An Engineering Perspective on UQ For Validation, Reliability, and Certification

Jane M. Booker*, Timothy J. Ross**, Amanda C. Rutherford*, Brian J. Reardon*, Francois M. Hemez*, Mark C. Anderson*, Scott W. Doebling*, Cliff A. Joslyn***

* Weapon Response Group, Engineering Sciences & Applications Division, **Los Alamos National Laboratory**

**Department of Civil Engineering
University of New Mexico

******* Knowledge Systems & Computational Biology Group, Computer & Computational Science Division, **Los Alamos National Laboratory**

ABSTRACT

The analysts in the Engineering Sciences & Applications (ESA) Division at Los Alamos have been developing and applying uncertainty quantification (UQ) tools and methods to model verification and validation (V&V) and engineering component/system certification for several years. We begin with dissecting the term UQ (uncertainty quantification), and demonstrating the roles of model V&V and engineering qualification in overall weapon system certification and reliability through various methodologies that have been developed.

Three examples of UQ applications are presented, showing various technical challenges and solutions in handling uncertainties. The first is a fundamental model-test comparison for a single component and a single output. However, several different models calculations are possible, including a crude "back of the envelope" estimate and three test points from a previous experiment. These are combined using an alternate mathematical theory for uncertainty, called possibility theory. Development of a total uncertainty (TU) metric is demonstrated for combining the possibility distribution formed from the models with a probabilistic distribution formed from the test data. The second example is an "end-to-end" V&V exercise supported by a test program. It illustrates the use of test planning and outlines a set of V&V uncertainty-focused program planning steps. The third example illustrates the various uncertainties involved with a complex modeling problem where the observable quantities are not the quantities of importance for the models. Here uncertainties of inference become important.

I. Definitions and Dissection of Terms and Concepts

In order to promote improved communications and establish reference points, the definitions and terminology used here are defined. Some of these definitions are becoming standardized by activities such as the American Society of Mechanical Engineers, Committee on Verification and Validation in Computational Solid Mechanics. Others, such as UQ, are specific to smaller groups at Los Alamos.

- **Uncertainty Quantification (UQ)** is the process of characterizing, estimating, propagating, and analyzing various kinds of uncertainty (including variability) for a complex decision problem. For complex computer and physical models UQ focuses upon measurement, computational, parameter (including sensitivities of outputs to input values), and modeling uncertainties leading to verification and validation. UQ is an assessment process/activity. Unfortunately the Quantification label implies providing a numerical statement. We maintain that a UQ process can be followed but that a system level or bottom line assessment need not be numeric. Statements such as "the system is certified for its requirements and life cycle" are the result of a UQ assessment but contain no numeric evaluation.

- **Verification** is the process of determining that a model implementation (code calculation) accurately represents the developer's conceptual description of the model and the solution to the model (AIAA, 1998).

- **Validation** (or Model Validation) is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model (AIAA, 1998). It is the act of quantifying how well and under what conditions a model or code matches real systems and thus can be trusted in a predictive capacity. Calculations or models are validated for a particular domain of applicability.

- **V&V** is combined processes of Verification and Validation. Verification should be the prerequisite of Validation. The boundaries between these two processes are not crisply defined. One may argue that the effect of mesh or grid sizing falls in either side. Uncertainties arise from incomplete validation and/or verification processes. Nitta and Logan (2004) characterize this risk in their QRC metric.

Metrics for V&V continue to develop. For example, Roache's Grid Convergence Index is becoming a standard for estimating grid convergence error in the verification community (Roache, 1998a). For validation, over the last several years the fundamental quantity for comparing numerical calculations (y^*) to test data (y) has been the difference or error between them, or simply $e = (y^*-y)$. Root mean square error, RMSE, or expected values, $E(e)$, of these errors have provided aggregate estimators for these differences. Chi-squared tests can also be used by designating either y values as observed and y^* as expected, or vice versa. However, these metrics are usually calculated without consideration of the uncertainties in y and y^*. If probability density functions (PDFs) are available for both y and y^*, then entropic-based metrics can be utilized at each response value, comparing the overlap of the distributions, f(y) to f(y*). Kullback Leibler and Jeffrey's J (Baroi, et al., 2004) are two such metrics. Non-parametric tests for empirical-based distributions on each y and y* such as Kolmorgov-Smirnoff type tests (Conover, 1971) are also common practice for determining differences between distributions. However, comparing PDFs at each response value, each (y_i, y_i*) pair, provides the comparison at each pair and is not a metric for all the values. Perhaps one could argue for a multivariate solution to accomplish this aggregation, but that is the topic of another paper. Our most recent approach is to develop a single validation *metric* focusing on the uncertainties involved, as discussed below in section II.3.

Developing a validation metric first implies consideration of the goals for the validation effort. At Los Alamos, that goal is to certify the nuclear package of the engineered weapon systems. It is the job of the weapons engineering community to deliver a nominal, working nuclear package to the physics community. Certification/qualification is therefore done first at the engineering level and then at the physics level, with the physics certification conditioned upon the engineering certification. The same is true of reliability; physics reliability is conditioned upon engineering reliability. The following definitions apply to this conditional breakdown at Los Alamos:

- **Certification:** An assessment of the overall system's ability to perform. Providing a warranty to customers. (Engineering certification ≠ physics certification).

- **Qualification:** An assessment of a component, subsystem, or individual function's ability to perform.

- **Reliability:** The classic textbook definition of reliability is the probability a system performs its intended function for a given period of time under specified conditions. The NNSA/DOE defines reliability as the probability that, in use, detonation with the specified yield occurs at the target for given specified conditions and for a specified time—a definition consistent with classic reliability.

A common use of reliability by decision makers is to view it as a *metric* for certification/qualification. Another such metric, common to the National Labs and their customers (DoD and DOE), is margin, which is defined as:

- **Margin** is the minimum of the distance from any assessed condition to its nearest performance threshold.

- **Performance Threshold** is the minimum acceptable condition necessary to perform an intended function. A performance requirement is a physical condition or configuration required of a product to function. If the product exceeds the performance requirement, it does not necessarily perform better, but it may be more likely to perform acceptably, thus instilling greater **confidence**. In general, the separation between margin and failure domains (i.e., PT) is defined not as the point where failure occurs, but rather as the point where we are no longer confident that the product performs its intended function. There may be more than one PT corresponding to upper and lower thresholds.

- **Confidence** is a commonly used term whose definitions include words like trust, belief, reliance, and certitude. It is the state of feeling sure (Webster, 1986). "Confidence comes from repetition, from the breath of many mouths," W.B. Yates.
  It is interesting to note that confidence also refers to the acts of a swindler, as in being duped by a confidence man. This definition originates from the trust elicited from and given to the swindler.
  It is also interesting to note that even the Greeks were unable to precisely (or mathematically) define what is meant by confidence. Outside of the statistical context discussed next, there is no modern day definition for the mathematical meaning or quantification of confidence. Therefore, we discourage its use in V&V and UQ studies unless defined using the statistical definitions. However, we are willing to note that confidence seems to have an inverse relationship to uncertainty, something we do focus upon quantifying.
  In the statistical realm, confidence has a specific meaning in sampling and inference when referring to a confidence interval for an unknown parameter (e.g., the mean). The interpretation of a confidence interval is difficult and often misused. Its meaning refers to a sampling process and the calculation of multiple confidence intervals for multiple repeated samples. For example, if one were to take 100 samples from a population and calculate 100 95% confidence intervals (one per sample) for the mean, then 95 of those confidence intervals would contain (cover) the true (unknown) value of the mean. The ASME (2004) guide on validation is considering using statistical confidence interval estimation as a validation metric.
  Another statistically based confidence definition is in common use. The so-called confidence level is defined as the complement of a significance level in statistical hypothesis testing. Confidence level is $1-\alpha$, where α is the chosen significance level or the Type I error[1].

Another *metric* specifically designed for engineering certification/qualification based upon margin and the uncertainties associated with determining margin is the Engineering Index (EI) (Dolin, et al., 2002):

- **Engineering Index** is a normalized representation of margin, which primarily measures the extent to which an engineered product exceeds its performance requirements.

---

[1] Type I error is the chance (e.g., 5%) that a null hypothesis is rejected when it should not have been rejected, i.e., the null hypothesis is true. This is a chosen, and therefore, controlled error in statistical inference.

Physicists at Los Alamos and Lawrence Livermore Labs have defined a similar concept for physics certification under the terminology of QMU (Quantification of Margin and Uncertainty) and the concept of *gates*. These will not be discussed further here. Instead we will focus upon the role of uncertainties and their characterizations, usually referred to as quantification. Uncertainty is therefore defined in the broadest sense as:

- **Uncertainty:** Everything that is not known absolutely. Aspects of uncertainty include variability, imprecision, vagueness, ambiguity, lack of knowledge, inconsistencies, conflict, non-specificity, entropy, multiple alternatives, inferences, prediction, and the unknown.

This definition highlights the various types of uncertainties present within a complex problem such as weapon system engineering certification. Many mathematical theories exist for handling these different types. These are collectively known as **Generalized Information Theories** (Klir and Wierman, 1999). There is an on-going development of a metric to quantify **total uncertainty** expressed as an aggregation of the various forms of uncertainty.

- **Total Uncertainty (TU)** is defined as the combination of the two general types of uncertainty: *irreducible* or natural *variability* (which cannot be reduced, but only quantified) and lack of specific information, *reducible,* (which can be reduced with the acquisition of more information). *Error*, whether numerical / computational or mistakes, could be interpreted in either category. Some may be familiar with the taxonomy of uncertainties using the terms *epistemic* (lack of knowledge and/or reducible) and *aleatory* (irreducible). This taxonomy specifically distinguishes statistical (or natural, random) variation from other, reducible forms of uncertainty.

- **Generalized Information Theories (GITs)** are the collection of mathematical theories for characterizing different kinds of uncertainties. These include measure-based theories such as are evidence theory, possibility theory, fuzzy set theory, random intervals, imprecise probabilities, Choquet capacities, and probability theory. These can further be classified based upon their foundations from either fuzzy sets or crisp sets. Each theory has its own set of axioms, specifying how to combine sets and measure functions.

Having established the use and meaning of the above terms, the sections that follow describe some uncertainty quantification (UQ) efforts in engineering V&V activities under the ASC program at Los Alamos.

II. LANL Engineering UQ Activities in V&V

We begin with describing the important roles of V&V and UQ in Los Alamos engineering reliability and certification tasks. These tasks support the ASC program mission to provide leadership in the development, implementation and improvement of scientifically rigorous methods for assembling evidence supporting the credibility of code calculations and the utility of these calculations for Los Alamos's mission of stockpile stewardship (Doebling, 2004). Two of the strategic goals for the ASC program at LANL are discussed here (Doebling, 2004):

- Uncertainty Analysis: development and deployment of rigorous methods for characterizing uncertainties in calculations and data.
- Validation Assessment: development and deployment of rigorous methods for aggregating evidence on the adequacy of computations with respect to physical phenomena.

In the absence of test data, the high level decisions about certification must be made utilizing all available knowledge and information. Heavy reliance is therefore placed on expert knowledge and experience, and upon validated calculations and models. What little test data is or can be made available assumes a duality of purpose: 1) using data itself for estimating reliability and other decision metrics and 2) using the data for validating calculations and models which, in turn, provide additional information for estimating reliability and other decision metrics. This places an additional burden upon the importance of V&V efforts. Care must be taken not to analyze the same data in both roles simultaneously, constituting double counting of the same data.

1. UQ and V&V in Engineering

Figure 1 illustrates a simplified version of UQ activities among the Los Alamos organizations and the processes leading to nuclear physics package (system) reliability assessment and certification. The detail under the physics side is left blank, denoted by three vertical dots, for focusing on the engineering side. The multiple, crossing arrows indicate the complexity of the use and flow of information and analyses. Each box contains activities requiring UQ and each arrow represents the propagation of those uncertainties. Methodologies and metrics interplay at all levels as well.

2. Methodologies for UQ

The metrics outlined in Section I are often difficult to calculate because:

1. Test/experimental data are lacking or sparse.
2. Systems are complex in structure and functions.
3. Environments known or tested do not correspond to requirements.
4. Performance (and therefore performance thresholds) is not well known or understood.
5. Physical models may not be adequate to describe behavior.
6. System level requirements are not easily translated to lower levels.
7. Multiple variable (characteristics of interest) interrelationships are not well known.
8. Uncertainties from lack of knowledge are difficult to characterize.
9. Heavy reliance is placed upon models (which must be validated) and expert knowledge (which must be properly elicited).

10. All information (e.g., models, knowledge, data, experience) must be utilized and combined with the appropriate uncertainties attached.
11. Performance changes with time (e.g., aging, degradation).

Some of the challenges posed in this list are being partially addressed, and in that process, methodologies are being developed. A few examples include:

- Enhanced Reliability Methodology (aka PREDICT) is an information integration methodology designed to address items 1, 2, 7, 8, 10, and 11 (Meyer, et al., 1999).
- Quantitative Reliability at Confidence (QRC) is a risk-based methodology for reliability and V&V designed to address items 1, 2, 3, 5, 6, 7, 9, and 11 as described in Nitta and Logan (2004).
- Engineering Index of Goodness for the Enduring Nuclear Stockpile (EIGENS) is another information integration methodology designed to address items 1, 3, 4, 6, 7, 8, 9, 10, 11 (Dolin, et al., 2002).

While much more development of these, and other, methodologies is required to address all the challenges, progress continues. One of the areas of progress is the concept of utilizing different mathematical theories for different kinds of uncertainties.

3. Predictive Modeling

One of the goals of model formulation (either physical, statistical or computational) is to utilize the model for prediction. Assessing the utility of a model involves addressing three interdependent issues (Hemez and Ben Haim, 2002):

1. Improving the fidelity of test data; the degree of matching between test data and calculations (predictions)
2. Studying the robustness of predictions to various kinds of uncertainties. This requires the concept or determination of a maximum or horizon uncertainty for which all models or information sources meet a given fidelity (matching) requirement.
3. Establishing the prediction capability of models in situations where testing is not possible. This is also called *prediction looseness*, referring to the range of predictions expected from a family of equally credible models or information sources.

A fundamental theorem developed by Ben Haim and Hemez proves that all three of these issues cannot be achieved simultaneously in an information-gap decision theory framework (Ben Haim, 2001); a non-measure based decision framework characterizing ignorance. At best, one can only achieve success on any two, while sacrificing the third. Thus an important tenet of predictive modeling is established, similar to the optimization problem of trading off “better,” “faster,” and “cheaper” in project planning and decision making. Trade offs include (Hemez and Ben Haim, 2004):

- *Robustness to uncertainty decreases as (model-to-test) fidelity improves.* Model-to-test fidelity refers to the degree of matching (“correlation) between test data and model / calculations results. Models (and associated) information chosen or tuned to better reproduce the available test data are more vulnerable to errors in modeling assumptions, errors in the functional form of the model, and uncertainty and variability in the model parameters.

- *Prediction looseness increases as robustness to uncertainty improves.* Prediction looseness is a concept related to the range or scatter expected from a family of equally credible models. Models / information made more immune to uncertainty provide a wider range of predictions, hence lessened predictive power.
- *Prediction looseness decreases as (model-to-test) fidelity improves.* Models / information chosen or tuned to better reproduce the available test data provide more consistent forecasts, leading to a false sense of assurance or decreased uncertainty.

Similar trade offs can be found in the measure-based quantification of uncertainties (the TU metric) described below in the first application, section III.1.

4. Development of a *TU* metric for Validation

One of the goals of the engineering activities at Los Alamos is to develop a metric (yardstick) for validation. This metric would incorporate the comparison of calculations (models/calculations) to test/experimental data comparing the uncertainties in both. The concept of *TU* involves the mathematical combination of uncertainties from the data and calculation sources. However models and test data involve more than comparing one quantity of interest. Often, important variable quantities are functions of each other (e.g., stress versus strain) or functions of time (pressure versus time). Therefore multivariate or multi-dimensional comparisons are necessary to understand the interrelationships among these variables, and are a subject of ongoing research and development.

Assessing modeling uncertainty is a daunting task, encompassing uncertainty evaluation associated with model choice, model relevance, model domain of applicability, parametric uncertainties, input to output sensitivities, multivariable relationships, and lack of knowledge. Some of these (e.g., parametric uncertainties) can be handled with probability-based variance. Others (e.g., model choice) represent a vagueness or ambiguity within the decision process that is not easily captured as variability. For these, another kind of mathematical theory is suitable, one that focuses on the possible value of model choices. For the metric development in this section, we shall focus only on two of the more prevalent types of uncertainty for the test data and for the model choice– *probability* theory and *possibility* theory, respectively.

The fundamental difference between these two theories of quantifying uncertainty is that in *probabilistic* bodies of evidence all the evidence is concentrated on the singletons of a universe of information, whereas in *possibilistic* bodies the evidence is located on collections of nested sets within the universe of information (see Figure 2). Both formalisms are uniquely represented by distribution functions, but their normalization requirements are different. The collection of values in a *probability* distribution are required to add or integrate to unity, while for *possibility* distributions the largest values are required to be unity (a condition called *normality*). Figure 3 shows an overlay of both distributions.

These differences in mathematical properties of the two theories make each theory suitable for modeling various types of uncertainty and less suitable for modeling other types. For example, probability theory is an ideal tool for formalizing uncertainty in situations where event frequencies are known or where evidence is based on outcomes of a large number of independent and repeatable trials, or where evidence can be summarized by a subjective *willingness to bet*. Possibility theory, by contrast, is ideal for formalizing incomplete information expressed in terms of vague or ambiguous terms, or where evidence supports conflicting events.

To assess the **total uncertainty** in the process of validation assessment, the hypothesis is formulated that **total uncertainty** metric should scale between two extreme conditions on uncertainty, i.e., between the case of *no-uncertainty* and the case of *maximum uncertainty*. If a prediction is made on the response of some structural system and the level of uncertainty expressed in that prediction is *close* to the extreme of *no-uncertainty,* then credibility[2] in that prediction exists. On the other hand, if the uncertainty is *closer* to the other extreme, the case of *maximum uncertainty*, then less credibility exists in the prediction. Of more importance, however, is the development of a "metric of credibility" or "confidence" that will scale linearly with the quantified level of uncertainty and, in a mathematical sense, measure the degree of *closeness*.

Suppose we are predicting the value of a variable of interest in a mechanics calculation, say the maximum stress in a metal bar. The case of *no uncertainty* is defined where all information and evidence support only one value of the variable of interest (stress) and there is no evidence on all other potential values of that variable. Probability is associated with a value equal to unity on one value of the variable (stress) and zero probability on all other potential values of that variable. The other extreme of *maximum uncertainty* is then defined as the case where all potential values of stress are completely possible (i.e., certain) and all potential values of the variable are equi-probable (the case of a uniform probability distribution).

In the literature, the forms of uncertainty associated with a *possibility distribution* are those called *non-specificity* and *discord*. Non-specificity refers to a kind of imprecision which is connected with sizes of relevant sets of alternatives. Discord means that there is conflict among the various sets of alternatives. The form of uncertainty associated with a probability distribution is a kind of *strife*, where again there is *conflict* among the various specific values of alternatives. *Probabilistic* strife (or conflict) is most often termed as *entropy*, ($\log_e$), and it is different from *possibilistic* discord. In the former, all evidence is nested on single values of the variable of interest, whereas the latter supports evidence that is nested on collections, or sets, of various values of the variable of interest. Hence, **total uncertainty,** as used in the context presented here, is defined as the combination of possibilistic *non-specificity* with the probabilistic entropy, or conflict. The new

---

[2] One could argue that using the term "credibility" is no better than using the word "confidence." We would agree. Its casual use (communication with Roger Logan) or colloquial use here connotes a qualitative assessment or even a warm, fuzzy feeling and is not meant to be quantified.

procedure presented here combines these uncertainties. Other combination procedures in the literature first reduce the various uncertainties to the bit-level ($\log_2$) before a combination is performed, or the various uncertainties are used to formulate a data-tuple (see Klir and Wierman, 1999).

The development of the total uncertainty begins by first defining an uncertainty matrix, A, which contains the *possibilistic* uncertainty vector in its first column and the *probabilistic* uncertainty vector in its second column. These vectors are formed from the possibility and probability distributions (Figure 3). Each row (π, p pair) corresponds to a particular value of the *quantity of interest*, meaning is it a discrete vertical cut of the overlaid distributions.

For a variety of compelling and intuitive reasons (Ross, et al., 2003) a procedure known as singular value decomposition (SVD) (Klema and Laub, 1980) was chosen as a means to calculate **total uncertainty**. Many of these reasons have to do with an analogy of this approach to the extraction of modal frequencies from a model of a structure undergoing dynamic motion. In this process modal extraction of frequencies are extracted using an eigenvalue analysis of the structure. The frequencies from the eigenvalue analysis represent the *total energy* of the structure during vibration in its normal modes (eigenvectors). In our analogy for the characterization of uncertainty, the singular values of our SVD analysis represent the *total energy of the uncertainty* in the matrix *A*, or more simply the total uncertainty.

To begin this development, start first with the decomposition of *A*, an m x 2 matrix of columns expressing the two types of uncertainty:

$$A = \begin{bmatrix} \pi_1 & p_1 \\ \pi_2 & p_2 \\ \pi_i & p_i \\ \pi_{i+1} & p_{i+1} \\ \pi_m & p_m \end{bmatrix} \text{ and } A = U\Sigma V^T \qquad (1)$$

where p_i's are probabilities, π_i's are possibilities, *U* is an orthonormal m x m matrix whose columns are the left singular vectors of *A*, Σ is an m x 2 matrix containing the singular values of *A*, and *V* is an 2 x 2 orthonormal matrix whose columns are the right singular vectors of *A*. Parameter m is the length of the two uncertainty vectors, the number of discrete values or vertical cuts of the quantity of interest, such as a peak acceleration.

What is of particular interest is the fact that the singular values in the matrix Σ contain the "energy" or the *total uncertainty* of the vectors quantifying different uncertainties expressed in *A*. The expression for ***Total Uncertainty, TU***, is then given by:

$$TU = 2\left\{\left[\lambda_{\max}\sum_{i=1}^{m}\sigma^2{}_i\right] - 1\right\}, \text{ for i = 1, 2,...,m} \quad (2)$$

where, $\lambda_{\max} = \left[\pi_{\max}^2 + p_{\max}^2\right]^{-1}$, and where σ_i is the ith singular value in Σ, where, π_{max} is the largest possibility value in the first column of *A*, and p_{max} is the largest probability value in the second column of *A*. Finally, the extreme cases for *no-uncertainty* and *maximum uncertainty* are given by the expressions, *TU* = 0, and *TU*= 2(m-1) (see Ross, et al., 2003). An example application of this *TU* metric is given in Section III below.

The matrix *A* can be expanded to include other columns, representing other types of uncertainties in the validation process, such as model parametric uncertainties. Research is ongoing to extend *TU* based upon this SVD approach to include other mathematical theories, such as Dempster Schafer belief functions, fuzzy membership functions, or random intervals—all of which can play important roles in the kinds of uncertainties experienced in a V&V process.

5. Outline of a Validation Methodology/Process Focusing on Uncertainties

There have been several validation programs successfully implemented at Los Alamos in structural dynamics engineering applications (see Hemez and Ben-Haim, 2002; Doebling et al., 2004a, Doebling et al., 2004b and, Anderson, 2002). The focus of these is the validation of structural dynamics models. Primarily the interest is in nonlinear response models, such as the transient response of structures with multiple component interfaces and nonlinear materials.

The engineering modeling and simulation communities have not standardized the steps involved for verification and validation. Many taxonomies and flowcharts are available (e.g., Roache, 1998b and Trucano et al., 2002), and there is no need to provide another here. Instead, Table 1 contains a brief outline of considerations and issues involved with associated uncertainties.

The background for interpreting Table 1 includes the following statements:

- The validation of a model can only be defined over a prescribed domain of the simulation input parameters.
- This domain needs to be specified whenever a statement about the model's validity is made.
- The validation process is here defined to begin after the computer code for that model has been properly verified.
- Numerical error evaluation, mesh or grid size and convergence issues, discretization issues are all encompassed within the verification process[3].

[3] Verification here refers to code verification and solution verification as described in the ASME (2004) guide.

Table 1. Outline of V&V Considerations with a Focus on Uncertainty

V&V TASKS	UNCERTAINTY ISSUES
Define desired phenomena for study (initial listing of measured variables, inputs and responses for models/calculations, covariates related to things measured and things calculated)	Errors of omission of related phenomena and variables Can measurements be made on all desired phenomena? Can models be formulated to completely capture all phenomena?
Define the hierarchical levels for models and tests	Can levels of tests and models be the same? Assumptions made about specific levels for both tests and models.
Define the domain for the phenomena plus modeling domain and test / experimental domain	How do these three domains overlap? Can tests be performed in the desired domain? Can calculations be done in the desired domain?
Model(s) selection	Model choice uncertainty: Can any model be truly representative of phenomena (aka inference uncertainty)? Can multiple models be representative (aka modeling uncertainty)? What are boundary and initial conditions for each model? Are these realistic? Are there faster running, less detailed, lower fidelity models? Can metamodels be used?
Computer code implementation of model(s).	What are numerical errors? What are discretization errors? What are approximations used? What are grid and mesh issues?
Computer code verification	Model selection and numerical error uncertainties
Input/output parameter sensitivity study of model(s) Design of computer experiments study for parsimonious code runs[4]	What inputs most influence the output responses? Variance-based uncertainties addressed.
Dimension reduction, trimming models, inputs or outputs to a parsimonious set for modeling	Information can be lost with such selections.

[4] Parsimonious code runs refers to the process of narrowing down the design space for computer runs to a manageable set by taking advantage of parameter and variable interrelationships ("correlations"). For example if there are 10 input variables, but the sensitivity study shows only 4 are most influential, then the suite of computer runs would be designed around changes in those 4.

Planning experiments/tests	Can the needed phenomena be measured (inference uncertainty)? What variables or effects can/cannot be controlled/measured? Must variable selection be done?
Design of experiments / tests utilized	Can the necessary tests/experiments be conducted to aid in model validation? Can sufficient tests be conducted to cover the variable space? Must interactions be sacrificed to limit number of tests?
Conduct experiments / tests	What are measurement errors? Missing or incomplete data.
Data analysis / feature extraction / dimension reduction of test data	Information can be lost by feature extraction and dimension reduction. Variance-based data analysis techniques (e.g., multivariate techniques) can reduce the variable space. Metamodeling can summarize data for model comparison.
Data analysis / feature extraction / dimension reduction of model inputs and outputs	Information can be lost by feature extraction and dimension reduction for models. Variance-based data analysis techniques (e.g., multivariate techniques) can reduce the variable space. Metamodeling can be used for analyzing model outputs.
Design code runs for validation	Matching calculations to test results for multiple variables; inputs and responses. Comparison of model(s) to test data; a validation metric. Uncertainties in both model(s) and data are combined.
Prediction—using models to draw conclusions where testing is limited.	Uncertainties of prediction; extrapolation and interpolation come into play.

The uncertainties inherent in the statements or in answering any of the questions in the right column may not be easily characterized and may not be conducive to probabilistic (entropic, variance-based) uncertainty alone. To proceed through any decision point indicated by these issues, decisions must be made concerning the use of the model(s) and data. These decisions may involve implementing component-level tests or running lower fidelity models to gain information (and reduce the uncertainty from lack of knowledge). Therefore, an iterative structure of model-test-model-test would be beneficial. This is especially true if the structure of the problem / phenomenon is a complex hierarchy of levels.

General uncertainties follow the same principles of implementation as entropic uncertainty from probability theory does. In handling uncertainties, one of these actions takes place:

- Control the uncertainty (as in a controlled effect in an experiment);
- Lacking control ability, measure the uncertainty either directly or by comparing changes in influential variables;
- Lacking measurement ability, estimate the uncertainty (based upon general knowledge, experience or first principles);
- Lacking the above, it may be possible and practical to assume the uncertainty is unimportant (because it's magnitude is in the noise level), and document that deliberate decision;
- Lacking all of the above, whether justified or not, the uncertainty is ignored, often without any consideration or documentation.

Unfortunately, the last bullet is the most common practice. To make an honest uncertainty assessment, if the last two bullets is chosen, then the entire V&V process and the results are **conditioned** upon the assumptions or omissions made. In other words, the results are **only** applicable given those assumptions being true or omissions being unimportant. While that statement sounds obvious, it is most often not explicitly stated or considered as part of the V&V activities.

A few terms in Table 1 are worth some discussion, such as metamodeling. Metamodeling is used for efficient parameter optimization in a finite element simulation. A metamodel can be a response surface, surrogate model, or any fast-running model built from polynomials, sinusoidal functions, splines, neural networks, multivariate analyses (e.g., principle components), or general linear/nonlinear model regressions. The primary use of metamodeling is to fill in gaps in the parameter when relatively few complex finite element (FE) model runs are feasible. Multidimensional metamodel results are easily displayed in three-dimensional graphs as two variables at a time with a response variable on the vertical. The steps for formulating a metamodel can be summarized as:

- Select candidate set of input parameters.
- Divide these parameters into a number of levels (numerical values).
- Perform FE simulation runs at several different combinations of these parameter values.
- Select a level for each parameter for each simulation run, resulting in a matrix of input values P (row = run #, column = input parameter). Selection is done by design of experiment techniques, utilizing expert knowledge, etc.
- Obtain output feature value from each simulation run, resulting in a vector of feature values (row = run #) F.
- Use regression techniques to estimate the metamodel coefficients.

In a complex physical problem, the number of phenomena that can be measured experimentally or modeled in a calculation is limited by physical problem constraints, measurement capability, first principles and theory, and time and money. To necessitate practicality, often variables must be screened and hence eliminated at various stages of planning the study. When there is some choice in the variable selection process either for the experiments or the modeling, a formalized selection process provides a logical and documented approach for choice. PIRT (Phenomena Identification and Ranking Table) is one such useful model development and variable screening/selection technique (Trucano, et al., 2002). It is based upon fundamental scoring and ranking principles often employed in risk and reliability assessments. The example in Table 2 (ASME, 2004) illustrates the evaluation of four phenomena, categorized by type and ranked according to importance to the response of interest and to the level of modeling capability.

Table 2. PIRT Example

Phenomena	Phenomena Type	Importance to Response of Interest	Level of Model Capability
A	Interface	High	Medium
B	Interface	Medium	High
C	Loads	Medium	Low
D	Materials	Low	Low

PIRT is not the only screening technique for variable selection. Others include:

1. Parsimonious general linear statistical modeling of responses
2. General sensitivity analysis using derivatives or variances
3. Bayesian variable screening

Those phenomena (as in Table 2) most important for the (multiple) responses of interest that also have good modeling capability become important for planning and designing the experiments. This set should ideally be the variables chosen as the treatments or quantities varied in the experimental design. Those not chosen should ideally either be controlled (held constant), measured (as covariates), estimated (by some other means), assumed to have minimal, unimportant effect on the response, or ignored. Unfortunately, often times the last of these occurs. The practice of good statistical experimental design is not as difficult to implement as one might believe. For example, while it takes $2^{12} = 4096$ tests for a full factorial experimental design for only 12 factors at two levels (high and low) each, just testing the 12 main effects at those two levels requires a minimum of 13 tests. This savings in test runs is at the expense of losing all interactions among the 12 variables. The trade off between test runs and number of effects (interactions) that can be tested is a decision that requires documentation of reasons, including the expert knowledge used (or assumed) in arriving at those decisions. These same experimental design issues are relevant to designing the number and types of model calculations as well as the experiments.

Design of experiments techniques such as screening designs, Taguchi arrays and augmented designs, permit ways of iterating for more tests and models and their comparisons when numerous variables and model choices are present. For example one could use a minimal design as described above to explore through the 12 variate space using a coarse finite element model to gain understanding and estimate which of the 12 is the most influential on the responses. Then one could augment that design by adding more levels of the most influential variables and perhaps use a finer mesh model as well. These ideas are often put into practice as described below in the second application. Again, proper documentation of the reasons behind the iteration choices is vital for peer review and for potential future updating.

III. Some Validation Applications

The three projects described in this section illustrate the various LANL engineering V&V activities, including the research challenges. The first is a fundamental model-test comparison for a single component and a single output. However, several different models calculations are possible, including a crude "back of the envelope" estimate and three test points from a previous experiment. The different model calculations are combined using an alternate mathematical theory for uncertainty, called *possibility theory*. Development of a **total uncertainty** (TU) metric is demonstrated for combining the possibility distribution formed from the models with a probabilistic distribution formed from the test data. The second example is an "end-to-end" V&V exercise supported by a test program. It illustrates the importance and use of test planning and outlines a set of V&V, uncertainty-focused, program planning steps. The third example illustrates the various uncertainties involved with a complex modeling problem where the observable quantities are not the quantities of importance for the models. Here uncertainties of inference become important.

1. Hyperelastic Foam

The first application is a case study involving a single component and the crushing of this hyperelastic foam from a simple impact loading. Of interest was a single response variable: the computational prediction of the peak acceleration (PAC) within the foam. Although experiments were performed on this example (see Figure 4.b), only the uncertainties involved in a computational prediction of the peak accelerations were considered. A finite element code was used to model the system.

Two types of analyses were performed on the uncertainties. First, *probability theory* was used to assess the uncertainties of the test results for the peak acceleration (PAC). Second, *possibility theory* was applied to various model configurations, whose parameters included such variables as friction coefficient, amount of preload applied, and impact angles of the drop table (see Figure 4.a). For example, uncertainties from previous test results on similar foam, simple first principle calculations, choices of material models, equation solvers, and boundary conditions were represented by intervals of possible values for PAC. Table 3 contains the interval values of PAC from the various models and sources of information. These six intervals are formulated into a π-DF (*possibility* distribution function) using a method developed in a dissertation by Donald (2003). This method permits weighting the different intervals according to the how each model best represents the real physics of the problem. It then determines values of possibility based upon the weights and the amount of overlaps among the six. That π-DF and the corresponding *probability* density function (PDF) of the test results for PAC are plotted in Figure 5.

Table 3. Intervals from Different Model / Information Sources

Model	Lower Value of PAC	Upper Value of PAC
SDOF, Material Models [2 cubics]	0.2480	0.2850
SDOF, Material Models [cubic, bilinear]	0.8570	0.1250
SDOF, HKS/"ABAQUS"	0.2850	1.6435
Preload in Bolt [min, max]	1.1943	2.4771
Hand calculation on I-mv [.01s, .001s]	0.3000	3.0000
Old test data range from 3 gages	1.2170	1.5940

In equations (1) and (2) above, the length of the probability and possibility distributions was m = 30[5]. Using these equations, the total uncertainty is expressed by the combination of probabilistic and possibilistic types and is equal to a value of $TU = 17.44$, which is 30% of the *maximum uncertainty* that the problem could contain (i.e., $TU_{\max} = 58$). What this means is that the problem does contain a level of uncertainty that is closer to the case of *no-uncertainty* than it is to the case of *maximum uncertainty*. More importantly, if additional analyses on this problem were to be performed, and the value for *TU* decreased below 17.44, this would equate to a situation of improved capability in the prediction of peak accelerations. If additional (more than 6) models or information became available and the new TU remained about 30% of its maximum, then this would be a situation of robustness to uncertainty.

The fidelity to data or matching of test data to calculations is not explicitly performed using the SVD approach. The overlaps or matching of the PDF and the π-DF distributions is explicitly (numerically) incorporated in the decomposition and singular value determination, which in turn, determines the value of TU. No additional measure of this overlap or matching is required. Thus the uncertainties and the matching are combined into the single TU metric, an advantage to our approach.

2. "End-to-End" V&V Study on the Threaded Assembly

This model validation application explores various techniques for the problem of propagating an explosive-driven shock through of a complex threaded assembly joint (Doebling et al., 2004). The steps of the study described below follow the outline of V&V uncertainty analysis issues listed in section II.4. The multi-year study was designed in two phases of testing and modeling. The first phase of experiments and modeling was designed to simulate acceleration response of component mass simulators under this kind of loading. The lessons learned from Phase I have been used to design the experiments and analyses for the next round of tests and modeling, Phase II.

The response of interest is the energy transmission of an explosively driven impulse, and

[5] It should be noted that results from simulations of TU indicate that m=30 is considered a small vector length for converging to asymptotic results.

the mechanics of interest are energy dissipation in various interfaces due to friction and preload. Two test suites of experiments were initially defined: shock response experiments and modal experiments. In the shock response experiments the explosive charge delivered an applied stress versus time impulses. Accelerometers measured the acceleration versus time impulse response. Measurement uncertainty bands were estimated according equipment specifications. The modal experiments were conducted to identify modal frequencies and mode shapes of the threaded assembly. The goal was to explore the interaction and dynamic behavior between the main components. The modal test consisted of 33 nodal inputs with seven accelerometer locations on the structure. A micro hammer was used to induce the excitation. Guidance on the frequencies and modes was learned from this test suite of experiments. There currently is no model for test to calculation comparison for the modal responses.

Only a small number of tests, four, were allocated for Phase I for the shock experiment suite; however, one of the goals was to determine the repeatability of experimental results. Therefore at least two of the tests were replicates. Even with only four test units, two factors could be explored: assembly tolerance (tight versus loose) and manufacturing tolerance (tight versus loose). The four tests were designed as: tests 1&2 as tight-loose, test 3 as loose-loose and test 4 as tight-tight. Responses for the shock experiments included time histories with corresponding temporal moments and moving averages, power spectral densities with spectral moments and shock response spectra with amplitude differences and spectral moments. Additional analysis included examining peak accelerations and fractal analysis of how a signal grows or decays in various time scales. Variability in the results was high enough to preclude definitive conclusions about the effects and the repeatability.

Another result came from the beginning of the Phase II study, involving the identification of phenomena judged to be important to the response, energy transmission through the threaded assembly. This was done using the PIRT technique with the initial table containing 50 parameters relating to friction, preload, material properties and load input to the structure. This parameter set was down-selected using engineering judgment (rooted in experience with the threaded assembly problem) to 11 parameters, involving friction and preloads, and will be further down-selected through a coarse sampling parameter sensitivity analysis. (Doebling et al., 2004). An independent suite of experiments was conducted to identify friction coefficients between several parts. Variability in the results from these tests was also high, but information was leveraged to generate friction uncertainty distributions. Preload tests were fairly inconclusive due to high variability.

At the time of the Phase I experiments (July 1999), the engineers anticipated that the modeling of this phenomena was at the limits of simulation ability. Indeed one of the lessons learned from the study was the limited ability to correctly simulate the flow of energy through some of the assembly components. Finite element models were created using ParaDyn. Analysis parameters included preloads, friction coefficients (static and kinetic), and loading characteristics. Modeling of all materials was done in the elastic region because the loads in the four experiments were low. In addition to the concern over accurate modeling of the energy flow mentioned above, concern was also for the nonlinearities at the interfaces. Energy flow is affected by these interfaces.

A statistical response surface metamodel was constructed using polynominals from the finite element model results for a reduced parameter space of six dimensions. This model was used for test – analysis comparisons in lieu of running more finite element models. An error metric was defined as the distance from each experimental data point to the response surface. The norm was taken over all four experiments as the measure of matching between the data and the model results. There is additional modeling uncertainty involved by substituting the response surface model for the finite element results from the estimation of the surface itself. In this variance-based uncertainty analysis, the RMSE from the response model fitting was used as an indicator of the uncertainty between the estimated surface and the FE models.

With the results and lessons learned from Phase I, the planned Phase II study is designed to provide:

- Updates to the variable and parameter screening were already done using information from the friction and preload testing.
- A finer sampling scheme of finite element model runs to develop the response surfaces.
- Updates to the finite element model.
- Results from additional experiments, specifically with different loads and parameter values, i.e., enough experiments to better characterize replicates.
- Better estimates for sensitive parameters.
- A more complete validation.

Because one of the main goals of using models is prediction, Phase II will apply the model to complex loads that cannot easily be implemented in the laboratory environment.

3. Inference Uncertainty—Studying Poorly Understood Physics with Poor Testing

Inertial Confinement Fusion (ICF) programs at Los Alamos and other laboratories have provided the experimental opportunities to better understand and move towards the goal of controlled nuclear fusion. The decades old idea of controlled fusion was to produce commercially available energy that would be "too cheap to meter." However, huge uncertainties from lack of knowledge of the first principles physics have limited mankind's success in reaching this promise for cheap power. Part of this uncertainty also stems from what we will call *inference uncertainty*. Here inference uncertainties arise either (1) when quantities that can be measured are not the same as the quantities that can be calculated/modeled, or (2) when experimental regimes that can be tested do not match those in the real problem where experimentation is prohibited. An example of the latter (2) would be where ICF experiments are used to make inferences about uncontrolled fusion. An example of the former (1) would be to use the diagnostic measurements from ICF experiments to define the number of neutrons.

Before tackling this difficult problem in an application where the first principles are poorly understood, let us examine how some inference uncertainties have been more traditionally handled. Classical, statistical inference is characterized by variances. For example, when an experiment is performed, or an observation made, or a sample taken, one makes an inference about the entire population of (or set of all possible) tests, observations, or samples by what is found in that single test, observation or sample. This inference is possible if that one trial is representative of the population. The more homogeneous the population, the more likely a single representative from it will share the same characteristics of the whole. One commonly utilized way of guaranteeing a good representative choice is to make a random selection. If there is some inhomogeneity, then that resulting variation dictates the number of randomly chosen tests, observations or samples that one must examine in order to capture the variation of the underlying population. The entire field of statistical sampling is based upon this idea of inference, and the uncertainty associated with that inference process is variance-based.

However, even in this fundamental scenario of understanding a population from a limited amount of sampling, we can produce the more elusive kinds of inference uncertainties described above. Specifically, one may not be able to measure exactly the quantities or conditions of interest in the population. For example, the status of the US economy is "measured" by looking at indices such as the GNP, employment rate, etc. However, these indices are only related to the status of the economy and do not provide a direct measure of it. That relationship (inference) is not known from theory and can only be somewhat (uncertainly) characterized statistically using decades worth of data analysis. Many statistical variance-based techniques have been employed to model this relationship and account for the variance of the inference from the measured to the desired (unmeasured). These include areas of econometric modeling employing variance estimate techniques for the inference uncertainties involved. Instrumental variables, variances in predictor variables in statistical models, and Bayesian methods are a few such techniques. However, in this example the data rich environment permits inference uncertainties to be characterized as variabilities, even when the underlying theory (for model construction) is not well known.

In the fusion example, the theory is not well understood, so the modeling is difficult, and the experimental data is not numerous. Like with the threaded assembly application, time histories and integration of those are the quantities that can be measured in ICF experiments. Here experiments that are possible to run do not directly measure the quantities of interest. In addition, they are not necessarily in the desired parametric regimes (time, temperature, etc.). Given these *substitutions*, the uncertainty of inference may not be so conducive to variance-based methods. We are currently engaged in understanding the uncertainties of inference in the ICF test regime and then relate those to other physics regimes where the first principles theory is not well understood. How to best characterize the uncertainty from inferring one quantity by measuring another and the uncertainty of inferring test results in the regime of interest using tests performed in another regime is not well established. If the knowledge about these inferences is so poor that only qualitative statements can be made, then perhaps Zadeh fuzzy sets could be incorporated. Perhaps some quantification is known about these inferences but only in the non-specific form of intervals of possible values, indicating a random interval approach for uncertainty characterization. This is a matter of additional research; however at this point in time it appears that what little is known must be properly elicited from the experts using formal elicitation and analysis methods such as from Meyer and Booker (2001). Once that knowledge is extracted, the appropriate mathematical theory for uncertainty quantification should be employed to tackle this difficult problem of inference uncertainties. If the inference uncertainty can be expressed as a distribution function within any of the General Information Theories (GIT) for uncertainties, then that function can be added as an additional column to the SVD matrix for calculating TU.

IV. Future Challenges

We have noted several important issues regarding uncertainty quantification:

- Uncertainties are of different types on the data and model sides; most of which are ignored in the validation process. We have specifically here pointed out modeling choice uncertainties and inference uncertainties. Prediction uncertainty is another important one for validation.
- Uncertainties may not all necessarily be best characterized by probability theory. However, many GITs exist and can be utilized.
- When utilizing different GITs for uncertainties, the research problem becomes how to merge these uncertainties within a validation study. Zadeh's latest proposal is for a unified theory of uncertainty—a challenging task that we have long advocated (Zadeh, 2004). Work has already been done on merging fuzzy membership functions with probability (Singpurwalla and Booker, 2004).
- Real world problems are complex in structure, with multiple levels, variables, parameters, models, tests, requirements, inputs and responses. More research work is needed to handle all the uncertainties associated with these in an integrated manner.

New uncertainty quantification analysis tools and methods are constantly being developed. We applaud these efforts. The TU metric presented here, while only in its development stage, provides some promise for combining differently characterized uncertainties along with incorporating the comparison of test and models. However, we caution that the theoretical development is not yet completed.

ACKNOWLEDGEMENTS

The authors wish to thank Roger Logan (LLNL) for his review and helpful comments, and Scott Doebling, Acting ASC Program Manager and the Los Alamos ASC Program for supporting this work.

This work was performed under the auspices of Los Alamos National Laboratory, operated by the University of California, for the U.S. Department of Energy under contract W-7405-ENG-36.

REFERENCES

American Institute of Aeronautics and Astronautics (AIAA) (1998) “Guide for the Verification and Validation of Computational Fluid Dynamics Simulation,” G-077-1998.

American Society of Mechanical Engineers (ASME) (2004), “Guide for the Verification and Validation of Computational Solid Mechanics,” PTC 60, Committee on V&V in computational solid mechanics, Draft 2004-03-22.

Baroi, A.C., Reardon, B.J., Booker, J.M. (2004) “The Use of Singular Value Decomposition as a Test Statistic,” Los Alamos National Laboratory report, in progress.

Ben-Haim, Y. (2001) *Information-Gap Decision Theory*, Academic Press, San Diego.

Booker, J.M., Ross, T.J., Hemez, F.M., Anderson, M.C., Reardon, B.J., and Joslyn, C.A. (2003) “ Quantifying Total Uncertainty in a Validation Assessment Using Different Mathematical Theories,” *Proceedings for the 14th Biennial Nuclear Explosives Design Physics Conference* (NEDPC), Los Alamos, NM, October 20-24, 2003, Los Alamos National Laboratory report, LA-UR-03-9001.

Conover, W.J. (1971) *Practical Nonparametric Statistics*, Wiley & Sons, New York.

Doebling, S. W., Hemez, F. M., Robertson, A. N., Maupin, R. D., Schultze, J. F., Cundy, A. L., Hylock, J. E. (2004) “ Validation of the Transient Structural Response of a Threaded Assembly: Phase I,” LA-14105-MS, Los Alamos report.

Doebling, S. W., (2004) “Overview of LANL Simulation Validation Program,” presentation at ASC Alliance V&V Workshop, La Jolla, CA (July, 2004).

Dolin, R., Booker, J., Faust, C., Reardon, B., Hamada, M. (2002) “Principles of an Engineering Index: A Metric for Quantifying Margin Relative to Performance,” Los Alamos report LA-13934.

Donald, S. (2003) “Development of Empirical Possibility Distributions in Risk Analysis,” Ph.D. dissertation, University of New Mexico, Albuquerque, NM.

Hasselman, T.K., Yap, K.C., Wathugala, G.W., and Anderson, M.C. (2003) “Uncertainty Quantification and Prediction Accuracy Assessment for Nonlinear Models,” under review for journal publication.

Hemez, F.M. and Ben-Haim, Y. (2002) “Information-Gap Robustness for the Test-Analysis Correlation of a Nonlinear Transient Simulation,” Proceedings of the 9th AIAA/ISSMO Symposium on Multi-disciplinary Analysis and Optimization, Paper number AIAA-2002-5420.

Hemez, F.M., Ben-Haim, Y. (2004) “The Good, The Bad, and The Ugly of Predictive Science”, 4th International Conference on Sensitivity Analysis of Model Output (SAMO-04), March 8-11, 2004, Santa Fe, New Mexico.

LA-UR-03-8492.

Klema, V.C. and Laub, A. J. (1980) “The Singular Value Decomposition: Its Computation and Some Applications,” *IEEE Trans. On Automatic Control*, Vol. AC-25, 164-176.

Klir, G. and Wierman, M. “Uncertainty-based Information: Elements of Generalized Information Theory,” Physica-Verlag/Springer-Verlag, Heidelberg and New York, (1999).

Klir, G. J., and Yuan, B. (1995) *Fuzzy Sets and Fuzzy Logic: Theory and Applications,* Prentice Hall, New Jersey.

Meyer, M.A., Booker, J.M., Bement, T.R. (1999) "PREDICT—A New Approach to Product Development," 1999 R&D 100 Award winner, *R&D Magazine*, Vol. 41, 161.

Meyer, M. A. and Booker, J.M. (2001), *Eliciting and Analyzing Expert Judgment*, SIAM, Philadelphia.

Nitta, C. K. and Logan, R. W. (2004) "Qualitative and Quantitative Linkages from V&V to Adequacy, Certification, Risk, and Benefit / Cost Ratio," Lawrence Livermore National Laboratory, Livermore, CA, UCRL-TR-205809Draft for Foundations '04 workshop, Tempe, AZ, Oct 13-15.

Roache P.J. (1998a) "Verification of Codes and Calculations," *AIAA Journal*, v36, 696-702.

Roache P.J. (1998b) *Verification and Validation in Computational Science and Engineering,* Hermosa Publishers, Albuquerque, NM.

Ross, T.J., Booker, J.M., Hemez, F.M., Anderson, M.C., and Joslyn, C.A., (2003) "Use of Total Uncertainty in Understanding Safety Margin and Model Validation in Structural Mechanics", Los Alamos National Laboratory, Los Alamos, NM, in preparation for journal submission.

Trucano, T. G., Piltch, M., Oberkampf, W. L. (2002) "General Concepts for Experimental Validation of ASCI Code Applications," Sandia Report, SAND2002-0341, Sandia National Laboratories, Albuquerque, NM.

Shinn R., Hemez F., and Doebling, S. (2003) "Estimating the Error in Simulation Prediction over the Design Space", *Proc. 44th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conference*, Norfolk, Virginia, April 7-10.

Singpurwalla, N.D. and Booker, J.M (2004) "Membership Functions and the Probability of Fuzzy Sets," to appear *Journal of the American Statistical Association*, 99, September.

Webster's Third New International Dictionary (unabridged) (1986), P.B. Gove, editor, Merriam-Webster Inc. Springfield, MA

Zadeh. L.A. (2004) "Towards a Unified Theory of Uncertainty," Keynote lecture presented at the International Conference on Information Processing and Management of Uncertainty, Perugia, Italy, July4-7, 2004

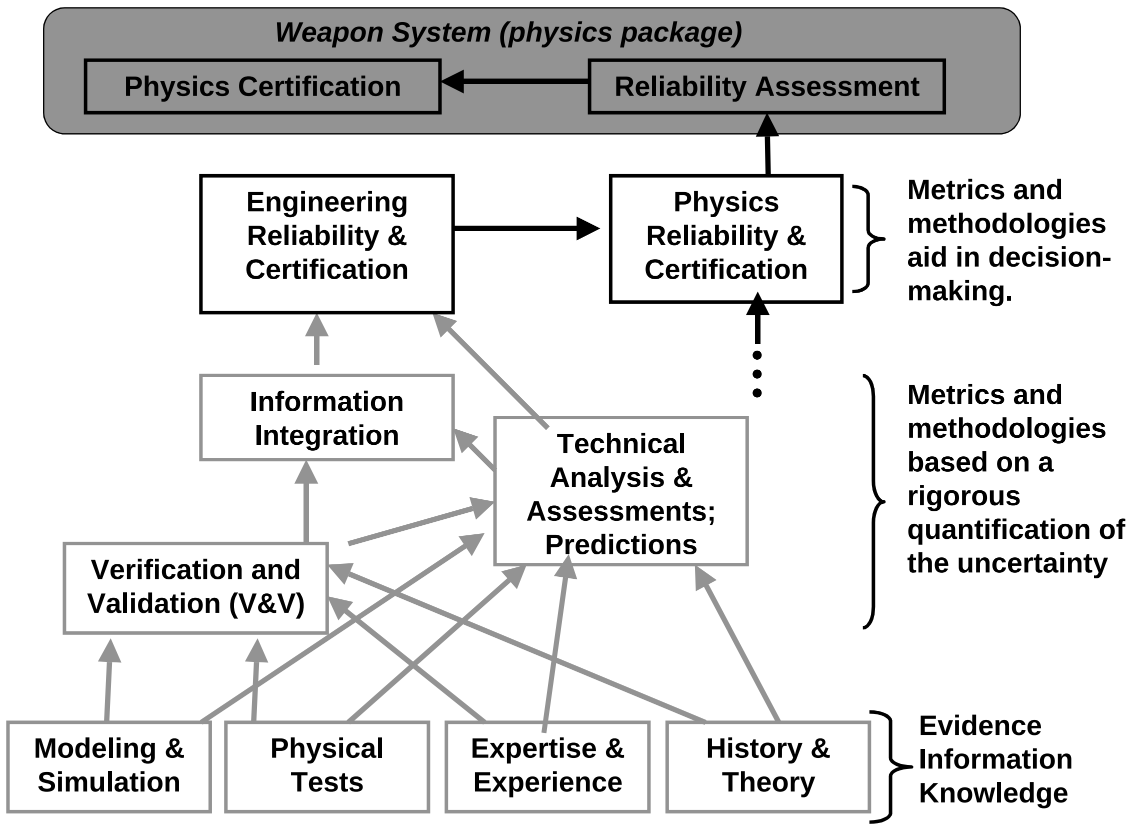
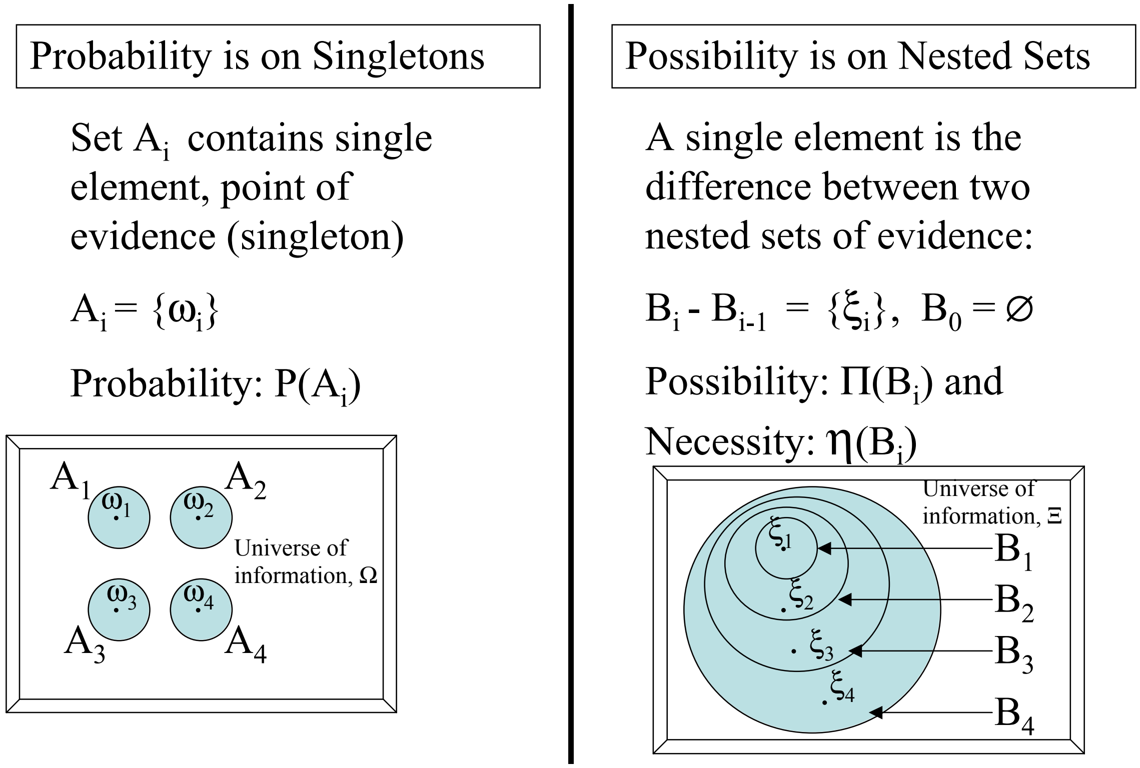


Figure 1. Los Alamos Weapon System UQ Activities



Cannot map evidence $A_2 = \{\omega_2\}$, a singleton, to evidence $B_2 = \{\xi_1, \xi_2\}$, nested space, using probability. ω_2 would either map on ξ_1, ξ_2, or both. This mapping requires necessity and possibility to accommodate the choices.

Figure 2. Singleton Versus Nested Sets: Probability Versus Possibility

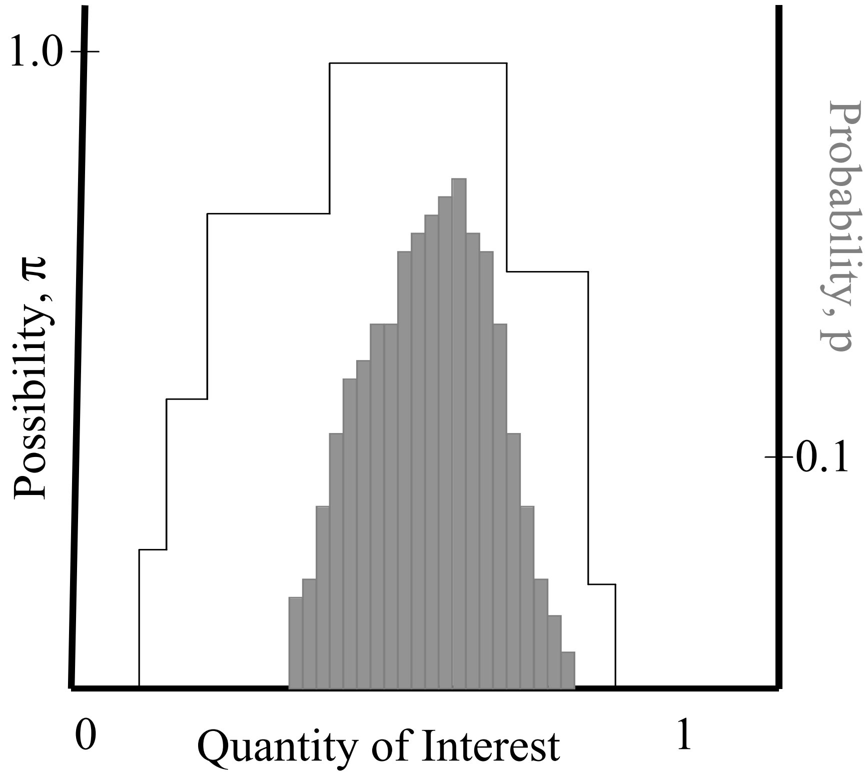
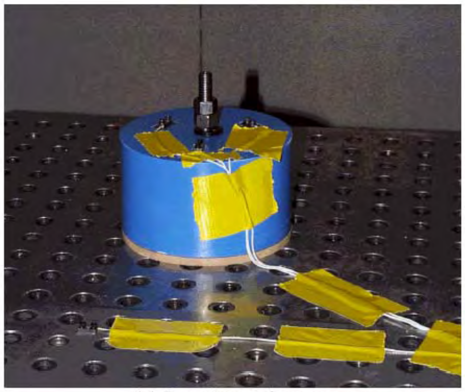
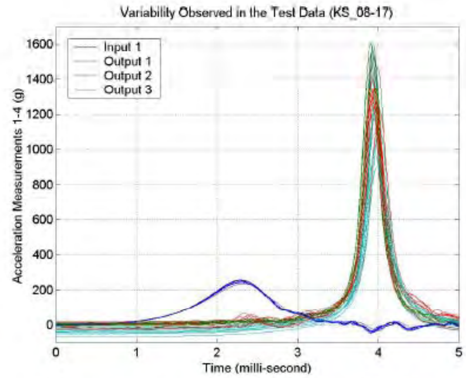


Figure 3. Probability histogram (grey) and possibility distribution for a quantity of interest.



(a) Experimental set-up.



(b) Signals measured during 10 replicate tests.

Figure 4. Shock wave through hyper-elastic foam, (a) the experimental set-up and (b) the input and output signals for 10 tests.

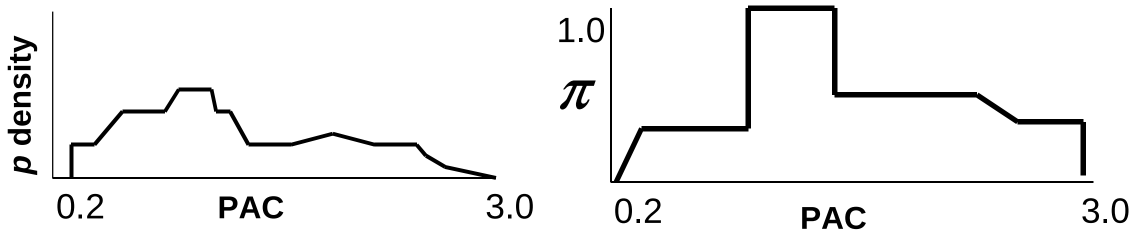


Figure 5. PDF for the tests and π-DF for the models (PAC is in 10^3 g's).