment, and with different techniques. This method, of course, is quite expensive and time consuming.

We have demonstrated our approach on four different sets of experimental data: three of the data sets were for surface pressure measurements and one was for body forces and moments. The method has been applied to three hypersonic wind tunnels: Tunnels A and B of the von Karman Gas Dynamics Facility at the U. S. Air Force Arnold Engineering Development Center, Tullahoma, Tennessee, and the Hypersonic Wind Tunnel Facility at Sandia National Laboratories, Albuquerque, New Mexico. The method showed that, even in these high quality flowfield facilities, the largest contributor to experimental uncertainty was due to flowfield nonuniformity. It was shown that the flowfield nonuniformity can be up to three standard deviations higher than the random (total instrumentation) uncertainty. In terms of the percent of total (random and bias) estimated uncertainty, the flowfield nonuniformity can be up to 90% of the total, whereas the random uncertainty is 10%.

The method relies on careful construction and execution of the wind tunnel run matrix so that combinations of runs yield information on both random and bias errors. For body force and moment measurements, we refer to the random uncertainty as the uncertainty caused by all of the following components and their interactions: strain gage hysteresis, nonlinearity, thermal sensitivity shift, and thermal zero shift; the analog data reduction system; the data recording system; model pitch, roll, and yaw alignment; model geometry imperfections; run-to-run variations in setting freestream conditions in the test section; and base pressure transducers and instrumentation for eliminating base drag. That is, the random uncertainty combines all uncertainty components in the entire experiment except those due to test section flow field nonuniformity. In Ref. 125 this is referred to as an end-to-end estimate of random uncertainty in the experimental result. To calculate the random uncertainty, one compares all possible combinations of body force and moment measurements that are made for the same physical location in the test section. We refer to this type run-to-run comparison as a repeat run comparison. Immediately repeating a particular case yields statistical information on short-term facility repeatability. Repeating runs in varying order, on different days, after the model has been disassembled and reassembled, and in separate facility entries can uncover subtle errors that are related to facility operations, specific personnel, time of day, etc. Repeat runs require careful introspection in their selection and sequence and are critical to an assessment of statistical precision of the data. Repeat runs are not afterthoughts in this approach; they are essential elements in the method and must be incorporated into the experimental plan.

For an experiment measuring surface pressures on a body we refer to the random uncertainty as that caused by all of the following random errors and their interactions with each other: pressure sensor hysteresis, nonlinearity, sensitivity drift, and zero shift; reference pressure variation; analog amplifier system variation; data digitizing and recording system variation; model pitch, roll and yaw alignment variation; variations in freestream Mach number and Reynolds number within a run; variations in freestream Mach number and Reynolds number from run to run. The random uncertainty combines all experimental uncertainty in the entire experiment, except that due to test section flow field nonuniformity and model geometry imperfection uncertainty. To calculate the random uncertainty, one compares pressure measurements for the same pressure port from different runs with the model at the same physical location and orientation in the test section. For the same angle of attack, roll angle, flap deflection angle, and axial location, each pair of port measurements compared will have the same location in the vehicle-induced flow field. When differences in pressure port measurements are made in this way the uncertainty due to flow field nonuniformity and model geometry imperfection cancels out.

Test section flow field non-uniformity uncertainty is uncertainty in surface pressure measurements caused by the following: nonuniformity of freestream flow in the test section; bias errors in the alignment of the model in pitch, roll, and yaw. The uncertainty in an experimental measurement due to a combination of test section flow field nonuniformity uncertainty and random uncertainty is computed by comparing measurements made at different locations in the test section. For example, on a surface pressure measurement experiment, the combined flow field nonuniformity and random uncertainty is calculated by comparing surface pressures for the same port on the body at the same relative location in the vehicle flow field, but at different locations in the test section. This procedure will not include any uncertainty due to model imperfections because by using the same ports for both comparisons, this uncertainty component cancels in taking the difference between the two measurements.

Imperfect model geometry uncertainty can only be determined by measuring local surface quantities on a body of revolution, i.e., a body that has an infinite number of symmetry planes. For example, imperfect model geometry uncertainty in a surface pressure experiment are those caused by the following: model geometry deviations (measurable deviations of the physical model from the conceptual, or mathematical, description of the model); and model/sensor installation imperfections (poorly fabricated or burred pressure orifice, and a pressure leak between the orifice and the transducer). Imperfect model geometry uncertainty, along with random uncertainty, is computed by comparing surface measurements for different transducers, with both transducers sensing at the same physical location in the test section and at the same relative location in the vehicle flow field. This requirement can only be met on bodies of revolution. This procedure will yield the combined model geometry and random uncertainty, but will not include any uncertainty due to flow field nonuniformity.

The dominant contribution of nonuniform flow to experimental uncertainty has been suspected by wind tunnel experimentalists,[125] but not until use of our procedure has it been quantified. We strongly suspect that the largest contribution to measurement uncertainty in most, if not all, near-perfect gas hypersonic wind tunnels is due to flow field nonuniformity. Although this technique has not been applied to any other wind tunnel to our knowledge, we believe the dominance of flow field nonuniformity error will also occur in other facilities. We believe the nonuniform flow contribution will be even a higher percentage of the total experimental uncertainty in transonic flow wind tunnels and shock tunnels. We encourage others to use the present statistical method to determine if this is the case. The critical assessment of one's own facility, however, will be viewed as a risk by many facility managers and owners. Some will choose to avoid the risk. We strongly believe that critical assessment of experimental uncertainty is just as important as critical assessment of computational error and uncertainty. Assessment of computational error and uncertainty will be discussed in the next section. In Section 5, it will be shown that the accuracy of the experimental uncertainty estimate is critical because it sets the limit on quantitative assessment of validation.

4. Computational Error and Uncertainty Quantification

In this section we will discuss the role of computational error and uncertainty in validation. There are two reasons for being concerned about the numerical accuracy of the computational solution of the model that is being subjected to validation. The first reason is obvious: non-zero computational error contributes to the discrepancy between a calculation and experimental data used in validation. The second reason is related to the first, but has more subtle implications. That is, numerical error needs to be distinguished from errors that arise from inaccurate conceptual models. In other words, ideally we should demonstrate that numerical error is small before we even begin to attempt to assess the adequacy of the conceptual model for the intended application. For example, in the absence of some kind of compelling evidence that the computational errors are small, a rational possibility is that a large computational error might cancel with a conceptual model error to give us evidence for the accuracy of the conceptual model in a validation study. Or, a significant discrepancy between experiment and computation for a particular validation study might be concluded to reside in the conceptual model, when in fact it lies in numerical error.

It is difficult, if not impossible, to *rigorously* demonstrate that numerical error is small for complex models. There are many reasons for this, which we will discuss briefly below. However, the implication of the absence of some kind of convincing proof that the conceptual model is being solved accurately by the code is that we are driven to an analysis that is quite similar philosophically to validation. That is, we aim to understand what the computational error is for given calculations or types of calculations for given codes. In many cases, this error will be understood to be small enough to make the pursuit of validation analyses sensible. Also, we expect that over time we will accumulate increasing amounts of evidence, through the testing and application process of the code, that the numerical error is small for important classes of computations.

In this section we briefly discuss some of the error and uncertainty estimation issues affecting our ability to perform validation. First, we will discuss what we specifically mean by error and uncertainty in calculations, including appropriate definitions, what we mean by estimating these quantities, and their role in validating calculations. Then, we turn our attention to a summary of approaches to estimating computational error. This discussion also makes it imperative to address some remarks to the problem of working with coarse grids. By coarse grids we means grids which are either known to be inadequate, or believed to be inadequate, to capture the continuum mathematics of the PDEs. We will end our discussion with a brief summary of methods that can be used to quantify uncertainty in the output quantities of computational models.

Finally, we want to emphasize one aspect of computational error that we will not discuss in this section. Given appropriate discretizations of the PDEs and algorithms for solving those discretizations, one must still implement those algorithms in a computer program (code). The computer implementation itself is a rich source of computational errors, especially as the computer program becomes more complex. Roache has pointed out that code generation and software testing are important considerations in verification.[130] Software errors contribute to computational error and must be eliminated to have high confidence in computational error estimates. As the software becomes more complex and general purpose, to find and eliminate this kind of error through software testing is increasingly a domain of software engineering.

4.1 Uncertainty and Error in Computation

As discussed in several contexts above, uncertainty in computations influences our ability to perform validation. However, the origins of uncertainty go well beyond the discretization, algorithm, and coding issues that see themselves properly at the center of the computational error problem. As defined in Ref. 131 and the AIAA Guide, *uncertainty* is "a potential deficiency in any phase or activity of the modeling process that is due to lack of knowledge." A complete discussion of uncertainty, also referred to as epistemic or reducible uncertainty, is well beyond the scope of this article. Rather, we will focus on issues of uncertainty that most directly influence validation assessment.

The first feature that is worth emphasizing in this definition of uncertainty is the word "potential," meaning that the deficiency may or may not exist. In other words, there may be no deficiency, say in the prediction of some event, even though there is a lack of knowledge if we happen to model the phenomena correctly. The second key feature of uncertainty is that its fundamental cause is incomplete information. Lack of knowledge commonly exists in the poor understanding of complex physical processes, imprecisely defined or vague knowledge of sequential scenarios (e.g., failure scenarios), quantities or parameters that are required for computational modeling but are not measured, and finally, conflicting evidence for quantities, parameters, and physical processes. Probability theory is commonly used to mathematically represent uncertainty, but more appropriate theories include evidence theory, possibility theory, fuzzy set theory, and imprecise probability theory.[132-135]

As in Ref. 131, we distinguish between uncertainty and variability. *Variability* is the inherent variation associated with the physical system being modeled. In the literature, variability is also referred to as stochastic or irreducible uncertainty. Sources of variability can be singled out from other contributors to computational nondeterminism by their representation as distributed quantities that can take on values in an established or known range, but for which the exact value is not known. For example, variability could be present in a calculation through the definition of a part with known manufacturing tolerance variability. Variability is mathematically quantified through the use of a probability distribution, assuming that there is sufficient information to permit this.

The importance of a particular source of variability is problem dependent. The role that variability plays in assessing computational error depends on the sensitivity of the model in general, and its results for particular calculations, and on the source of variability. A macroscopic calculation of two bodies in sliding contact will not be dependent on a stochastic field representing the variability of the contacting surfaces. However, a microscopic model of friction certainly will be sensitive to this representation. When the computational model is sensitive to the variability then the role of variability must be quantified and its contribution to validation determined. This requires stochastic techniques, which we will briefly discuss in Section 4.3.

Ref. 131 and the AIAA Guide define *error* to be: "A recognizable deficiency in any phase or activity of modeling and simulation that is not due to lack of knowledge." This definition stresses the feature that the deficiency is identifiable or knowable upon examination; i.e., the deficiency is not caused by lack of knowledge. From these definitions one can see that variability and uncertainty are somewhat related, but error clearly has different characteristics. Variability and uncertainty are normally thought to produce stochastic, or non-deterministic effects, whereas errors commonly yield a reproducible, or deterministic, bias in the simulation. Error, of course, is the focus of classical numerical analysis in regard to computational modeling. The interested reader should consult Ref. 136 to see an illustration of a general methodology that addresses variability, uncertainty, and error in one framework.

The computational error we discuss in this paper, *acknowledged errors*, is characterized by an approach or ideal condition that is considered to be more accurate. Examples of acknowledged errors are finite precision arithmetic in a computer, approximations made to simplify the modeling of a physical process, and conversion of PDEs into discrete equations. Therefore, this error can certainly be quantified or measured in principle. This measurement could proceed by comparison with a test problem or series of test problems, for example, or by a careful convergence assessment in a given application. This latter approach will be discussed in greater detail in Section 4.2. It might also be the case that quantification of the error can not be achieved for reasons that have nothing to do with our ability to recognize that the error is present. For example, we know that errors are introduced by a discretization. However, if we cannot perform a detailed grid convergence study, we may have no quantitative basis for measuring that error.

If divergence from the correct or more accurate approach is observed by one means or another, the divergence may be either corrected or allowed to remain in the model. It may be allowed to remain because of practical constraints, such as the error is acceptable given the requirements, or the cost to correct it is excessive.

Examples of *unacknowledged errors* are blunders or mistakes. There are no straightforward methods for estimating, bounding, or ordering the contribution of unacknowledged errors. The most common techniques for detecting unacknowledged errors are procedural methods. For example, formal software inspections are a procedure for detecting unacknowledged errors in code creation. A similar inspection of code input data performs the same purpose for the user of the software.

4.2 Error Estimation Methods

We now briefly turn to the issue of estimating computational error or, more precisely, the error that results from the discretization of the underlying partial differential equations. As stated above, a large part of Ref. 130 addresses this issue as a significant component in calculation verification. Although we wish to emphasize a point of view that is strongly validation centered, we will make essentially the same assumptions that Roache does in his discussion:

1) That the software expressing the numerical algorithms has been "verified," or that we at least have substantial evidence to suggest that the software is performing properly.

2) That the discretization is formally consistent with the partial differential equations. In other words, in the asymptotic limit as temporal and spatial discretization spacing tends to zero, the discrete equations converge to the proper partial differential equations.

Consistency by itself does not guarantee that the solution of the discrete equations will converge to the correct solution of the partial differential equations. Additional information is required for consistency to yield convergence.[137,138] We restrict our attention to initial-value problems for linear partial differential equations so that we can state the following famous result:

> Lax Equivalence Theorem: Given a properly posed linear initial-value problem for an evolutionary partial differential equation, and a finite-difference approximation to it that satisfies consistency, then the approximation converges to the correct solution if and only if the approximation is stable.

The assumption of linearity is critical, because it already hints at the weaknesses of *a priori* error estimates for CFD calculations. Richtmeyer and Morton[137] go on to show that consistent discretizations have positive order of convergence. In other words, for an appropriate norm, and under technical domain assumptions on the initial data, in the limit of spatial and temporal discretization increments tending to zero we have $\| u_{exact} - u_{disc} \|$ bounded by a term on the order of a positive power of the spacing. We write this as:

$$\left| u_{exact} - u_{disc} \right| = O(\Delta t^p, \Delta h^q) \qquad \text{Eq. (1)}$$

where p > 0 and q > 0. In Eq. (1) we have not specified what the norm is. It could be an L^2 norm or a supremum norm, for example. Formal analysis of discretizations of partial differential operators typically yields estimates of the type given in Eq. (1). In fact, one seeks methods in which p and q are actually large, for example equal to two or higher. Such estimates are *a priori* estimates of computational error, assuming of course that the convergence implied by the conditions of the Lax Equivalence Theorem actually takes place. We will refer to this specifically as *solution convergence*, as opposed to grid convergence.

Recall that the scheme must be stable as well as consistent for solution convergence. Even assuming that an estimate such as Eq. (1) is true for a given numerical approximation to the CFD equations of interest, it does not provide enough information for us to sufficiently characterize the computational error to facilitate validation. We have also had to assume that the grid is uniform for Eq. (1). This is virtually never true in real calculations. For unsteady flow calculations, the temporal spacing is not uniform either. Formal truncation error estimates become very difficult to derive on nonuniform meshes, and so practically they are viewed as local error estimates in such cases. We want to stress this point: validation requires solution convergence, not simply an error analysis such as Eq. (1).

Thus, an *a priori* estimate of the accuracy of a particular calculation is problematic for playing a useful role in providing sufficient error determination to support validation. This does not mean that developing these estimates, even in very seriously simplified problems, is not important. Stability and consistency implying solution convergence are crucial sanity checks for discretizations of PDEs. There is no reason to believe that any scheme that does *not* converge to the correct solution for linearized, constant coefficient, homogeneous mesh discretizations of the CFD equations will ever work on the complete forms of the equations. But the actual performance of discretizations in terms of stability and effective convergence to the correct solution is ultimately almost always decided from empirical *a posteriori* analysis of calculations.

This latter point introduces one further problem in attempting to actually infer what the error of a calculation is. Knowing the calculation converges as the spatial and temporal discretization resolution limit to zero, even empirically, still does not necessarily inform us as to the error of the calculation. Solution convergence is fundamentally asymptotic. From this perspective, the only real confidence we can have in a calculation is if we actually achieve sufficient discretization resolution, both temporal and spatial, to directly confirm solution convergence. In other words, from an asymptotic convergence point of view, the only useful statement about computational error that we can make is that:

$$\left| u^{r1} - u^{r2} \right| \le \varepsilon \qquad \text{Eq. (2)}$$

where r1 is one level of temporal and spatial refinement of the discretization, while r2 is a more refined level, and ε is the accuracy required of the calculation. It is common to expect ε to vary from one validation problem to another. Eq. (2), of course, assumes that the convergence implied by Eq. (1) when the discretization is stable is eventually monotonic and that we have entered the asymptotic regime. Being able to always converge a given computation in the sense of Eq. (2) would solve all of our practical problems. Unfortunately, the full convergence of a numerical computation implied by Eq. (2) can hardly ever be obtained for complex simulations, simply because the computational resources required to achieve convergence are still not available.

A posteriori error development – where error is assessed "after the fact" – are discussed in some detail in Refs. 130,138, especially in the context of grid convergence extrapolation. A concise discussion of this topic within a larger context of characterizing modeling uncertainties can also be found in Ref. 139. We will follow the line of argument presented in this report for the most part. Consider for simplicity a steady state computational problem in one spatial dimension with uniform mesh spacing h. Under the assumptions of:

1) u is a smooth solution (existence of all the derivatives is necessary to justify the application of a Taylor expansion in the mesh spacing).

2) The formal convergence order p of the discretization method is known *a priori*.

3) The mesh spacing is small enough that the leading order error term dominates the total discretization error. This is also called the asymptotic range of the discretization.

Then we expand the exact solution of the partial differential equation as:

$$u_{exact} = u_{disc}(h) + \alpha h^p + O(h^{p+1}) \qquad \text{Eq. (3)}$$

It is implicit that Eq. (3) is a local expansion – we have simply suppressed the explicit dependence on x to simplify the appearance of the equation. α is a constant independent of h and is also independent of the spatial coordinate under the assumption that the convergence order is uniform over the spatial domain.

Applying Eq. (3) to two different grid resolutions (assume that $h_2/h_1 < 1$) and one has:

$$\begin{aligned} u_{exact} &= u_{disc}(h_1) + \alpha h_1^p + O(h_1^{p+1}) \qquad \text{Eq. (4)} \\ u_{exact} &= u_{disc}(h_2) + \alpha h_2^p + O(h_2^{p+1}) \end{aligned}$$

We then find that

$$u_{exact} = \left[\frac{h_2^p u_{disc}(h_1) - h_1^p u_{disc}(h_2)}{h_2^p - h_1^p} \right] + O(h^{p+1}) \qquad \text{Eq. (5)}$$

The term in brackets in Eq. (5) represents an extrapolation of the discrete solution toward the exact solution. For example, in the special case of p=2 and with the use of centered difference schemes, it turns out the this extrapolation is fourth order. This is the original h-extrapolation method of Richardson.[140]

The *a posteriori* error estimate resulting from Eq. (5) is

$$E(h) = \left[\frac{u_{disc}(h_2) - u_{disc}(h_1)}{h_2^p - h_1^p} \right] h^p \qquad \text{Eq. (6)}$$

In Eq. (6), $E(h) \equiv u_{disc}(h) - u_{exact}$. This approximation neglects the higher order terms, which is allowed under our assumption of being in the asymptotic range. The development of Eq. (6) can be extended to multiple spatial dimensions and time-dependent problems.

Although a Richardson extrapolation error estimate as above can also be applied to finite element calculations, specific *a posteriori* error estimators are also available in the finite element literature. Their explicit discussion is beyond the scope of this paper. *A posteriori* error estimates are of significant importance for finite element adaptivity, where both the spatial grid density (h-adaptivity) and the order of the finite element scheme (p-adaptivity) can be adapted. For validation, however, the real issue is error estimation, not adaptivity of finite elements.

The first assumption underlying Eq. (6) to discuss (and stress) is that you must demonstrate that the grid is fine enough to be in the asymptotic region of the solution convergence. We believe that Roache does not emphasize this practical matter sufficiently. Another of our stated assumptions in Eq. (6) is that the grid is uniform. For complex multi-dimensional problems this assumption is simply never satisfied in practice. It was also assumed that the formal order of accuracy of the method was *correctly* known *a priori*. There are several reasons that this need not be true. One of them is that grid nonuniformity is known to influence the effective order of a discretization. As emphasized in our discussion about *a priori* error estimates, the true order of accuracy of complicated numerical methods for complex problems is often assessed via empirical means. One example is assessing the performance of shock capturing methods. Another example is assessing the performance of multidimensional upwind schemes for advection.

A more serious issue is the assumption of smoothness of the solution. Essentially for all engineering fluid dynamics problems this assumption is not true in practice. This will reduce the formal *a priori* order of the discretization error estimate even in regions well removed from the singularity. The "pollution" of particular regions of a calculation by the presence of singularities such as shock waves or geometrical singularities is a subject of grave concern, even for application of an error estimate like Eq. (6) in a local sense (in other words, for adaptivity). Often the only gauge of this pollution is through empirical assessment. The reader should consult a series of papers by Babuska and colleagues[141-143] as well as the paper by Oden[144] for a discussion of this problem from a finite element point of view. Recent work of Zhang and his colleagues[145] discusses the effect of the presence of a shock wave structure on *a posteriori* error estimates for the Euler equations.

Finally, we will stress that there may be practical issues in performing calculations on suitably refined grids for developing an estimate like Eq. (6). Consider a large calculation, one that is at the limit of available computational resources. It will be impossible to refine the grid for such a problem. Roache has pointed out that to use Eq. (6) one does not need to necessarily refine the grid. Rather, if one is in the situation described, one could coarsen the grid instead, thereby achieving the needed two refinement levels. However, suppose that the large calculation resulted from the requirement to resolve an important flow structure (such as a reacting flow front or a shock wave) with the *minimal* needed resolution. This is the case in certain compressible flow problems, where all of the overall grid resolution is driven by the need to resolve a captured shock with four (say) grid points. In this case, one can not coarsen the grid, because one encounters a qualitative change in the computed flow upon doing this. One is essentially constrained to work with one grid. As a result, one is not able to make a reasonable error estimate needed for validation.

There are other issues associated with computational error that go beyond error estimates. We can ask fundamental questions of how well the computational world matches the true dynamics of the continuum PDEs. For example, how well are the dynamical system features of the underlying differential equations matched by the computational dynamics. In the case of steady flows, this question is concerned with the stability and bifurcation behavior of the computed solutions. Attempting to measure the computational error in this case is much more difficult. For example, it is only in relatively simple cases that the true stability and bifurcation behavior may be well understood for the conceptual model. An entire taxonomy of potential threats to computational accuracy arises in this point of view. The dangers are nicely summarized in Ref. 146 and the work referenced there.

In conclusion, we emphasize that addressing the goal of validation of computational fluid dynamics requires significant evidence that the equations of the underlying conceptual model are solved accurately. Otherwise, attempts at validation via comparison with experiments, no matter how refined, are dangerous. We have argued that *a priori* error estimates are inadequate for this task because of the typical assumptions that underlie them. Instead, we recommend that serious effort be made to understand computational error in an *a posteriori* sense, especially through the use of grid refinement studies.

4.3 Uncertainty Quantification in Computations

We now turn our attention to a discussion of treating stochastic parametric uncertainty within the context of validation. We will simply refer to "uncertainty" in this discussion, but we will always be considering this restricted type of uncertainty. The main purpose is clarify the need for nondeterministic methods in CFD validation.

It is typical when simulating experiments that one encounters physical parameters in the partial differential equation or in the initial or boundary conditions that are not known precisely for an experiment, or series of experiments. Common examples of such parameters are thermal conductivity, surface roughness, flow rates in complex systems, inflow nonuniformities, and thermochemical transport properties. Whatever the parameter or set of parameters is, we make the assumption that a value of the parameter is required to perform the needed computational simulation.

One standard approach is to estimate, by one means or another, a single value of such a parameter and compute with that selected value. This might be a fully adequate way of dealing with this uncertainty, especially if experience suggests that the range of potential parameter values is very small and if the calculation is known to not be extremely sensitive to the parameter in question. The resulting calculation intrinsically is interpreted as "typical" or "representative" for that parameter. The statistical issues underlying the uncertainty of the parameter are never explicitly treated in the CFD computations.

The difficulty with this straightforward approach begins to be noticeable when the range of variation of the parameter is large or when the calculation is known to be sensitive to the parameter values. If multiple parameters having physical origins that are relevant in a computation of a validation experiment are also uncertain, we claim that it is entirely possible that their interaction in the calculation may magnify the influence of their uncertainty on the final results of the calculation. This latter statement is fairly subtle, but the need to be alert to this possibility is perhaps obvious upon consideration.

When parameter uncertainty is important we claim that performing a single calculation with best estimates of single value(s) of the parameter(s) is not an appropriate method of dealing with this uncertainty. We believe that this statement is especially important when performing calculations that are intended for direct quantitative comparison with validation experiments. Instead, it is important to incorporate the uncertainty directly into the computational analysis.

The simplest strategy for incorporating uncertainty of this kind directly into computation is performed in three steps. First, assume that the uncertainty in the parameters of interest is characterized by probability distributions. Sometimes such distributions can be directly estimated from experimental data related to the parameters. Sometimes such distributions must be simply assumed. At any rate, these probability distributions must be specified.

In the second step, values from these input probability distributions are selected using statistical sampling procedures, such as Monte Carlo sampling methods or more complex methods (see, for example, Ref. 147 for details about Monte Carlo methods). These sampled values are then used in a set of computations. This is important, so we will state it again with emphasis. The assumed prior probability distributions for the uncertain parameters are used to generate a *set* of calculations. This is also sometimes called an *ensemble* of calculations, and so this approach is also called *ensemble computing*.

The key issue is that a single calculation is no longer sufficient; a set of calculations must be performed. Obviously, this need is disturbing – where once one might have performed a single calculation, now one must perform a potentially large number of calculations. We have not raised nor answered the question of whether sufficient computational resources are available to execute more than the one calculation. However, the constraints enforced by availability of computing are formidable.

Upon completion of this set of calculations the third step is to analyze this set of calculations, typically using statistical inference, to estimate a probability distribution for the output variable(s) of interest that results from the given input parameter distributions. In general, we can not deduce the exact output distribution that results from the assumed form of the parameter distributions that generate the computational input associated with those parameters. Instead, it is common practice to determine estimates of important statistical parameters associated with that output distribution through the use of statistical procedures.

These statistical estimates have considerable interest for analyzing computational predictions. For example, the mean of the output calculations provides us with an estimate of the expected value of the output quantity of interest, given the uncertainty structure specified by the input distributions. This estimate is certainly of interest when comparing computations with experimental data that are likely to have their own uncertainty. Another statistic of interest is the estimated variance of the output distribution, which can be interpreted as a measure of computational output uncertainty, given the input uncertainty.

For readers unfamiliar with this methodology, we stress that it is not true that the mean of the output given the input uncertainty can be determined by performing a calculation for a single set of inputs that is chosen to be the mean of the input distributions. We must perform the ensemble of calculations to develop a statistically rational estimator for the mean. The same statement is true for estimating other output statistics.

To summarize, step one is called *characterizing the source* of uncertainty in the computational study. Step two is sometimes called *uncertainty propagation*. Step three is called *uncertainty quantification* of the output. The general methodology we are thus advocating for incorporating parameter uncertainty into CFD computations is to execute all three steps in the manner suggested above. This will clearly be nontrivial for hard computational problems simply because of the computational burden imposed. We have also failed to state certain subtleties, such as whether complex structure in the resulting output probability distribution can actually be discovered using such a crude approach. We will simply state that extracting intricate output distribution structure will either require a large number of sample input values, or considerably more sophisticated approaches for performing the methodology.

Because we are unaware of any study in CFD that pursues this methodology, we will reference the work of Ref. 148 for an interesting summary of issues associated with this methodology in computational solid mechanics. The work discussed in Ref. 136 provides a discussion of the intricacies of characterizing the sources of uncertainty. The book by Kleijnen[149] is an interesting reference that provides a broad summary of methodologies that go beyond Monte Carlo for assessing output distributions statistically. Finally, a recent interesting reference that discusses the assessment of modeling uncertainty in climate modeling is Ref. 150.

An issue is stressed in Ref. 148 that is worth repeating here because it emphasizes a subtly in performing uncertainty quantification. As Red-Horse and his colleagues note, if we want to estimate, for example, the mean of the output quantity of interest using the methodology above, the estimated statistic that we determine is actually a *conditional mean*. In other words, the estimated mean of the output given the uncertain input parameters that we calculate in this way assumes that the computational model is "correct," or valid, in a fundamental way. This is in addition to the operational assumption that the input uncertainty distributions are "correct," or accurate enough, for the intended application of the computations. We thus have a coupling of the validation problem to the quantification of uncertainty in the computational model. To validate our computational model, we must characterize uncertainty quantitatively, using the methodology proposed above or something similar. However, for this characterization to be fully accurate requires a valid computational model. One way of looking at this coupling of the understanding of parameter sensitivity and validation is through the use of Bayesian inference. We briefly comment on this below.

It is also possible to remove the intrinsic foundation of a validated model in the estimation of output statistics by including fundamental model uncertainty, sometimes called *structural uncertainty*, in the methodology.[117] But can structural uncertainty, which is far more general than our simple parameter uncertainty described previously, even be captured in a probabilistic framework? Non-probabilistic approaches, referenced above with regard to epistemic uncertainty, are currently of interest for characterizing structural uncertainty.

We now return to our mention of Bayesian inference. Step 2, the problem of propagating input uncertainty through a computational model to understand the resultant output as described above, is sometimes referred to as the *forward uncertainty problem.*[151] There is an associated *inverse uncertainty problem*, or *backward problem*, which is conceptually much more difficult but equally important when performing validation. This problem asks whether, when given the input parameter uncertainty, an estimate of output uncertainty, and the additional knowledge one acquires by performing validation comparisons with experiments, we can improve our estimated output uncertainty? In particular, might we be able to improve our original prior distributions that characterize the parameter uncertainty? This problem can be cast as a problem in Bayesian statistical inference.[151] (See Ref. 152 for an introduction to Bayesian inference.)

Finally, we comment that the subject of uncertainty and its role in assessing the predictive content of intricate computational models of complex physical phenomena (turbulence, climate, economic systems, biological systems, etc.) has become of increasing interest for formal study in the past few years. This subject is not the same as the study of so-called "complex systems," although complex systems could certainly be one of the targets for wishing to be predictive with a computer calculation. Rather, the focus here is to extend the simple discussion we have provided above about error, variability, and uncertainty estimation to harness real mathematical and computational control of these quantities on behalf of providing *quantitative* assessments of the true predictive accuracy of calculations. This is a dramatically difficult task. We recommend to the reader the proceedings of a workshop held at Los Alamos National Laboratory in May 1998 for recent papers on this topic.[153]

Part of the difficulty is to provide better understanding of computational accuracy when it is known that we are in an under-resolved grid or time-step situation. This is an outstanding problem for research. Recent work attacks this problem and illustrates the formidable difficulties in two distinctly different applications areas: porous flow[154] and dynamical systems.[155-157] While the applications and specific technical attacks of these two groups of researchers are distinctly different, we find it fascinating that a common deep thread in their work is the treatment of insufficient information using probabilistic methods. We should stress, however, these authors do not discuss one problem of interest to us. That is, the problem of validation of under-resolved computational models, which remains of outstanding importance for pragmatic validation approaches for realistic applications of large-scale computational models.

The most important point we wish to emphasize in concluding this section is the inadequacy of single (point) solutions of computational fluid dynamics when confronting variability or uncertainty in the intended application. We must be willing and prepared to perform sets of calculations to develop even primitive insight into the influence of the uncertainty on the calculated results.

5. Comparisons of Computation and Experiment

5.1 Hypothesis Testing

As discussed in Section 2.3, validation quantification requires the determination and evaluation of a specified metric for measuring the consistency of a given computational model with respect to experimental measurements. One approach that has been traditionally taken in statistical analyses is hypothesis testing.[36,39,158] Hypothesis testing is a well developed statistical method of choosing between two competing models of an experimental outcome by using probability theory to minimize the risk of an incorrect decision. Hypothesis testing formulates the validation quantification measure as a "decision problem" in which one wishes to decide whether or not the hypothesized model is consistent with the experimental data. This technique is regularly used in the operations research community for testing mutually exclusive models, i.e., the model is either true or false. For example, suppose the hypothesis is made that a coin is fair. That is, in the toss of the coin it is equally likely that "heads" will appear as often as "tails." The competing hypothesis is that the coin is unfair. Experimental data are then obtained by tossing the coin a given number of times, say N, and recording what percentage of outcomes are heads and what percentage are tails. Hypothesis testing then allows one to probabilistically determine the confidence of a fair coin. The confidence in the determination will depend on N, that is, as N increases, the confidence in the conclusion increases.

Hypotheses testing has not been used to any significant degree in validation quantification of computational physics. It seems there are two reasons for lack of interest in this approach. First, validation of computational models of physical processes do not fit into the category of true or false hypotheses. For example, we would not expect to see it proclaimed: "Computer code xyz has been proven false!" Hypothesis testing would be more appropriate, for example, for testing whether Newtonian mechanics is true versus relativistic mechanics. In other words, model validation in the computational sciences is fundamentally an estimation process, *not* a true or false issue. For example, the appropriate questions in validation are: (1) What is the measure of agreement between the computational result and the experimental result? (2) How much does the numerical error in the computational solution affect the measure of agreement? and (3) How much does the experimental uncertainty affect the measure of agreement?

Second, if hypothesis testing is used to prove a hypothesis true, given the available evidence, then the hypothesis, i.e., the model, can be used to replace the fiducial measure, i.e., the experiment. In the computational sciences, validation is properly understood to mean that the measure of agreement attained for one comparison case is an inference of validity for future cases, i.e., prediction. The accuracy of the prediction depends on many additional factors, such as, the range of applicability of all of the submodels that comprise the complete computational model, the change in coupling of the various physical processes from the validation case to the prediction case, the skill of the analyst in computing the prediction, and any additional uncertainties that may enter into the prediction that were not present in the validation.

Even though the hypothesis testing approach does not appear to be a constructive route forward for validation quantification, the approach has developed the concept of error types for incorrect conclusions drawn from hypothesis testing.[36,39,158] A type 1 error, also referred to as *model builder's risk*, is the error in rejecting the validity of a model when the model is actually valid. This can be caused by errors on both the computational side and the experimental side. On the computational side, for example, if a grid is not sufficiently converged and the computational result is in error, then an adverse comparison with experimental data is misleading. That is, a poor comparison leads one to conclude that a submodel, such as a turbulence or reacting flow model, needs to be improved or "re-calibrated" when the source of the poor comparison is simply an under-resolved grid. On the experimental side, the model builder's risk is most commonly caused by a poor comparison of computation and experiment that is due to an unknown bias error in the experimental data. Examples of these in wind tunnel testing are the following: a bias error exists in the calibrated freestream Mach number in the test section, a pressure reference value used in a differential pressure transducer drifts due to temperature of the transducer, and bias error due to flowfield nonuniformity and model geometry imperfections (discussed above). We believe that unknown bias errors in experimental results are the most damaging in validation because if the experimental measurement is accepted, then it is concluded that the computational result is consistently in error; whereas in reality, the experiment is consistently in error. If the error is believed to be in the computation, then a great deal of effort will be expended trying to find the source of the error. Or worse, a computational submodel will be re-calibrated using the biased experimental data. This results in transferring the experimental bias into the computational model and then biasing all future computations with the code.

The type 2 error, also referred to as *model user's risk*, is the error in accepting the validity of a model when the model is actually invalid. As with type 1 error, this can be caused by errors on both the computational side and the experimental side. On the computational side, the logical reverse of the type 1 error described above can occur. That is, if a grid is not sufficiently converged and the computational result agrees well with the experiment, then the favorable comparison is also misleading. For example, if a finer grid is used one can find that the favorable agreement can disappear. This shows that the original favorable agreement has compensating, or cancelling, errors in the model. We believe that compensating errors in complex simulations is a common phenomenon. Only the tenacious user of the code, or an uncommonly self-critical code developer, will dig deep enough to uncover the compensating errors. In a competitive or commercial code development environment, such users or code developers as these can be very unpopular, and even muffled by co-workers and management. On the experimental side, the logical reverse of the type 1 error described above can occur. That is, if an unknown bias error exists in the experiment, and a favorable comparison between computation and experiment is obtained, the implication of code validity is incorrect. Similar to the type 2 error on the computational side, only the self-critical experimentalist will continue to examine the experiment in an attempt to find any experimental bias errors.

Type 1 and type 2 errors are two edges of the same sword. In the OR literature, however, it is well known that model user's risk is potentially the most disastrous. The reason, of course, is that an apparently correct model (one that has experimental evidence that it is valid) is used for predictions and decision-making, when in fact it is incorrect. Type 2 errors produce a false sense of security. In addition to the potentially disastrous use of the model, we contend that the model user's risk is also the more likely to occur in practice than the model builder's risk. The reason for this is that with experimental evidence that the model is valid, there is little or no interest by analysts, experimentalists, managers, or decision makers to expend any more time or resources pursuing possible problems in either the computations or the experiments. Everyone is enthused by the agreement of results and "Victory" is declared. Anyone who questions the results can risk loss of personal advancement and position within his or her organization.

5.2 Validation Metrics

Most of the authors reviewed in Section 2.3 addressed validation quantification from a parametric uncertainty estimation viewpoint, that is, a probability distribution of a system response measure (due to input parameter uncertainty) is compared with a probability distribution due to experimental uncertainty. As discussed in Section 4.4, the probabilistic response measure is normally computed using Monte Carlo sampling so that a comparison with the experimental data can be made. Coleman and Stern[17] take a sharply different approach. First, they include in their analysis an estimate of numerical solution error and how it affects the validation comparison. This, of course, is of major concern in CFD solutions and in all of computational physics. Second, they do not deal with the propagation of input probability distributions, but instead address the question of estimating the total uncertainty due to computation and experiment. And third, they define a metric for validation. If the difference between the experimental measurement and the computational result is less than the validation metric, then the result is declared validated. They define their validation metric as the square root of the sum of the squares of the experimental uncertainty, the numerical solution error, and the modeling uncertainty arising from using previous comparisons with experimental data. Written in their nomenclature, one has

$$U_V = \sqrt{U_D^2 + U_{SN}^2 + U_{SPD}^2}$$

where U_V is the validation metric, U_D is the experimental uncertainty, U_{SN} is the numerical solution error, and U_{SPD} is the modeling uncertainty arising from using previous comparisons with experimental data.

Coleman and Stern's article added a new direction of thinking in validation quantification and validation metrics. However, we believe their approach is conceptually flawed for three reasons. First, they deal with numerical solution error as a statistical quantity, when it has no practical connection to probability theory. That is, they deal with numerical solution error in exactly the manner as experimental precision error. Numerical errors such as spatial and time-step discretization, artificial dissipation, approximate factorization, intra-step iterative convergence, and global iterative convergence are *not* related to probability theory. Recent work of Refs. 154-157 attempt to use statistical methods to estimate the numerical error on coarse grids, but this is very tenuous at this point. Numerical solution errors are more closely related to bias errors in experimental error estimation than precision errors.

Second, their validation metric U_V specifically excludes the uncertainty due to modeling assumptions, U_{SMA}. They state "...we define the validation uncertainty U_V as the combination of all uncertainties that we know how to estimate (i.e., all but U_{SMA})". In their nomenclature,

$$U_V = \sqrt{U_D^2 + U_S^2 - U_{SMA}^2}$$

where U_S is the uncertainty in the simulation. Ignoring the uncertainty due to modeling assumptions defeats the primary purpose of validation. Their argument for ignoring U_{SMA} is analogous to ignoring an experimental bias error because we do not know how to estimate it. The modeling error must be evaluated through a comparison of simulation with experiment.

Third, their validation metric increases as the experimental uncertainty U_D increases, and as the numerical solution error U_{SN} increases, and as the modeling uncertainty arising from using previous comparisons with experimental data U_{SPD} increases. In other words, it becomes easier to achieve validation as the experiment and the simulation become poorer. Coleman and Stern recognize this paradox with the explanation "...since the greater the uncertainties in the data and the code predictions, the greater the 'noise level' U_V." Concerning their validation metric, Roache[159] comments "besides the evident potential for misinterpretation in the use of U_V, a more basic problem exists with their proposal; it fails to account for an acceptable error tolerance in the validation." We agree with Roache's criticism and we add two additional criticisms. On first glance it appears to make sense that as the spread in the experimental data gets larger, then the criterion for validation should also get larger. This logic is actually erroneous because the spread in the data is not the key issue. The important issue is the *mean* of the data and the level of confidence in the mean. The final criticism is analogous to the first; however, it deals with the increase in the spread of the computational simulation. That is, the computational simulation becoming poorer (larger numerical error) is not the issue; it is how an *accurate* numerical solution compares with the experimental data.

We commend Coleman and Stern for initiating thinking in validation quantification metrics. We argue that the validation metric should have the following features. First, we agree with Coleman and Stern that the metric should incorporate an estimate of the numerical error in the computational simulation. However, we do not agree that it should be grouped with the experimental uncertainty, and also, it should not be represented probabilistically. Second, the metric should not exclude any modeling assumptions or approximations used in the computation of the simulation result. That is, the computational result must reflect all uncertainties and errors incurred in the modeling and simulation process. Third, we agree with Coleman and Stern that the metric should incorporate an estimate of the random errors in the experimental data, e.g., an estimate of the variance of an assumed Gaussian distribution. In addition, we believe the metric should also include an estimate of the correlated bias errors in the experimental data. Fourth, the metric should depend on the number of experimental replications of a given measurement quantity. That is, the metric should reflect the level of confidence in the experimental mean that has been estimated. And fifth, the metric should be able to incorporate uncertainty in the computation that is due to both random uncertainty in experimental parameters and any uncertainty that is due to lack of experimental measurement of needed computational quantities. That is, the metric must use nondeterministic methods to propagate uncertainty through the computational model.

In the following we suggest various validation metrics that address some of these recommended features, but much work needs to be done.

5.3 Examples of Validation Metrics

To clarify our views on validation quantification and how we believe validation metrics should be constructed, we will discuss an example whose physics and mathematics are much simpler than fluid dynamics. We consider the example of a boundary value problem described by a second order, nonlinear, ordinary differential equation (ODE), Fig. 6. Let the general form of the ODE be given by

$$\frac{d}{dx}\left[p(x,y)\frac{dy}{dx}\right] + q(x,y)\frac{dy}{dx} + r(x,y) = 0 \qquad \text{Eq. (7)}$$

where p(x,y), q(x,y), and r(x,y) are arbitrary functions of the independent variable, x, and the dependent variable, y. Let B_0 and B_L represent arbitrary boundary conditions at x=0 and x=L, respectively. This ODE and and its boundary conditions can represent many types of steady state heat conduction in a one-dimensional solid. Examples are heat conduction in heterogeneous and anisotropic solids, temperature dependent thermal conductivity, internal heat generation, and convection and radiation heat loss along the solid. This ODE, of course, is intended to be analogous to multidimensional boundary value problems in fluid dynamics.

Figure 6: Domain of Boundary Value Problem

The physical modeling parameters for this problem are incorporated into the functions p, q, and r, and also incorporated into B_0 and B_L. For example, in a heat conduction problem, if the thermal conductivity is a constant, then p is equal to the constant conductivity. Also for a heat conduction problem, if the boundaries are given as a specified temperature, then y(0) = constant and y(L) = constant are the boundary data for B_0 and B_L, respectively.

Let $Y(x_i)$ be the experimental measurements corresponding to the dependent variable, y, at $x = x_i$ for i = 1, 2, 3, ... I. That is, a total of I locations are experimentally measured along the solid. Given this mathematical model and experimental data, we will consider two examples that emphasize different issues in validation quantification. For each example we suggest different types of validation metrics that would be useful for various situations. For both examples we assume the following:

1) The numerical solution to the ODEs is without error. For an actual numerical solution this practically means that the numerical solution error is carefully quantified, and it is shown that the relative error of the output quantities of interest is very small.

2) There is no parametric uncertainty in the model. Stated differently, experimental measurements of p, q, r, B_0, and B_L have been made, they are assumed to be without measurement error, and they are used as input to the model.

3) The model and the experiment do not exhibit bifurcation or any chaotic phenomena.

Given assumption (1), we are avoiding the issue of how does one conduct validation when numerical solution error is significant. As discussed in Section 4, we do not believe true validation can be conducted when no estimate is available for the numerical solution error in the system responses that are compared with experiment . With assumption (2) we are restricting our examples to the issue of validation of physical modeling fidelity and avoiding two issues: parametric fidelity in the model and modeling when required code input parameters were not measured in the experiment. With techniques for propagation of uncertain parameters through the CFD model, discussed in Section 4, we are certain validation can still be conducted. However, validation becomes more complex because one must then deal with probability distributions for the computational inputs and responses. In the second example we address the issue of uncertainty in the experimental measurements.

<u>Example 1: Experimental Measurement Error is Zero</u>

Since the experimental measurement error is zero, $y(x_i)$ and $Y(x_i)$ can be directly compared in different ways to generate different validation metrics. One useful validation metric based on comparison of each of the individual measurements and the computations at the measurement locations is

$$V = 1 - \frac{1}{I}\sum_{i=1}^{i=I} \tanh\left|\frac{y(x_i) - Y(x_i)}{Y(x_i)}\right| \qquad \text{Eq. (8)}$$

where V is the validation metric. This type of metric has the following advantages. First, it normalizes the difference between the computation and the experimental data. As a result, a relative error norm is computed. Second, the absolute value of the relative error only permits the difference between computation and experiment to accumulate, i.e., positive and negative differences cannot offset one another. And third, when the difference between the computation and experiment is zero at all measurement locations, then the validation metric is unity, i.e., perfect agreement between computation and experiment. When the summation of the relative error becomes large, the validation metric approaches zero.

Figure 7 shows how the validation metric given in Eq. (8) varies as a function of constant values of the relative error at all spatial locations. As can be seen from Fig. 7, if the summation of the relative error is at a value of 100% of the experimental measurement, then the validation metric would yield a value of 0.239. If functions different than tanh were used in Eq. (8), then the value of the metric for a constant error of 100% would, of course, be different. However, we do not believe the quantitative value of any metric is important in an absolute sense. We have chosen here to require that the metric be unity when perfect agreement is attained and zero for very poor agreement of computation and experiment. That is, the metric should simply reflect a measure of the agreement between computation and experiment. Validation should not be viewed as binary issue, e.g., is the hypothesis true or false, or, is the computation within the scatter of the data, or not.

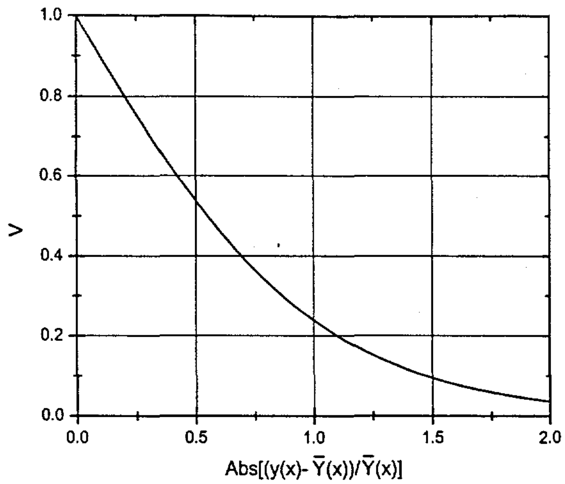

Figure 7: Proposed Validation Metric as a Function of Relative Error

To make it clearer why a hypothesis testing view point is not as constructive, consider the following. Let the validation metric be defined as

$$\left|\frac{y(x_i) - Y(x_i)}{Y(x_i)}\right| \begin{Bmatrix} \le \varepsilon \ \ for\ all\ i = 1, 2, 3, \ldots I \Rightarrow valid \\ > \varepsilon \ \ for\ any\ i = 1, 2, 3, \ldots I \Rightarrow invalid \end{Bmatrix} \quad \text{Eq.(9)}$$

where ε is a specified accuracy criteria. If a metric such as Eq. (9) is used, then the result is only "pass" or "fail." From an engineering point of view, validation is an estimation problem. From a scientific point of view, validation could be considered as truth issue (see Ref. 103 and Appendix C of Ref. 130). An additional argument against the type of metric given in Eq. (9) is that it merges the issue of validation accuracy and the question "Is the simulation adequate for the intended uses of the model?" These are clearly separate issues. A useful validation metric should *only* measure the agreement between computation and experiment. Whether the measure of agreement is adequate for the intended uses must be viewed as a separate question. And indeed, the adequacy of the model for the intended uses, i.e., prediction, is the *most important* question.

If a sufficient number of measurements are made along the solid, i.e., I is large in some sense, then one could construct a continuous function to represent Y(x) along the solid. For example, one could use a cubic spline to interpolate $[x_i, Y(x_i)]$ along the solid. Then one could construct an improved global level metric analogous to Eq. (8):

$$V = 1 - \frac{1}{L}\int_0^L \tanh\left|\frac{y(x) - Y(x)}{Y(x)}\right| dx \qquad \text{Eq. (10)}$$

If a sufficient number of experimental measurements are available to accurately construct the interpolation function, then this metric is more advantageous than Eq. (8) because it accounts for error in locations where experimental data are not available.

<u>Example 2: Varying Number of Experimental Measurements</u>

Let there be N experimental measurements made at each of the x_i locations along the solid. Assume that the experimental random error is normally distributed, i.e., Gaussian, and assume that the measurement bias error is zero. The mean value of the N measurements at the position x_i is given by

$$\bar{Y}(x_i) = \frac{1}{N}\sum_{n=1}^{N} Y_n(x_i) \qquad \text{Eq. (11)}$$

As N becomes large, the mean, $\bar{Y}(x_i)$, approaches the true value, $\tilde{Y}(x_i)$, of the Gaussian distributed measurements, that is,

$$\tilde{Y}(x_i) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} Y_n(x_i) \qquad \text{Eq. (12)}$$

Consider first the case when a large amount of data are available, N large. Then the estimated mean and the true mean are effectively identical. $\tilde{Y}(x_i)$ is what should be compared with the computation, not its estimated uncertainty, *regardless* of the spread of the measurement distribution. Therefore, essentially the same validation metrics as given above in Example 1, where the data are perfect, can be used for this case. For example, the metric given in Eq. (10) would now be written as

$$V = 1 - \frac{1}{L}\int_0^L \tanh\left|\frac{y(x) - \tilde{Y}(x)}{\tilde{Y}(x)}\right| dx \qquad \text{Eq. (13)}$$

$\tilde{Y}(x)$ would be the interpolated function constructed using $[x_i, \tilde{Y}(x_i)]$.

Now consider the case when a limited quantity of data are available. Assume that at each measurement position along the solid there are the same number of experimental measurements, N. The following suggested metric incorporates both the estimated variance of the data and the number of measurements at each station, N. Using the triangle inequality and taking expectations, one can extend the metric given in Eq. (13). One obtains

$$V = 1 - \frac{1}{L}\int_0^L \tanh\left[\left|\frac{y(x) - \bar{Y}(x)}{\bar{Y}(x)}\right| + \int_{-\infty}^{+\infty} \frac{s(x)}{\sqrt{N}}\left|\frac{z}{\bar{Y}(x)}\right| f(z)\, dz\right] dx$$

Eq. (14)

where *s(x)* is the sample standard deviation as a function of position along the solid, and $f(z)$ is the probability density function for a student's t-distribution with N-1 degrees of freedom. The integral inside the brackets is the expected absolute relative error of the experiment. The quantity $s(x)$ is approximated through the interpolated function constructed using $s(x_i)$, where

$$s^2(x_i) = \frac{1}{N-1}\sum_{n=1}^{N}\left(Y_n(x_i) - \bar{Y}_n(x_i)\right)^2 \qquad \text{Eq. (15)}$$

It is obvious that no estimate of experimental uncertainty can be made if only one data point at each location, N=1, is available.

The probability density function for the student's t-distribution is given by

$$f(z;\nu) = \frac{\Gamma\left((\nu+1)/2\right)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\left[1 + \frac{z^2}{\nu}\right]^{-(\nu+1)/2} \qquad \text{Eq. (16)}$$

where ν is the number of degrees of freedom, N-1, and $\Gamma(\alpha)$ is the gamma function, given by

$$\Gamma(\alpha) = \int_0^{\infty} \xi^{\alpha-1} e^{-\xi} d\xi \qquad \text{Eq. (17)}$$

Figure 8 shows the validation metric from Eq. (14) as a function of constant relative error, for a coefficient of variation, $s(x)/Y(x) = 0.1$, and for various values of N.

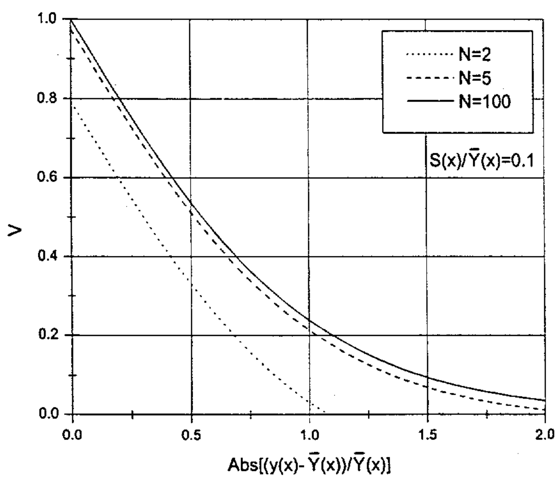

Figure 8: Validation Metric as a Function of Relative Error and Data Quantity

The advantageous features of the validation metric given by Eq. (14) are the following. First, when a large amount of data are available, i.e., N is large, the second integral in Eq. (14) approaches zero. As a result, Eq. (14) approaches the metrics defined in Eqs. (10) and (13). Second, the student's t-distribution replaces the Gaussian distribution for Gaussian measurement errors when N is small. When N becomes large, the student's t-distribution approaches the Gaussian distribution. This makes the validation metric consistent with the assumed probability distribution for the random measurement error. Third, when N is small, the second integral in Eq. (14) increases the magnitude of the integrand of the first integral. This results in decreasing the magnitude of the validation metric. That is, when small quantities of data are available, the validation metric should reflect the fact that confidence in the validation measure is decreased. And fourth, as the coefficient of variation in the data increases, for a given number of measurements N, the magnitude of the validation metric decreases. That is, as the spread of the data increases, the confidence in the validation measure is decreased.

6. Conclusions and Recommendations

In this paper we address the three facets of validation: the design, execution, and analysis of validation experiments; error and uncertainty estimation in the computational simulation; and the quantitative comparison of computation and experiment. We accentuate the issues that must be addressed by the CFD analyst and the experimentalist in validation assessment. The concept of the construction of a validation hierarchy for complex engineering systems is relatively new, but we believe it is critical to the validation of multidisciplinary, interacting, computational simulation capabilities. The understanding of the unique characteristics of validation experiments, as opposed to traditional types of experiments, is slowly developing. We believe many of the requirements needed for validation experiments will be difficult to accept by some testing facilities. A few of the requirements put the facility at risk with respect to their competitors, and many of the requirements will necessitate a change in the culture of experimental investigations. In addition, the statistical uncertainty estimation procedure we recommend will more critically quantify experimental shortcomings, specifically correlated bias errors. In previous work we have demonstrated that the dominant contributor to hypersonic wind tunnel experimental uncertainty is the nonuniformity of the flow in the test section. We believe this will also be true for all other speed ranges of wind tunnels, especially for transonic flow and shock tunnels.

Error and uncertainty estimation for computational simulations are a necessary requirement for validation. Error estimation for the numerical solution of nonlinear partial differential equations is in its early stages of development. We draw a distinction between error estimation and error analysis. Error analysis is based on linear PDEs, commonly with constant coefficients, simple initial and boundary conditions, and uniform grids. None of these niceties are true in real engineering analyses and, hence, error estimation on real problems is difficult, expensive, and filled with risks. We also believe that CFD must begin to adopt nondeterministic simulation strategies. Validation experiments at the benchmark and subsystem level will necessarily involve missing data or uncertain data that are needed for computational simulations. This will necessitate probabilistic treatment of parameters in the CFD submodels or in the initial conditions or boundary conditions for the PDEs. Propagation of uncertain parameters or conditions through the CFD model will likely rely on methods such as Monte Carlo or Latin Hypercube sampling. On this topic, we believe that CFD can learn much from probabilitistic analyses and risk assessment methods in structural dynamics.

Implementation of most of the procedures recommended here, for both experiment and computation, will be neither inexpensive nor easy. Some may be technically or economically impractical in particular situations. With each included step, however, the quality of the code validation process will be improved. We firmly believe that validation is a process, not a product. To achieve the level of maturity in CFD characterized by "value exceeds expectation" and "most analysis done without supporting experimental comparisons," will require a much deeper understanding of mathematics, physics, computation, and experiment, and their relationships.

Acknowledgements

The authors thank Brian Rutherford, Robert Easterling, Thomas Paez, and Martin Pilch of Sandia National Laboratories for many helpful suggestions and discussions during preparation of this paper.

References

1. DoD, "DoD Directive No. 5000.59: Modeling and Simulation (M&S) Management," Department of Defense, available: www.dmso.mil/docslib/, 1994.
2. DoD, "Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide," Defense Modeling and Simulation Office, Office of the Director of Defense Research and Engr., available: www.dmso.mil/docslib, 1996.
3. Cohen, M. L., Rolph, J. E., and Steffey, D. L., eds. *Statistics, Testing, and Defense Acquisition: New Approaches and Methodological Improvements*, National Acadey Press, Washington, DC, 1998.
4. Chapman, D. R., Mark, H., and Pirtle, M. W., *Computer vs. Wind Tunnels*, in *Astronautics & Aeronautics*, 1975, pp. 22-30.
5. Bradley, R. G., "CFD Validation Philosophy," North Atlantic Treaty Organization, AGARD-CP-437, *Fluid Dynamics Panel Symposium: Validation of Computational Fluid Dynamics*, Lisbon, Portugal, 1988.
6. Marvin, J. G., "Accuracy Requirements and Benchmark Experiments for CFD Validation," AGARD, AGARD-CP-437, *Fluid Dynamics Panel Symposium: Validation of Computational Fluid Dynamics*, Lisbon, Portugal, 1988.
7. Neumann, R. D., "CFD Validation-The Interaction of Experimental Capabilities and Numerical Computations," American Institute of Aeronautics and Astronautics, AIAA Paper No. 90-3030, 1990.
8. Mehta, U. B., "Some Aspects of Uncertainty in Computational Fluid Dynamics Results," *Journal of Fluids Engineering*, Vol. 113, No. 4, 1991, pp. 538-543.
9. Oberkampf, W. L., and Aeschliman, D. P., "Joint Computational/Experimental Aerodynamics Research on a Hypersonic Vehicle: Part 1, Experimental Results," *AIAA Journal*, Vol. 30, No. 8, 1992, pp. 2000-2009.
10. Dwoyer, D., "The Relation Between Computational Fluid Dynamics and Experiment," American Institute of Aeronautics and Astronautics, *AIAA 17th Ground Testing Conf.*, Nashville, TN, 1992.
11. Bertin, J. J., Martellucci, A., Neumann, R. D., and Stetson, K. F., "Developing a Data Base for the Calibration and Validation of Hypersonic CFD Codes - Sharp Cones," American Institute of Aeronautics and Astronautics, AIAA Paper No. 93-3044, *24th AIAA Fluid Dynamics Conf.*, Orlando, FL, 1993.
12. Cosner, R. R., "Issues in Aerospace Application of CFD Analysis," AIAA Paper No. 94-0464, *32nd Aerospace Sciences Meeting & Exhibit*, Reno, NV, 1994.
13. Marvin, J. G., "Perspective on Computational Fluid Dynamics Validation," *AIAA Journal*, Vol. 33, No. 10, 1995, pp. 1778-1787.
14. Oberkampf, W. L., Aeschliman, D. P., Henfling, J. F., and Larson, D. E., "Surface Pressure Measurements for CFD Code Validation in Hypersonic Flow," American Institute of Aeronautics and Astronautics, AIAA Paper No. 95-2273, *26th AIAA Fluid Dynamics Conf.*, San Diego, CA, 1995.
15. Aeschliman, D. P., and Oberkampf, W. L., "Experimental Methodology for Computational Fluid Dynamics Code Validation," *AIAA Journal*, Vol. 36, No. 5, 1998, pp. 733-741.
16. AIAA, "Guide for the Verification and Validation of Computational Fluid Dynamics Simulations," American Institute of Aeronautics and Astronautics, AIAA-G-077-1998, Reston, VA, 1998.
17. Coleman, H. W., and Stern, F., "Uncertainties and CFD Code Validation," *Journal of Fluids Engineering*, Vol. 119, 1997, pp. 795-803.
18. Popper, K. R., *The Logic of Scientific Discovery*, Basic Books, New York, 1959.
19. Popper, K. R., *Conjectures and Refutations: The Growth of Scientific Knowledge*, Routledge and Kegan, London, 1969.
20. Carnap, R., "Testability and Meaning," *Philosophy of Science*, Vol. III, 1963, .
21. Conway, R. W., "Some Tactical Problems in Digital Simulation," *Management Science*, Vol. 10, No. 1, 1963, pp. 47-61.
22. Naylor, T. H., and Finger, J. M., "Verification of Computer Simulation Models," *Management Science*, Vol. 14, No. 2, 1967, pp. 92-101.
23. Churchman, C. W., *The Systems Approach*, Dell,

New York, 1968.
24. Naylor, T. H., *Computer Simulation Experiments with Models of Economic Systems*, Wiley, New York, 1971.
25. Shannon, R. E., *Systems Simulation: The Art and Science*, Prentice-Hall, Inc., 1975.
26. Zeigler, B. P., *Theory of Modelling and Simulation*, 1st ed., John Wiley & Sons, New York, 1976.
27. Schlesinger, S., "Terminology for Model Credibility," *Simulation*, Vol. 32, No. 3, 1979, pp. 103-104.
28. Checkland, P. B., *Systems Thinking, Systems Practice*, John Wiley & Sons, New York, 1981.
29. Oren, T. I., "Concepts and Criteria to Assess Acceptability of Simulation Studies: A Frame of Reference," *Communications of the ACM*, Vol. 24, No. 4, 1981, pp. 180-189.
30. Landry, M., Malouin, J.-L., and Oral, M., "Model Validation in Operations Research," *European Journal of Operational Research*, Vol. 14, 1983, pp. 207-220.
31. Sargent, R. G., "Simulation Model Validation," in *Simulation and Model-Based Methodologies: An Integrative View*, T.I. Oren, B.P. Zeigler, and M.S. Elzas Eds., Springer-Verlag, Berlin, 1984, pp. 537-555.
32. Oren, T. I., Zeigler, B. P., and Elzas, M. S., eds. *Simulation and Model-Based Methodologies: An Integrative View*, Springer-Verlag Berlin Heidelberg, 1984.
33. Banks, J., and Carson, J. S., II, *Discrete-Event System Simulation*, 1st ed., Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1984.
34. Sargent, R. G., "An Expository on Verification and Validation of Simulation Models," *1985 Winter Simulation Conference*, Sacramento, CA, 1985.
35. Balci, O., and Nance, R. E., "Formulated Problem Verification as an Explicit Requirement of Model Credibility," *Simulation*, Vol. 45, No. 2, 1985, pp. 76-86.
36. Neelamkavil, F., *Computer Simulation and Modelling*, 1st ed., John Wiley & Sons, New York, 1987.
37. Bratley, P., Fox, B. L., and Schrage, L. E., *A Guide to Simulation*, 2nd ed., Springer-Verlag, New York, 1987.
38. Sargent, R. G., "Validation of Mathematical Models," *Symposium on Validation of Geosphere Flow and Transport Models*, Stockholm, Sweden, 1990.
39. Law, A. M., and Kelton, W. D., *Simulation Modeling and Analysis*, 2nd ed., McGraw-Hill, New York, 1991.
40. Fossett, C. A., Harrison, D., Weintrob, H., and Gass, S. I., "An Assessment Procedure for Simulation Models: A Case Study," *Operations Research*, Vol. 39, No. 5, 1991, pp. 710-723.
41. Hodges, J. S., and Dewar, J. A., "Is it You or Your Model Talking? A Framework for Model Validation," RAND, R-4114-AF/A/OSD, Santa Monica, 1992.
42. Davis, P. K., "Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations," RAND, R-4249-ACQ, Santa Monica, 1992.
43. Bailey, M. P., and Kemple, W. G., "The Scientific Method of Choosing Model Fidelity," *1992 Winter Simulation Conference Proceedings*, Arlington, VA, 1992.
44. Wise, J. A., Hopkin, V. D., and Stager, P., eds. *Verification and Validation of Complex Systems: Human Factors Issues*, Springer-Verlag Berlin, Berlin, 1993.
45. Miser, H. J., "A Foundational Concept of Science Appropriate for Validation in Operational Research," *European Journal of Operational Research*, Vol. 66, 1993, pp. 204-215.
46. Gass, S. I., "Model Accreditation: A rationale and process for determining a numerical rating," *European Journal of Operational Research*, Vol. 66, 1993, pp. 250-258.
47. Landry, M., and Oral, M., "In Search of a Valid View of Model Validation for Operations Research," *European Journal of Operational Research*, Vol. 66, 1993, pp. 161-167.
48. Dery, R., Landry, M., and Banville, C., "Revisiting the Issue of Model Validation in OR: An Epistemological View," *European Journal of Operational Research*, Vol. 66, 1993, pp. 168-183.
49. Oral, M., and Kettani, O., "The Facets of the Modeling and Validation Process in Operations Research," *European Journal of Operational Research*, Vol. 66, 1993, pp. 216-234.
50. Leijnse, A., and Hassanizadeh, S. M., "Model Definition and Model Validation," *Advances in Water Resources*, Vol. 17, 1994, pp. 197-200.
51. Kleijnen, J. P. C., "Verification and Validation of Simulation Models," *European Journal of Operational Research*, Vol. 82, 1995, pp. 145-162.
52. Caughlin, D., "Verification, Validation, and Accreditation (VV&A) of Models and Simulations through Reduced Order Metamodels," *1995 Winter Simulation Conference*, Arlington, VA, 1995.
53. Kleijnen, J. P. C., "Statistical Validation of Simulation Models," *European Journal of Operational Research*, 1995, pp. 21-34.
54. Sargent, R. G., "Verifying and Validating Simulation Models," *1996 Winter Simulation Conference*, Coronado, California, 1996.
55. Pace, D. K., "Fidelity Considerations for RDE Distributed Simulation," 1, *1997 Fall Simulation Interoperability Workshop Papers*, 1997.
56. Pace, D. K., "Dimensions and Attributes of Simulation Fidelity," 1, *1998 Fall Simulation Interoperability Workshop Papers*, 1998.
57. Box, G. E. P., "Sampling and Bayes' Inference in Scientific Modeling and Robustness," *Journal Statist. Soc. A*, Vol. 143, No. A, 1980, pp. 383-430.
58. Oberkampf, W. L., Martellucci, A., and Kaestner, P. C., "SWERVE Surface Pressure Measurments at

Mach Numbers 3 and 8," Sandia National Laboratories, SAND84-2149, SECRET Formerly Restricted Data, Albuquerque, NM, 1985.

59. Roache, P. J., "Need for Control of Numerical Accuracy," *Journal of Spacecraft and Rockets*, Vol. 27, No. 2, 1990, pp. 98-102.
60. Blottner, F. G., "Accurate Navier-Stokes Results for the Hypersonic Flow Over a Spherical Nosetip," *Journal of Spacecraft and Rockets*, Vol. 27, No. 2, 1990, pp. 113-122.
61. Mehta, U. B., "Computational Requirements for Hypersonic Flight Performance Estimates," *Journal of Spacecraft and Rockets*, Vol. 27, No. 2, 1990, pp. 103-112.
62. Martellucci, A., "The Challenging Process of Validating CFD Codes," American Institute of Aeronautics and Astronautics, AIAA-90-1402, *AIAA 16th Aerodynamic Ground Testing Conference*, Seattle, WA, 1990.
63. Desideri, J. A., Glowinski, R., and Periaux, J., eds. *Hypersonic Flows for Reentry Problems, Vol. I: Survey Lectures and Test Cases for Analysis*, Springer-Verlag, Berlin, 1991.
64. Desideri, J. A., Glowinski, R., and Periaux, J., eds. *Hypersonic Flows for Reentry Problems, Vol. II: Test Cases-Experiments and Computations*, Springer-Verlag, Berlin, 1991.
65. Walker, M. A., and Oberkampf, W. L., "Joint Computational/Experimental Aerodynamics Research on a Hypersonic Vehicle: Part 2, Computational Results," *AIAA Journal*, Vol. 30, No. 8, 1992, pp. 2010-2016.
66. Marvin, J. G., "CFD Validation Experiments for Hypersonic Flows," AIAA, AIAA-92-4024, *AIAA 17th Aerospace Ground Testing Conference*, Nashville, TN, 1992.
67. Lin, S. J., Barson, S. L., and Sindir, M. M., "Development of Evaluation Criteria and a Procedure for Assessing Predictive Capability and Code Performance," *Advanced Earth-to-Orbit Propulsion Technology Conference*, Marshall space Flight Center, Huntsville, AL, 1992.
68. Singhal, A. K., "Validation of CFD Codes and Assessment of CFD Simulations," *Fourth International Symposium of Transport Phenomena and Dynamics of Rotating Machinery*, Honolulu, Hawaii, 1992.
69. Dolling, D. S., "Problems in the Validation of CFD Codes Through Comparison with Experiment," North Atlantic Treaty Organization, AGARD, AGARD-CP-514, *Fluid Dynamics Panel Symposium: Theoretical and Experimental Methods in Hypersonic Flows*, Torino, Italy, 1992.
70. Graves, R., Jr., "Software Validation: The Bridge from R&D to Industrial Application," American Institute of Aeronautics and Astronautics, AIAA-92-0587, *30th Aerospace Sciences Meeting & Exhibit*, Reno, NV, 1992.
71. Abgrall, R., and Desideri, J. A., "The European Hypersonic Data Base: A New CFD Validation Tool for the Design of Space Vehicles," AIAA, AIAA-93-3045, *AIAA 24th Fluid Dynamics Conference*, Orlando, FL, 1993.
72. Oberkampf, W. L., "A Proposed Framework for Computational Fluid Dynamics Code Calibration/Validation," American Institute of Aeronautics and Astronautics, AIAA Paper No. 94-2540, *18th AIAA Aerospace Ground Testing Conference*, Colorado Springs, CO, 1994.
73. Bussoletti, J. E., "CFD Calibration and Validation: The Challenges of Correlating Computational Model Results with Test Data," American Institute of Aeronautics and Astronautics, AIAA-94-2542, *18th AIAA Aerospace Ground Testing Conference*, Colorado Springs, CO, 1994.
74. Settles, G. S., and Dodson, L. J., "Supersonic and Hypersonic Shock/Boundary-Layer Interaction Database," *AIAA Journal*, Vol. 32, No. 7, 1994, pp. 1377-1383.
75. AGARD, "A Selection of Experimental Test Cases for the Validation of CFD Codes," NATO Advisory Group for Aerospace Research & Development, AGARD-AR-303-Vol. I, 1994.
76. AGARD, "A Selection of Experimental Test Cases for the Validation of CFD Codes," NATO Advisory Group for Aerospace Research & Development, AGARD-AR-303-Vol. II, 1994.
77. AGARD, "Quality Assessment for Wind Tunnel Testing," NATO Advisory Group for Aerospace Research & Development (AGARD), AGARD-AR-304, 1994.
78. Mehta, U. B., "Guide to Credible Computational Fluid Dynamics Simulations," American Institute of Aeronautics and Astronautics, AIAA Paper No. 95-2225, *26th AIAA Fluid Dynamics Conference*, San Diego, CA, 1995.
79. Roache, P. J., "Verification of Codes and Calculations," American Institute of Aeronautics and Astronautics, AIAA Paper No. 95-2224, *26th AIAA Fluid Dynamics Conference*, San Diego, CA, 1995.
80. Aeschliman, D. P., Oberkampf, W. L., and Blottner, F. G., "A Proposed Methodology for CFD Code Verification, Calibration, and Validation," ICIASF, Paper 95-CH3482-7, *16th International Congrerss on Instrumentation for Aerospace Simulation Facilities*, Dayton, OH, 1995.
81. Cosner, R. R., "CFD Validation Requirements for Technology Transition," American Institute of Aeronautics and Astronautics, AIAA Paper No. 95-2227, *26th AIAA Fluid Dynamics Conference*, San Diego, CA, 1995.
82. Oberkampf, W. L., Blottner, F. G., and Aeschliman, D. P., "Methodology for Computational Fluid Dynamics Code Verification/Validation," American Institute of Aeronautics and Astronautics, AIAA Paper No. 95-2226, *26th AIAA Fluid Dynamics Conf.*, San Diego, CA, 1995.
83. McDonald, W. W., "Introduction to the Submarine Damage Mechanisms Project," Naval Surface Warfare Center, IHTR 1824, Indian Head, 1995.

84. Mehta, U. B., "Guide to Credible Computer Simulations of Fluid Flows," *Journal of Propulsion and Power*, Vol. 12, No. 5, 1996, pp. 940-948.

85. Sindir, M. M., Barson, S. L., Chan, D. C., and Lin, W. H., "On the Development and Demonstration of a Code Validation Process for Industrial Applications," American Institute of Aeronautics and Astronautics, AIAA Paper No. 96-2032, *27th AIAA Fluid Dynamics Conf.*, New Orleans, LA, 1996.

86. Dolling, D. S., "Considerations in the Comparison of Experimental Data with Simulations Consistency of Math Models and Flow Physics," American Institute of Aeronautics and Astronautics, AIAA-96-2030, *27th Fluid Dynamics Conference*, New Orleans, LA, 1996.

87. Cosner, R. R., "The Role of Validation in the CFD Process at McDonnell Douglas/St. Louis," American Institute of Aeronautics and Astronautics, AIAA-96-2273, *19th AIAA Advanced Measurement and Ground Testing Technology Conference*, New Orleans, LA, 1996.

88. Barber, T. J., "The Role of Code Validation and Certification in the Design Environment," American Institute of Aeronautics and Astronautics, AIAA Paper No. 96-2033, *27th AIAA Fluid Dynamics Conference*, New Orleans, LA, 1996.

89. Benek, J. A., Kraft, E. M., and Lauer, R. F., "Validation Issues for Engine/Airframe Integration," American Institute of Aeronautics and Astronautics, AIAA 96-2031, *27th AIAA Fluid Dynamics Conference*, New Orleans, LA, 1996.

90. Rizzi, A., and Vos, J., "Towards Establishing Credibility in CFD Simulations," American Institute of Aeronautics and Astronautics, AIAA Paper No. 96-2029, *27th AIAA Fluid Dynamics Conference*, New Orleans, LA, 1996.

91. Van Wie, D. M., and Rice, T., "Quantification of Data Uncertainties and Validation of CFD Results in the Development of Hypersonic Airbreathing Engines," American Institute of Aeronautics and Astronautics, AIAA 96-2028, *27th AIAA Fluid Dynamics Conference*, New Orleans, LA, 1996.

92. Oberkampf, W. L., Aeschliman, D. P., Henfling, J. F., Larson, D. E., and Payne, J. L., "Surface Pressure Measurements on a Hypersonic Vehicle," American Institute of Aeronautics and Astronautics, AIAA Paper No. 96-0669, *34th Aerospace Sciences Meeting*, Reno, NV, 1996.

93. Haynes, T. S., Reed, H. L., and Saric, W. S., "CFD Validation Issues in Transition Modeling," American Institute of Aeronautics and Astronautics, AIAA-96-2051, *27th AIAA Fluid Dynamics Conference*, New Orleans, LA, 1996.

94. Holden, M. S., Moselle, J. R., Sweet, S. J., and Martin, S. C., "A Database of Aerothermal Measurements in Hypersonic Flow for CFD Validation," American Institute of Aeronautics and Astronautics, AIAA Paper No. 96-4597, *AIAA 7th International Space Planes and Hypersonic Systems and Technologies Conf.*, Norfolk, VA, 1996.

95. Porter, J. L., "A Summary/Overview of Selected Computational Fluid Dynamics (CFD) Code Validation/Calibration Activities," American Institute of Aeronautics and Astronautics, AIAA Paper No. 96-2053, *27th AIAA Fluid Dynamics Conference*, New Orleans, LA, 1996.

96. Mair, H. U., "Preliminary Compilation of Underwater Explosion Benchmarks," SAVIAC, 1, *67th Shock and Vibration Symposium*, 1996.

97. Beck, M. B., "Water Quality Modeling: A Review of the Analysis of Uncertainty," *Water Resources Research*, Vol. 23, No. 8, 1987, pp. 1393-1442.

98. Tsang, C.-F., "A Broad View of Model Validation," OECD, Paris, France, *Proceedings of the Symposium on Safety Assessment of Radioactive Waste Repositories*, Paris, France, 1989.

99. Ketelle, R. H., Lee, R. R., Bownds, J. M., and Rizk, T. A., "Model Validation Lessons Learned: A Case Study at Oak Ridge National Laboratory," Oak Ridge National Laboratory, CONF-89085406/DE89 015900, Oak Ridge, TN, 1989.

100. LeGore, T., "Benchmark Test Cases for Computational Fluid Dynamics," American Society of Mechanical Engineers, FED-Vol. 93, *1990 Spring Meeting of the Fluids Engineering Division*, Toronto, Ontario, 1990.

101. Davis, P. A., Olague, N. E., and Goodrich, M. T., "Approaches for the Validation of Models Used for Performance Assessment of High-Level Nuclear Waste Repositories," Sandia National Laboratories, NUREG/CR-5537; SAND90-0575, Albuquerque, NM, 1991.

102. Sheng, G., Elzas, M. S., Oren, T. I., and Cronhjort, B. T., "Model Validation: A Systemic and Systematic Approach," *Reliability Engineering and System Safety*, Vol. 42, 1993, pp. 247-259.

103. Oreskes, N., Shrader-Frechette, K., and Belitz, K., *Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences*, in *Science*, 1994, pp. 641-646.

104. IEEE, "ANSI/IEEE Std 100-1984: IEEE Standard Dictionary of Electrical and Electronics Terms," 1984.

105. IEEE, "IEEE Standard Glossary of Software Engineering Terminology," IEEE, IEEE Std 610.12-1990, New York, 1991.

106. Knepell, P. L., and Arangno, D. C., *Simulation Validation: A Confidence Assessment Methodology*, 1st ed., IEEE Computer Society Press, Washington, 1993.

107. Rakitin, S. R., *Software Verification and Validation*, Artech House, Boston, MA, 1997.

108. ANS, "American Nuclear Society: Guidelines for the Verification and Validation of Scientific and Engineering Computer Programs for the Nuclear Industry," ANSI/ANS-10.4-1987, 1987.

109. NRC, "Regulatory Guide 1.168: Verification, Validation, Reviews, and Audits for Digital Computer Software Used in Safety Systems of

Nuclear Power Plants," US Nuclear Regulatory Commission, Available: www.nrc.gov/NRC/RG/01/01-168.htmt, Washington, 1997.

110. ISO, "ISO 9000-3: Quality Management and Quality Assurance Standards - Part 3: Guidelines for the application of ISO 9001 to the development, supply and maintenance of software," International Standards Org., Geneva, Switzerland, 1991.
111. Schmauch, C. H., *ISO 9000 for Software Developers*, ASQC Quality Press, Milwaukee, WI, 1994.
112. Schwer, L. E., *Personal Communication*, 1999.
113. Sindir, M. M., and Lynch, E. D., "Overview of the State-of-Practice of Computational Fluid Dynamics in Advanced Propulsion System Design," American Institute of Aeronautics and Astronautics, AIAA Paper No. 97-2124, *28th AIAA Fluid Dynamics Conference*, Snowmass, CO, 1997.
114. Saaty, T. L., *The Analytic Hierarchy Process*, McGraw-Hill, New York, 1980.
115. Lee, L. H., and Poolla, K., "Statistical Validation for Uncertainty Models," Springer-Verlag, Lecture Notes in Control and Information Sciences, Vol. 202, *Feedback Control, Complexity, and Identification: A festschrift for Professor George Zames*, Montreal, Canada, 1994.
116. Lee, L. H., and Poolla, K., "On Statistical Model Validation," *Journal of Dynamic Systems, Measurement and Control*, Vol. 118, 1996, pp. 226-236.
117. Draper, D., "Assessment and Propagation of Model Uncertainty," *Journal of the Royal Statistical Society B*, Vol. 57, No. 1, 1995, pp. 45-97.
118. Laskey, K. B., "Model Uncertainty: Theory and Practical Implications," *IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans*, Vol. 26, No. 3, 1996, pp. 340-348.
119. McKay, M. D., "Evaluating Prediction Uncertainty," Los Alamos National Labs., NUREG/CR-6311 LA-12915-MS, Los Alamos, NM, 1995.
120. Hanson, K. M., "A Framework for Assessing Uncertainties in Simulation Predictions," *Physica D*, Vol. 133, 1999, pp. 179-188.
121. Hills, R. G., "Statistical Validation of Engineering and Scientific Models with Application to CTH," Sandia National Labs., Albuquerque, NM, 2000.
122. Wilson, G. E., and Boyack, B. E., "The Role of the PIRT in Experiments, Code Development and Code Applications Associated With Reactor Safety Assessment," *Nuclear Engineering and Design*, Vol. 186, 1998, pp. 23-37.
123. Oberkampf, W. L., Aeschliman, D. P., Tate, R. E., and Henfling, J. F., "Experimental Aerodynamics Research on a Hypersonic Vehicle," Sandia National Labs., SAND92-1411, Albuquerque, NM, 1993.
124. Oberkampf, W. L., and Blottner, F. G., "Issues in Computational Fluid Dynamics Code Verification and Validation," *AIAA Journal*, Vol. 36, No. 5, 1998, pp. 687-695.
125. AIAA, "Assessment of Wind Tunnel Data Uncertainty With Application to Wind Tunnel Testing," American Institute of Aeronautics and Astronautics, S-071A-1999, Reston, VA, 1999.
126. Coleman, H. W., and Steele, W. G., Jr., *Experimentation and Uncertainty Analysis for Engineers*, Second Edition ed., John Wiley & Sons, New York, 1999.
127. Box, G. E. P., Hunter, W. G., and Hunter, J. S., *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, Wiley, New York, 1978.
128. Milliken, G. A., and Johnson, D. E., *Analysis of Messy Data: Vol. 1; Designed Experiments*, Lifetime Learning Pubs., Belmont, CA, 1984.
129. Youden, W. J., "Enduring Values," *Technometrics*, Vol. 14, No. 1, 1972, pp. 1-11.
130. Roache, P. J., *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers, Albuquerque, NM, 1998.
131. Oberkampf, W. L., Diegert, K. V., Alvin, K. F., and Rutherford, B. M., "Variability, Uncertainty, and Error in Computational Simulations," ASME, ASME-HTD-Vol. 357-2, *AIAA/ASME Joint Thermophysics and Heat Transfer Conference*, Albuquerque, NM, 1998.
132. Klir, G. J., and Folger, T. A., *Fuzzy Sets, Uncertainty, and Information*, 1st ed., Prentice Hall, Englewood Cliffs, NJ, 1988.
133. Walley, P., *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, 1991.
134. Klir, G. J., and Wierman, M. J., *Uncertainty-Based Information: Elements of Generalized Information Theory*, Vol. 15, Physica-Verlag, Heidelberg, 1998.
135. Pal, N. R., "On Qunatification of Different Facets of Uncertainty," *Fuzzy Sets and Systems*, Vol. 107, 1999, pp. 81-91.
136. Oberkampf, W. L., DeLand, S. M., Rutherford, B. M., Diegert, K. V., and Alvin, K. F., "Estimation of Total Uncertainty in Computational Simulation," Sandia National Laboratories, SAND2000-0824, Albuquerque, NM, 2000.
137. Richtmeyer, R. D., and Morton, K. W., *Difference Methods for Initial-Value Problems*, Interscience, New York, NY, 1967.
138. Morton, K. W., *Numerical Solution of Convection-Diffusion Problems*, CRC Press, Boca Raton, FL, 1996.
139. Alvin, K. F., Oberkampf, W. L., Rutherford, B. M., and Diegert, K. V., "Methodology for Characterizing Modeling and Discretization Uncertainties in Computational Simulation," Sandia National Laboratories, SAND2000-0515, Albuquerque, NM, 2000.
140. Richardson, L. F., and Gaunt, J. A., "The Deferred Approach to the Limit," *Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences*, Vol. 226, 1927, pp. 299-361.

141. Babuska, I., and Oh, H.-S., "Pollution Problem of the p- and h-p Versions of the Finite Element Method," *Communications in Applied Numerical Methods*, Vol. 3, 1987, pp. 553-561.
142. Babuska, I., Strouboulis, T., Upadhyay, C. S., and Gangaraj, S. K., "A Posteriori Estimation and Adaptive Control of the Pollution Error in the h-Version of the Finite Element Method," *International Journal of Numerical Methods in Engineering*, Vol. 38, 1995, pp. 4207-4235.
143. Babuska, I., Ihlenburg, F., Strouboulis, T., and Gangaraj, S. K., "A Posteriori Error Estimation for Finite Element Solutions of Helmholtz' Equation - Part II: Estimation of the Pollution Error," *International Journal of Numerical Methods in Engineering*, Vol. 40, 1997, pp. 3883-3900.
144. Oden, J. T., Feng, Y., and Prudhomme, S., "Local and Pollution Error Estimation For Stokesian Flow," *International Journal of Numerical Methods in Fluids*, Vol. 27, 1998, pp. 33-39.
145. Zhang, X. D., Pelletier, D., Trepanier, J. Y., and Camarero, R., "Verification of Error Estimators for the Euler Equations," American Institute of Aeronautics and Astronautics, AIAA-2000-1001, *38th AIAA Aerospace Sciences Meeting*, Reno, NV, 2000.
146. Yee, H. C., and Sweby, P. K., "Aspects of Numerical Uncertainties in Time Marching to Steady-State Numerical Solutions," *AIAA Journal*, Vol. 36, No. 5, 1998, pp. 712-724.
147. Gamerman, D., *Markov Chain Monte Carlo*, Chapman & Hall, London, 1997.
148. Red-Horse, J. R., Paez, T. L., Field, R. V., and Romero, V., "Nondeterministic Analysis of Mechanical Systems," Sandia National Laboratories, SAND2000-0890, Albuqueque, NM, 2000.
149. Kleijnen, J. P. C., *Statistical Tools for Simulation Practitioners*, 1st ed., Marcel Dekker, Inc., New York, 1987.
150. Palmer, T. N., "Predicting Uncertainty in Forecasts of Weather and Climate," *Reports on Progress in Physics*, Vol. 63, 2000, pp. 71-116.
151. Glimm, J., and Sharp, D. H., "Stochastic Methods for the Prediction of Complex Multiscale Phenomena," Los Alamos National Laboratory, LAUR-97-3748, 1997.
152. Earman, J., *Bayes or Bust?*, The MIT Press, Cambridge, MA, 1992.
153. Chen, S., Margolin, L., and Sharp, D. H., eds. "Predicatability: Quantifying Uncertainty in Models of Complex Phenomena," 133, *Physica D*, 1999.
154. Glimm, J., Hou, S., Kim, H., Sharp, D. H., and Ye, K., "A Probability Model for Errors in the Numerical Solutions of a Partial Differential Equation," Los Alamos National Laboratory, LAUR-99-5352, Los Alamos, NM, 1999.
155. Chorin, A. J., Kast, A. P., and Kupferman, R., "On the Prediction of Large-Scale Dynamics Using Unresolved Computations," Lawrence Berkeley National Laboratory, LBNL-42283, Berkeley, CA, 1998.
156. Chorin, A. J., Kast, A. P., and Kupferman, R., "Optimal Prediction of Underresolved Dynamics," *Proceedings of the National Academy of Sciences*, Vol. 95, 1998, pp. 4094-4098.
157. Chorin, A. J., Kast, A. P., and Kupferman, R., "Unresolved Computation and Optimal Prediction," *Communications in Pure and Applied Mathematics*, Vol. 52, 1999, pp. 1231-1254.
158. Lehmann, E. L., *Testing Statistical Hypotheses*, Wiley, New York, 1986.
159. Roache, P. J., "Discussion: Uncertainties and CFD Code Validation," *Journal of Fluids Engineering*, Vol. 120, 1998, pp. 635-636.