SANDIA REPORT

Description of the Sandia Validation Metrics Project

Timothy G. Trucano, Robert G. Easterling, Kevin J. Dowding, Thomas L. Paez, Angel Urbina, Vicente J. Romero, Brian M. Rutherford, and Richard G. Hills

Sandia National Laboratories

Description of the Sandia Validation Metrics Project

Timothy G. Trucano
Optimization and Uncertainty Estimation

Robert G. Easterling
Nuclear Weapons Program Integration and Studies

Kevin J. Dowding
Thomas L. Paez
Angel Urbina
Vicente J. Romero
Validation and Uncertainty Quantification Processes

Brian M. Rutherford
Independent Surveillance Assessment & Statistics

Sandia National Laboratories
P. O. Box 5800
Albuquerque, New Mexico 87185-0819

Richard G. Hills
Department of Mechanical Engineering
New Mexico State University
Las Cruces, New Mexico 88003

Abstract

This report describes the underlying principles and goals of the Sandia ASCI Verification and Validation Program Validation Metrics Project. It also gives a technical description of two case studies, one in structural dynamics and the other in thermomoechanics, that serve to focus the technical work of the project in Fiscal Year 2001.

Acknowledgements

Contents

List of Illustrations

Executive Summary

Recent developments have dictated fundamental changes in the processes associated with the design, certification, and management of the United States nuclear weapons stockpile. For example, one of the most important effects of these changes is the accelerated growth of the application of large-scale computational modeling and simulation in nuclear weapons certification processes. The DOE Accelerated Strategic Computing Initiative (ASCI) program is broadly tasked with developing the methodologies, computing hardware, and computational software for enabling computational science and engineering (CS&E) that can be confidently used for high-consequence U. S. nuclear stockpile applications.

When considering the role of computational modeling and simulation in stockpile applications, an issue of paramount concern is our ability to assess and forecast the expected accuracy of complex computational predictions. The response of Sandia National Laboratories to this issue has been to initiate the **Validation Metrics Project** (VMP). The VMP is funded under the DOE Accelerated Strategic Computing Initiative (ASCI) Verification and Validation (V&V) program at Sandia. The goal of the VMP is to develop and apply quantitative model validation methodologies to measure and improve confidence in ASCI CS&E. As such, we view the VMP as an important component of achieving success for the ASCI program at Sandia. The purpose of this document is to describe the underlying framework for the general problem of validation of code applications, the approach taken to address the issues comprising it, and the project goals for the VMP.

There are four essential elements in the process of applying CS&E in high-consequence applications.

(1) We must compare computational models with experimental data, either existing or to be gathered in a directed manner. Such comparisons, of necessity, require dealing with uncertainty in both the experimental data and the computational model.

(2) We must develop rigorous, quantitative measures (metrics) for assessing the confidence, or lack thereof, that results from the experimental data – model comparison element (1). This element is also directly associated with model qualification activities, where the canonical question that must be answered is: "Is the comparison of the model with the data good enough to support the intended application of the model?"

(3) We must predict model error in application regimes different than those covered by the data comparisons associated with elements (1) and (2). Prediction of error is clearly important for high-consequence applications.

(4) We must develop methodologies for forecasting uncertainty in our prediction of the model error in element (3).

A more extensive discussion of features of these elements is given in the main body of this report. In our view, each of these elements requires research and development. The VMP emphasis is on the first two elements – comparison and confidence metrics – for

fiscal year (FY) 2001. The latter two elements – predicting error and forecasting uncertainty– will receive greater attention as the project evolves, in particular in FY 2002.

We have chosen to focus the work of FY 2001 on two weapons-related case studies. Such case studies are natural vehicles for researching, prototyping, applying, and generalizing methodologies. The objectives of both case studies is to develop useful, quantitative comparison methodologies for the associated experimental data and computational models in the presence of uncertainty; and to quantify model application confidence resulting from these types of comparisons.

The first case study centers on the subject of structural random vibration that is of interest in the current Stockpile-to-Target-Sequence (STS) normal environment requirements. This class of application offers a number of advantages, including the fact that there is a long history of certification using random vibration STS criteria makes it a well-understood problem, and quantitatively characterized validation data are available or can be measured.

The second case study involves analyses surrounding the time-dependent thermal decomposition of hydrocarbon foams and their influence on STS abnormal environment analyses. This application is notable in that it has complex, less well-characterized validation data associated with it than is the case for the structural dynamics case study.

These case studies usefully encompass a variety of issues associated with validating the application of CS&E models for stockpile management applications at Sandia. Complete descriptions of these case studies are provided in the main body of this report.

The major deliverables for the VMP in FY 2001 are documentation and recommendations. This report represents the first of the expected deliverables. We will also document the approaches and results of the work on the two case studies. In addition, we plan to document a set of guidelines for designing and performing validation experiments in CS&E that resonate with the technical aspects of this project. Finally, we will document lessons-learned from the work of FY 2001, as well as a set of recommendations for the next phases in the project based on the work performed in FY 2001. These recommendations will stress, of course, factors that we believe have the greatest influence on generalizing the knowledge and experience developed under this project to the larger ASCI CS&E capability at Sandia.

1. Introduction

The Department of Energy's (DOE's) Accelerated Strategic Computing Initiative (ASCI) is designed to develop high performance computational tools and models to help manage the safety and reliability of the enduring nuclear weapon stockpile. This effort is the essence of the Science-Based Stockpile Stewardship (SBSS) program. For new computational tools to be used with confidence by weapon designers and decision-makers, in lieu of or as a supplement to physical testing, a sound and viable verification and validation program (V&V) is required. The goal of the ASCI V&V program at Sandia (Pilch *et al*., 2000a) is to characterize the predictive capability of ASCI computational tools in a credible manner while remaining within the constraints of available funding resources.

The ***Validation Metrics Project*** (VMP), one of While several elements of the Sandia V&V program, focuses on validation. Validation is also the emphasis of this report. *Validation* of a computer model for a given application is defined in the DOE Defense Programs (DOE/DP) ASCI Program Plan (DOE, 2000) as:

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

The design and implementation of the validation process are not well defined, either in the ASCI Program Plan (DOE, 2000) or the Sandia V&V planning guidance (Pilch, *et al*., 2000a). The Sandia ASCI V&V program created the VMP to develop methods for gauging "the degree to which a computer model is an accurate representation of the real world …."

In the near term, the VMP is directly aimed at supporting achievement of the ASCI FY 2002 Level 1 Milestone, VV-2.1 (DOE, 2000): "Demonstrate initial validation methodology …for normal and abnormal Stockpile-to-Target-Sequence (STS) environments behavior." This is meant to deliver one or more technical methodologies that support scientific assessment of the predictive capability of one or more Sandia ASCI codes *for their defined applications*. The primary focus of Sandia ASCI code applications is technical support of weapon systems projects for normal, hostile, and abnormal environments.

The validation assessment methodologies developed and recommended as a result of this project are intended to withstand critical scrutiny. Such scrutiny will likely include peer review processes originating within weapon system programs at Sandia. Deciding whether the predictive capability of codes for specific applications measured by the means developed in this project is appropriate or sufficient for the required application is beyond the scope of this project. Another element of the Sandia V&V program is

required. This element, ***code qualification***, along with the peer review process formulated by the Sandia V&V program (Pilch, *et al*., 2000b), the weapons engineering certification campaign (Campaign 6), as well as other weapons program elements, will ultimately decide when a code has sufficient predictive capability for a specific application. Development of useful validation metrics, whether by the VMP or by other means, is therefore a *necessary* condition to perform code qualification, but not a sufficient condition.

The VMP builds on a body of previous work (Paez, *et al*., 1996; Barney, *et al*., 1997; Hunter, *et al*., 1997; Perez, *et al*., 1997), as well as work performed in FY 2000 under a Sandia V&V program element called "Validation methodologies" (Blackwell *et al*., 2000; Dowding, 2001a,b; Easterling, 2001a,b,c; Hills and Trucano, 1999, 2001a, and 2001b). Major differences between the Validation Methodologies project and the current VMP are (1) increased focus on the FY 2002 ASCI V&V milepost; and (2) a more deliberate basis for generalizing the executed work and broadening its technical impact on the Sandia V&V program. The FY 2001 work performed under the VMP must lead to additional work to be performed in FY 2002 to achieve the stated milepost.

In our opinion, previous work has established some basic ideas and candidate approaches to use in assessing a code's predictive capability. The next step is to test these ideas, and either refine and more generally apply them, or reject them in favor of other approaches. Experience applying these conceptual ideas will allow us to refine them or reject them. It is with this in mind that the core FY 2001 VMP tasks are two case studies. The first is related to weapons simulations in normal STS vibration environments. We call this the ***Structural Dynamics Case Study***. In the second case study, we will study the problem of foam decomposition under thermal states related to fire relevant to abnormal STS environments. We call this work the ***Foam Decomposition Case Study***.

The case studies focus the technical content of this project. Differences in the type and quality of the data in each case study, as well as differences in the anticipated use of the computational models in each case, provide wide scope for formulating and implementing technical validation methodologies. Both case studies are believed to offer appropriate opportunities for technical progress. Results of work on these case studies are the most important product of this project and will be carefully documented.

Additionally, elements of this project are aimed at generalizing the case studies. This white paper lays out principles and constraints that are likely to be required for any technically focused validation study for application of Sandia ASCI codes. We will also develop a more precise picture of the broad range of data that must be encompassed in performing code application validation for STS environments applications. This information will be documented in a more rigorous work product that will develop Sandia V&V program guidelines for appropriate validation data characteristics and requirements for comparison of codes with such data.

The path forward for the VMP is also important. Our recommended path forward will be based on a detailed accounting of lessons-learned from FY 2001. We will specifically discuss the applicability of the methodologies investigated in the case studies to further

validation studies based on data presented in our larger picture of validation data available at Sandia.

In the remainder of this report we will first develop a motivation for studying the subject of validation metrics in Section 2. In Section 3 we will present a discussion of the technical challenges that lie at the core of successful execution of this project. It is important for the reader to note that this material does not serve as a blueprint for the technical approaches that are most favored by the VMP. Rather, our emphasis here has been on elucidating ideas that convey the rationale for viewing work in the area of validation metrics as having a research and development component, as well as an applications component. Sections 4 and 5 describe our planned case studies, including a detailed description of the planned work and the intended outcomes. We conclude in Section 6 with a description of the success metrics for evaluating this project and the anticipated path forward of the VMP. A summary and some conclusions are presented in Section 7.

Note to the reader: Unless otherwise stated, "model" is this report always means the computational code and the physics models it implements, as well as the input necessary for performing calculations (such as the mesh definition, choice of material parameters, and computational parameter settings).

2. Goals and Approach

2.1 Goals

The main goal of the ASCI V&V program is to establish confidence in modeling predictions of ASCI simulation codes for application to the Stockpile Stewardship Program. In the context of two case studies, the VMP supports achieving this goal by developing and testing systematic means by which (1) code predictions should be compared with data and (2) assessment of code application predictive capabilities rooted in these comparisons should be performed and communicated. The project will generalize the results of the case studies to provide guidelines for:

- the design and conduct of suites of model validation experiments and computations;

- the comparison of experimental and computational results to evaluate predictive capability; and

- the extension of these results to statements about the predictive capability of computational models in untested situations.

The decision to apply these guidelines systematically in the Sandia ASCI V&V program is, of course, beyond the scope of this project.

2.2 Motivation.

The emphasis of the ASCI program is on computational prediction. Therefore, users of computational predictions, from weapon designers to weapons program decision-makers, need to be provided with information on how accurate the prediction is and on what basis.

In short, the most important, and likely hardest, question we must ask is: What is the quantitative assessment of our confidence in the prediction? An example of a desirable statement that illustrates a useful way of answering this need is: "Based on our understanding of the underlying physics, our ability to translate that understanding into a computational code, and our comparative analysis of an extensive suite of experiments and computations, we are confident that actual system response will differ from the computational prediction by no more than 20%." Whether such a statement can ever be made, or whether a soft approximation to such a statement can ever be made, is at issue and certainly a major subject of study for the VMP. Such statements of predictive capability, or approximations to them, provide the necessary frame of reference against which a computational prediction can be compared to a weapon program requirement. It is very important to understand how close we might be able to come to providing such information for high consequence applications of ASCI codes. Confidence in the predictions derived from simulations of weapon performance in normal, abnormal, and hostile environments supporting STS requirements demands a credible, convincing basis for quantifying predictive capability. It is hard to imagine a significant weakening of this position while still claiming high impact for ASCI computational predictions.

2.3 Philosophy.

Confidence we have in computational predictions comes predominantly from comparisons of computations with experiment and/or test data. We use the term ***model validation*** to describe this comparison. Directed experimental programs must be conducted more or less explicitly for this purpose, along with historical data when available. Model validation experiments may range from single-physics, tightly controlled laboratory-scale experiments for a single phenomenon, through of combined or coupled physical tests, to very complex and expensive system-level multi-physics tests.

As we progress through this spectrum of potential validation experiments, test units may vary from simple geometric shapes of single materials to complex assemblies with many materials. At the same time, the acquired experimental test data vary from being very simple and well characterized to very complex and poorly characterized. At each level of complexity, however, the intent for model validation is that comparisons of computational predictions to experimental results provide useful information on current accuracy and future predictive capability of the model. The basic underlying philosophy is that the more satisfactory the comparisons are observed to be, the greater the resulting confidence one will have in the model in specified applications. While this general philosophy and approach are easily recognized to those who have performed model validation, constructive and quantitative guidelines for the systematic implementation of the process are not well characterized.

The Sandia V&V planning guidelines (Pilch, *et al*., 2000a) distinguished three increasingly complex categories of model validation, along with a proposed model accreditation category. These categories reflect increased complexity of the validation activity from a single phenomenon, through simply coupled phenomena, to fully coupled phenomena of the complexity of the intended application. In this report, we find it convenient to distinguish qualitatively different validation problems that lie at opposite

ends of the experimental spectrum. The first problem is ***phenomenon-centric*** validation (another useful choice of words is ***unit-centric validation***). The second problem is ***application-centric*** validation (or ***system-centric validation***). Success in application-centric validation requires success in phenomenon-centric validation as a necessary precondition. Quantitative validation methodology to support the solution of both validation problems is needed, as we will now explain.

The goal of phenomenon-centric validation is to address the degree to which a model adequately represents a single physical phenomenon for the application of interest. The phenomenon itself may be well characterized experimentally or it may not. The key point is that validation experiments supporting phenomenon-centric validation are designed to isolate that particular phenomenon. It is important to ensure consistency between the experiment and the model so that the experiment satisfies the basic assumptions and application conditions of the model. Generally, such validation experiments are also designed to involve relatively simple geometries and materials. The acquired data are expected to be simpler than those associated with the driving application. The validation process and needed metrics in this case are focused on determining how accurately a model predicts the isolated phenomenon as represented by the defined validation experiments and their resulting data.

Application-centric validation measures the accuracy with which a model represents an intended realistic application. The applications of interest to Sandia typically have several phenomena of interest that are coupled to a greater or lesser degree. Validation experiments in this case typically must be multi-phenomena experiments or tests, with more complex geometries and materials than the experiments that support phenomenon-centric validation. We expect fewer useful data to be available for application-centric validation and these data will be more complex than arise in phenomenon-centric validation (possibly exhibiting complex space-time correlations that may not be present in simpler validation problems). The model calculations required for comparison with such experiments are also typically far more complex than those required for performing phenomenon-centric validation. The application-centric validation process and needed metrics focus on characterizing how accurately the model predicts the complicated and coupled phenomena representative of a driving application. This is a far more daunting task than phenomenon-centric validation, which is not to claim that phenomenon-centric validation is easy. It properly includes phenomenon-centric validation as one of its tasks. Unfortunately, it is often the case that the large-scale experiments or tests that serve the role of validation experiments for application-centric validation are intended only to support qualitative system design or qualification decisions. In such a case, these tests are sometimes not even properly *defined* as model validation experiments (for example, the data that may be collected could be impossible to compare with the model under study), let alone capable of providing the kind of quantitative data that are required for model validation.

A key issue for research is how to properly integrate these two distinct validation problems. For example, the two levels of associated validation experiments, along with any intermediate levels, cannot be performed in isolation. Certainly for the physical phenomena underlying the application phenomenon-centric validation is a critical

precursor to application-centric validation. Each level of validation contributes to the assessment of the predictive capability of the model for the intended application. Phenomenon-centric validation should be defined by the ultimate application need for the model and the resulting demands of application-centric validation. This view has been stressed in the Sandia V&V guidelines (Pilch, *et al*., 2000a). Effective links between phenomenon-centric and application-centric validation, such as specification of application relevant parameter ranges, conditions, and accuracy requirements that constrain the associated phenomenon-centric validation activities, are important for ultimate success of application-centric validation. For example, while it may be reasonable from a phenomenon-centric perspective to conduct experiments pertaining to material performance at 300°C, if the application requirement pertains to performance at 750°C, then the phenomenon-centric validation experiments may have little value relative to application-centric validation.

It is our observation that a majority of current Sandia directed validation efforts are phenomenon-centric, despite the emphasis on application-centric validation that is in the overall Sandia V&V program (Pilch, et al., 2000a). While this emphasis is consistent with a philosophy of performing tasks that are technically simpler prior to performing more complex tasks, the fact remains that an approach that is focused only on phenomenon-centric validation will not likely be successful if the resulting efforts are divorced from the required applications. We also note that there are situations where phenomenon-centric validation experiments can be performed while applications-centric validation experiments cannot, as well as vice versa. An important goal of the VMP is to develop understanding of how to properly link these two canonical validation problems and to explore methodologies for achieving application-centric validation that can be generalized.

2.4 Process

The general goal of the VMP, as stated above, is to define the process by which the predictive capability of computational models can be assessed from model-experiment comparisons. To reiterate, by *predictive capability*, we mean the potential difference between a computationally predicted outcome for a given situation (*e.g.*, weapon subjected to a specified shock at some point in the Stockpile-to-Target Sequence) or for a given class of situations (*e.g.*, weapon deliveries within defined velocity and attitude envelopes) and the outcome of the actual event or events. Obviously, prediction cannot be perfect and we have no expectations that it has to be in any case we care about. A large number of random and systematic effects, as well as the inherent approximations of mathematical models, intervene to make nature differ from computation. Phenomena occur in nature that are not captured in a computational model. But, with the right set of experiments and computations and with suitable analysis methods, supported by an understanding of the underlying phenomena, we *may* be able to measure and bound the cumulative magnitude of these effects and thus measure predictive capability.

Figure 1 displays our view of the overall process of measuring or estimating model predictive capability *and* using that information to make stockpile stewardship decisions. The top ellipse – "System Environments and Performance Requirements" – encompasses

15

the intended application of the computational model. There, system requirements specify various performance goals and outcomes for given scenarios and environments. The computational model will be applied to predict these system performance outcomes in the specified scenarios. The simplicity of the diagram masks an elaborate and complex interaction between the weapon program customer that must use the computational prediction and the personnel that develop the prediction. Comparing the prediction to the system requirements fundamentally demands a yardstick – a metric – for measuring uncertainty of the prediction, as well as suitably quantitative requirements for the application of this metric. Neither of these is a trivial task. Model uncertainty is depicted in Figure 1 by the shaded ellipse surrounding the prediction ellipse. Requirements and scenarios are depicted as crisply defined, a rather strong simplification of the reality in some applications.
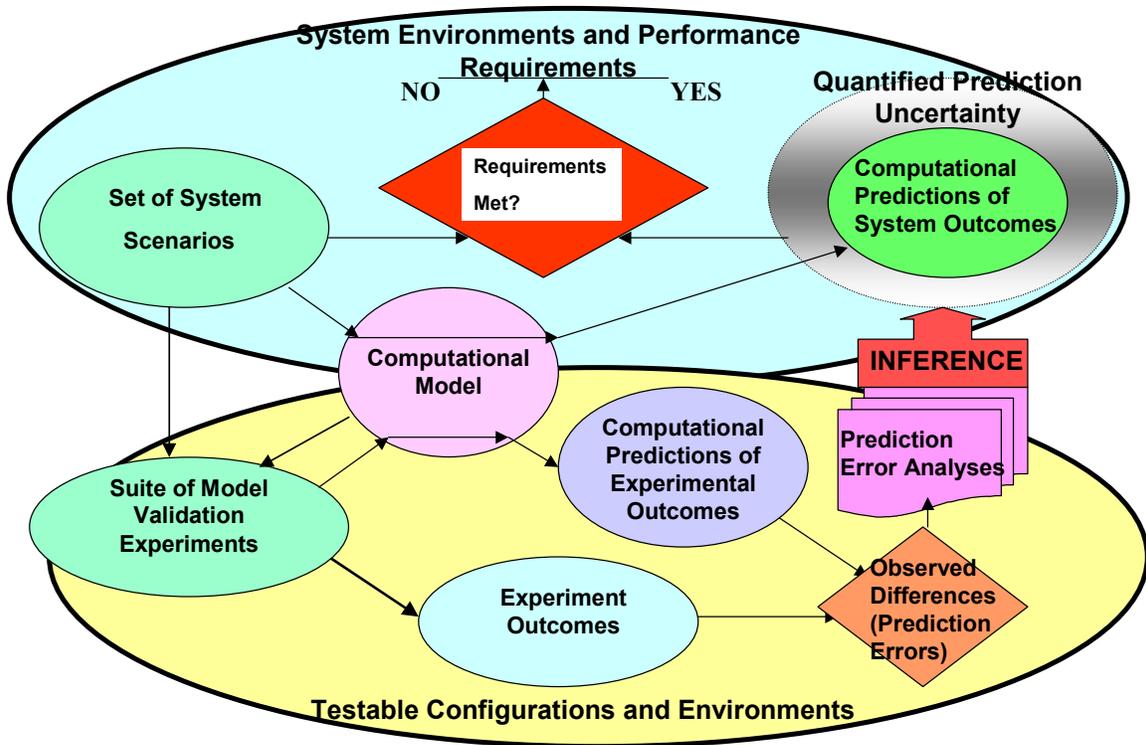
Figure 1. A system view of the process of quantification of predictive capability for complex computational models illustrates the underlying difficulty.

To develop a yardstick for comparing computational predictions to system requirements, a suite of validation experiments and their attendant uncertainty, corresponding computations, and their comparison must be conducted. The bottom ellipse in Figure 1 – "Testable Configurations and Environments" – suggests the key elements required to do this. The state of our current information is included in this region. In addition, the process for generating additional information and an appropriate collection of validation

data is schematically depicted. This information provides, first, an evaluation of modeling fidelity in the tested situations. (How well predictive capability is evaluated is clearly dependent on the definition and conduct of the validation process.) Second, and most importantly, the ensemble of observed differences that constitute current experience must serve as the *quantitative* basis for an inference about model prediction capability in the system applications of interest in the upper ellipse. This, in a nutshell, is the major technical problem underlying any effort to develop useful "validation metrics."

A more detailed view of the process elements centered in Fig. 1 around the transition from "Inference" to applying the computational predictions is shown in Figure 2. System requirements, such as that the system shall function successfully under specified delivery conditions, lead to scenarios that capture those requirements. Based on our scientific understanding and computational capabilities, we develop a computational model for predicting the outcomes of these scenarios. To evaluate the predictive capability of this model, we need to conduct a validation test program. This validation program, consisting of a suite of experiments (as discussed in section 2.3) and corresponding computational simulations of those experiments, is constrained and shaped by the scenarios, the capabilities of the model, and the experimental capabilities.

For each experiment in the program, calculations mirroring the experiment must be performed. There can be major compatibility issues to resolve in accomplishing this, even at the level of phenomenon-centric validation experiments. We will refer to this issue as **alignment**, as in "proper alignment of model computations and validation experiments." While it is beyond the scope of the current report to discuss this issue in much detail, we will provide additional insight into what this means below.

Assuming that the alignment between experiment and calculation is good enough we can meaningfully compare and analyze the differences between the calculations and the experiment. This process provides quantitative information about the predictive capability of the model for the phenomena conditions and application constraints at which the validation experiments have been conducted. We must then seek to draw inferences about the model predictive capability for the scenarios by which system performance relative to requirements can be judged, typically not fully covered (or addressed at all) in existing experiments. A cycle of inference is thus created. If predictive capability is judged satisfactory, by criteria whose specification may be difficult in particular applications, the model is deemed ready to contribute to the decision processes. If predictive capability is judged unsatisfactory, we must begin again and examine everything from system design and underlying requirements to computer model to validation test program and discover opportunities for improvement. Applying this process, it may well be determined that the model predictive capability is insufficient for the intended application or, possibly, that an understanding that the predictive capability is sufficient cannot be formed. The early phases of the VMP, in particular the case studies described in Sections 4 and 5, do not directly address the question of whether prediction capability associated with the models under study is sufficient for a specific task. However, this question is of implicit concern to the project.
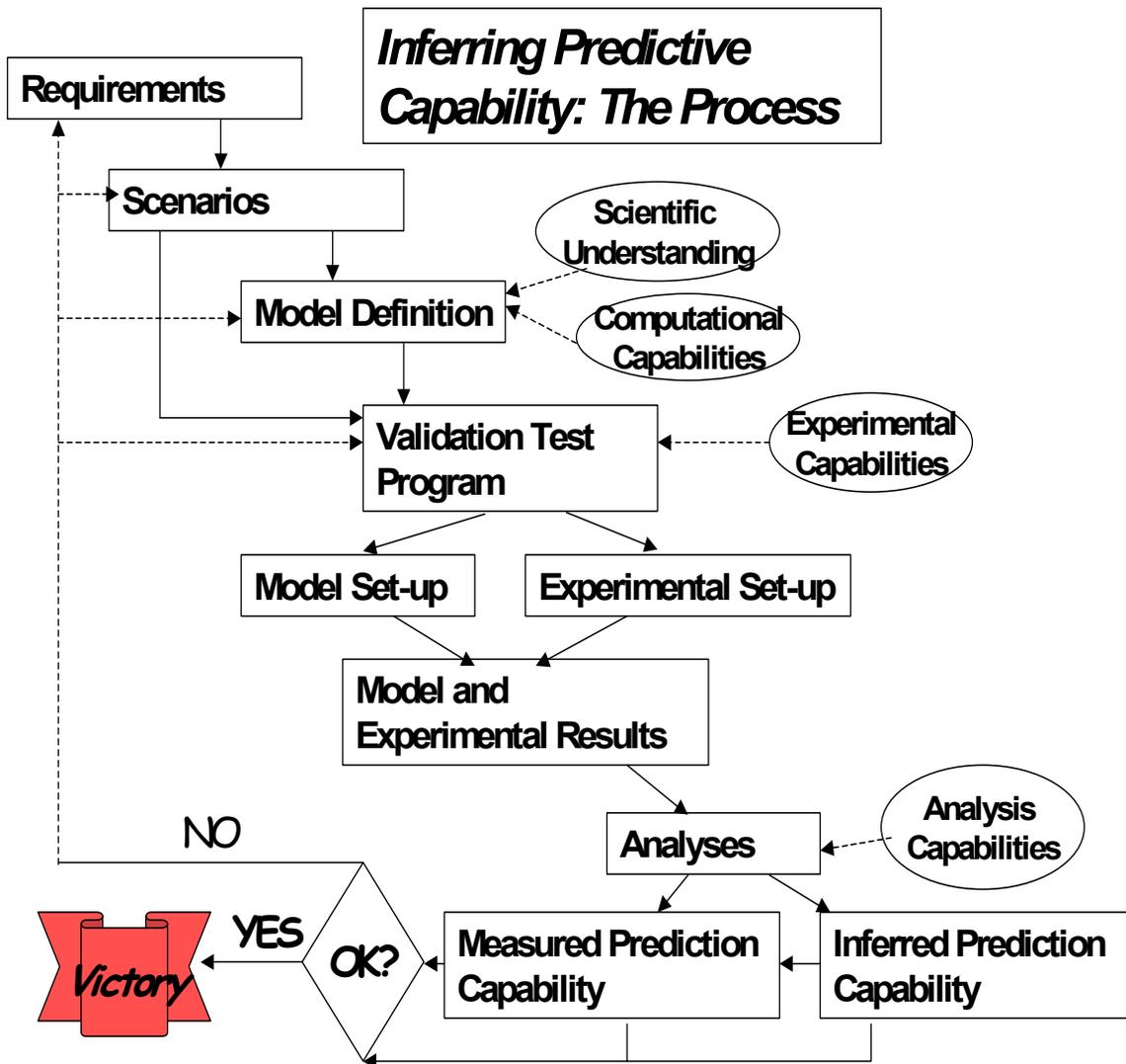
Figure 2. The process steps and decision elements required for quantifying predictive capability from validation activities.

2.5 Requirements and Constraints

The most basic requirements and constraints of the VMP originate in the details underlying Figure 2. In particular, case studies chosen to investigate quantitative validation methodologies and resulting metrics of confidence must appropriately capture the process elements of Fig. 2 and satisfy the following constraints:

1. The case studies must be chosen so that their origins in the Sandia weapons program are apparent and appropriate.

2. Sandia Defense Programs scenarios governing the intended application of the models in the case studies must be defined.

3. Models in the case studies must be at a sufficient level of sophistication that the data comparison exercises will make sense. We assume that this requirement is implicitly satisfied by the choice of ASCI codes as appropriate models for us in the case studies.

4. Experiments compatible with calculations will be (have been) performed within the timeframe of the VMP – *i.e.*, during the first three quarters of FY 2001.

5. Experimental data appropriate for comparison to model predictions must be available or acquired during the course of the case study.

6. Appropriate comparison methods and methodologies for assessing the resulting confidence must be available for investigation in the case studies.

7. While the decision box "OK" in Figure 2 requires more than validation, as argued in the Introduction, the case studies should provide suitable opportunities for determining what elements of validation metrics are necessary for model qualification processes.

In this report we will discuss the impact of these requirements on the VMP. Items four, five, and six above are intended to enforce a sufficient level of model – experiment alignment to insure that the process of comparison is not vacuous. The individual case study descriptions in Sections 4 and 5 are specifically designed around these constraints. An expanded discussion of the issues involving model, data, and comparisons is given below in Section 3.

We now emphasize one more time that underlying all of the work proposed for the VMP are the elements in Figure 2 labeled "Requirements" and "Scenarios." These are critical to our ultimate ability to qualify our predictive confidence in a model for a defined application. A great deal of work is required to develop these requirements and scenarios to the level of detailed needed for the process shown in Figure 2. In Appendix A we give a preliminary discussion of some of the issues involved in mapping requirements and scenarios defined by the weapons program at Sandia into a usable set of requirements and constraints for the predictive confidence assessment process in Figure 2 for an abnormal environments application. This is primarily aimed at introducing the reader to the difficulties.

3. Technical Issues

3.1 Introduction

Figures 1 and 2 illustrate the role of the model validation in the overall ASCI V&V program. Most of the physical and computational experimentation designed to support model-based prediction occurs in this stage of analysis. A number of objectives are addressed through this experimentation, including those listed below: 1) Comparison of results from the two sources can lead an iterative process of refinements to the models,

experimental methods, or even to requirements and/or expectations. Two major concerns in assessing the model are accuracy and predictive capability. 2) Comparison of the information generated through validation experimentation to requirements can provide feedback clarifying the needs of, and directions for, additional experimentation. This would include determination of whether or not sufficient information has been obtained for the decisions required in this stage of analysis. 3) For situations where model performance appears adequate for the application, all information should be used to evaluate the model's predictive capabilities. A mathematical framework is introduced in this section that can facilitate these objectives. Challenges involved with and methods of analysis for addressing these objectives are discussed.

To discuss the wide variety of technical issues associated with validation metrics, we utilize a simple mathematical formalism to frame the discussion. To start with, we represent a prediction generated by a computational simulation as

$$y^*(x) = M(x, \phi), \tag{3.1}$$

where $M(x, \phi)$ represents the computational model of the phenomenon of interest; x is a vector of model input variables, some of which could be fields (a function of space and/or time) in the most general case; ϕ is a vector of numerical parameters, some of which could also be fields; y^* is the model output or *prediction*, which could potentially be a vector of quantities that themselves could be fields. The variable y used below is the experimentally determined response of which y^* is the model prediction.

In general, the model's input vector x describes a physical entity and the environment to which it is subjected for purposes of both computation and experiment. The input vector x is what enforces alignment of the model with the experiment. For the case of some quantities, such as dimensions, geometries, initial velocities, or specified mechanical loadings, the variables expressing these details clearly enter into x. Thus, x will include physical dimensions and environmental variables, as well as variables defining initial and boundary conditions.

But, we also want to place less obvious parameters, such as material model parameters, in this vector. By doing this we are not assuming that the material model in the computation is correct in any given application. Rather, we are stating the obvious fact that if a material model allows us to choose between copper and iron, for example, then we should not choose copper when iron is used in the experiment. This is an *alignment* problem. We would certainly not expect to agree with the experiment if we allowed this confusion. (But the reader should note that an additional problem arises if a computational model happens to agree with an experiment that is not aligned with the calculation.) The material model quantities that allow this kind of distinguishing of materials that are also distinguishable in the experiment must be included in the vector x.

In general, we expect that any material variables that are truly constitutive should be placed in the input vector x. Such variables uniquely identify materials both computationally and experimentally (or they would not be constitutive!). How accurate the model is that utilizes these variables is, of course, a *validation* question, not an *alignment* question.

The numerical model parameter vector ϕ contains all parameters that are necessary for performing calculations that do not influence alignment of the calculation with the experiment. In short, ϕ contains all other numerical model parameters that are not placed in x. These parameters include all parameters used to define numerics, such as grid definitions and numerical algorithm specifications. Examples of such parameters are those controlling grid placement and size, time steps and their controls, convergence criteria, accuracy controls on adaptive difference schemes, and many other numerical features. In short, ϕ certainly contains the parameters that are used to influence the accuracy of a numerical calculation, but that have no effect on whether or not one is calculating the correct experiment. More confusingly, ϕ may also contain material parameters that are non-constitutive, or at least not obviously constitutive. An example of such a parameter could be a molecular relaxation time that has no experimental analog. There may also be similar parameters involved in complex boundary conditions, such as radiant heating sources.

The most essential point about x is that the choice of x in a model prediction *defines* the corresponding validation experiment. This is worth repeating, so we do so: *The validation experiment is defined in terms of the input vector x, not the computational parameter vector ϕ.* An existing experiment, in turn, defines the choice of input vector x for an appropriately aligned calculation. The experiment has no explicit influence on the choice of ϕ, however.

As we suggested above, there are gray areas where the decision to place an input variable in x or ϕ is not clear. And there are other subtle issues that should suggest to the reader why an extended discussion of the topic of alignment is complex and beyond the scope of the present report. For example, what if a critical dimension is measured incorrectly in an experiment? What if a critical dimension is improperly constructed in the numerical model? What if a measurement gauge in an experiment is improperly located in space? What if the computational model records synthetic data at an incorrect location? What if the vector ϕ is chosen so improperly that the resulting calculation will not simulate the associated experiment with any accuracy? In all of these cases, and an effectively infinite number of other similar situations, are the experiment and the calculation still aligned? Can they be aligned, at least in principle? How would we recognize that the source of the error that will surely result if we compare calculations and experiments in these situations is due to alignment, rather than some problem of validation? In the following we basically assume that the experiments and models we discuss are in good alignment, thus begging these questions. This is for simplicity of presentation in the current report, not because these questions are not important.

Returning to our basic discussion, the computer model $M(x, \phi)$ written above is an operator that transforms input variables x and numerical parameters ϕ into the predicted result $y*$. This transformation is assumed to be deterministic in this report in the sense that for a given and fixed specification of the vectors x and ϕ the code always gives the same $y*$.

We also assume in this discussion that the code has been verified (thus resolving one of the issues we raised above by definition). One way of stating this is that the code is "validation ready." In practice this simplistic assumption is likely to be wrong. What we really assume by this statement is that sufficient effort has been devoted to verification of the model that pursuing validation makes sense to begin with.

The reader should keep in mind that the simple formalism of Eq. (3.1) can disguise all of the mathematical effort included in an ASCI code at Sandia. For example, $M(x,\phi)$ could represent the approximate solution (operator) of a system of partial differential equations. As an illustration of this point consider conservation of energy for heat conduction in solid material, which is modeled by the solution of the single partial differential equation on a finite one-dimensional domain Ω

$$\frac{\partial}{\partial X}(k\frac{\partial \hat{y}}{\partial X}) - \rho C_p \frac{\partial \hat{y}}{\partial t} = 0 \text{ for } X \in \Omega. \qquad (3.2)$$

In this case, the components of the input vector x include the thermal physical properties of the material and the definition of the spatial domain Ω and its boundary $\partial\Omega$, as well as any initial and boundary condition specifications. We write $x = (\rho, k, C_p, \Omega, \partial\Omega)$, where we have hidden possibly complex initial and boundary conditions in the symbols $\{\Omega, \partial\Omega\}$. The experimental response y is the dependent variable temperature. We have written it as $\hat{y}$ in (3.2) because we wish to emphasize our *assumption* that what is measured experimentally is well enough approximated by (3.2) for this comparison to make sense to begin with (the alignment problem again). The numerical solution of (3.2), whether by finite elements or finite differences or some other method, is the generic model operator that we have written in (3.1). This numerical solution yields the model prediction y^*. We should point out that it is not necessarily the case that a dependent variable from a model like the solution of the partial differential equation (3.2) is what is compared with experiment for model validation purposes. In many cases attributes of the dependent variables given by functionals, such as maximum values, are used instead.

Now, consider an experiment conducted at a specified x, having outcome $y(x)$. The ***prediction error*** of the model at x is defined as

$$e_x = y(x) - y^*(x) \qquad (3.3)$$

Note that e_x contains all of the bias and uncertainty associated with both the experiment and the model. We symbolically neglect the dependence of e_x on ϕ in the following discussion. First, varying ϕ certainly does not influence the experimental outcome. The dependence of $y^*(x)$ on ϕ is implicit – we simply choose not to write it out. If the reader prefers, one can interpret our assumption of "verified code" above to mean that we have also selected ϕ to provide the optimal (and sufficient!) numerical accuracy for simulating the experiment(s) under discussion. Evaluating model predictive capability requires selecting a sequence $\{x_i, y(x_i), y^*(x_i) : i = 1,...,n\}$ and evaluating e_x at each element of this sequence. (This is not possible if validation experiments cannot be aligned with model predictions.)

3.2 Statistical Model

Prediction error e_x contains random and systematic contributions. We will represent the relationship between $y(x)$ and $y*(x)$ in the form

$$y(x) = y*(x) + e_x \qquad\qquad (3.4)$$

where e_x is now specifically understood to be a random variable (or random field) with an unknown probability distribution that generally depends on x. Equation (3.4) is a statistical model of the relationship between experiment y and model prediction $y*$. The prediction error e_x represents the combined effects of both model and experimental uncertainty, including measurement error and more general uncertainty associated with the experimental outcome at x, as well as model prediction uncertainty. An alternative formalism could make this more explicit, but the current discussion is sufficient for our purposes. The term e_x also reflects any hidden lack of alignment between experiment and model, in the sense we discussed it in Section 3.1, between the model and experiment, although this contribution may be very difficult to elucidate.

Comparing experiment and model for the purpose of evaluating model accuracy and predictive capability involves investigating e_x over a range of x-values determined through the experimental design (see section 3.3.2 below). There are at least two specific concerns about model performance to be evaluated. First is the concern that there are regions of the input space where the expectation $E(e_x)$ is not (even approximately) zero, indicating possible biases in the model. A second concern is that the model is not adequately predictive; there are inputs, either not included in x (and hence not utilized in the modeling) or included yet not fully utilized by the model. Models that lack predictive capability yield high variability in e_x, even after compensation has been made for measurement errors associated with the physical experimentation.

Evaluating the information available through model validation involves trying to make an assessment of both the information generated through completed experimentation, and the information that might be gained through further testing and computational analysis. This evaluation is best accomplished after model deficiencies discussed in the previous paragraph have been addressed and, to the extent possible, cleared up so that reasonable models for e_x are available. Comparison of this information to requirements can also help establish the needs and directions for further experimentation and computation. An important goal of the VMP that plays a role in each of the key objectives stated in Section 3.1 is to determine guidelines for evolving the suite of x – points at which to conduct experiments and computations and for comparing and analyzing results.

Viewing the differences between experiment and model as statistical has engineering precedent. For example, in bridge design, civil engineers use a mathematical model for "scour" – the erosion of soil around a bridge's foundation due to river flooding (Johnson 1995). This model is a function of soil type, flood magnitude, river velocity and other pertinent variables. For predictions civil engineers incorporate an additional "modeling factor" to represent the deviation of actual scour depths from the theoretical model. This

modeling factor corresponds to e_x in our notation and is statistical. A more complex application where differences between observations and model results are statistical is climate prediction (Epstein, 1985; Wilks, 1995).

A key assumption underlying the present discussion is that e_x can be modeled as a random variable or random field. One purpose of the VMP is to investigate methodology that can be used to model e_x and to test the viability of these assumptions. It is important to realize that information about the probability distribution of e_x at a particular x-point can be obtained from the suite of experiments and calculations, not just the experiment(s) solely at that x − point. For example, it may be reasonable to assume, based on the ensemble of data, that the variance of e_x is constant over some x-region. Thus, all the e_x data in that region can be pooled to estimate this common variance. More generally, spatial statistical methods (Sacks *et al.*, 1989; Cressie, 1993; Rutherford *et al.*, 2001) can be used to link the ensemble of data and provide estimates of the distribution of e_x at selected x − points. That is, e_x distributions at nearby x-points are likely to be more similar than e_x distributions at widely separated x-points. Spatial models capture and capitalize on such relationships. By whatever modeling method is considered or adopted, the general objectives of the analysis of the e_x data are the following. First, estimate the probability distribution of e_x at the x − points at which computations and experiments are conducted. Second, estimate the probability distribution of e_x at x − points pertaining to physical entities and environments that have not, cannot, or will not be subjected to physical testing. This estimated distribution can be applied to estimate e_x for the relevant application. Because estimation of predictive capability in this sense must be nondeterministic in our view, we also deem it important to therefore characterize the reliability of the estimate by some means. Whether the estimate of e_x which results from this process is "good enough" is not a question that can be answered by this project. (And, whether a statistical approach can be used to derive *useful* predicted limits on e_x is part of the work of the VMP.) Some recent work of interest in this regard is found in Field and Red-Horse (2001).

In summary, our objectives include answering the questions: "Is the model adequate for the application?" "What is the predictive capability of the code?" and "How well is that capability understood?" Answering the latter two questions directly asserts whatever confidence we might have in predictions based on the application of the model. Their answer also calls for estimating uncertainty associated with the predictions made by the code, even when those predictions may be deterministic, and for quantifying the uncertainty of this estimate. The objectives of the model validation phase of analysis are addressed using data from computational simulations and physical testing. Methods for specifying the data to use – experimental design and methods for the data analysis are the topics of the two remaining subsections of Section 3.

3.3 Experimental Design Challenges

3.3.1 Introduction

The VMP will develop and test methods for assessing model performance and evaluating predictive capability of computational models through case studies where the nature of (3.4) will be investigated for these applications. This section summarizes some of the challenges that are likely to be encountered in this effort and indicates the directions that will be taken. It will act to further inform the reader as to our rationale for selecting the case studies. It is anticipated that in the course of this project's work on these case studies other issues will emerge and novel ways of resolving them may be developed.

In broad terms, validation experimental design means selecting a set of $x-$points that defines appropriate validation experiments as well as defines suitable computational predictions. In terms of physical experiments, this includes determining experimental plans that specify the test hardware, methods, conditions, instrumentation, data collection, and post-processing techniques used to obtain information required for subsequent data analyses. All of these elements have different nuances for experiments that are designed for model validation studies as opposed to phenomena exploration or calibration. We feel that this point must be emphasized. A number of specific experimental design-related issues are important for the model validation applications anticipated for the ASCI program. The headings "Interpolation and Extrapolation," "Experimental Objectives," and "Constraints and Sensitivity," describe the remaining topics to be covered in Section 3.3.

3.3.2 Interpolation and Extrapolation

ASCI applications may involve interpolation or extrapolation from experimental experience. Problems involving interpolation require predictions that are based on results from the region of $x-$space where the experimentation occurs. Those involving extrapolation require predictions based on regions of $x-$space where there are no physical experiments. Intuitively, interpolation among test data should be easier and could still be the purpose of the intended modeling application. Whether extrapolation or interpolation are required in a given case, and what types of prediction will eventually be required, is very dependent on the specific application that defines the validation efforts. (See, for example, the case study descriptions in Sections 4 and 5.) The role of experimental design is illustrated schematically in Fig. 3. For ease of illustration the space of validation experiments and applications is there defined by two meta-variables, configuration and environment. (Other meta-variables may be appropriate for different applications.) In the context of Defense Programs needs at Sandia, because of treaty, regulatory, or economic reasons it may not be possible to test hardware configurations in their required environments. For this reason, Fig. 3 depicts a situation where we seek to extrapolate from test data.

Our basic problem is to extend what we can learn about model predictive capability (represented by the prediction errors $e_x = y(x) - y^*(x)$ in Fig. 3) at the selected $x-$points where we can evaluate it to an inference about predictive capability where we

cannot evaluate it experimentally. A necessary condition for attempting to do this is that we have enough data – enough experimental x – points – to be able to reasonably accomplish this. It may well be that making this decision can not be separated from the predictive inference in complex model applications. Given this presumption, we can then focus on the technical problems of how to do it. If we believe that we don't have enough data, this question as we have framed it then becomes moot. Therefore, we will assume that we have enough data to make the investigation worthwhile.

This type of inference requires an extension of the model itself plus an extension of what we know about unmodeled phenomena, which may be only partially represented or suggested by the observed prediction errors. Making this extension successfully and credibly surely requires subject-matter knowledge about the axes along which we can make such extensions. To illustrate this point, consider the design of a new neutron generator that must function properly when subjected to some range of radiation environments. The Comprehensive Test Ban Treaty precludes the historical means of testing a neutron generator in its threat environment via an underground nuclear explosion. Lesser, yet more experimentally controlled, environments, however, can be obtained from various above-ground radiation sources. On the configuration axis initial computational predictions and experiments may be done for simple geometries, such as a flat plate or an aluminum cylinder, rather than for a neutron generator with its complex assembly of diverse parts and materials. The challenge is to integrate all these results to obtain statements about system-level predictive capability. Appendix A discusses this problem in greater detail for a weapon-in-a-fire scenario.

Choosing the validation experimental design and conducting the experiments leads to several questions that must be successfully addressed. As well, the mechanics of the solution of the inference problem of predicting e_x beyond the region where data has been acquired involves the solution, or approximate solution, of a host of problems. The following subsections will freely speculate about some of the ideas that we feel are germane to attacking these problems. But there are other problems that we will not even speculate about in detail in this report. An example of one such problem is the recognition that subject-matter expertise alone will not dispose of model uncertainty due to unknown unknowns, such as code bugs or a fundamentally incorrect conceptual model. Failure to resolve this problem contributes to e_x in its most general sense. However, not only is such a problem beyond the scope of this report, but its full (or partial) resolution lies beyond the short-term scope of the VMP itself. Clearly, some methodologies for coping with uncertainties not uncovered by comparison with experimental data, whether due to poorly chosen x – points or to an insufficient number of them, are called for. These methodologies appear to us to also be especially important for the problem of usefully bridging phenomenon-centric validation to application-centric validation.

3.3.3 Experimental Objectives

In principle, one of the most useful problems that the VMP might hope to "solve" in some sense is to optimize the design of a validation experiment program. Meaningful validation experiments should be designed to meet one or more explicit objectives. In general, the experiments conducted (1) should provide a sufficient test of model bias and

predictive capability and (2) the collective set of experiments and associated computational predictions should provide a basis for making the desired inference of predictive confidence for the targeted applications in application-centric validation. Of specific concern is that experimental procedures are required that can contribute to the extrapolation of our inference from one region of parameter space to another. What exactly (1) and (2) mean is properly part of the research of this project, especially the case studies. We will simply take the statements at face value in the following in order to help illustrate conceptually some of the required thinking.

There are various ways to translate these objectives into a basis for experimental design. Consider one simple. One measure of predictive capability at x is the standard deviation of prediction error, e_x, at that point. One might define the objective to estimate this standard deviation within P% and then derive the number of experiments required to achieve that precision. These experiments could either be n replications at the selected x-point or n total experiments at different x – points within a region within which it is reasonable to expect an approximately constant standard deviation of e_x. The experiments should also provide the data with which to test such an assumption.
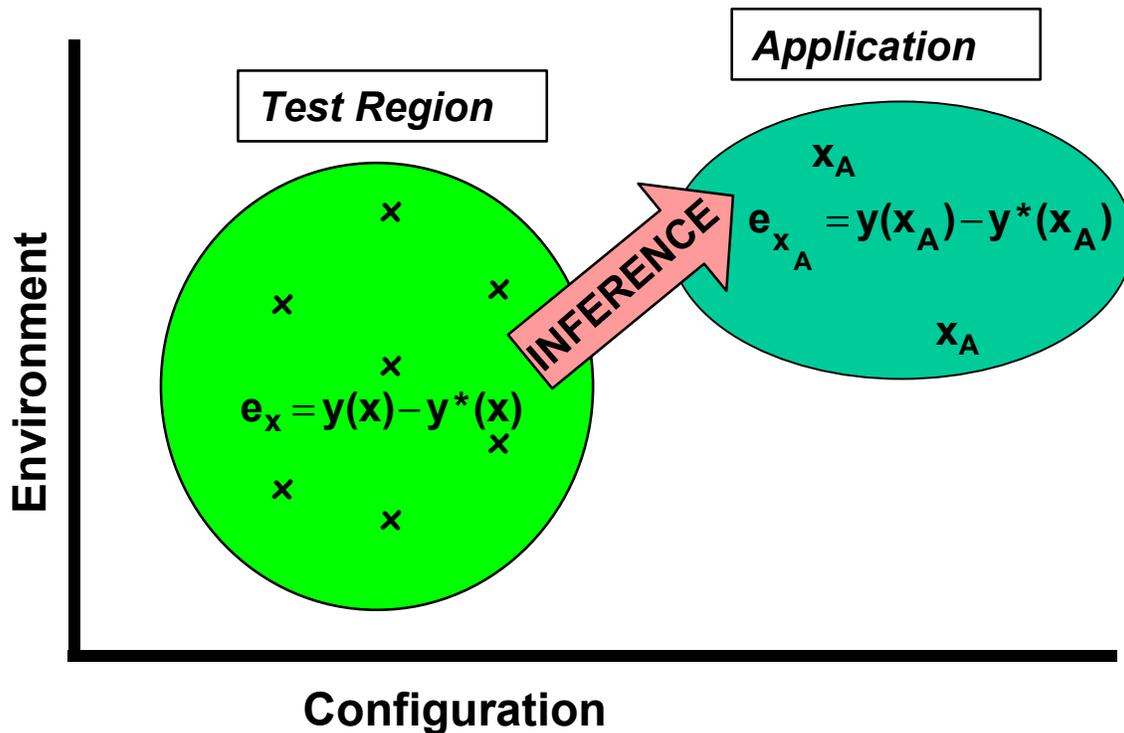

Figure 3. Inferring predictive capability from the model validation process.

The required precision with which to measure predictive capability could also be used in setting experimental goals. Some percentage of experiments might be selected strictly on the basis of subject matter expertise. Another experimental objective might be ***model breaking***. In that case, x − points are chosen for which there is reason, either past experience or subject-matter expertise or both, to believe that the computational model may be inadequate. This kind of validation testing may also be justified by a bounding approach to characterizing prediction capability. In other words, rather than attempt to characterize prediction capability at each of several x − points, one would instead pick situations for which it can be argued that prediction capability elsewhere could be no worse. This is actually quite similar to past practice where are computational predictions have been used to guide the early phases of design and qualification activities. What we are talking about here, though, is a significantly greater formalization of such practice.

The conduct of a validation experiment influences how well predictive capability can be measured and thus addresses the second objective. But whether the basis is sufficient is not resolved by this alone. A variety of random and systematic factors can differentiate computational prediction and nature. Validation experiments need to be conducted in ways that allows these factors to be manifested as they would in an application of interest, if at all possible. For example, predictive capability measured in a tightly-controlled, pristine lab environment may not be appropriate for inferring predictive capability in a much less controlled, noisier application environment. The objective of assessing predictive capability in a specific application influences experimental design in terms of both what is controlled and what is not controlled in the experiments.

One of the work products of the VMP in FY 2001 is to provide general guidance for the design of a set of validation experiments. The hope is that such guidance can serve as a minimal starting point, even though many details will be application dependent. We expect that suitable general guidance can be developed that will prove to be useful.

3.3.4 Constraints

Time, resources, and experimental capability substantially constrain validation experimental design and conduct. Such constraints must be balanced against the experimental objectives in arriving at a plan for model validation experimentation. A difficult decision will have to be made as to whether a meaningful evaluation of predictive capability is possible under existing constraints in any given situation.

. The set of selected x − vectors also needs to be meaningful to both the experimentalist and modeler in order that both computational predictions and experiments at selected x − points can be performed and compared. This is the intuitive heart of our concern with alignment. Our discussion so far has implicitly assumed that the full x − vector can be controlled or accurately measured in a validation experiment, which is an alignment problem. If the modeler's x − vector contains variables that have no experimental meaning, then our hoped-for alignment is likely not true and it may not be possible to make meaningful comparisons. If the modeler's x − vector requires measurements that cannot or will not be made, or if the model output includes measurements that can't be made, then the result will be increased prediction uncertainty.

3.3.5 Sensitivity

The objective of characterizing predictive capability in some quantitative sense over some high-dimensional $x-$ space can quickly require an experimental design for the validation experiments that exceeds available or reasonable resources. It is important to develop rational approaches for simplifying the design. One way to achieve this is to vary only a subset of the variables in x while holding the others fixed at nominal, or at least bounding, values. This may be easier said than done, especially for more complex tests associated with application-centric validation. Typically, one would like to apply sensitivity analysis to this problem to determine those elements of x that the output depends most sensitively on. Unfortunately, resolution of the sensitivity problem is also very resource intensive. In some sense, we also need to reduce the number of elements of x in order to perform sensitivity analysis, a vicious circle. As it is, performing sensitivity analysis for complex models with many dimensional input vectors x remains an important problem, the solution of which will positively influence our ability to perform validation.

Because of the potential expense and difficulty of performing systematic sensitivity analysis, both model simplification and experimental simplification are attractive elements in model validation. However, care must be taken that this type of simplification does not acquire unacceptably *ad hoc* characteristics in complex model validation problems. At its worst, for example, model simplification may simply assume facts that are instead supposed to be established by rigorous validation methodologies. Similarly, experimental simplification may rely on assumptions about physical processes that cannot be fully justified.

3.4   Analysis.

3.4.1 Introduction

After conducting a suite of experiments and computational predictions we assume that the next reasonable task is to analyze the resulting data, $\{x_i, y(x_i), y*(x_i) : i = 1,...,n\}$, where the subscript is used to index the $x-$ points at which experiments and computations were conducted. It is important to note that this subscript is assumed to refer to distinct experiments. Given the computational and experimental outcomes from the suite of experiments, the objective of the analysis of these results is to assess the model for bias and measure and/or estimate predictive capability. The following subsections address additional challenges that arise in this analysis and require some kind of resolution. The topics "Metrics," "Choice of Variables," "Point Prediction," and "Distributional Prediction" are discussed.

3.4.2 Metrics.

Here we provide an example discussion of validation metrics to clarify the technical meaning of the concept. Because we have insufficient evidence of which metrics might be successful for applications of interest to the ASCI program at Sandia, we claim no intent in this discussion to constrain the work of the VMP to only these metrics.

A model of predictive capability at an x – point, such as Eq. (3.4), can be characterized by a variety of "parameters" (in the statistical sense of being a characteristic of a probability distribution) of the probability distribution of e_x. The expected value and the standard deviation of e_x are two important examples. Other parameters might be the square root of the expected squared error $[E(e_x^2)]^{1/2}$, where $E(\cdot)$ denotes expectation; the expected absolute error times three; the 99th percentile of the distribution of absolute error; the lower and upper 95th percentiles on the distribution of e_x; and others that have traditional importance in statistics, including correlations. If the computational model was designed to be conservative on the high-side (i.e., e_x is intended to be negative), the metric of interest might be $P(e_x > 0)$, where $P(\cdot)$ is probability. When e_x has a normal distribution all of these distributional characteristics are functions of the two parameters that characterize a normal distribution, the mean value and the standard deviation, an particularly simple situation. We make no claims that we expect this wonderful simplification in general.

None of these measures of predictive capability are known *a priori*; they can only be estimated from the experimental and computational results. The uncertainty of such estimation will be proportional to the amount of data that is available. It will likely be quite large for very limited data, but this remains to be illustrated in specific investigations. Statistical methods can account for estimation uncertainty by methods such as confidence limits (Brownlee, 1965). For example, with 90% confidence the upper 95th percentile of the distribution of e_x is no more than UL$_{90/95}$. (UL$_{90/95}$ is upper statistical tolerance limit that reflects the amount of data used to estimate a distribution's parameters, in this case the distribution's 95th percentile. See Hahn and Meeker, 1991). Other methods may also be candidates for forecasting the reliability (accuracy and precision) of estimated measures of e_x. The essential point is that any "metric" of predictive capability derived from a model validation process – the "validation metrics" in the title of our project – will be a statistical estimate and the reliability of that estimate must therefore also be considered.

Other metrics that may be of interest in assessing prediction confidence can be obtained by treating the model validation problem as a hypothesis test. In terms of the statistical model Eq. (3.4), a hypothesis that might be tested is $E(e_x) = 0$, which is the hypothesis of no bias at a particular x or perhaps within some x-region. From this approach, the resulting validation metric is either a pass/fail decision or a measure of the degree to which the hypothesis is contradicted by the data. Such metrics, or rather their underlying logic, are not direct measures of model predictive capability, such as an accurate prediction of e_x. One must also apply them carefully. For example, the poorer the quality of predictive or experimental results (in other words, the greater the associated uncertainty), then the harder it will be to reject hypotheses of model invalidity. However, a finding of significant bias, for example, is a essential statement of inconsistency between experimental data and prediction. As a consistency test hypothesis-test based

metrics thus play a clear role in the analysis of the computational and experimental results.

Logically, of course, we view consistency analysis of the comparison of model and experiment as a necessary precursor to any serious effort to predict e_x beyond the established experimental base. For one thing, evidence of inconsistency may be one of our prime methods for determining that model and experiment are actually not aligned, and thus *should probably not be compared at all*. We also note that Hills and Trucano (2001a) analyze a case where e_x is relatively small, but it is also known that the model is inconsistent with the data. What one makes of this discovery in the case of complex application-centric validation is not clear. Such an eventuality at least suggests that any attempt to extrapolate e_x from the data discussed in that report is dangerous.

3.4.3 Choice of Analysis Variables.

In both experiments and computations there are a large number of variables that can be observed and compared. We have already suggested that resource constraints may make simplification via sensitivity analysis or model simplification desirable. But, making the analysis manageable and the results meaningful and communicable also requires a careful simplification of the variables *y* for which to actually evaluate predictive capability.

In our view, the selection of the variables *y* should first be driven by requirements of the motivating application. If the requirement is that peak strain at a given location should not exceed some value, for example, then the model validation objective is to quantify the predictive capability pertaining to calculated peak strain at that location. While it would add confidence in application of the computational model to know that the complete strain versus time history at various sites in the test device can be reasonably well predicted, it is really not technically clear how appropriate it is to devote a lot of analysis to measuring predictive capability over an extensive time and space grid that is complementary to a key requirement. We do know that human confidence in a particular calculation appears to increase when more than one prediction from that calculation is found to agree with experiment.

A requirements focus may be a way to also reduce the dimensionality of the data, which in general may be histories of responses such as acceleration, strain, or temperature in time and space, to a small number of 'integral' variables such as peak acceleration, peak-to-peak strain, the 'area-under-the-curve,' or the time to reach critical temperatures at selected points in a system or component. This is particularly important for models as general as ASCI codes, where an overwhelming amount of output can be generated. Danger lurks, however. We also know that it is often true that a model be in good agreement with one channel of data in an experiment but not in agreement with another data channel. If the requirement is concentrated in the first data channel, it is then doubtful that the confidence stimulated by the sole agreement with that channel is actually well-founded.

3.4.4 Inference

Suppose, as an example, that at n selected $x-$points, $\{x_i : i = 1,...,n\}$ we have conducted and compared the results of model validation experiments to model predictions. Purely for illustrative purposes, suppose in addition that we have performed enough experiments or have other sources providing estimates of the standard deviation of the prediction error distribution of e_x at each point: $\{s_i : i = 1,...,n\}$. Suppose further that we have measures of the precision of these estimated standard deviations. (In a conventional statistical setting, these measures would be the "degrees of freedom" associated with the estimates.) Let the point $x_A \neq x_i$ define a point prediction that is required for the application of interest, as in Fig. 3. The inference problem of concern to us is then how to use the collection of $\{x_i, s_i : i = 1,...,n\}$ results to determine s_A, the *estimated* standard deviation of e_x at x_A, *and* to also obtain a measure of the precision of that estimate.

This, in general, is a difficult and philosophically deep problem. We have no ready solutions or approximations thereof for this problem at the time of writing of this report. However, we can make several observations about this problem. For example, we note that the ability to satisfactorily solve such an inference problem depends on a number of considerations. First, the definition of the $x-$space is critical in order that x_A and the set of x_i be comparable within the goals of the inference. We emphasize again that the definition of the variables in the $x-$vector is not just a modeling issue. For application-centric validation, the experimenter(s), the originator(s) of the requirements, and the decision-maker(s) have to ultimately be able to operate and communicate in terms of this $x-$space.

Next, the ability to infer s_A with precision depends on the location of x_A relative to that of the x_i. If x_A is, in some sense, surrounded by the x_i, the problem is intuitively one of interpolation. If x_A lies beyond the x_i (as in Fig. 3), then inference apparently requires extrapolation. Even if the underlying scientific relationships are known to extend over a region containing both the x_i and x_A, there is no such basis for extrapolating the probability distribution of e_x which, after all, reflects factors in nature not captured by the scientific model. Any kind of inference about e_x, whether based upon interpolation or extrapolation, will be dependent on empirical trends and expert interpretation of those trends, expecially for applications involving extrapolation. It is clearly worthwhile to test at some $x-$points that are as close as possible to the intended application. This means that some system-level testing will be highly desirable for application-centric validation by definition. The VMP seeks to define refinements of the inferential approach that will help to define more precisely what these expensive system-level tests should be. If it is impossible to perform such near application tests, the VMP can still hope to provide value by defining a useful methodology for describing the uncertainty in e_x in the results.

As an example, suppose that radiation effects testing of aluminum cylinders in various above-ground radiation environments and the corresponding computational predictions indicate that peak stress at various locations and orientations can be computationally

predicted under these conditions with a relative standard deviation of about 10%. Suppose that above-ground radiation tests of and computations for a sub-scale partial mock-up of a weapon component indicate a relative standard deviation of about 20%. We now want to make a credible, defensible statement about predictive capability in terms of calculated peak stress at critical points in a realistic component subjected to a radiation environment that is different from those achievable in conventional test facilities. Our ability to formulate such a statement depends on our ability to link the test configurations and environments, $\{x_i : i = 1,...,n\}$, to the application, x_A. Such a linkage may have to be achieved through expert judgement rather than mathematical formalism. And significant uncertainty in this linkage may result from inability to perform a test near application conditions. This uncertainty, of course, should be quantified if at all possible.

In passing, we note that the spatial representation of the experimental design (in x-space) and inference problems involving e_x suggests that spatial statistical methods, such as kriging, may be of use to model a metric which is dependent on e_x, such as the estimated standard deviation of e_x at x, as a function of x. They would then serve as the basis for developing an estimate the value of that metric at x_A and also provide an estimate of the uncertainty of that estimate (Cressie, 1993).

We conclude by observing that current practice is not unfairly characterized by stating that inference of e_x follows from intuitive validation metric (for example, via curve overlays of experimental data and model predictions). If the apparent agreement looks good, then one is more justified in using the model predictions as a surrogate for nature than if the apparent agreement looks bad. We have no illusions that we will eliminate subjectivity from inferences of predictive capability, but we do hope to express it in more useful ways that are especially relevant as guides for the validation experiment process and that then allow us to gain more value out of it for characterizing confidence in models.

3.4.5 Distributional Predictions and Uncertainty Quantification

A deterministic code calculates a prediction for a single, completely specified situation. Predictions of application interest, though, are often distributional predictions, not single point predictions. Weapon systems are not identical and delivery and target conditions vary from mission to mission. In such situations the objective may be to predict the resulting probability distribution of some characteristic of weapon performance, such as maximum shock on a key component, over some probability distribution of system variables and environmental conditions. Features of this distribution, such as the probability of exceeding a failure threshold or expected system performance, are generally of particular application interest. The question is: "How should information about predictive capability, as assessed via validation metrics, be included in model distributional predictions?"

Here is an illustration of the issue. Suppose that a subset $\{x_{R1},...,x_{Rp}\}$ of the x-points in the collection $\{x_i : i = 1,...,n\}$ is to be treated as random to obtain an output distributional prediction. Suppose further, as a starting point, that the probability distribution of

$\{x_{R1},...,x_{Rp}\}$ is specified. There are a number of possible approaches that can be utilized here depending on the assumptions that are made. Under our assumed statistical model (3.4), the law of total variance yields the result (Easterling, 2001) that if one fixes the values of the variables in x that are outside the subset $\{x_{R1},...,x_{Rp}\}$ then

$$V_R[y(x)] = V_R[y*(x) + e_x] = V_R[y*(x) + E(e_x)] + E_R[V(e_x)] \qquad (3.5)$$

where $V(\cdot)$ denotes variance, $E(\cdot)$ denotes expectation, and the subscript R denotes that the indicated variance or expectation is with respect to the distribution of $\{x_R\}$. This expression relates the variance of y in nature (the left-hand side of (3.5)) to the model-based variance of $y*$ and the bias and variance of the prediction-error distribution.

Suppose, to simplify things for this discussion, that $E(e_x) = 0$ for all x in the x-region of interest. This means that predictions are unbiased (a condition one might hope for after the iterative process of checking the model is complete).. Then (3.5) becomes

$$V_R[y(x)] = V_R[y*(x)] + E_R[V(e_x)] \qquad (3.6)$$

Stochastic propagation of the assumed distribution of $\{x_{R1},...,x_{Rp}\}$ through the model $M(x,\phi)$ (typically called uncertainty propagation) provides an estimate of the first right-hand term in (3.6). Model validation experiments and data analysis, if successful, provide an estimate of the second right-hand term. Their sum estimates the left-hand side. Given information about the precision of each right-hand estimate (degrees of freedom in a conventional setting) a measure of the precision of the combined estimate of $V_R[y(x)]$ can be obtained. Further analyses can be done, for example, to estimate failure probability and provide confidence limits on failure probability. This analysis is similar to that done in the civil engineering example mentioned earlier, the difference being that the distribution of e_x estimated here, rather than assumed to be known.

The application of uncertainty quantification (UQ) in the prediction of model distributions facilitates inference of predictive capability from more controlled validation experiments to the less controlled motivating applications. That is, in validation experiments certain x's may be tightly controlled whereas in an application they will vary, perhaps considerably. Predictive capability in the application can then be evaluated by using UQ methods to propagate the application's $x-$ distribution through the model. As long as the extra-model variability observed in the validation experiments and computations also applies to the application, the UQ and the validation results can be combined as in eq. 3.6. Such synthesis allows us to minimize the need to execute validation experiments at the application level, at least in order to develop a credible prediction of e_x and its uncertainty. In situations where some of the assumptions above do not hold or where more complicated features of the response are to be predicted, more involved methods for modeling the distribution (spatially in x-space) are available; see Sacks, *et al.* (1989), Rutherford, *et al.* (2001) for examples.

UQ has other roles in validation. For example, some of the components of the model input vector x might be specified by measurements from the experiment for which a

computational prediction is to be compared, and thus contain measurement error. (Note that this situation precludes performing a true model prediction, and this is not particularly desirable from the viewpoint of model validation.) This measurement error can be propagated into the e_x data. UQ methods can then be used to estimate and remove both the bias and the variance introduced into e_x through the treatment of measurement error in x. Similarly, estimation error associated with the model parameter, ϕ, is transmitted into the e_x data. UQ methods can be used to evaluate the effect of parameter-estimation error on predictive capability.

One prime objective of the UQ effort at Sandia is to provide tools that are needed to perform the computations and aggregate the information necessary to predict probability distributions from the model $M(x,\phi)$. Uncertainty analyses of this type for ASCI-scale models and applications can be computationally expensive. The computational complexity of an ASCI computational model, as well as the management of a potentially large number of model predictions, requires application of parallel computing tools.

In that vein, efforts have been undertaken in FY 2001 to further develop a software environment to perform UQ, in particular distribution prediction via forward propagation of probability distributions through the large-scale computational models $M(x,\phi)$. This particular effort is designed to leverage existing software architecture components from the DAKOTA (Design Analysis Kit for OpTimizAtion) framework (Eldred, 1998). The intent is to develop discipline independent tool capability (that is, equally applicable to models arising in fluid, thermal, solid, and electrical analyses) that utilizes a variety of non-intrusive techniques, including analytical reliability methods and sampling methods. This work also is intended to provide a foundation for future research and development of further UQ methods as the tools gain maturity.

The VMP will likely generate additional needs for UQ efforts at Sandia. Consequently, coordination with UQ software tool development is an essential part of this project. This coordination allows for the rapid incorporation, extension, and evolution of capabilities required to help meet our stated validation tasks. These needs will involve not only UQ techniques applicable to the computer models, but also those more suited to the analysis of experimental data.

4. Structural Dynamics Case Study Description

4.1 Introduction

Improvements in our ability to mathematically model structural physical systems have led to increased reliance on system models for design and performance assessment. Yet it is generally acknowledged that most mathematical models do not perfectly simulate real structural system behavior. Some reasons for this, on the side of the mathematical model, are that (1) there is always uncertainty about the correct values for some mathematical model parameters, and (2) the form of the governing equations and the manner in which the parameters enter only approximate physical reality. On the experimental side, (1) all

measurements of system behavior include noise, (2) measurements are made at only a finite (sometimes small) number of points, (3) measurements provide average behavior over a region, and many other reasons.

In view of these things, it is desirable to define a technique for comparing mathematical model behavior to experimental outcomes that accommodates both analytical and experimental uncertainty. Again, on the analytical side, there are several techniques for assessing the probabilistic character of response measures from mathematical models of physical systems. Some of these techniques are designed to wrap around deterministic models of physical systems. Examples are the Monte Carlo method, first and second order reliability methods, and the advanced mean value method. (See, for example, Wu and Wirsching, 1987.) Other techniques are specifically designed to model stochastic systems. An example is the stochastic finite element method. (See Ghanem and Red-Horse, 1999.)

On the experimental side, the classical methods of statistics make it difficult to estimate probabilistic models of arbitrary system behaviors because of their common reliance on assumptions of Gaussian distribution. However, modern, computer intensive, methods for statistical analysis like the bootstrap (discussed below) can be very useful for assessing the probabilistic character of response measures from experimental physical systems. (See Efron and Tibshirani, 1993.) The technique performs well on measures of system performance that are non-Gaussian, and where measured data are very limited. It yields estimates of bias, standard error and confidence intervals on experimental measures of interest.

This investigation exploits our abilities to develop probabilistic models of measures of behavior of mathematical models of structural dynamic systems as well as the corresponding experimental measures of behavior of the physical systems being modeled. We use this information in a probabilistic/statistical framework to determine the non-acceptability or acceptability of the mathematical model. This can be done for arbitrary measures of system behavior, however, we choose to focus on a metric of the spectral density of system response. Spectral density is the fundamental and most widely employed measure of second order behavior in systems excited with stationary random inputs. An example is presented.

4.2 Stochastic Structural Dynamics

Mathematical models for the analysis of structural dynamic systems are important because they permit the approximate relation of input excitations, system parameters, initial conditions and boundary conditions to the response behavior of structures. This activity is, of course, essential for the assessment and design of structural systems. (For a detailed description of the mathematics of structural dynamics and random vibration, see Wirsching, Paez and Ortiz, 1995.) Mathematical models of passive structural dynamic systems approximate system behavior with matrix equations. When a system is modeled as linear, then the matrix equations are linear. When the excitation is a vector of stationary random processes, the problem is one of random vibration. When the system parameters are random variables or random fields, the problem is one of stochastic

system dynamics. These problems are often characterized as problems of probabilistic structural dynamics.

A system excited by random input executes response that is a random process. When the excitation and response are stationary random processes, they are usually characterized in terms of first and second order averages of motion. The most important second order measure of behavior of a system excited by stationary random excitation is the spectral density. The autospectral density of a random process describes the distribution of mean square signal content in the frequency domain. The integral of the autospectral density of a stationary random process equals the mean square of the random process over all frequencies. When the excitation is modeled as a vector of mean zero stationary random processes it is denoted $\{Q(t), -\infty < t < \infty\}$ and its spectral density characterizes it. Its spectral density is an NxN matrix of functions denoted $S_{QQ}(\omega)$, where ω denotes frequency. The diagonal elements are autospectral densities of the individual excitation random processes at the system degrees of freedom, and the off-diagonal elements are the cross-spectral densities between distinct input random processes.

In practical situations, structural systems excited by stationary random inputs quickly attain a stationary state of response and, therefore, they possess their own spectral densities. The response is a stationary random process and is denoted $\{X(t), -\infty < t < \infty\}$. The auto- and cross-spectral density functions involving random responses at the degrees of freedom of the analyzed system are contained in:

$$S_{XX}(\omega) = H(\omega)S_{QQ}(\omega)H^{*T}(\omega) \qquad -\infty < \omega < \infty \qquad (4.1)$$

where $H(\omega)$ is the matrix of frequency response functions for the system. The i(th) row–j(th) column element in $H(\omega)$ is the frequency response function at degree of freedom i associated with excitation at degree of freedom j. The (i)th row – (j)th column element in the matrix $S_{XX}(\omega)$ is the cross-spectral density between the random responses at the (i)th and (j)th degrees of freedom. The diagonal elements in $S_{XX}(\omega)$ are the autospectral densities of the responses. When the structural system characteristics are deterministic quantities this is a deterministic expression relating the second order averages (in the frequency domain) of the excitation to those of the response.

When the structural system characteristics are random variables or random fields then the relation in Eq. (4.1) is probabilistic. That is, each function in the matrix $H(\omega)$ is a (complex valued) random process, and, in general, all the random processes are statistically dependent. By specifying the probabilistic character of some or all the structural system parameters and using them in a probabilistic analysis we are able to specify the probabilistic character of the response spectral density matrix $S_{XX}(\omega)$. Some methods for accomplishing this are the Monte Carlo method, first and second order reliability methods, the advanced mean value method, and the stochastic finite element method. To denote the potential for expression of the frequency response function matrix and the spectral density of the response random processes to be dependent on random parameters in the structural system, we write:

$$S_{XX}(\omega, A) = H(\omega, A)S_{QQ}(\omega)H^{*T}(\omega, A) \qquad -\infty < \omega < \infty \qquad (4.2)$$

where A is a vector of random variables that describes randomness in the structural system.

In a following section on validation of structural dynamic models, we will propose that the validation metric be defined in terms of a windowed autospectral density. Let $\{X(t), -\infty < t < \infty\}$ denote the response at a point on the structure, and let $S_{XX}(\omega, A), -\infty < \omega < \infty$, denote its spectral density. (This is one element from the diagonal of the matrix defined in Eq. (4.2).) A mean square response metric that is a windowed function of $S_{XX}(\omega, A), -\infty < \omega < \infty$, can be defined:

$$\Sigma_{XX}^2(\omega_c, A) = \int W(\omega, \omega_c) S_{XX}(\omega, A) d\omega \qquad (4.3)$$

where $W(\omega, \omega_c)$ is a suitably defined spectral window centered at the frequency ω_c. The probability distribution of this quantity is a function of the joint probability distribution of the elements of A. Some structural system parameters will be taken as random variables in the numerical example to follow, and the probability distribution of $\Sigma_{XX}^2(\omega_c, A)$ will be estimated using the Advanced Mean Value Method.

4.3   The Bootstrap

Our objective in mathematical model validation is to compare the computed and the experimentally estimated spectral densities for a structural system in a probabilistic framework. The previous section showed how the probability distribution for a metric of the computed spectral density might be obtained. We now show how the statistical character of a metric of spectral density estimated from experimental results might be defined. The method to be used is the bootstrap.   The bootstrap is a technique for statistical analysis of measured data. The data may be non-Gaussian and not Gaussian-related, as is often the case for extreme responses of mechanical systems. The data may be limited in quantity, though clearly, only those extremes of response represented in the data can be assessed, and the probabilities associated with rare events can only be accurately assessed to the degree permitted by the amount of data available. The statistical measure of interest may be one of the classical measures of random data behavior or something more complicated (as in the present application).

The fundamental idea behind the bootstrap is that data measured from a random source are treated, for purposes of statistical estimation, as though they completely describe the random source. For example, when the data $x_j, j = 1, ..., n$, are measured from a scalar random source we may be interested in estimating some parameter θ of the probability distribution from which the data emanate. We might use a formula $\hat{\theta} = g(x)$ (where $x = \{x_1, x_2, ..., x_n\}$, and the formula is obtained, for example, through maximum likelihood estimation) to estimate an estimator of the parameter θ. Normally, we are interested in assessing the quality of the estimator $\hat{\theta}$, by approximating its bias, standard error, or we may wish to use it to estimate confidence intervals on θ. When the data

$x_j, j = 1,...,n$, come from a Gaussian source, and when the parameter is compatible with classical analysis, this assessment can be done using the classical theory of statistics. In other situations, the bootstrap can be useful.

In the current application we seek to extend the use of the bootstrap to the assessment of statistics of the spectral density of a stationary random process. We do so in the following way. Assume that one realization of a scalar random process $\{X(t), -\infty < t < \infty\}$ has been measured and denote it $x(t), 0 \leq t \leq T$. Sample the signal in discrete time at a time increment Δt, separate the measured realization into M segments of equal length, n, window each segment, if desired, using for example, a Hanning window, and denote the results $x_j, j = 1,...,M$. Each x_j is a column vector of length n. The segments represent separate measured realizations of the stationary random process. Based on these segments, the spectral density estimator for the random process is:

$$\hat{G}_{XX}(f_k) = \frac{2C\Delta t}{Mn} \sum_{j=1}^{M} |\xi_{kj}|^2 \qquad k = 0,...,n/2 \tag{4.4}$$

where $\xi_{kj}, k = 0,...,n-1, j = 1,...,M$, is the discrete Fourier transform of the vector x_j, and C is a constant that accounts for the windowing of the measured signal segments.

The statistical behavior of the spectral density estimator can be approximated using the bootstrap. To do so, create a bootstrap sample of the random process realizations by selecting M integer-valued random numbers from the interval [1,M]. Denote these numbers $j_i, i = 1,...,M$, and use them as indices. (The sequence of numbers $j_i, i = 1,...,M$, may contain duplicates of one or more digits in [1,M], and some of the values in [1,M] may not be present.) Create a bootstrap sample based on these indices.

$$x_b^* = \{x_{j_1}, x_{j_2},...,x_{j_M}\} \tag{4.5}$$

Now use the elements of the bootstrap sample x_b^* as in Eq. (4.4) to obtain a bootstrap replicate of the spectral density estimate, $(G_{XX}(f_k))_b^*$. Repeat this procedure B times to generate an ensemble of bootstrap replicates of the spectral density estimator $(G_{XX}(f_k))_b^*$, b=1,...,B. These functions form the basis for the approximate statistical analysis of the spectral density of the stationary random process.

We will propose later in this paper that the validation metric be defined in terms of a windowed autospectral density. Such a quantity can be defined in terms of the bootstrap replicates of the autospectral density:

$$\left(\hat{\Sigma}_{XX}^2(f_c)\right)_b^* = \sum_k W_k(f_c)\left(\hat{G}_{XX}(f_k)\right)_b^* \qquad b = 1,...,B \tag{4.6}$$

where $W_k(f_c), k = 0,...,n/2$, is a suitably defined spectral window function. For each value of b, the quantity on the left is a realization of a scalar valued random variable, $\hat{\Sigma}^2_{XX}(f_c)$. Let us refer to it as windowed spectral density. We take the random variable as being completely characterized by the $\left(\hat{\Sigma}^2_{XX}(f_c)\right)^*_b, b = 1,...,B$. The bootstrap uses these realizations to approximate bias, standard error, and confidence intervals for the windowed spectral density. The replicates $\left(\hat{\Sigma}^2_{XX}(f_c)\right)^*_b, b = 1,...,B$, can be used to completely describe the sampling distribution of the windowed spectral density, $\hat{\Sigma}^2_{XX}(f_c)$. This is done in the example that follows.

4.4 Validation of Structural Dynamics Models

The previous two sections have defined analytical and experimental versions of a windowed spectral density metric. The former is a random variable $\Sigma^2_{XX}(\omega_c, A)$ which depends for its stochastic nature on a vector of random variables, A, relating to structural system randomness. The latter is a random variable $\hat{\Sigma}^2_{XX}(f_c)$ which reflects randomness connected to execution of a (single) random vibration experiment (on a single structure). (Randomness connected to system-to-system variability would need to be handled differently.) Let $\omega_c = 2\pi f_c$ and assume that the spectral windows, $W(\omega, \omega_c), -\infty < \omega < \infty$ and $W_k(f_c), k = 0,...,n/2$, used to define the two random variables have equivalent properties. Then the two random variables describe equivalent measures of the mean square response of the analytically described and the experimentally described random process. Define:

$$Z(f_c) = \Sigma^2_{XX}(\omega_c, A) - \hat{\Sigma}^2_{XX}(f_c) \tag{4.7}$$

This is a random variable, and it might be called the margin of conservatism of the mathematical model in the neighborhood of the frequency f_c and at a single point on the system. (The following development considers only validation of a model with respect to its predictions at a single point. Multiple validations will be necessary to assess model sufficiency with regard to its various predictions.) Once we estimate its probability distribution, we can assess the level of conservatism of the mathematical model. As mentioned previously, there are several ways to accomplish this. In the following section we do so using the Advanced Mean Value Method. Because the analytically and experimentally derived metrics of windowed mean square response are statistically independent, the process of estimating the probability distribution of $Z(f_c)$ based on the probability distributions of $\Sigma^2_{XX}(\omega_c, A)$ and $\hat{\Sigma}^2_{XX}(f_c)$ is greatly simplified.

A probabilistic/statistical validation criterion can be defined in terms of the random variable $Z(f_c)$, or in terms of a collection of variables $Z(f_{c,k}), k = 1,...,N$, if indeed the random variable is defined at multiple center frequencies, $f_{c,k}, k = 1,...,N$. There are many possible criteria. Here are a few.

When we are only interested in knowing whether the mathematical model of a system is conservative, we might define a validation criterion as follows:

- Define a collection of frequencies $f_{c,k}, k = 1,...,N$, at which validity of the mathematical model is to be tested.
- Estimate the probability distribution of $Z(f_{c,k}), k = 1,...,N$, at each frequency.
- For all frequencies below a cutoff value, f_{cutoff}, require that a fraction equal to or greater than $1 - \beta_0$ (where $0 < \beta_0 << 1$) of the probabilities $P(Z(f_{c,k}) > 0)$ exceed the level $1 - \alpha_0$ (where $0 < \alpha_0 << 1$).
- For all frequencies above the cutoff value, f_{cutoff}, require that a fraction equal to or greater than $1 - \beta_1$ of the probabilities $P(Z(f_{c,k}) > 0)$ exceed the level $1 - \alpha_1$.

If the model satisfies the criterion, then it is validated with respect to its mean square behavior, over the frequencies tested, at the (compound) level specified at the structural location under consideration.

When we are interested in the accuracy of a mathematical model we might define a validation criterion as follows:

- Define a collection of frequencies $f_{c,k}, k = 1,...,N$, at which validity of the mathematical model is to be tested.
- Estimate the probability distribution of $Z(f_{c,k}), k = 1,...,N$, at each frequency.
- For all frequencies below a cutoff value, f_{cutoff}, require that a fraction equal to or greater than $1 - \beta_0$ of the probabilities $P\left(\left|Z(f_{c,k})\right| \leq \kappa_0 \sqrt{V[Z(f_{c,k})]}\right)$ (where κ_0 is a constant and $V[\cdot]$ denotes the variance) exceed the level $1 - \alpha_0$.
- For all frequencies above the cutoff value, f_{cutoff}, require that a fraction equal to or greater than $1 - \beta_1$ of the probabilities $P\left(\left|Z(f_{c,k})\right| \leq \kappa_1 \sqrt{V[Z(f_{c,k})]}\right)$ exceed the level $1 - \alpha_1$.

If the model satisfies the criterion, then it is validated with respect to its mean square behavior, over the frequencies tested, at the (compound) level specified at the structural location under consideration.

In addition to simply requiring that the mathematical model be conservative or accurate, we may also seek to control the accuracy of the analysis and/or the accuracy of the experiment. In addition to requiring all the elements defined above, in either case, we might also require:

- For all frequencies below a cutoff value, f_{cutoff}, require that a fraction equal to or greater than $1 - \gamma_0$ (where $0 < \gamma_0 << 1$) of the standard deviations of the random margins of conservatism satisfy $\sqrt{V[Z(f_{c,k})]} \leq \lambda_0$ (where λ_0 is a limit on the magnitude of the standard deviation of the random variable $Z(f_{c,k})$).

- For all frequencies above a cutoff value, f_{cutoff}, require that a fraction equal to or greater than $1 - \gamma_1$ of the standard deviations of the random margins of conservatism satisfy $\sqrt{V[Z(f_{c,k})]} \leq \lambda_1$.

This requirement has the effect of imposing a standard of accuracy on the physical experiments performed to obtain the probabilistic description of $\hat{\Sigma}^2_{XX}(f_c)$ and the results of the analysis that leads to the probabilistic description of $\Sigma^2_{XX}(\omega_c, A)$. The latter requirement, though it flows from analysis, is actually also a requirement related to experimentation. The means to diminishing the degree of unpredictability in the random variable $\Sigma^2_{XX}(\omega_c, A)$, to the degree this is possible, is to diminish the degree of uncertainty in the random vector A. This can be done through increased duration of experimentation, improved quality of experimental hardware, etc. That is, it is achieved by spending more money on experiments.

As emphasized in Section 2 above, the general approach proposed here for model validation is useful in the sense that it permits the comparison of analytical to experimental results. However, the analyst might be interested in characterizing the validity of a mathematical model for prediction of system responses that were not tested. In that case, a scheme for performing model validation starts with the performance of a sequence of model validation experiments. These should bound and inhabit the parameter space of potential interest in one or more applications. The mathematical models that correspond to the validation experiments are then constructed, and "point" validation analyses, like the ones described above, are performed. The results of these comparisons might then be used to infer the degree of validity of mathematical models without corresponding validation experiments. The inference might be accomplished through interpolation of the point results.

4.5 Example

Elements of the techniques described above are now applied to a simple system for purposes of demonstration. A schematic of the physical system whose mathematical model was tested for validity is shown in Figure 4. It is a monolithic structure (no joints). It is a steel bar with material properties assumed to equal standard values.

A finite element model (FEM) of the system was constructed in MSC/PATRAN Version 8.5. The finite element analysis was performed in Salinas, a Sandia-developed structural dynamics finite element analysis program. The beam was modeled with free-free boundary conditions and the modal characteristics of the system (i.e. mode shapes and frequencies) and mass and stiffness matrices were obtained. Convergence of the model was examined with respect to modal frequencies and windowed spectral density. The model was deemed converged based on this examination. Post-processing of these results was done in Matlab. It is critical to note that the analyst is required to specify modal damping values for the system in order to obtain the computed frequency response functions referred to in Eq. (4.2). Further, it is desirable to specify these damping values in a predictive manner, without reference to the measured data, because it is intended that

the mathematical model be available for use where its accuracy cannot be confirmed with experiments. Based on expert opinion and prior experience testing systems with no joints, we estimate (conservatively, it is hoped) that the system under consideration has modal damping factors with a mean value of 0.001, and a coefficient of variation of 0.2. We further assume that the modal damping factors are independent random variables, and that they have lognormal probability distributions. These characteristics were used in all the following calculations.
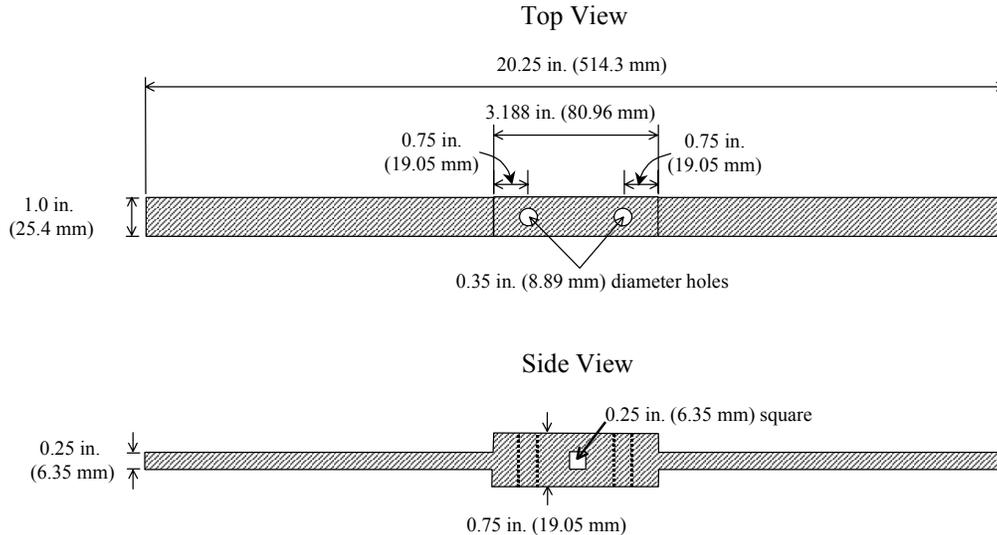
Figure 4: Schematic of physical system

Experimental data were also obtained for this system by testing it in random vibration. Test parameters were chosen as typical values for a test of this type. The spectral density of the force excitation was estimated using standard procedures.

The estimated spectral density of the force excitation was assumed to precisely represent the spectral density of the random process underlying all force realizations for this physical system. It was used, along with results obtained from the finite element model described above and the assumed mean value of the modal damping factors, in Eq. (4.2) to estimate the response spectral densities. In addition, spectral density of the response was estimated using measured data.

The probabilistic character of the windowed metric of the calculated response spectral density was estimated over four frequency bands. A Hanning window was used. (This is simply a haversine.) The center frequencies and widths are listed below:

- Window center frequencies: 100, 178, 316, 562 Hz
- Window frequency widths: 150, 267, 474, 844 Hz

The probability distribution of each of the metrics is not a standard form.

The bootstrap was used to estimate realizations of the experimentally-based spectral density estimate. Each realization was windowed using the windows described above.

The resulting areas were used as the basis for inferring the statistical character of the experimentally established windowed metric of the spectral density.

To judge the validity of the finite element model, it was assumed that the windowed metrics of the spectral density follow a normal probability distribution in both the analytical and experimental cases. (A more sophisticated numerical probabilistic analysis could have been performed.) A simple form of the first validation procedure listed in the section entitled "Validation of Structural Dynamics Models" (the one related to conservatism of the model) was used to validate the present finite element model. It was required that the probability that the margin of conservatism defined in Eq. (4.7) exceed 0 for all four windowed metrics be equal to or greater than 0.90. The probabilities were calculated, and are plotted for the fours spectral windows in Figure 5.

It is clear that, in this case, the finite element model passes the test, and is judged conservative. It is most likely, though, that the mathematical model would not pass any reasonable test of accuracy. In other words, not only is the model conservative, it is probably overly conservative by any reasonable standards.
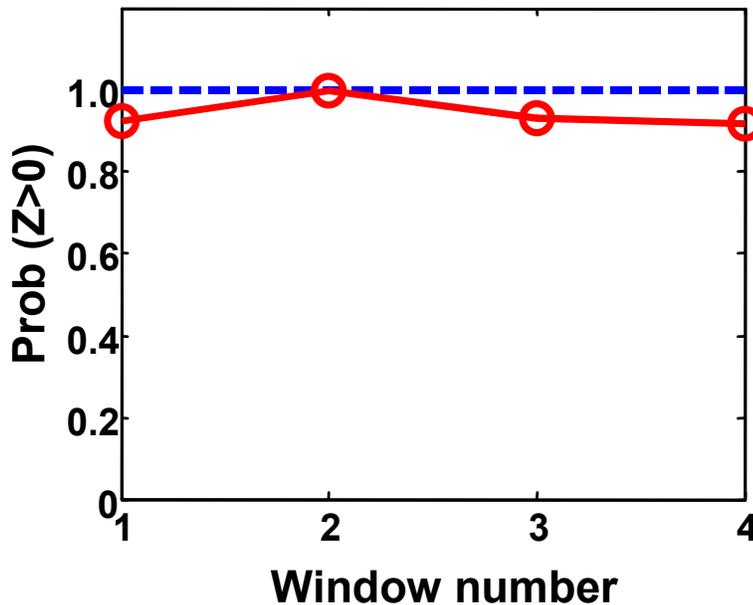
Figure 5. Probability that the margin of conservatism defined in Eq. (7) exceeds 0.

One more fact was recognized in this analysis. The standard error of the estimated spectral density is relatively small, in this application. Therefore, the standard deviations of the windowed metrics of the experimentally estimated response spectral density are also small. One reason is that the physical system is a monolithic beam. There are no joints in the system whose characteristics might vary as the physical test progresses. A

second reason is that the system was only tested once to obtain the results. That is, the experimental setup was not dismantled then reset. Variation in the placement of transducers and differences among transducers would have caused increased variation. Third, only one system was tested. System-to-system variation would have caused increased variation.

4.6 Final Comments

An approach for the practical validation of mathematical models of structural dynamic systems has been outlined. It is based on a windowed metric of the spectral density of the response at one or more points on the system, and it is related to the mean square response of the system within user-defined frequency bands. Because response spectral density and mean square response are useful, fundamental measures of system behavior, the validation of a mathematical model using the proposed approach can yield real confidence that finite element models are truly useful for prediction of structural response, at least in the region of parameter space occupied by the validated mathematical model. Rigorous methods for inferring the validity of the particular mathematical model under discussion in other regions of parameter space closer to realistic applications and based on multiple validation comparisons similar to the one described here need to be developed.

Of course, other metrics for validation of finite element models might be specified. They must, however, satisfy the requirements that they provide useful information to the designer and analyst, and that they present the reasonable possibility of being satisfied.

5. Thermal Foam Decomposition Case Study Overview

5.1 Introduction

In a weapon system rigid foam is used as an encapsulant to thermally isolate and support critical components. In an abnormal environment, typically a hydrocarbon fuel or propellant fire, the rigid foam will decompose when the temperature exceeds 300 C and expose critical components to the harsh thermal environment. Historically, based on a single test, radiation parameters in weapon models have been "tuned" to reflect the protective effect of foam. More recently conducted experiments and comparisons with existing weapon models (Dobranich, 1999) have shown that this approach is not adequate for predictive models, thus suggesting that physics-based models are needed for foam decomposition in an abnormal thermal environment. Subsequently, a model has been under development and a program has been initiated to validate this foam decomposition model (Hobbs *et al*., 1999). Several aspects of this program make it suitable for a case study in the VMP. Among other things, this program provides for:

- Quantitative comparison of model predictions with experimental measurements for a case with complexity relevant to nuclear weapon models.

- Mapping the requirements for the phenomenon-centric validation from the (weapon) application-centric requirements.

- Quantifying the coverage of the present validation experiments for the foam decomposition model.

- Studying the interaction between validation experiments at different complexity levels -- the interaction between purely foam decomposition experiments with experiments having foam decomposition interacting (nonlinearly) with thermal diffusion and radiation.

5.2   Foam Decomposition Model

An extensive study of the chemical structure of the foam and its evolution due to a thermal insult has been used to develop a thermal decomposition model (Hobbs *et al.*, 1999). This decomposition model (termed CPUF) relies on four fundamental models:

1. Foam structure –initial population of the most probable structures that comprise the chemical structure of the foam.

2. Chemical kinetics – how the chemical structure evolves in a thermal environment.

3. Lattice statistics – the relationship between the chemical structure and the macro structure of the foam.

4. Vapor equilibrium – determines the phase (vapor or condensed) of the decomposition products.

The foam decomposition model is coupled to thermal diffusion through the energy equation (conservation of energy) via two physical mechanisms. The first mechanism is through mass loss. Mass is removed from consideration in conservation of energy as the foam model predicts the loss of mass due to decomposition. The second coupling mechanism is through energy removal from the chemical (decomposition) reactions. Consequently, the prediction of mass loss, or rate of mass loss, and of thermal effects are appropriate to consider for validating the foam model, if they can be measured experimentally.

5.3   Validation Data

Two types of experiments have been conducted to support validation of the CPUF model for application to abnormal thermal environment scenarios. First, experiments to study the basic model and calibrate specific model parameters were conducted. These were TGA (ThermoGravimetric Analysis) experiments. Second, a series of experiments with foam dimensions comparable to those in a weapon model were conducted. Each experimental study and the available physical data are summarized below.

5.3.1   TGA (ThermoGravimetric Analysis) Experiments

A detailed discussion of these experiments is given in Erickson et al. (2001) and briefly summarized here. TGA experiments were conducted on foam samples with a mass of nominally 5 mg. The small foam samples were exposed to a measured and controlled thermal environment while its mass is monitored. These experiments were used to demonstrate that the foam model was based on appropriate physical mechanisms and to

calibrate aspects of the model. The TGA experiments studied response to various thermal environments

1. Isothermal experiments at 270, 300, and 330 C. The experiment ramped the temperature quickly (in approximately 10-15 minutes) to the indicated temperature and held.

2. Dual isothermal experiments. The experiments started with an isothermal experiment at 300 C that was held for a period then increased to 380, 400, and 420 C and held for the duration of the experiment. The process began as outlined for the isothermal experiment, after approximately 2 hours when mass loss was minimal the temperature was ramped and held at the final isothermal temperature.

3. Non-isothermal experiments. Temperature was continuously ramped from its initial value at a constant rate of 5, 20, or 50 C/minute.

Because the decomposition is known to depend on the pressure, additional experiments investigating the effect of pressure for the same thermal environments identified previously was studied. Pressures of 1-70 atm were investigated. The data available from TGA experiments are the time-resolved solid mass fraction and foam temperature.

Experiments were conducted at two laboratories (SNL and BYU) with TGA apparatuses with (quite) different designs. The data at pressures greater than ambient were collected at BYU and ambient pressure data were mainly collected at SNL. Some ambient pressure experiments were conducted at BYU allowing for laboratory-to-laboratory comparison.

Representative data for experiments at ambient pressure are shown in Fig 6. TGA data were used to calibrate parameters in the foam decomposition model. Many model development activities have roots similar to this activity. Fundamental experiments are conducted to develop a model and identify pertinent physical mechanisms and then the experiments are used to calibrate the model. After a certain level of confidence is established in the model, the focus shifts from model development to model validation. How to utilize the experimental data, some or all of which may have been used to calibrate parameters in the model is an open issue to be address in this project.

A significant database of TGA experiments has been collected; 10s to potentially 100s of experiments may be utilized. The database has sets of repeated experiments and investigates various aspects of the experimental procedure. Furthermore, experiments were repeated at different laboratories to identify potential laboratory bias. Hence there is necessary data to quantitatively estimate experimental uncertainty and independently (of the model) study the experimental error.
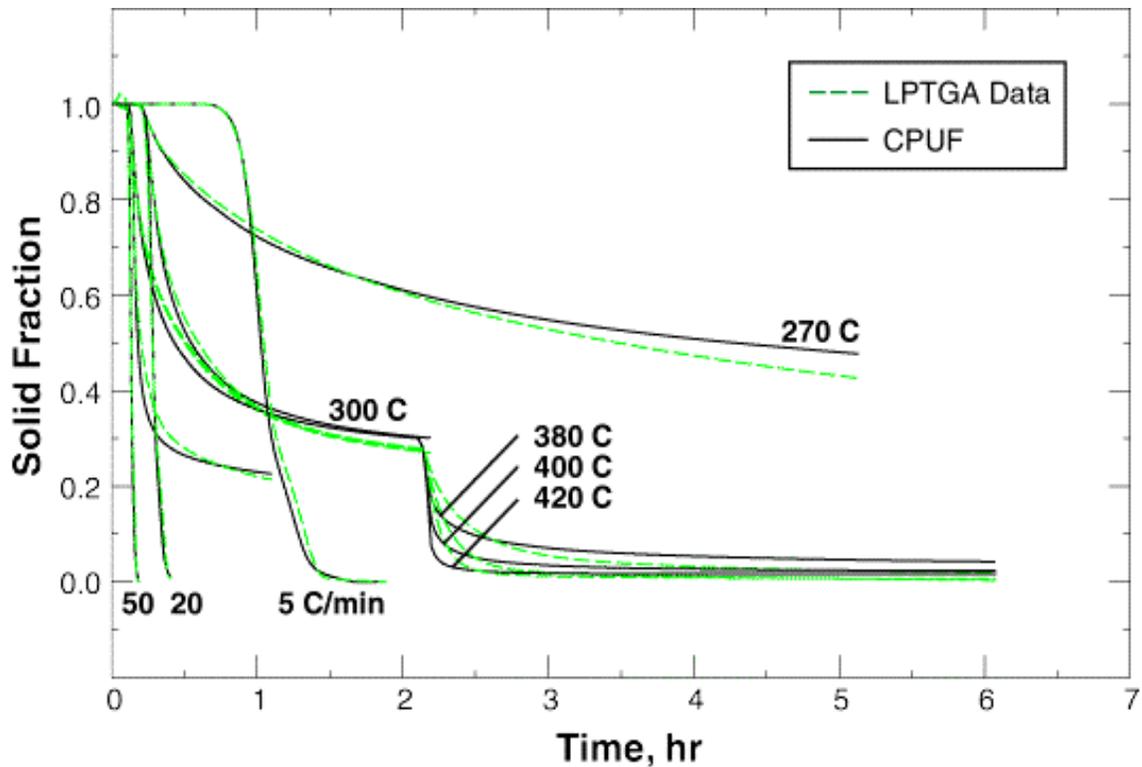
Figure 6. Representative data for TGA experiments

5.3.2 Component Scale Experiments

A series of component-scale experiments on cylindrical foam specimens (nominally 8.8 cm in diameter and 16.4 cm long) have been conducted. The experiment and data are summarized here; see Bentz and Pantuso (1999) for more details. A schematic of the experiment is shown in Fig 7. The foam is contained in a thin-walled stainless steel sleeve that is instrumented with thermocouples along its axial length. A plate of 0.375 inch thick stainless steel (304) is welded to the sleeve to form a "cup" container for the foam specimen. Heating lamps control the temperature of the plate to maintain a fixed value. At a plate temperature comparable to an abnormal thermal environment the foam will begin to decompose near the plate. The decomposition front will recede away from the plate.

A matrix of 15 experiments has been conducted. The experimental matrix addresses the effects of:

1. Plate temperature (600, 750, and 900 C).

2. Foam density (5.7 and 22.7 lb/ft^3).

48

3. Heating orientation (with respect to gravity) of the plate (bottom, side, and overhead).

4. Component embedded in the foam (steel slug or aluminum shell).

The (spatially varying) sleeve temperature and plate temperature are measured during the experiment. Additional available data include: i) time-resolved temperature measurements from thermocouples within the foam specimen, ii) transient X-ray images showing the location of the foam front as the experiment proceeds, iii) time-resolved temperature response of internal components (when applicable). Typical data for the sleeve and plate temperature and a x-ray image are shown in Fig 8 (a model prediction is shown with the x-ray image).

The experimental data has several uncertainties associated with it. For quantitative values of the front location, the x-ray is digitized and analyzed with image processing software. Assembling a time-series of x-rays provides front location as a function of time. Judgement is required to identify the location of the front in the digitized x-ray. The sleeve and plate temperatures are measured with thermocouples, as are temperature measurements within the foam. Random measurement errors can be estimated in the data, but systematic (bias) errors may be difficult, possibly impossible, to quantify with any confidence.

5.4   Model Predictions

5.4.1   TGA Experiments

Model predictions for the TGA experiments only involve the CPUF model (Hobbs et al. (1999). With the measured temperature history of the foam as an input and 18 parameters describing the decomposition chemistry, CPUF calculates the solid mass fraction. These inputs represent the more obvious components of x; see Eq. (3.1). Additionally, inputs defining the (initial) chemical structure of the foam, macro-structure, and the equation of state are required.

The process of calibrating the model uses the measured foam temperature and selects the remaining (18) parameters to minimize the difference between the measured and predicted solid fraction. Calibration can be performed on all TGA experiments, subsets, or even a single experiment. The appropriate balance between insight to the parametric fitting and insight to the validity of the model must reached. This balance will be studied to some extent. At minimum, a subset of experiments will <u>not</u> be used for calibrating, allowing for the comparing model predictions with the experimental data.
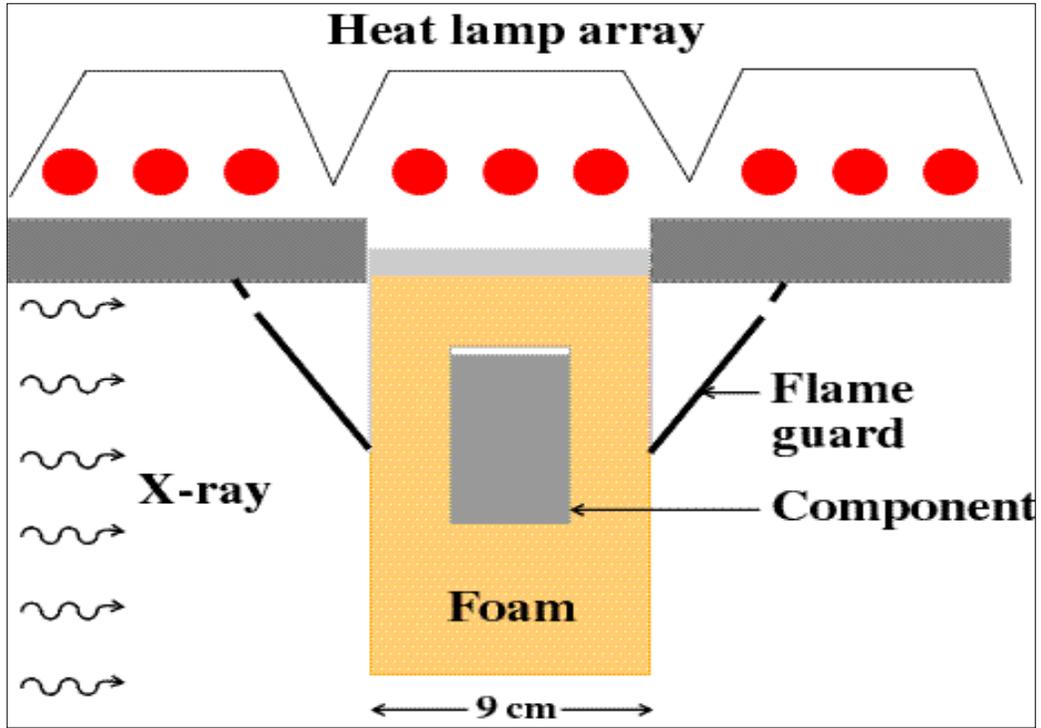
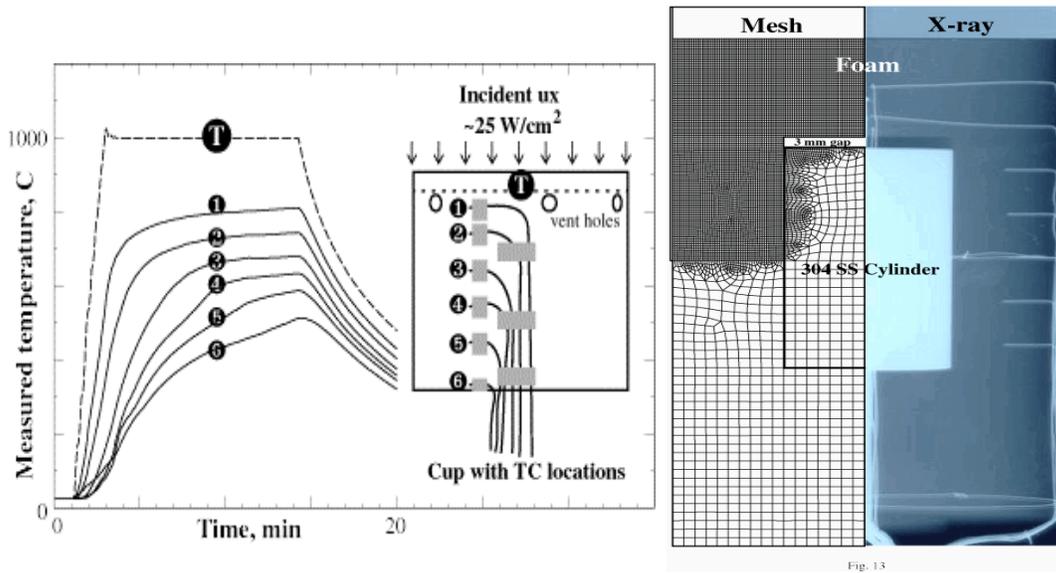**Figure 7. Component scale experiment.**

Figure 8. Data from component scale experiments.

The potential sources of uncertainty in the predicted solid fraction from CPUF are:

1. chemistry parameters (18)
2. measured foam temperature
3. initial foam structure
4. vapor-equilibrium model.

Since the chemistry parameters are being estimated, confidence intervals on the resulting estimates can be established during the calibration. The remaining sources of uncertainty, to a certain extent, are convolved into the calibration constants.

5.4.2 Component-scale Experiments

The computational model of the component-scale experiment has coupled models of thermal diffusion, radiation, and foam decomposition (CPUF). The coupled models are implemented in the finite element code Coyote (Gartling et al., 1994a,b). The complete computational capability is not yet available in the ASCI code Calore, but will be by the end of FY 2001. There are two important issues to realize in this validation study. Using this experiment to study the validity of the CPUF model, depends on the validity of other models, mainly thermal diffusion and enclosure radiation. Although the validity of these additional models (diffusion and radiation) has not been quantitatively studied, the models are generally accepted as valid physics models. Second, the (necessary) approximations to model the experiment may impact the outcome of the validation study. We need to reconcile to what extent the modeling approximations are relevant to the application.

The computational model is summarized. The apparatus is approximated in the computational model as 2D axis-symmetric. One can visualize the model as a cup filled with foam. The measured temperature along the sleeve is input as the boundary temperature at the outer radius of the cup. The measured plate temperature is input as a boundary condition on the end of the cup. As the foam front recedes from the end of the cup a cavity develops. The model assumes the energy within this cavity is exchanged by only radiation. Furthermore, the decomposition gasses that occupy the cavity do not participate.

Several parameters in the computational model reside in ϕ; see Eq. (3.1). Recall that these parameters only impact the model predictions. At least two parameters are possibly important. The first is the element size, particularly at foam decomposition front. A grid resolved solution requires elements on the order of 25-50 μm. Because the element size is not reasonable for large-scale models, a bias-correction was developed to allow for larger element sizes. The second parameter is a density threshold for removing elements from the mesh. The typical value removes an element from the model when the condensed phase of the foam is less than 1% of its initial value.

Addressing the model uncertainty and alignment between the experiment and model for this experiment is more involved than the previous (TGA) experiment. Identifying the sources of uncertainty and potential misalginment, and approaches to quantify these are effects are to be addressed in this study.

5.5   Validation Issues

The goal of the validation analysis for this case study is to analyze sets of computational predictions and experimental outcomes for these foam decomposition experiments in order to characterize the current predictive capability of the computational models. Different validation metrics will be tested and evaluated to address some of the challenges that were discussed in Section 3. Additional methodologies to quantify confidence will be studied and documented.

The case study will in particular address several questions, including the following:

1. Are the phenomenon-centric validation activities, such as the TGA experiments, adequately linked to the applications for which the model will be ultimately used?

2. What is the appropriate metric, as well as the rationale and process for selecting the metric, by which to judge predictive capability of the foam model for the intended application? What measurements and predictions do we compare?

3. Is the parameter range over which the model must be validated for the application compatible with the parameter range of the existing experiments?

4. Do the existing experiments provide an adequate basis for characterizing predictive capability at selected experimental conditions and to infer predictive capability at additional conditions? This question may be specifically addressed by dividing the available set of experiments into two sets. On one set of experiments we could develop an appropriate validation metric and a method for predicting it beyond that set of data. We could then test the performance of the prediction method by comparing the prediction with the second set of data.

5. Can we integrate or otherwise relate different levels of experimental complexity? The TGA experiments address essentially single physics phenomena pertaining to the thermal decomposition of the foam. The component experiments involve multi-physics phenomena, including foam decomposition, radiation exchange, and thermal diffusion. The two experiments provide a testbed to begin to study the interplay among a hierarchy of experiments. In particular, we will investigate to what extent information about predictive capability at the single-physics level can be carried forward to evaluate predictive capability at higher levels.

6.   Success Metrics and Path Forward

First and foremost, success for the VMP in FY 2001 is defined by documented work on the case studies described in Sections 4 and 5. Our hope is that the planned work on these case studies will be successful in the terms defined in these sections. If this is true, then the work on the case studies should extend naturally in future years. Since these case studies are also oriented towards ASCI applications mileposts at Sandia directed at Normal and Abnormal environments (DOE, 2000), any VMP success in FY 2001 will increase the likelihood that the case studies will be a successful component of these mileposts.

As well, a major ASCI V&V program milepost in FY 2002 (DOE, 2000) is to provide a first demonstration of validation methodology for Normal and Abnormal environments modeling. We are supporting progress toward achieving the FY 2002 V&V milepost by working on these case studies in FY 2001. Successful work on the FY 2001 case studies should facilitate future work directed to this milepost.

Finally, an ASCI milepost exists in FY 2004 (DOE, 2000) that targets first application of uncertainty quantification to support stockpile analyses utilizing ASCI computational models. Our work on the VMP can be considered a prototype of the kind of work that will be required to achieve the UQ milepost. Success in the VMP in FY 2001 should provide input into the type of work we should be performing to successfully achieve that milepost.

If the proposed technical work on one or both of the case studies fails to achieve the success defined for that particular case study, then the documentation of the work is still important. As part of the path forward for FY 2002 emerging from the FY 2001 work on the VMP we intend to document our lessons-learned, whether positive or negative. This helps suggest alternatives and reduces the likelihood of repeating mistakes in later work. Success must be measured by the final delivery of these documented work products.

The documentation of a useful path forward from FY 2001, including the FY 2001 lessons-learned, is an important component of our definition of success. Aside from the specific information that can be documented from the work on the case studies, we intend to have the following components in such a document:

- Generalization of the results from the case studies. It is important to widen the scope of the technical methodologies we believe to be useful as rapidly as possible, given the wide scope of the Defense Programs workload at Sandia. Included in this will be our best approach to a roadmap for model validation to guide other participants in ASCI and experimental programs at Sandia. Practices recommended in this roadmap should be practical and relatively simple.

  While model validation per se has been recognized to be key to increased confidence in the use of computational predictions, little guidance has been developed for just how the process should work. Inquiring minds want to know: What do we do and how do we do it? How will we know if we're successful? The VMP should provide guidelines that are as definitive as possible as quickly as possible. Success will be measured by utility of the road map we develop in the path forward document.

  As an independent element in the service of this goal, we also intend to deliver a second document that has guidelines for appropriate validation data and model – data comparisons. These guidelines must constrain the validation data and model – data comparisons that will be elements of the proposed roadmap. For example, we have distinguished "validation experiments" from "phenomena exploration and discovery" experiments in this report. The distinction is indeed critical, but we have not devoted detailed attention to defining it clearly. The detailed exposition of this topic is part of the content

of our planned paper discussing guidelines for validation experimental data and comparisons of numerical models with such data. This document will also include the results of an effort to understand and document the full scope of potential validation experimental data that the ASCI models at Sandia expect to deal with. Success will be measured by the utility and robustness of the guidelines as they are used by future model validation projects at Sandia.

- The focus of the current case studies is on phenomenon-centric validation. We are not likely to have the opportunity this fiscal year to integrate information on predictive capability for different single-phenomenon computational predictions to obtain statements about predictive capability of a combined multi-phenomena computational model. Thus, the path forward is to develop ideas to extend these case studies in this direction, and to identify subsequent case studies that require this integration. Presumably these further case studies would be appropriate for the next phase of the VMP in FY 2002.

  This integration objective will depend on model structure and the ability to decompose the overall model into sub-models and relationships among them. With 'suitable' (which remains to be defined) relationships and information on sub-model predictive capability, we have the opportunity to achieve the desired integration. As we come to understand this problem, there may be important implications for model-building.

- We have hinted here at attendant difficulties associated with resolving the fundamental question of "How much validation is enough?" This is properly a model *qualification* issue. As illustrated in the Appendix, properly defining Defense Programs requirements for the intended model application and translating them into model validation requirements is critical and difficult. The VMP will devote little attention to this model qualification problem in FY 2001. This is a natural area for increased attention in FY 2002. The path forward should define how to best focus attention in FY 2002 on this problem.

Finally, the VMP is properly a single component of the overall ASCI V&V program at Sandia. Intrinsically, therefore, VMP project success must be linked to overall success of this program. V&V Program success metrics at Sandia are currently being defined. Alignment of the FY 2001 VMP with selected program metrics will be discussed in the path forward document.

7. Conclusions

Validation of numerical models encompasses both smaller scale phenomenon-centric methodologies and larger scale applications-centric methodologies. The passage from a smaller scale validation effort, which tends to have better characterized data and model constraints, to more complicated applications of the type of interest to Defense Programs at Sandia is a complex task that is worthy of ongoing research and development. The overall goal of model validation, as we see it, is to provide a credible and defensible

quantified level of confidence in the predictive capability of the model for specific targeted applications of complexity comparable to that of important Defense Programs applications. Whether this goal can ever be achieved is a fair question. And, determining whether the deduced confidence is sufficient for the ultimate task, which is typically the use of the model in a weapon exploration, design or qualification program, is very important. But this determination lies well beyond the specific scope of model validation projects in general, and certainly of the VMP. Model validation provides necessary information for making this determination, but is insufficient by itself.

We believe that validation of models for specific applications requires a complicated blend of analytical, numerical, statistical, and experimental skills. This speaks to the difficulty of the process. The basic elements are presented in this report. The difficulty lies in the details. There is not a unique set of steps one follows to establish model validity for a complex application. Validation is a continuous process that must adapt as it progresses. The goal of the VMP is to demonstrate and develop one or more validation processes, including the necessary support tools that have promise of being more generally applicable to other model applications tasks.

We have stressed that the validation process involves several key elements. The detail and level to which each element is addressed will potentially influence whatever level of final confidence we may succeed in determining for the model. Focusing on only a single element will likely not benefit the overall model validation process. The elements we have discussed to some degree in this report include (1) the definition of the model itself, and most certainly the application requirements that drive it; (2) the requirements for and availability of appropriate validation experiment data; (3) quantitative confidence in the selected validation experimental data, typically as expressed through quantification of uncertainty of the validation experiments; (4) quantitative understanding of the uncertainty imbedded in the model predictions that are compared to the validation experiment data; (5) specific methodologies for comparison of model predictions and experimental data that properly treat the uncertainties in both the model and the data; and (6) methods for inferring predictions of model error that are based on these comparisons and which assess our uncertainty in this prediction.

In conclusion, we stress that in our view the proper way to understand model validation is as the assembly and application of a set of tools and methodologies. These include (1) sensitivity analysis; (2) methods of uncertainty quantification for both model and experiment; (3) statistical tools for inference in the presence of uncertainty; and (4) elements of formal experimental design.

The VMP is completely centered on researching, developing, and applying technical methodologies that enable the quantitative characterization of confidence in model predictions based on the ***canonical validation activity*** – the comparison of experimental data with model predictions. The intent of the project is to document the nature of these methodologies and our experience applying them to relevant case studies. The project will also perform work deemed to be necessary to provide the greatest opportunity for the generalization of these methodologies to other pressing Sandia validation activities. Finally, the project will perform work that supports ongoing efforts to define and begin to apply to qualification processes that are required to determine whether the level of

confidence in the model achieved through the application of validation metrics approaches is sufficient in particular applications.

References

1. P. Barney, C. Ferregut, L. Perez, N. Hunter, and T. Paez (1997), "Statistical Validation of System Models," Proceedings of the 30th Annual Hawaii International Conference on System Sciences, Maui, Hawaii.

2. J. Bentz and John Pantuso (1999), "Letter Report for the Thermal Degradation of Polyurethane Foam at Radiant Heat Facility," internal memo, Sandia National Laboratories, Albuquerque NM, November 1999.

3. B. F. Blackwell, W. Gill, K. J. Dowding, and R. G. Easterling (2000), "Uncertainty Estimation in the Determination of Thermal Conductivity of 304 Stainless Steel," Proceedings of the 35th National Heat Transfer Conference.

4. K. A. Brownlee (1965), Statistical Theory and Methodology in Science and Engineering, John Wiley & Sons, New York.

5. N. A. C. Cressie (1993), Statistics for Spatial Data, John Wiley & Sons, New York.

6. U.S. Department of Energy Defense Programs (2000), "Accelerated Strategic Computing Initiative (ASCI) Program Plan," DOE/DP-99-000010592.

7. D. Dobranich and Walter Gill, (1999), "Thermal Simulations and Experiments for the W76 AF&F Assemblyl – An Integrated Approach," Sandia National Laboratories, SAND99-1925.

8. K. Dowding (2001a), "Quantitative Validation of a Thermal Model," to be presented at 2001 International Mechanical Engineering Conference and Exposition, November 11-16, 2001, New York.

9. K. Dowding (2001b), "Validation Study of a Thermal Contact Conductance Model," to be presented at 2001 International Mechanical Engineering Conference and Exposition, November 11-16, 2001, New York.

10. R. G. Easterling (2001a), "Measuring the Predictive Capability of Computational Models: Principles and Methods, Issues and Illustrations," Sandia National Laboratories, SAND2001-0243.

11. R. G. Easterling (2001b), "Statistical Analysis of Transient Thermocouple Measurements," to appear in Proceedings of the 2001 National Heat Transfer Conference.

12. R. G. Easterling (2001c), "Quantifying the Uncertainty of Computational Predictions," to appear in the Proceedings of the Soc. for Experimental Mechanics Annual Conference on Experimental Mechanics.

13. B. Efron and R. Tibshirani, (1993), An Introduction to the Bootstrap, Chapman and Hall, New York.

14. Eldred, M.S. (1998), "Optimization Strategies for Complex Engineering Applications," Sandia National Laboratories, SAND98-0340.

15. E. S. Epstein (1985), <u>Statistical Inference and Prediction in Climatology: A Bayesian Approach</u>, Meteorological Monograph No. 42, American Meteorological Society.

16. K. L. Erickson, J. N. Castaneda, T. A. Ulibarri, D. K. Derzon, and A. M. Renlund (2001), "Thermal Decomposition Chemistry of Rigid Polyurethane (RPU) Foams," Draft SAND report, Sandia National Laboratories, Albuquerque NM.

17. R. V. Field and J.R. Red-Horse (2001), "A Probabilistic Approach to Model Uncertainty," Proceedings of the 8th International Conference on Structural Safety and Reliability (ICOSSAR), Newport Beach, CA, June, 2001, CD proceedings.

18. R. Ghanem and J. Red-Horse (1999), "Propagation of Probabilistic Uncertainty in Complex Physical Systems Using a Stochastic Finite Element Approach," Physica D, Vol. 133, pp. 137-144.

19. D. K. Gartling, R. E. Hogan, and M. W. Glas (1994a), "Coyote – A Finite Element Computer Program for Nonlinear Heat Conduction Problems Part I – Theoretical Background," Sandia National Laboratories, SAND94-1173.

20. D. K. Gartling, R. E. Hogan, and M. W. Glass (1994b), "Coyote – A Finite Element Computer Program for Nonlinear Heat Conduction Problems Part II – User's Manual," Sandia National Laboratories, SAND94-1179.

21. G. J. Hahn and W. Q. Meeker (1991), <u>Statistical Intervals</u>, John Wiley & Sons, Inc., New York.

22. R. G. Hills and T. G. Trucano (1999), "Statistical Validation of Engineering and Scientific Models: Background," Sandia National Laboratories, SAND99-1256.

23. R. G. Hills and T. G. Trucano (2001a), "Statistical Validation of Engineering and Scientific Models with Application to CTH," Sandia National Laboratories, SAND2001-0312.

24. R. G. Hills and T. G. Trucano (2001b), "Statistical Validation of Engineering and Scientific Models: A Maximum Likelihood Based Metric," Sandia National Laboratories, forthcoming report.

25. M. Hobbs, Kenneth Erickson, and Tze Y. Chu, "Modeling Decomposition Unconfined Rigid Polyurethane Foam," Sandia National Laboratories, SAND99-2758.

26. N. Hunter, P. Barney, C. Ferregut, L. Perez, and T. Paez (1997), "Statistical Validation of Stochastic Models," <u>Proceedings of the 15th International Modal Analysis Conference</u>, SEM, Orlando, Florida.

27. Johnson, P.A. (1995). "Comparison of Pier Scour Equations Using Field Data," *ASCE Journal of Hydraulic Engineering*, v. 121, 626-629.

28. T. Paez, P. Barney, N. Hunter, C. Ferregut, and L. Perez (1996), "Statistical Validation of Physical System Models," <u>Proceedings of the 67th Shock and Vibration Symposium</u>, SVIC, Monterey, California.

29. L. Perez, C. Ferregut, C. Carrasco, T. Paez, P. Barney, and N. Hunter (1997), "Statistical Validation of Plate Finite Element Model for Damage Detection," <u>Proceedings of the SPIE Conference on Smart Structures</u>, San Diego, California.

30. M. Pilch, T. Trucano, J. Moya, G. Froehlich, A. Hodges, and D. Peercy (2000a), "Guidelines for Sandia ASCI Verification and Validation Plans – Content and Format: Version 2.0," Sandia National Laboratories, SAND2000-3101.

31. M. Pilch, T. Trucano, D. Peercy A. Hodges, E. Young, and J. Moya (2000b) "Peer Review Process for the Sandia ASCI V&V Program: Version 1.0," Sandia National Laboratories, SAND2000-3099.

32. B. Rutherford, T. Trucano, and L. Lehoucq (2001), "Application of Response-Modeling Analysis to Problems Involving Model Validation," Sandia National Laboratories report, to be published.

33. J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989), "Design and Analysis of Computer Experiments," Statistical Science, Vol. 4, No. 4, 409-435.

34. D. S. Wilks (1995), <u>Statistical Methods in the Atmospheric Sciences</u>, Academic Press, San Diego, California.

35. P. H. Wirsching, T. Paez, and K. Ortiz (1995), <u>Random Vibrations: Theory and Practice</u>, John Wiley & Sons, New York.

36. Y.-T. Wu and P. H. Wirsching (1987), "A New Algorithm for Structural Reliability Estimation," Journal of the Engineering Mechanics Division, ASCE, Vol. 113, No. 9, pp. 1319-1334.

Appendix - Weapon Stockpile Requirements Underlie Model Validation Requirements

A.1 The Weapon-in-a-Fire Safety Assessment Problem

A number of probabilistic risk assessments have concluded that fire, either alone or in combination with other factors (mechanical/electrical) is a dominant contributor to risk of inadvertent explosion of weapons. To support weapon design and certification efforts with regard to weapon safety in fire environments, several ASCI codes are presently being developed. The FUEGO/SYRINX (Moen *et al.*, 2001; Burns *et al.*, 2001) fire modeling capability will reflect the current state of understanding of hydrocarbon fuel fires. Joined with the ASCI heat transfer code CALORE (Lober *et al.*, 2001) for predicting weapon thermal response, the fully integrated weapon-in-a-fire (WIAF) simulation capability promises to significantly surpass the capabilities in previous large-scale simulation efforts on behalf of weapon design and certification for abnormal thermal environments. Predictions of the thermal response of a weapon in a fire allow identification of weapon explosion risk in specific Stockpile-to-Target Sequence (STS) heating environments and conditions that the weapon might credibly encounter. Calculated risk in these environments can be compared to set certification criteria for the acceptable level of risk, which can then serve as a basis for deciding whether the existing design is good enough or if design improvements or additional deployment controls are required to reduce risk to acceptable levels.

Validation requirements for the WIAF modeling capability are ultimately tied to weapon risk certification requirements in the applicable STS environments. This appendix traces these top-level requirements into explicit and implicit requirements on model validation and the model validation process.

The scenario space of interest for this stockpile driver involves all credible thermal accident scenarios that can occur over the entire life span of the given weapon from creation to disassembly. As such, the STS-driven validation space for the FUEGO/SYRINX/CALORE WIAF modeling capability is very broad. Figure A1 shows the most important parameters that characterize this space for the weapon-in-a-fire problem. Important parameters not evident from the figure include axial extent of the fire along the weapon, and rotational orientation of the weapon relative to the fire. Also, the barrier drawn in the figure is meant to generically represent any physical objects, barriers, packaging (e.g. a shipping container), and/or envelopment (i.e. immersion or partial immersion in mud, snow, or water) that shield or partially shield the weapon from the fire. Selective shielding of limited portions of the weapon can cause preferential heating of certain safety-critical components, which can exacerbate the vulnerability of the weapon to a fire of given severity (as defined by its characterizing parameters such as fuel type, fuel spillage volume and spatial distribution, ground topography and permeability, ambient environmental conditions, and the amount of time to effective fire suppression or self-extinction).
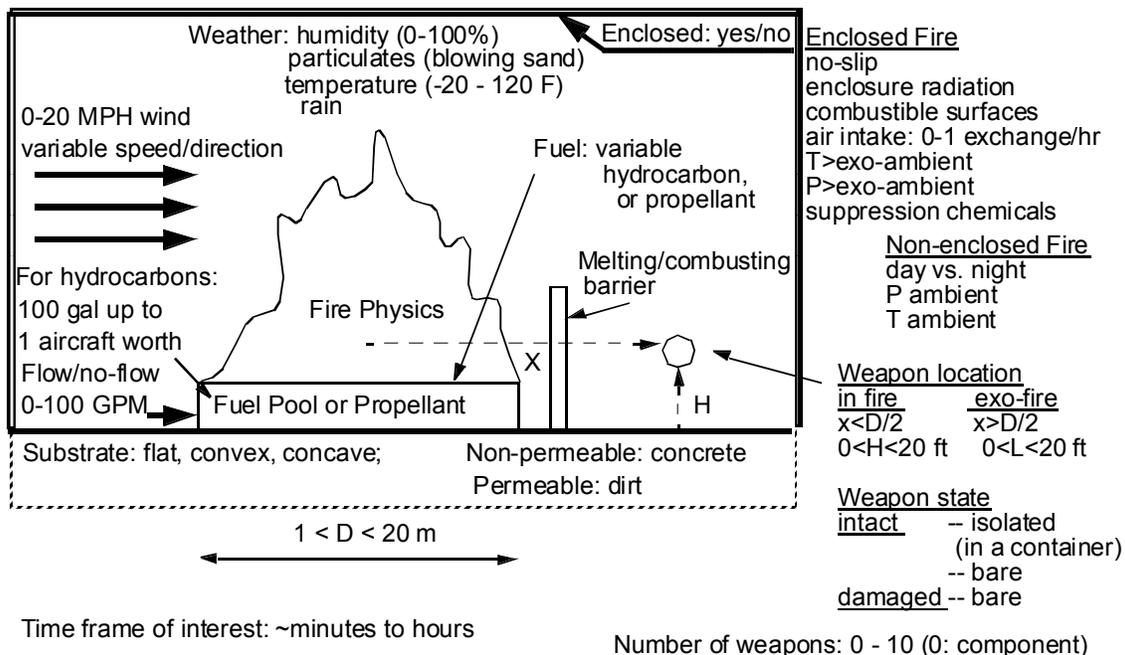
Figure A1. Generic space for the Weapon-In-A-Fire (WIAF) problem.

The entire parameter space of interest is too broad to address without prioritization. Prioritization is the first step in any attempt to define reasonable model validation requirements. Expert opinion is normally the most cost effective way to globally screen and pare down the applicable parameter space to an affordable prioritized subset for quantitative analysis. The expertise should span the technical and programmatic aspects of the problem being addressed and be applied with a system-theoretic philosophy (Saaty, 1990). For the WIAF problem, this involves weapon designers at both the system and component levels, as well as safety engineering and risk assessment representatives, and modelers, experimentalists, and code developers, as well as subject matter experts in probability and statistics.

For the sake of illustration, assume that expert opinion identifies (through a consideration of perceived severity and likelihood) the following scenario classes for further investigation: 1) an open hydrocarbon pool fire without any wind or weather, 2) an open hydrocarbon pool fire with wind but no weather, 3) a facility/enclosure with a hydrocarbon fuel fire with moderate venting, and 4) an open propellant fire with no wind or weather. Further, within each scenario class, several weapon orientations relative to the fire will need to be considered, including: A) weapon fully engulfed by the flame volume; B) weapon partially engulfed by the flame volume with and without partial immersion in cooling media; C) weapon non-engulfed (a "stand-off" fire) with and without selective partial shielding (i.e. "smart fires"). In each of these categories, quantitative analysis and optimization are then used to determine significant weapon vulnerabilities over applicable

parameters that define shielding and immersion conditions and axial, elevational, and rotational orientations of the weapon relative to the fire. Model validation issues are intimately involved in the quantitative identification, assessment, and resolution of these vulnerabilities as described in the next section.

A.2   Translation of Stockpile Requirements to Model Validation Requirements

The top-level programmatic requirement for safety embodies an integral numerical requirement or "constraint" that the assembled elements of the design and certification program must meet. All contributors to the risk assessment "forward problem" and the design-for-safety "reverse problem" are constrained by this requirement. The specific preferences of the system and component designers, based primarily on cost, schedule, and performance issues, ultimately determine the specific path(s) taken toward satisfying this constraint, with input from physics analysts and safety engineers. This constraint, along with others that the designers must meet or approximately satisfy, dictates the feasible space of design solutions from which the final qualifying design will emerge.
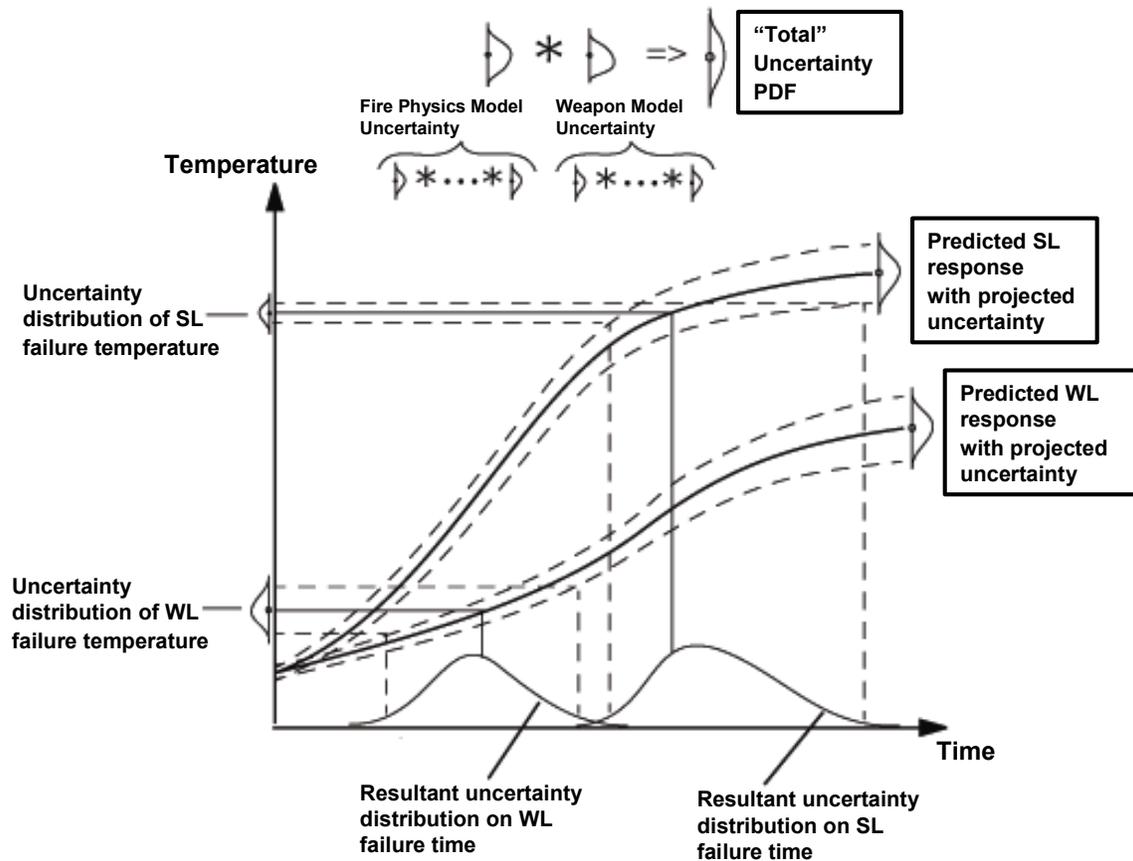

Figure A2. Illustration of Stronglink/Weaklink thermal failure race in risk assessment "Forward Problem."

Figure A2 distills the factors ultimately of importance for weapon safety determination. For simplicity, a simple "stronglink/weaklink" safety theme is considered in which, in order to maintain positive assurance against inadvertent detonation, the weapon's stronglink component must outlast the weaklink component. (More complex cases are considered in the risk assessment methodology reports of Carlson *et al*., 1991 and Bohn, 1996.) If the stronglink does not maintain normal function at least up until the time the weaklink is rendered inoperable ("fails"), then a Loss Of Assured Safety (LOAS) condition exists. The top-level programmatic requirement is that the risk probability of LOAS must be less than 10-Y per weapon lifetime, where Y is a positive number that varies for different STS environments.

Figure A2 shows hypothetical temperature response curves (the solid curves) for the weak and strong links in a weapon subjected to STS abnormal heating conditions identified in the screening and prioritization process discussed above. Since fire is a very complex process that exhibits stochastic behavior to greater or lesser extents depending upon the presiding conditions, the components will experience somewhat different thermal response trajectories in different physical tests having nominally the same parameters characterizing the fire. Unit-to-unit weapon system hardware variability over several such tests would also contribute to thermal response variability, as would measurement uncertainty and test-to-test variability. Hence, some variability and uncertainty in component response profiles arises from physical sources. Additionally, economic constraints force us in general to use computational simulation to run "virtual" weapon-in-a-fire tests to investigate weapon vulnerability in the priority STS scenarios. In these investigations, modeling error and uncertainty combines with physical sources of variability and uncertainty to yield total uncertainty. The dashed curves in Figure A2 indicate corresponding uncertainty distribution contours (say at 5% and 95% cumulative distribution levels) of predicted component response. The solid curves are nominal "best estimates" of component response derived with physics models that are bias-adjusted or "calibrated" from model validation data. We assume that the associated uncertainty distributions or bands about these best-estimate predictions have been estimated in the model validation process. (The methods discussed in the main body of this report are aimed at estimating these uncertainties.)

The identification of model-form bias error within the resolution allowed by the model validation experiments is the initial requirements of model validation. Further, if model bias and uncertainty must be reduced to meet system-level objectives, then model validation elements also play a key role in deciding the best route toward reducing these, and in defining what it might take in terms of modeling and analysis, experiments, and submodel/code development activities. This latter "reverse mode" usage of model validation information for active improvement of the model is distinguished from, and has somewhat different aims and requirements than, the "forward mode" usage for model assessment and characterization. The distinctions will become more apparent and well defined as we move forward with application of model validation methodologies to real problems.

As Figure A2 shows, for set temperature thresholds (exemplified in the figure by the solid horizontal lines depicting nominal component failure temperatures), uncertainty in predicted component thermal trajectories maps into uncertainty in the times at which the components are projected to fail. As explained later, the greater the modeling uncertainty the greater the calculated and perceived safety risk attributed to the weapon. Decreasing the modeling uncertainty can therefore decrease the perceived vulnerability of a weapon in a given heating scenario. This in turn reduces the corresponding "thermal hardness" margins that must be designed into the weapon for the particular scenario examined. Relaxation of such thermal hardness requirements can in turn translate into significant cost savings in weapon design and production efforts, as well as considerable improvements in weapon attributes such as cost, performance, weight, and so on.

The graphic at the top of Figure A2 indicates that "total" modeling uncertainty in predicted component response is the result of several constituent effects. Consider the system-level model validation task of quantifying modeling uncertainty in simulations of weapon thermal response to a particular type of fire heating. If certain conditions are met (see Romero, 2001), then system-level uncertainty is the convolution product of the individual contributions, allowing a clean decomposition of the system-level validation/UQ problem into separate decoupled subproblems. Assuming that such complexity separation is allowed in the WIAF problem, Sandia's radiant heat facility could be used to subject the weapon to heat fluxes that are either nominal or bounding heat fluxes representative of those from a real fire of this type (where the mean flux of the real stochastic fire is determined from available measurement data and fire simulations). Since flux rates imposed on the weapon can be measured and controlled fairly well in the facility, the applied heating boundary condition on the weapon can be represented reasonably well in a simulation. Hence, this testing situation provides a good basis for isolated validation of the weapon thermal model in the applicable thermal regime. Weapon thermal model bias error, and resolution uncertainty thereof, can thus be characterized in isolation from the fire. Then, to predict weapon response (with associated uncertainty) in a real fire of this type, realistic heat flux boundary conditions must be imposed on the weapon model. If available, measured stochastic radiative and convective fluxes (including measurement uncertainties thereof) from a prototypical fire can be translated into uncertainty distributions of imposed heat flux on the model. Alternatively, if the fire itself is being modeled, then modeling bias and uncertainty from applicable model validation experiments are combined with estimated stochastic uncertainty to determine the uncertainty of the imposed flux on the weapon model. The integration of the bias and uncertainty information gained from isolated characterization of the weapon thermal response model and fire heating conditions into a total modeling uncertainty is then a relatively straightforward process as described by Romero (2001).

If the necessary conditions are not met for the fire/weapon complexity separation and reconvolution tactic above, then various layers of complication (see Romero, 2001) will exist, both in determining the uncertainty information (thus the model validation methodology requirements and process becomes more complex), and in propagating the information upward to determine impact at the system level. In the most extreme case, if interaction effects between the weapon and fire are so large that the basis for complexity

separation completely breaks down, then the system-level model validation problem is not separable into parts, and the inextricably coupled problem must be considered a non-separable single entity for model validation purposes. This means that a statistically significant number of tests must be performed at the full system level to adequately characterize system-level modeling ability. In fact, depending on the specific heating conditions and weapon characteristics, the full spectrum from completely separable to fully nonseparable cases applies for the generic WIAF problem, as specific examples in (Romero, 2001) illustrate.

When we can appeal to complexity separation in system-level model validation, it is usually cost effective to do so. At lower levels of complexity, isolated subsets of physics and/or hardware are generally more affordably and effectively characterized because validation experiments are typically easier to perform, control, and interpret as complexity decreases. Complexity separation should in fact be applied recursively where legitimately allowed, in order to produce a cascading tree of successively simpler aspects or elements in arriving at sufficiently simple subsystems (of physics and/or geometry) to allow effective uncertainty characterization (whether through active testing or from suitable data available from the literature). Thus, as the graphic at the top of Figure A2 indicates, even after separating the weapon from the fire we would strive for further complexity decomposition of the weapon response subproblem into aspects and tasks that we can reasonably expect to address in isolation, such as the foam decomposition/ablation model validation task described in Case Study #2 in Section 5 of this document. This case study will not only provide an opportunity for canonical model validation, but also for studying the effectiveness of invoking the complexity decomposition paradigm for isolated characterization then upward propagation of the findings for inference at higher modeling levels.

Several requirements associated with systems level model validation are immediately apparent: 1) characterize the separability regime of the current application; 2) separate the problem as possible in a cascading manner, decoupling separable elements through "linking variables" that are maximally "orthogonal" with respect to the decoupling (see Romero, 2001 for examples); 3) propagate downward through inverse analysis individual "operational space" requirements (e.g. environmental boundary conditions) from the system level; 4) search the literature and design/perform experiments to obtain suitable validation data for the separated elements of the bigger problem; 5) use validation methodology and metrics to characterize modeling bias, and resolution uncertainty thereof; 6) propagate this uncertainty information upward to make inferences at higher levels. (As described in Romero, 2001, this last step can be quite convoluted, involving mapping out and inverting sometimes highly nonlinear functional relationships across separation interfaces, and sometimes transforming measured variables in the experiment into variables more conducive to upward propagation of the uncertainty information gained in the experiment.)

 From this discussion, it is clear that systems engineering plays a very prominent role in systems-level model validation. We also make the observation that unvalidated models will unavoidably be used in the above procedures (for separability characterization, downward operational space mapping, experiment design and interpretation, bias and

uncertainty characterization in the separate model validation activities, and upward propagation of obtained uncertainty information). However, since many of these operations depend only on relative accuracy (accurate *trend* information with respect to the parameters in play) rather than on *absolute* predictive accuracy over the parameter space, the use of unvalidated models in the above operations may not be a significant detraction from the process.

Returning to Figure A2, uncertainty distributions are also indicated for the threshold failure temperatures of the weak and strong links. These map through component response trajectories into uncertainties in their failure times. Failure temperature uncertainty can be caused by random unit-to-unit variabilities that can be correlated with the particular batch and/or manufacturer they come from. Failure temperatures can also depend significantly on heating rates, so the failure distributions must take this into account. Accordingly, experimentally and/or computationally determined component heating profiles in STS weapon heating environments of concern should be used to identify the correct boundary conditions in component failure characterization tests. Validated simulation models can also be used to design testing approaches and apparatus capable of applying the necessary component heating profiles, and in translating test results into measurable metrics for failure criteria, such as component outer case temperatures (see Romero and Thomas, 1993). Romero (1996) shows that realistic uncertainties in component failure temperatures can map to relatively large uncertainties in the failure times of the weak and strong links. As with uncertainty in thermal response profiles, decreasing the uncertainty of component failure thresholds can decrease the perceived vulnerability of a given weapon and therefore its thermal hardness requirements, which can result in significant cost savings in weapon design, testing, and production, as well as considerable improvements in weapon attributes. Hence, it is critical to use accuracy-validated thermal response models (both at system and component levels) in the planning and interpretation of component failure characterization tests.

Uncertainty in the component response profiles and in their failure temperatures combine into resultant distributions in their failure times (depicted on the time axis in Figure A2). If the stronglink and weaklink failure-time distributions overlap to any extent, then there is some probability that the stronglink will fail before the weaklink (see the risk assessment methodology reports of Carlson et al. (1991) and Bohn (1996) for risk calculation procedures). Thus, given a particular heating scenario, a particular weapon system, associated component failure characteristics in the relevant heating regime, and associated physical and modeling uncertainties, the probability of LOAS can be estimated such that the weapon system can nominally be said to meet or not meet the safety qualification criterion of less than 10-Y probability of LOAS in the particular STS environment.

 The estimation of risk probability is really the "forward" problem in a larger "reverse" or "inverse" design problem. The forward problem determines how "hard" the system is in a particular environment, whereas the reverse problem is to explicitly design the system to most cost effectively meet the 10-Y hardness requirement across all credible STS environments.

If an initially proposed design is analyzed and found not to surpass hardness requirements with sufficiently comfortable margins in one or more environments, then system and component designs can be iterated until the system does sufficiently exceed requirements. Alternatively, less modeling uncertainty in the problem translates into smaller projected widths of the weak and strong link time-to-failure distributions, and hence into smaller projected risk probability. Optimal design practice suggests that system and component design tradeoffs be considered along with uncertainty management and reduction possibilities in determining the most promising path toward cost effective satisfaction of problem constraints, requirements, and objectives. Analysis can reveal the reduction in projected risk that can be achieved by, say, a 10% reduction in modeling uncertainty. This benefit is then measured against projected costs for achieving this reduction in modeling uncertainty, as are similar-costing benefits associated with manipulating the system and/or component thermal hardness characteristics. Thus, tradeoffs in system/component design and characterization, and in M&S uncertainty reduction, are explicitly (numerically) assessed in top-level engineering planning and resource allocation to ultimately enable more rational management of the project toward project goals.

Though in the past, no specific numerical accuracy requirements per se have been put on modeling and simulation accuracy in supporting weapon design and qualification, it is envisioned that in the future accuracy requirements may come in the form foreshadowed above: "if we can set a goal to contain modeling uncertainty to within some negotiable percentage about the nominal thermal response predictions, then we can include this tolerance in risk projections when iterating over system and component hardware options in early design stages." Currently, at the very least, a standing requirement exists to estimate simulation uncertainty in the forward problem of risk assessment. Toward this requirement, ASCI V&V plans (Romero *et al*., 2001; Tieszen *et al*., 2001) are being written to coordinate and integrate V&V and UQ activities with traditional M&S technology to precisely state how M&S uncertainty will be quantified and controlled in the WIAF problem. The implicit requirement beyond this is to reduce modeling uncertainty over time through wise application of resources, as determined through systems analysis that incorporates V&V and UQ technologies, to optimally identify, plan, and direct beneficial experimental, modeling, and code/algorithm development and assessment activities.

Appendix References

1.   Bohn, M.P. (1996), "An Integral Formulation for Race Cutset Quantification," Sandia Laboratories internal memorandum.

2.   Burns, S.P., and Houf, B. (2001), "SIERRA/SYRINX User's Manual," Sandia National Laboratories report in development.

3.   Carlson, D.D., A.S. Benjamin, R.J. Breeding, D.M. Kunsman (1991), "Probabilistic Risk Assessment Methods for Nuclear Weapon System

Applications, Volume 2 – Methods," Sandia National Laboratories, SAND-91-0323/2, August, 1991.

4.   Lober, R.R., S.W. Bova, R.J. Cochran, K.J. Dowding, M.W. Glass (2001), "CALORE, A Thermal Analysis Program, Vol. 2: User Manual," Sandia National Laboratories report in development.

5.   Moen, C.D., Evans, G.H., Aro, C.J., Domino, S.P. (2001), "SIERRA/FUEGO User Manual," Sandia National Laboratories report in development.

6.   Romero, V.J. (1996), "Making Use of Optimization, NonDeterministic Analysis, and Numerical Simulation to Assess Firing Set Robustness in a Fire," Proceedings of the 14th Intn'l. System Safety Conf., Aug. 12-17, 1996, Albuquerque, NM, published by the System Safety Society, Sterling, VA.

7.   Romero, V.J. (2001), "Pragmatic Model Validation Metrics and Methodology from a Systems Perspective," to be presented at the Aero- and Hydroacoustic Facilities, Experimental Techniques, and Validation of Computational Methods Symposium of the Intn'l. Mech. Engrng. Congress and Expo. in New York, Nov. 11-16, 2001.

8.   Romero, V.J., and Thomas, M.H. (1993), "Thermal Failure Characterization of the MC4109 Capacitor: Methodology and Test Results," Sandia National Laboratories, SAND93-0036, May, 1993.

9.   Romero, V.J., S.R. Tieszen, R.R. Lober, R.J. Cochran, D.D. Dobranich, T.Y. Chu, W.C. Moffatt (2001), "CALORE Verification and Validation Plan for Weapon-in-a-Fire Simulations," Sandia National Laboratories report, to be published.

10. Saaty, T.L. (1990), Multicriteria Decision Making – The Analytical Heiarchy Process: Planning, Priority Setting, Resource Allocation, RWS Publications, Pittsburgh, PA.

11. Tieszen, S.R., A.R. Lopez, C.D. Moen, T.Y. Chu, V.F. Nicolette, W. Gill, S.P. Burns, W.C. Moffat (2001), "SIERRA/FUEGO/SYRINX Verification & Validation Plan, Version 2," Sandia National Laboratories report, to be published.

Distribution

Hugh Coleman (2)
Department of Mechanical &
          Aero. Engineering
University of Alabama/Huntsville
Huntsville, AL 35899

Raymond Cosner (2)
Boeing-Phantom Works
MC S106-7126
P. O. Box 516
St. Louis, MO 63166-0516

Thomas A. Cruse
398 Shadow Place
Pagosa Springs, CO 81147-7610

P. Cuniff
U.S. Army Soldier Systems Center
Kansas Street
Natick, MA 01750-5019

Frank Dean (2)
Strategic Systems Programs
Nebraska Avenue Complex
287 Somers Court NW, Suite 10041
Washington, DC 20393-5446

Department of Energy (4)
Attn: Kevin Greenaugh, DP-153
            Paul Messina, DP-51
            Juan Meza, DP-51
            William Reed, DP-51
Department of Energy
1000 Independence Ave., SW
Washington, DC 20585

U. M. Diwekar (2)
Center for Energy and
          Environmental Studies
Carnegie Mellon University
Pittsburgh, PA 15213-3890

David Dolling
Department of Aerospace Engineering
          & Engineering Mechanics
University of Texas at Austin
Austin, TX 78712-1085

Isaac Elishakoff (2)
Dept. of Mechanical Engineering
Florida Atlantic University
777 Glades Road
Boca Raton, FL 33431-0991

Joseph E. Flaherty (2)
Dept. of Computer Science
Rensselaer Polytechnic Institute
Troy, NY 12181

John Fortna
ANSYS, Inc.
275 Technology Drive
Canonsburg, PA 15317

Roger Ghanem
Dept. of Civil Engineering
Johns Hopkins University
Baltimore, MD 21218

Mike Giltrud
Defense Threat Reduction Agency
DTRA/CPWS
6801 Telegraph Road
Alexandria, VA 22310-3398

James Glimm (2)
Dept. of Applied Math & Statistics
P138A
State University of New York
Stony Brook, NY 11794-3600

James Gran
SRI International
Poulter Laboratory AH253
333 Ravenswood Avenue
Menlo Park, CA 94025

Bernard Grossman (2)
Dept. of Aerospace &
          Ocean Engineering
Mail Stop 0203
215 Randolph Hall
Blacksburg, VA 24061

Sami Habchi
CFD Research Corp.
Cummings Research Park
215 Wynn Drive
Huntsville, AL 35805

Raphael Haftka (2)
Dept. of Aerospace and Mechanical
          Engineering and Engineering Science
P. O. Box 116250
University of Florida
Gainesville, FL 32611-6250

Achintya Haldar (2)
Dept. of Civil Engineering
       & Engineering Mechanics
University of Arizona
Tucson, AZ 85721

Tim Hasselman
ACTA
2790 Skypark Dr., Suite 310
Torrance, CA 90505-5345

George Hazelrigg
Division of Design, Manufacturing
       & Innovation
Room 508N
4201 Wilson Blvd.
Arlington, VA 22230

David Higdon
Institute of Statistics and Decision Science
Duke University
Box 90251
Durham, NC 27708-0251

Richard Hills (20)
College of Engineering, MSC 3449
New Mexico State University
P. O. Box 30001
Las Cruces, NM 88003

F. Owen Hoffman (2)
SENES
102 Donner Drive
Oak Ridge, TN 37830

G. Ivy
Logicon R&D Associates
P.O. Box 92500
Los Angeles, CA 90009

Ralph Jones (2)
Sverdrup Tech. Inc./AEDC Group
1099 Avenue C
Arnold AFB, TN 37389-9013

Leo Kadanoff
Research Institutes Building
University of Chicago
5640 South Ellis Ave.
Chicago, IL 60637

George Karniadakis (2)
Division of Applied Mathematics
Brown University
192 George St., Box F
Providence, RI 02912

Alan Karr
Institute of Statistics and Decision Science
Duke University
Box 90251
Durham, NC 27708-0251

J. Keremes
The Boeing Company
Rocketdyne Propulsion & Power
P. O. Box 7922
6633 Canoga Avenue
Canoga Park, CA 91309-7922

Hyoung-Man Kim
Boeing Company
M/S: ZC-01
502 Gemini Ave.
Houston, TX 77058

K. D. Kimsey
U.S. Army Research Laboratory
Weapons & Materials Research
Directorate
AMSRL-WM-TC 309 120A
Aberdeen Proving Gd, MD 21005-5066

B. A. Kovac
The Boeing Company
Rocketdyne Propulsion & Power
P. O. Box 7922
6633 Canoga Avenue
Canoga Park, CA 91309-7922

P. Krysl
Department of Computer Science
California Institute of Technology
1200 E. California Blvd./MS 256-80
Pasadena, CA 91125

W. K. Liu (2)
Northwestern University
Dept. of Mechanical Engineering
2145 Sheridan Road
Evanston, IL 60108-3111

Robert Lust
General Motors, R&D and Planning
MC 480-106-256
30500 Mound Road
Warren, MI 48090-9055

Sankaran Mahadevan (2)
Dept. of Civil &
               Environmental Engineering
Vanderbilt University
Box 6077, Station B
Nashville, TN 37235

Hans Mair
Institute for Defense Analysis
Operational Evaluation Division
1801 North Beauregard Street
Alexandria, VA 22311-1772

W. McDonald
Naval Surface Warfare Center
Code 420
101 Strauss Avenue
Indian Head, MD 20640-5035

Gregory McRae (2)
Dept. of Chemical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Michael Mendenhall (2)
Nielsen Engineering & Research, Inc.
510 Clyde Ave.
Mountain View, CA 94043

Sue Minkoff (2)
Dept. of Mathematics and Statistics
University of Maryland, Baltimore Co.
1000 Hilltop Circle
Baltimore, MD 21250

Max Morris (2)
Department of Statistics
Iowa State University
304A Snedecor-Hall
Ames, IW 50011-1210

Paul Muessig
Naval Air Warfare Center
Joint Accreditation Support Activity
Weapons Division, Code 418000D
1 Administration Circle
China Lake, CA 93555-6100

R. Namburu
U.S. Army Research Laboratory
AMSRL-CI-H
Aberdeen Proving Gd, MD 21005-5067

NASA/Ames Research Center
Attn: U. B. Mehta
MS: T27 B-1
Moffett Field, CA 94035-1000

NASA/Glen Research Center
Attn: Chris Steffen, MS 5-11
Cleveland, OH 44135

NASA/Langley Research Center (5)
Attn: Michael Hemsch, MS 280
                Jim Luckring, MS 280
                Ahmed Noor, MS 369
                Sharon Padula, MS 159
                Jim Weilmuenster, MS 408A
Hampton, VA 23681-0001

C. Needham
Applied Research Associates, Inc.
4300 San Mateo Blvd., Suite A-220
Albuquerque, NM 87110

A. Needleman
Division of Engineering, Box D
Brown University
Providence, RI 02912

Robert Nelson (2)
Dept. of Aerospace &
Mechanical Engineering
University of Notre Dame
Notre Dame, IN 46556

Efstratios Nikolaidis (2)
MIME Dept.
4035 Nitschke Hall
University of Toledo
 Toledo, OH 43606-3390

Tinsley Oden (2)
Texas Institute of Comp. Mechanics
University of Texas at Austin
Austin, TX 78712

Michael Ortiz (2)
Graduate Aeronautical Laboratories
California Institute of Technology
1200 E. California Blvd./MS 105-50
Pasadena, CA 91125

Dale Pace
Applied Physics Laboratory
Johns Hopkins University
111000 Johns Hopkins Road
Laurel, MD 20723-6099

Alex Pang
Computer Science Department
University of California
 Santa Cruz, CA 95064

Allan Pifko
2 George Court
Melville, NY 11747

Cary Presser (2)
Process Measurements Div.
National Institute of Standards
and Technology
Bldg. 221, Room B312
Gaithersburg, MD 20899

P. Radovitzky
Graduate Aeronautical Laboratories
California Institute of Technology
1200 E. California Blvd./MS 105-50
Pasadena, CA 91125

W. Rafaniello
DOW Chemical Company
1776 Building
Midland, MI 48674

Chris Rahaim (2)
SAIC
14 East Washington St., Suite 401
Orlando, FL 32801-2320

Pradeep Raj (2)
Computational Fluid Dynamics
Lockheed Martin Aeronautical Sys.
86 South Cobb Drive
Marietta, GA 30063-0685

J. N. Reddy
Dept. of Mechanical Engineering
Texas A&M University
ENPH Building, Room 210
College Station, TX 77843-3123

John Renaud (2)
Dept. of Aerospace & Mechanical Engr.
University of Notre Dame
Notre Dame, IN 46556

E. Repetto
Graduate Aeronautical Laboratories
California Institute of Technology
1200 E. California Blvd./MS 105-50
Pasadena, CA 91125

Patrick J. Roache
1108 Mesa Loop NW
Los Lunas, NM 87031

A. J. Rosakis
Graduate Aeronautical Laboratories
California Institute of Technology
1200 E. California Blvd./MS 105-50
Pasadena, CA 91125

Tim Ross (2)
Dept. of Civil Engineering
University of New Mexico
Albuquerque, NM 87131

J. Sacks
Institute of Statistics and Decision Science
Duke University
Box 90251
Durham, NC 27708-0251

Sunil Saigal (2)
Carnegie Mellon University
Department of Civil and
                Environmental Engineering
Pittsburgh, PA 15213

Len Schwer
Schwer Engineering & Consulting
6122 Aaron Court
Windsor, CA 95492

Paul Senseny
Factory Mutual Research Corporation
1151 Boston-Providence Turnpike
P.O. Box 9102
Norwood, MA 02062

E. Sevin
Logicon RDA, Inc.
1782 Kenton Circle
Lyndhurst, OH 44124

Mark Shephard (2)
Rensselaer Polytechnic Institute
Scientific Computation Research Center
Troy, NY 12180-3950

T. P. Shivananda
Bldg. SB2/Rm. 1011
TRW/Ballistic Missiles Division
P. O. Box 1310
San Bernardino, CA 92402-1310

Y.-C. Shu
Graduate Aeronautical Laboratories
California Institute of Technology
1200 E. California Blvd./MS 105-50
Pasadena, CA 91125

Don Simons
Logicon
222 W. Sixth St.
P.O. Box 471
San Pedro, CA 90733-0471

Munir Sindir
Boeing Company, MC IB39
Rocketdyne Propulsion & Power
P. O. Box 7922
6633 Canoga Avenue
Canoga Park, CA 91309-7922

Ashok Singhal (2)
CFD Research Corp.
Cummings Research Park
215 Wynn Drive
Huntsville, AL 35805

R. Singleton
Engineering Sciences Directorate
Army Research Office
4300 S. Miami Blvd.
P.O. Box 1221
Research Triangle Park, NC 27709-2211

W. E. Snowden
DARPA
7120 Laketree Drive
Fairfax Station, VA 22039

Bill Spencer (2)
Dept. of Civil Engineering
           and Geological Sciences
University of Notre Dame
Notre Dame, IN 46556-0767

D. E. Stevenson (2)
Computer Science Department
Clemson University
442 Edwards Hall, Box 341906
Clemson, SC 29631-1906

Tim Swafford
Sverdrup Tech. Inc./AEDC Group
1099 Avenue C
Arnold AFB, TN 37389-9013

Kenneth Tatum
Sverdrup Tech. Inc./AEDC Group
740 Fourth Ave.
Arnold AFB, TN 37389-6001

Ben Thacker
Southwest Research Institute
6220 Culebra Road
Postal Drawer 28510
San Antonio, TX 78228-0510
 Ren-Jye Yang
Ford Research Laboratory
MD2115-SRL
P.O.Box 2053
Dearborn, MI 4812

Simone Youngblood
DOD/DMSO
Technical Director for VV&A
1901 N. Beauregard St., Suite 504
Alexandria, VA 22311

M. A. Zikry
North Carolina State University
Mechanical & Aerospace Engineering
2412 Broughton Hall, Box 7910
Raleigh, NC 27695

FOREIGN DISTRIBUTION:

Yakov Ben-Haim (2)
Department of Mechanical Engineering
Technion-Israel Institute of Technology
Haifa 32000
ISRAEL

Gert de Cooman (2)
Universiteit Gent
Onderzoeksgroep, SYSTeMS
Technologiepark - Zwijnaarde 9
9052 Zwijnaarde
BELGIUM

Graham de Vahl Davis (2)
CFD Research Laboratory
University of NSW
Sydney, NSW 2052
AUSTRIALIA

Luis Eca (2)
Instituto Superior Tecnico
Department of Mechanical Engineering
Av. Rovisco Pais
1096 Lisboa CODEX
PORTUGAL

Charles Hirsch (2)
Department of Fluid Mechanics
Vrije Universiteit Brussel
Pleinlaan, 2
B-1050 Brussels
BELGIUM

Igor Kozin
Systems Analysis Department
Riso National Laboratory
P. O. Box 49
DK-4000 Roskilde
DENMARK

K. Papoulia
Inst. Eng. Seismology & Earthquake
            Engineering
P.O. Box 53, Finikas GR-55105
Thessaloniki
GREECE

Malcolm Wallace
National Engineering Laboratory
East Kilbride
Glasgow
G75 0QU
UNITED KINGDOM

Peter Walley
36 Bloomfield Terrace
Lower Hutt
NEW ZEALAND

Dr. Max Ruppert
UniBw Munich - BauV 2.2
Inst. Engng.Mech. & Struct.Mech.
D - 85577 Neuibiberg
GERMANY

Los Alamos National Laboratory (41)
Mail Station 5000
P.O. Box 1663
Los Alamos, NM 87545
    Attn: Peter Adams, MS B220
    Attn: Mark C. Anderson, MS D411

Attn: Thomas Bement, MS F600
Attn: Terrence Bott, MS K557
Attn: D. Cagliostro, MS F645
Attn: David Crane, MS P946
Attn: John F. Davis, MS B295
Attn: Helen S. Deaven, MS B295
Attn: Scott Doebling, MS P946
Attn: S. Eisenhawer, MS K557
Attn: Dawn Flicker, MS F664
Attn: George T. Gray, MS G755
Attn: Ken Hanson, MS P940
Attn: R. Henninger, MS D413
Attn: Brad Holian, MS B268
Attn: Kathleen Holian, MS B295
Attn: Darryl Holm, MS B284
Attn: James Hyman, MS B284
Attn: Michael E. Jones, MS B259
Attn: Cliff Joslyn, MS B265
Attn: James Kamm, MS D413
Attn: Jeanette Lagrange, MS D445
Attn: S. Keller-McNulty, MS F600
Attn: Elizabeth Kelly, MS F600
Attn: Ken Koch, MS F652
Attn: Len Margolin, MS D413
Attn: Harry Martz, MS F600
Attn: Mike McKay, MS F600
Attn: Mark P. Miller, MS P946
Attn: John D. Morrison, MS F602
Attn: Karen I. Pao, MS B256
Attn: M. Peterson-Schnell,
    MS B295
Attn: Doug Post, MS F661 X-DO
Attn: William Rider, MS D413
Attn: Tom Seed, MS F663
Attn: David Sharp, MS B213
Attn: Richard N. Silver, MS D429
Attn: Ronald E. Smith, MS J576
Attn: Christine Treml, MS H851
Attn: Daniel Weeks, MS B295
Attn: Morgan White, MS F663
Attn: Alyson G. Wilson, MS F600

University of California (20)
Lawrence Livermore National Laboratory
7000 East Ave.
P.O. Box 808
Livermore, CA 94550
    Attn: T. F. Adams, MS L-095
    Attn: Steven Ashby, MS L-561
    Attn: John Bolstad, MS L-023
    Attn: Peter N. Brown, MS L-561
    Attn: T. Scott Carman, MS L-031
    Attn: R. Christensen, MS L-160
    Attn: Evi Dube, MS L-095

Attn: Richard Klein, MS L-023
Attn: Roger Logan, MS L-125
Attn: C. F. McMillan, MS L-098
Attn: C. Mailhiot, MS L-055
Attn: J. F. McEnerney, MS L-023
Attn: M. J. Murphy, MS L-282
Attn: Daniel Nikkel, MS L-342
Attn: Cynthia Nitta, MS L-096
Attn: Peter Raboin, MS L-125
Attn: Peter Terrill, MS L-125
Attn: Charles Tong, MS L-560

Argonne National Laboratory
      Attn: Paul Hovland
MCS Division
Bldg. 221, Rm. C-236
9700 S. Cass Ave.
Argonne, IL 60439

SANDIA INTERNAL

1	MS 1152 1642	M. L. Kiefer
1	MS 1186 1674	R. J. Lawrence
1	MS 0525 1734	P. V. Plunkett
1	MS 0525 1734	R. B. Heath
1	MS 0525 1734	S. D. Wix
1	MS 1393 1902	J. R. Garcia
1	MS 0457 2001	W. J. Tedeschi
1	MS 0429 2100	J. S. Rottler
1	MS 0453 2104	D. L. McCoy
1	MS 0475 2105	R. C. Hartwig
1	MS 1393 2106	F. F. Dean
1	MS 0447 2111	J. O. Harrison
1	MS 0447 2111	P. D. Hoover
1	MS 0479 2151	P. A. Sena
1	MS 0479 2151	M. H. Abt
1	MS 0482 2161	V. J. Johnson
1	MS 0482 2161	R. S. Baty
1	MS 0481 2167	M. A. Rosenthal
1	MS 0481 2167	W. C. Moffatt
1	MS 0481 2168	K. D. Meeks
1	MS 0481 2168	K. Ortiz
1	MS 0509 2300	M. W. Callahan
1	MS 0769 5800	D. S. Miyoshi
1	MS 0759 5845	I. V. Waddoups
1	MS 0759 5845	M. S. Tierney
1	MS 0782 5861	R. V. Matalucci
1	MS 0737 6114	P. A. Davis
1	MS 0751 6117	L. S. Costin
1	MS 0718 6141	C. D. Massey
1	MS 0708 6214	P. S. Veers
1	MS 0747 6410	R. L. Camp
1	MS 0747 6410	G. D. Wyss
1	MS 0746 6411	R. M. Cranwell

1	MS 0746 6411	D. J. Anderson
1	MS 0746 6411	J. E. Campbell
1	MS 0746 6411	D. G. Robinson
1	MS 0746 6411	L. P. Swiler
1	MS 1140 6500	J. K. Rice
1	MS 0974 6522	G. D. Valcez
1	MS 0977 6524	W. R. Cook
3	MS 0977 6524	S. M. DeLand
1	MS 1138 6533	E. Shepherd
1	MS 1138 6534	L. M. Claussen
1	MS 1137 6534	A. L. Hodges
1	MS 1138 6534	M. T. McCornack
1	MS 1137 6534	S. V. Romero
1	MS 1137 6535	G. K. Froehlich
1	MS 0771 6805	P. G. Kaplan
1	MS 1395 6821	J. W. Garner
1	MS 1395 6821	P. Vaughn
1	MS 0779 6849	J. C. Helton
1	MS 0779 6849	R. P. Rechard
1	MS 0779 6849	M. J. Shortencarier
1	MS 0778 6851	G. E. Barr
1	MS 0778 6851	R. J. MacKinnon
1	MS 9051 8351	C. A. Kennedy
1	MS 9202 8418	W. P. Ballard
1	MS 9202 8418	R. M. Zurn
1	MS 9405 8700	T. M. Dyer
1	MS 9042 8725	W. A. Kawahara
1	MS 9405 8726	R. E. Jones
1	MS 9161 8726	P. A. Klein
1	MS 9161 8726	E. P. Chen
1	MS 9042 8728	C. D. Moen
1	MS 9405 8743	R. A. Regueiro
1	MS 9003 8900	K. E. Washington
1	MS 9012 8920	P. E. Nielan
1	MS 9217 8950	P. T. Boggs
1	MS 9217 8950	P. D. Hough
1	MS 9217 8950	M. L. Koszykowski
1	MS 1110 8950	L. J. Lehoucq
1	MS 9217 8950	K. R. Long
1	MS 0841 9100	T. C. Bickel
1	MS 0836 9100	M. R. Baer
1	MS 0828 9100	T. Y. Chu
1	MS 0828 9100	C. W. Peterson
1	MS 0835 9100	D. K. Gartling
1	MS 0834 9112	A. C. Ratzel
1	MS 0826 9113	W. Hermina
1	MS 0826 9113	T. J. Bartel
1	MS 0827 9114	J. E. Johannes
1	MS 0827 9114	K. S. Chen
1	MS 0827 9114	L. A. Mondy
1	MS 0827 9114	R. R. Rao
1	MS 0827 9114	P. R. Schunk
3	MS 0825 9115	W. H. Rutledge
3	MS 0825 9115	S. J. Beresh
3	MS 0825 9115	F. G. Blottner

1	MS 0828 9115	K. V. Chavez	1	MS 0847 9142	M. L. Blanford
3	MS 0825 9115	D. W. Kuntz	1	MS 0847 9142	M. W. Heinstein
1	MS 0825 9115	M. A. McWherter-Payne	1	MS 0847 9142	S. W. Key
3	MS 0825 9115	J. L. Payne	1	MS 0847 9142	G. M. Reese
3	MS 0825 9115	D. L. Potter	1	MS 0835 9142	J. R. Weatherby
1	MS 0825 9115	K. Salari	1	MS 0826 9143	J. D. Zepper
1	MS 0825 9115	L. W. Young	1	MS 0827 9143	K. M. Aragon
1	MS 0825 9115	W. P. Wolfe	1	MS 0827 9143	H. C. Edwards
1	MS 0836 9116	E. S. Hertel	1	MS 0847 9143	G. D. Sjaardema
1	MS 0836 9116	S. R. Tieszen	1	MS 0826 9143	J. R. Stewart
1	MS 0827 9117	R. O. Griffith	1	MS 0321 9200	W. J. Camp
1	MS 0367 9117	R. J. Buss	1	MS 0310 9209	G. S. Heffelfinger
1	MS 0827 9117	R. B. Campbell	1	MS 1111 9209	S. J. Plimpton
1	MS 0827 9117	D. Dobranich	1	MS 1110 9211	D. E. Womble
1	MS 0836 9117	R. E. Hogan	1	MS 1110 9211	R. Carr
1	MS 0827 9117	T. E. Voth	1	MS 1110 9211	S. Y. Chakerian
1	MS 0847 9120	H. S. Morgan	1	MS 0847 9211	M. S. Eldred
1	MS 0555 9122	M. S. Garrett	1	MS 0847 9211	A. A. Giunta
1	MS 0847 9123	A. F. Fossum	1	MS 1110 9211	W. E. Hart
1	MS 0847 9124	D. R. Martinez	1	MS 1110 9211	A. Johnson
3	MS 0847 9124	K. F. Alvin	1	MS 1110 9211	V. J. Leung
1	MS 0847 9124	T. B. Carne	1	MS 1110 9211	C. A. Phillips
1	MS 0847 9124	J. L. Dohner	1	MS 0847 9211	J. R. Red-Horse
1	MS 0847 9124	R. V. Field	20	MS 0819 9211	T. G. Trucano
1	MS 0557 9124	T. Simmermacher	1	MS 0847 9211	B. A. vanBloemen
1	MS 0847 9124	D. O. Smallwood			Waanders
1	MS 0847 9124	S. F. Wojtkiewicz, Jr.	1	MS 0318 9212	G. S. Davidson
1	MS 0557 9125	T. J. Baca	1	MS 1109 9212	R. J. Pryor
1	MS 0553 9126	R. A. May	1	MS 1110 9214	J. DeLaurentis
1	MS 0828 9132	J. L. Moya	1	MS 1110 9214	R. B. Lehoucq
1	MS 0828 9132	S. N. Burchett	1	MS 0321 9220	A. L. Hale
3	MS 0828 9133	M. Pilch	1	MS 1110 9223	N. D. Pundit
1	MS 0828 9133	B. F. Blackwell	1	MS 0321 9224	J. A. Ang
20	MS 0828 9133	K. J. Dowding	1	MS 0321 9224	R. E. Benner
1	MS 0828 9133	A. R. Lopez	1	MS 0321 9224	J. L. Tompkins
1	MS 0828 9133	K. E. Metzinger	1	MS 0847 9226	R. W. Leland
1	MS 0828 9133	W. L. Oberkampf	1	MS 0847 9226	B. A.Hendrickson
20	MS 0828 9133	T. L. Paez	1	MS 0847 9226	P. Knupp
1	MS 0828 9133	C. Romero	1	MS 0318 9227	P. D. Heermann
20	MS 0828 9133	V. J. Romero	1	MS 0318 9227	C. F. Diegert
20	MS 0828 9133	A. Urbina	1	MS 0310 9230	P. Yarrington
1	MS 0828 9133	W. R. Witkowski	1	MS 0819 9231	E. A. Boucheron
1	MS 1135 9134	S. Heffel	1	MS 0819 9231	K H. Brown
1	MS 0555 9134	J. T. Nakos	1	MS 0819 9231	K. G. Budge
1	MS 0835 9140	J. M. McGlaun	1	MS 0819 9231	D. E. Carroll
1	MS 0835 9140	J. A. Fernandez	1	MS 0819 9231	M. Christon
1	MS 0835 9140	A. Gurule	1	MS 0819 9231	R. R. Drake
1	MS 0835 9140	R. Garber	1	MS 0819 9231	A. C. Robinson
1	MS 0835 9141	S. N. Kempka	1	MS 0819 9231	R. M. Summers
1	MS 0835 9141	S. P. Burns	1	MS 0819 9231	M. K. Wong
1	MS 0835 9141	R. J. Cochran	1	MS 0820 9232	P. F. Chavez
1	MS 0835 9141	B. Hassan	1	MS 0820 9232	R. M. Brannon
1	MS 0835 9141	C. Roy	1	MS 0820 9232	M. E. Kipp
1	MS 0835 9142	J. S. Peery	1	MS 0820 9232	S. A. Silling
1	MS 0847 9142	S. W. Attaway	1	MS 0820 9232	P. A. Taylor

1	MS 0316 9233	S. S. Dosanjh
1	MS 1111 9233	D. Gardner
1	MS 1111 9233	A. G. Salinger
1	MS 1111 9233	J. N. Shadid
1	MS 0316 9235	J. B. Aidun
1	MS 1111 9235	H. P. Hjalmarson
1	MS 0660 9519	D. S. Eaton
1	MS 0660 9519	M. A. Ellis
20	MS 0419 9800	R. G. Easterling
1	MS 0421 9814	J. M. Sjulin
1	MS 0423 9817	R. A. Paulsen
1	MS 0423 9817	S. E. Dingman
1	MS 9003 9900	D. L. Crawford
1	MS 9003 9904	J. E. Kelly
1	MS 0139 9905	R. K. Thomas
1	MS 0428 12300	D. D. Carlson
1	MS 0490 12331	J. A. Cooper
1	MS 0829 12323	F. W. Spencer
1	MS 0829 12323	M. L. Abate
20	MS 0829 12323	B. M. Rutherford
1	MS 0638 12326	M. A. Blackledge
1	MS 0638 12326	D. E. Peercy
1	MS 0638 12326	D. L. Knirk
1	MS 0490 12331	P. E. D'Antonio
1	MS 0492 12332	D. R. Olson
1	MS 0405 12333	T. R. Jones
1	MS 0434 12334	R. J. Breeding
3	MS 0829 12335	K. V. Diegert
1	MS 1221 15002	R. D. Skocypec
1	MS 1179 15340	J. R. Lee
1	MS 1179 15341	L. J. Lorence
1	MS 0301 15400	J. L. McDowell
1	MS 9018 8945-1	Central Technical Files
2	MS 0899 9616	Technical Library
1	MS 0612 9612	Review & Approval Desk For DOE/OSTI