SAE International
SURFACE VEHICLE STANDARD
J2940 NOV2011
Issued 2011-11

Use of Model Verification and Validation in Product Reliability and Confidence Assessments

RATIONALE

SAE has numerous standards relating to the use of models[65-67,70,71], and product reliability[60-69]. Other professional organizations (AIAA[1], ASME[5,6], DoD[50], NASA[58], etc) have recent standards for Model Verification & Validation (V&V). Lacking, however, is a standard relating the increasing use of numerical and computer Model V&V to quantitative design assessments of product reliability. It is the intent of SAE J2940 to provide such a standard.

FOREWORD

Model-based product reliability assessments are usually made with limited component or system test data, and are heavily reliant on results of the model. As such, the value of the model's assessment of product reliability is always in question. The purpose of Model V&V is to add confidence, both qualitative and quantitative, to these assessments. At this point we note the following:

1. It is not rigorous to have *qualitative* confidence in a model-based product reliability assessment unless the same model can produce a *quantitative* assessment of Confidence C, i.e., a number such that $0 \le C \le 1$.

2. The most modern standards for Model V&V[6] stress the need to quantify Confidence C, $0 \le C \le 1$, during the V&V process.

3. All that is, therefore, lacking is a standard describing accepted practice and methods to link Confidence C, from Model V&V, with quantified model assessed product Reliability R, where $0 \le R \le 1$ is obtained from a combination of product requirements, relevant experimental testing, and use of the Model that has undergone V&V for quantitative Confidence C.

4. This standard (SAE J2940) prescribes that a when quantitative assessment of Reliability R, $0 \le R \le 1$ is based on a model claimed to be *validated*[1,5,6,39,50,58], the assessment of Reliability R should be accompanied by a quantitative assessment of Confidence C, such that $0 \le C \le 1$.

The message and requirement of this standard is simple: To assert compliance with this SAE J2940 standard, the model-based reliability assessment must:

1. Undergo quantitative Model V&V per procedures consistent with those listed in the V&V Reference 4.1 of this standard.

2. Contain statements as the output of the V&V process that quantify confidence, such as[49] “I am 90% confident that if I build and test the product, the results will be on the model’s predicted (point or line or region) ±12%”, etc.

3. During subsequent assessments of product reliability with the model, each Reliability R, 0≤R≤1, at the component or system level, shall be paired with a quantitative Confidence C, 0≤C≤1.

The following outlines methods that can be used to meet this standard. The referenced methods are all peer reviewed and published in the literature or as standards. They provide a graded implementation from very simple to very complex, depending on the organization’s assessment of the risk associated with a given reliability assessment. The list is not all inclusive and may be expanded in the future. The list intentionally covers a wide range of possible methods and implementations so as to remain flexible enough to accommodate various risk levels and organizational cultures.

1. SCOPE

This SAE standard outlines the steps and known accepted methodologies and standards for linking Model V&V with model based product reliability assessments. The standard’s main emphasis is that quantified values for Model-based product reliability must be accompanied by a quantified confidence value if the users of the model wish to claim use of a “Verified and Validated” model, and if they wish to further link into business and investment decisions that are informed by quantitative second-order risk and benefit cost considerations.

1.1 Purpose

The purpose of this standard is to provide a process for linking demonstrated and accepted standards for Model V&V with peer reviewed and published methods and standards for model based assessments of product reliability. The result will be a standard that enables a design process providing validated model based assessments of product reliability at a quantified confidence level. This will justify and enhance the success of investments and business decisions that are based on model assessments of product reliability.

1.2 Applicability and Intended Use

The intended use of this standard is in design and assessment situations where the reliability of a physical product is being assessed using a combination of a numerical model and relevant physical test data. This is not the same as either (a) reliability assessment of a product via “go versus no-go” testing, or (b) reliability assessment of software through algorithmic computer simulation. In the situation for which this standard was written, there is usually limited test data, and therefore, it is a fair consensus that under such constraints, a numerical value for assessed reliability has little credibility without an accompanying numerical value for assessed confidence. It is the goal of this standard to provide the basic essential steps toward such an assessment – use of verified and validated models to quantitatively assess reliability at a specified quantitative confidence. To enable implementation, the standard references numerous published and peer reviewed procedures that can meet the specifications of this standard.

1.3 Tailoring

This standard does not specify or mandate a particular numerical procedure to meet its requirements. Instead, this document provides a list (which may evolve in the future) of methods for both Model V&V and Product Reliability (at confidence) assessment that can be tailored to the risk and business considerations of a particular design or product.

A complete but simple procedure is given in Appendix A and B. Expanded procedures are discussed in 5.1.

2. REFERENCES

There are two procedural components needed to satisfy this standard (V&V and R at C). The following references contain procedures that vary in complexity, enabling implementation of each component on a graded scale.

2.1 Procedures and Issues for Quantitative Model V&V

(See procedure in Appendix A and B and Appendix C, General References, especially those noted below and in 5.1)

The most recent consensus documents on Model V&V include the following guides and standards:

[1]AIAA, 1998): AIAA, "Guide for the Verification and Validation of Computational Fluid Dynamics Simulations", AIAA-G-077-1998, Reston, VA, 1998.

[5]ASME, 2006): ASME, "Guide for Verification and Validation in Computational Structural Mechanics", ANSI Standard, ASME V&V-10, Dec 2006.

[6]ASME, 2009): ASME, "Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer", ANSI Standard, ASME V&V-20, Dec 2009.

[50](DoD, 2009): DoD, "DoD Modeling & Simulation (M&S) Verification, Validation, and Accreditation (VV&A)", DoD Instruction 5000.61, Dec 2009. http://www.dtic.mil/whs/directives/corres/pdf/500061p.pdf

[58](NASA, 2008): NASA, "Standard for Modeling & Simulation", NASA-STD-7009, Jul 2008. http://standards.nasa.gov/documents/detail/3315599

As a specific first-level implementation example, Appendix A contains a copy of the major sections of SAE 2009-01-0569[56,57] and provides a step by step example of an extension of the procedure in ASME V&V-20[6], compatible with the principles of the other recent V&V guides and standards. This is done such that interpolation or extrapolation (with appropriate caveats) can be done to the real-world application domain and then linked to model assessments of product reliability at confidence. Appendix B contains additional comments and clarification regarding model form uncertainty.

2.2 Procedures and Issues for Model-Based Assessment of Product Reliability at Confidence

(See procedure in Appendix A and B, and Appendix C, General References, especially those noted in 5.2)

There is no standard implementation of a linkage between Model V&V and assessment of Reliability at Confidence.

The first level procedure in Appendix A is offered as the simplest of peer reviewed and published starting points. This implementation and example may be coupled in series or parallel using methods such as discussed by (Epstein and Liao, 2005)[51] or (SAE, 2010c)[73] for a complete hierarchical system level procedure. A simple procedure such as Appendix A is recommended as a first but not necessarily sufficient step. Its greatest value in a high risk scenario is to provide a timely roadmap so the simplifications, assumptions, and need for more elaborate methods are visible and compelling.

At that point, a shift to the procedures in the remaining references (see 5.2) may (or may not) be necessary depending on the balance of risk, benefit, cost, and schedule. Some of these references offer partial and complete procedures and examples for model assessed reliability at confidence. In some cases (some Bayesian formulations), the term "Credibility" is used instead of "Confidence", but still on a scale such that Credibility=C, $0 \le C \le 1$, and the values of Credibility C trade off directly with values of Reliability R.

2.3 General References

The General References appear in Appendix C, and provide a supplemental list of relevant references on the topics of Model V&V and model based assessment of reliability (with confidence specified or unspecified).

3. DEFINITIONS

3.1 This standard is consistent with the definitions given in SAE JA1000[61], and the V&V standards from AIAA[1], ASME[5,6], DoD[50], and NASA[58], with clarifications and additions per below:

Confidence – A number C, 0≤C≤1 or 0%≤C≤100%, expressing the credibility of the model to make assessments and predictions of product reliability. A model can be used with high confidence (e.g., C=95% or C=99%) or point estimate confidence (i.e., low confidence or C=50% 1-tail or C=0% 2-tail) depending on the risks in the subsequent business and safety decisions. However, it must always be possible to assess and report a value of 0≤C≤1 from the Model V&V process. This value of C is paired with the model assessed value of product reliability R.

Reliability - A number R, 0≤R≤1 or 0%≤R≤100%, expressing, (see SAE JA1000[61]) "The ability of a product to perform a required function, under stated conditions, for a stated period of time". Also, the "period of time" may involve a one-time, one-shot use of the product that may occur during any point in a given time interval (e.g., airbag deployment) or it may involve periodic or continual usage (e.g., gun barrel life or valve spring life). Reliability as discussed in this standard is assessed using numerical or computer models of the product that have undergone the Model V&V process so that Reliability can always be reported with an accompanying level of Confidence.

4. PROGRAM REQUIREMENTS

Regarding compliance, this standard is directly compatible with the procedures in SAE JA1000 (SAE 1998a)[61], and also (AIAA 1998)[1], (ASME 2006)[5], (ASME 2009)[6], (DoD 2009)[50], and (NASA 2008)[58].

The new requirement of this standard is that assessments conforming to this standard shall:

4.1 Undergo Model V&V compatible with peer reviewed and published V&V standards, including at this time:

a. (AIAA, 1998)[1]

b. (ASME, 2006)[5]

c. (ASME, 2009)[6]

d. (DoD, 2009)[50]

e. (NASA, 2008)[58]

f. Other V&V Standards pending availability and compatibility

4.2 Provide a quantitative confidence 0≤C≤1 for model simulations used in the assessment of product reliability.

4.3 Provide quantified model based assessments of product reliability R, 0≤R≤1, where each R is paired with an assessed value of Confidence C, 0≤C≤1, obtained from Model V&V. The method used shall simplify to and form a continuum with a large sample formulation in the presence of sufficient data, even though in most cases this amount of data will never be obtained.

5. PROGRAM ELEMENTS

5.1 Model V&V

Model V&V shall involve the assessment and quantification of four general types of uncertainty terms. In (ASME, 2009)[6], these can be stated as standard uncertainties without specification of a confidence level 0≤C≤1. However, in this standard, it is a requirement to quantify confidence, since this confidence will be linked to assessed product reliability. Therefore, the four types of uncertainty terms must be assessed as a function of Confidence C, 0≤C≤1. Following is an outline of the four types of terms of uncertainty in the model output quantity of interest to be assessed, along with relevant references to aid with the assessment.

5.1.1 $u_{D:}$ Data Uncertainty

Obtain and assess relevant experimental data[4,6,7,11,39,71,72]

5.1.2 $u_{N:}$ Numerical Uncertainty

Assess Model Verification, to include software reliability, Code verification, and solution verification (grid refinement convergence, etc.) [6,11,14,27,28,39]

5.1.3 $u_{P:}$ Parametric Uncertainty:[5,6,22,24,72]

Assess the degree to which parametric variations in the inputs result in uncertainty u_P in the output quantities of interest.

5.1.4 $u_{M:}$ Model Form uncertainty:[26,28,35,70]

Assess the degree to which model physics choices, possible ranges of "tuning knobs" or "fitting parameters", and choices of distribution forms cause model form uncertainty u_M.in the output quantities of interest.

5.2 Model-Based Product Reliability at Confidence Assessment

Model based product reliability at confidence assessment can take one or more steps as follows:

5.2.1 Component Level

Appendix A-B provides a *recommended but not mandatory* first level procedure. More complex procedures are given in the Appendix C References noted here[8,10,16,17,18,28,30,33,35,38,41,66,48,72,73].

5.2.2 System Level

As a first step, procedure in Appendix A and B may be extended to the system level with methods such as (Epstein and Liao, 2005)[51] or procedures in (SAE 2010c)[73]. Success will depend on judgment and proper choice of independent inputs and minimization of cross-term interactions. More complex procedures can help mitigate these caveats but are less transparent and more resource intensive. These begin with methods for reliability[55,72-73,75-78], and then expand to include procedures for reliability at confidence. The references provide a range of such procedures[19,20,55,31,59,66,70-74].

5.3 Documentation and Reporting

5.3.1 Will document Model V&V per (SAE, 1998a)[61], (AIAA, 1998)[1], (ASME, 2006)[5], (ASME, 2009)[6], (DoD, 2009)[50], or (NASA, 2008)[58].

5.3.2 Will supply results in a form so that paired values of R and C and the component and system level can be used in subsequent second-order quantitative risk and benefit/cost assessments.

6. NOTES

6.1 Marginal Indicia

A change bar (l) located in the left margin is for the convenience of the user in locating areas where technical revisions, not editorial changes, have been made to the previous issue of this document. An (R) symbol to the left of the document title indicates a complete revision of the document, including technical revisions. Change bars and (R) are not used in original publications, nor in documents that contain editorial changes only.

APPENDIX A - A BEGINNING RECIPE LINKING V&V AND RELIABILITY AT CONFIDENCE

Simple, First-Level Procedures for Quantitative Model V&V Coupled with Reliability at Confidence.

In order to satisfy SAE J2940, any model claiming V&V shall report its V&V as discussed in the text of J2940, and shall also assess Reliability and Confidence as paired quantities in any subsequent Reliability analysis using that validated model. The example procedure below in Appendix A and B is a *recommended but not mandatory* first-level assessment process to meet J2940. See also more complex procedures referenced in Section 5.2, and see Appendix C, General References. These more complex procedures may be used in lieu of or together with the example procedures of Appendix A and B.

The text of Appendix A is taken verbatim from portions of SAE 2009-01-0569 (later in SAE International Journal of Materials and Manufacturing)[56], so as to preserve its peer-reviewed and published content without changes. It is a discussion of topics presented at the Panel Session "Evaluation of Studies on Non-Deterministic Approaches (NDA) for Complex Systems" at the 2007 SAE World Congress, and is presented as a first level example of the entire process from Model V&V through assessment of Reliability at Confidence. The caveats are discussed during the example along with suggestions to consider more complex procedures such as those referenced in Section 5.2 for the Reliability at Confidence assessment.

A.1 SUMMARY OF MAJOR POINTS

- Credible integration of models and design requires quantitative Model V&V.
- Quantitative Model Verification & Validation must quantify terms of four types as discussed below.
- Quantitative Model V&V couples to the design to give bounds on reliability, i.e. the model assessed reliability as a function of confidence.
- The process described next is compatible with the newest V&V Standards, and couples readily into reliability based design.

A.2 SIMPLE BUT ILLUSTRATIVE PROCEDURE TO LINK MODEL V&V TO PRODUCT RELIABILITY ASSESSMENT

Following is a simplified step by step methodology for a quantified linkage between Models, V&V, and Reliability in design. Detail, depth, and alternatives are provided by reference to existing standards, guides, and reviewed, published literature in this area. The example is relatively easy to follow, and is compatible with SAE[42-44,66], AIAA[1], and ASME[4,5,6] guides and standards in this area. However, the quantitative path it illustrates from Model V&V to Reliability assessment has been greatly simplified for illustrative purposes.

Our goal will be to assess Reliability, using a combination of models (sometimes but not always finite element models) and data, with the data typically sparse and sometimes not of the pedigree we might desire. The challenge of cost is always present, so that while higher fidelity simulations and higher quality, more abundant validation-specific test data are always desirable, we must have a methodology that lets us say what we can say with limitations on our computational and test resources, and then make that methodology compatible with the case where we have all the computational and test resources we want.

The discussion that follows is simplified, but can be linked to more elaborate treatments of each topic already in existence. In other words, it is a procedure to go from a "potpourri of models and test data", to quantitative V&V, then on to a quantitative assessment of reliability at a specific confidence level.

A.2.1 Simple Formula for Reliability Assuming A Normal PDF

Consider (Fig. 1) a bolt with fixed load L and random variable Strength S, with mean μ_S.

Then the *Central* Margin

$$M = \mu_S - L \quad [1]$$

And, the *Central* Factor of Safety

$$FOS = \mu_S / L \quad [2]$$

Assume that L is known and fixed, and that the *mean* μ_S of Strength S is also known with 100% Confidence, and fixed, and that the random variable S has a known standard deviation σ_S.

If the distribution of S is normal, we assess the Reliability Index RI, or β, as:

$$\beta = (M/\sigma_S) \quad [3]$$

And we have Reliability R such that:

$$R = \Phi(\beta) \quad [4]$$

Table 1 shows the resulting 1-tail Reliability R vs. Reliability Index β:

TABLE 1- RELIABILITY INDEX β AND ASSESSED RELIABILITY R, NORMAL DISTRIBUTION

β	R
0.0	.5000
1.0	.8413
2.0	.9772
3.0	.9986

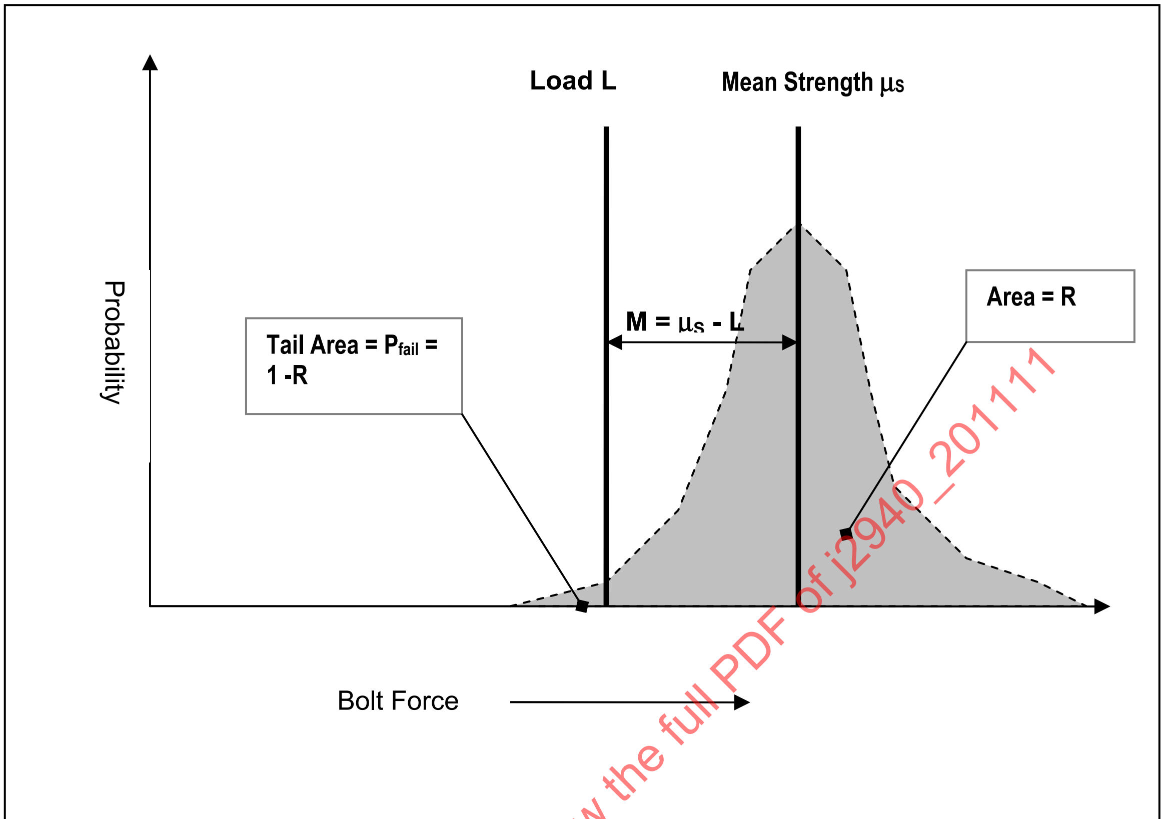

FIGURE 1 - QUANTITIES FOR SIMPLE RELIABILITY ASSESSMENT, NORMAL DISTRIBUTION.

A.2.2 Simple Formula for Reliability Assuming A Uniform PDF

Assume again that μ_S, the mean of S, is known at 100% Confidence and fixed, and that S has a standard deviation σ_S but with a *uniform* PDF (Fig. 2). Let the uniform PDF vary from a to b, where a is the lower limit and b is the upper limit.

Given this, the mean μ_S is equal to (a+b)/2 and the standard deviation σ_S is equal to (b-a)/(2[√3]).

Now, the distribution of strength S varies uniformly from $\mu_S - \sigma_S[\sqrt{3}]$ to $\mu_S + \sigma_S[\sqrt{3}]$.

Let

$$U = \sigma_S[\sqrt{3}] = (b-a)/2 \quad [5]$$

Then if we construct the Uniform Reliability Index, B_u, as the ratio of margin M (mean strength – mean load) normalized by half the range of the strength:

$$B_u = (M/U) \quad [6]$$

Note carefully that $B_u \neq \beta$. We now have 1-tailed Reliability R_u such that:

$$R = R_u = 0.5 + 0.5 * B_u \quad [7]$$

$$(0 \leq R \leq 1)$$

Table 2 shows the resulting 1-tail Reliability R_u vs. (Uniform) Reliability Index B_u:

TABLE 2 - RELIABILITY INDEX B_U AND ASSESSED RELIABILITY R, NORMAL DISTRIBUTION

B_u	R_u
0.0	.5000
0.25	.6250
0.50	.7500
1.00	1.000

The above is an oversimplified depiction of reliability assessment; however it will suffice in order to make our points in linking V&V with Reliability.

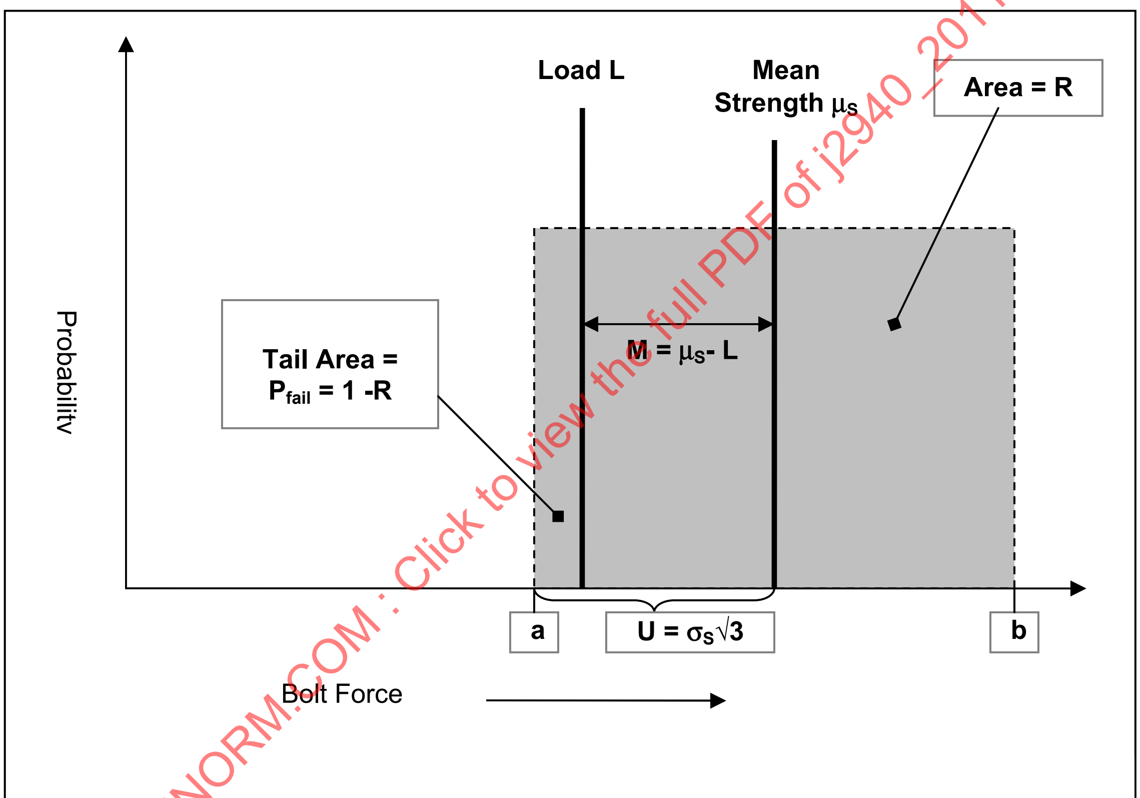

FIGURE 2 - QUANTITIES FOR SIMPLE RELIABILITY ASSESSMENT, UNIFORM DISTRIBUTION.

A.2.3 Details of the Model V&V to Reliability at Confidence Process

Consider the case above for the Normal Distribution:

$$\beta = (M/\sigma_S) \quad [3]$$

$$R = \Phi(\beta) \quad [4]$$

This looks simple enough, but we still must describe a procedure to obtain “M” and “σ_S”.

First, we will change notation from σ_S to σ_T, and denote this as the Standard (T=Total, from all contributors) Uncertainty of the Quantity Of Interest (QOI). The QOI is the focus of the Model V&V, and also the focus of our Reliability assessment equations above. We must now link our computer model results to these terms in Reliability, and we will do so using the products of Model Verification and Validation (V&V). This discussion borrows heavily from the language in the AIAA Guide for V&V[1], the ASME V&V-10 Guide for V&V in Computational Structural Mechanics[5], and the ASME V&V-20 Standard for V&V in Computational Fluid Dynamics and Computational Heat Transfer[6].
Where possible, we follow the ASME V&V-20 nomenclature for Standard Uncertainty (that is, all types of relevant uncertainty terms “i” expressed as an estimate of the standard uncertainty, u_i (not corrected for small sample size). Each of these u_i terms is comprised of a bias term b_i and a random uncertainty term s_i (the standard deviation). To obtain the standard uncertainty u_i, the terms b_i and s_i are added in quadrature. In practice, for design, we will later want to do two things:

1. “Dial out” the bias term b_i to calibrate our model to test data. Great caution is required here.
2. Assess our design at a numerical value of confidence, which may be different than the numerical value of confidence inherent in the standard uncertainty u_i (this is a function of the assessed or assumed distribution type).

Both [1] and [2] require special treatment of each u_i. For example, [1] “dialing out” a bias might mean we have added just that – a “dial” or “free parameter”, and we have to count that. On the other hand, if we want to use u_i at say “2 sigma” instead of “1 sigma” coverage (95% vs. 68% confidence for the normal distribution) we might want to multiply s_i by two but leave b_i unaltered. So we have to separate these terms back out b_i, multiply s_i by two, and then recombine. But then, it is subsequently still a good idea to multiply this new U_i by a small-sample correction factor as we will see below in Section 3.

In its simplest form, we must quantify the following four types of uncertainty terms during the Model V&V process:

u_D: The uncertainty in the experimental data and measurement of the output QOI[4,6,7].

u_N: The numerical uncertainty in the spatial, temporal, and iterative solution domains, obtained from Solution Verification[12,27,28,39].

u_P: The parametric uncertainty in the model solution output, due to known or estimated uncertainties in the input parameters (geometry, material properties, boundary and loading conditions, etc.) This is usually obtained via sensitivity coefficient analysis or Monte Carlo style sampling methods to get *model* sensitivity to each input and their combinations, multiplied by input distributions to get u_P. In the process of obtaining u_P we also get Importance Factors[5,6,22,56,26,67].

u_M: The model form uncertainty. This elusive[6] term contains all the remaining uncertainty not accounted for or underestimated in the above terms. Model form uncertainty may be due to missing physics, simplistic approximations, underestimating the width of the distribution of one of the other uncertainty terms, etc. It may be possible to assess u_M by Integral[26,59] or Top Down Validation[36], or by Cross-Validation[53] or assessment of Predictive Capability[28,37].

It seems legitimate by definition to view the total standard uncertainty as some combination of these four terms, since the last term, u_M, is a subjective “catch all” term called “model form”. The meaning and quantification of u_M in particular continue to challenge and frustrate the community, but some means of estimating u_M are at least possible if not universally accepted. A generalized combination of these four terms into a total standard uncertainty is given by:

$$\sigma_T = \sigma_T(\sigma_D, \sigma_N, \sigma_P, \sigma_M) \quad [8a]$$

In the ideal case where these four terms are separable and independent, we can express the total standard uncertainty in our model’s characterization of the reality of interest (e.g. the product performing its job) as σ_T, where:

$$\sigma_T^2 \approx \sigma_D^2 + \sigma_N^2 + \sigma_P^2 + \sigma_M^2 \quad [8b]$$

Or, as used without correcting for small sample size, in the ASME V&V-20 nomenclature:

$$u_T = u_T(u_D, u_N, u_P, u_M) \quad [8c]$$

And again, *in the ideal case* where these four terms are separable and independent,

$$u_T^2 \approx u_D^2 + u_N^2 + u_P^2 + u_M^2 \quad [8d]$$

In reality, the situation is rarely this simplistic, and even when it is this simple and independent, the Model Validation procedure to assess values for the terms in Eqn. [8] is complex. However, the forthcoming ASME V&V-20 document provides methods and relationships for assessing these terms whether or not they are independent or separable. These Standard Uncertainties may be obtained without making assumptions as to the form of their distribution (Normal, Uniform, etc.). It will be necessary to assess or assume forms for the distributions in order to make numerical assessments of confidence in reliability. For the moment, let’s assume the simple case, where the terms in Eqn. [8] are independent, and furthermore normally distributed, and develop the basic set of equations linking V&V with Reliability.

At this point, u_T is the term in the denominator of the Reliability Index β, so that

$$\beta = (M/u_T) \quad [3]$$

$$R = \Phi(\beta) \quad [4]$$

This treatment reduces conveniently to the case where, in a data-rich situation, the denominator contains only the product variability, that is, the term σ_D and an exactly correct assessment of the term σ_P. And if we assume for now that our model is exactly correct in its assessment of σ_P, and our assumption of normal distributions throughout is also correct, and if this is the one and only failure mode to be considered, we will obtain the correct Reliability of the product if the entire production run were to be tested in use.

It would appear that we are done. And, in the simplest case, where we neglect to assess the Confidence in our Reliability, and given the many other assumptions in this simplified situation, we are in fact done.

A.2.3.1 Assessing Confidence via Model V&V

Can we assume C=100% Confidence in our estimate of central margin $M = \mu_S - L$ and σ_T? We would be a bit foolish to assume that at this point. If we consider the realm of confidence to be expressed in terms equivalent to the statistics of probability, then our Confidence could range from $0\% < C < 100\%$. Our reliability so far contains no assessment of confidence, so although it is our best estimate of R and a very precise estimate, so far, it is a deterministic value, a point estimate of reliability. Whether the true reliability is above or below this assessed estimate is a matter of a coin flip. To simplify, this is the same as saying we have a one-tailed Confidence C_1=50% (a coin toss) or equivalently a two-tailed Confidence C_2=0%. That is, we estimate our reliability as a point value, but we know there is only an infinitesimal possibility it will take on that exact assessed value; even our model’s assessment only claims that the true reliability is either above or below our point estimate.

One of the Panel Discussion participants and later a reader of the draft of this narrative offered the following text that may be helpful in distinguishing Reliability, where we can assess that we approach C~100% Confidence, to the more common case where we know we do not have C=100% Confidence, but we may not know a way, let alone a unique way, to quantify the confidence that we do have:

> *"Since reliability is a probability, the frequentistic (number of occurrences of an event in n independent experiments divided by n as n tends to infinity) and subjective interpretations of the probability of an event (a decision maker's highest buying price of a lottery ticket that pays $1 if the event occurs and zero otherwise) may be helpful to the reader. Many reliability studies rely on the subjective interpretation to construct models of uncertainty."*

In complex NDA, we will probably never escape subjectivity. However, we can hope that subjectivity becomes synonymous with expert judgment. Our goal in this portion of the narrative is to construct a simple quantitative roadmap from Model V&V to Reliability. Hopefully it is a way to quantify the process by which the decision maker chose his or her highest buying price for the "lottery ticket". Although the assessment of confidence is often a *subjective* process, we feel it is important that assessment of confidence in reliability also be a ***quantitative*** albeit ***subjective*** process. Until we can find a way to quantify the confidence in our model assessed reliability, our procedure is missing a key component and remains open ended. Even our simplified procedure is going to take quite a few more steps to close this important loop from Model V&V to Reliability.

How do we assess our confidence in our estimate of central margin $M = \mu_S - L$ and $u_T \sim \sigma_T$, and what effect will this have on our point estimate of reliability? This is where the linkage between Model V&V and Reliability is essential. Without a validated model that quantifies its validation output at a specified confidence level, we cannot make this linkage.

First of all, we have to establish confidence in our estimate of each standard uncertainty term in eqn. [8d], $u_T^2 \approx u_D^2 + u_N^2 + u_P^2 + u_M^2$. What is our confidence in u_T? Usually, we have not measured the population standard deviation, of any of the terms u_i, let alone their possible covariance. Instead, we have typically measured a sample standard deviation s_i. We want a confident *estimate* of σ_i instead, at whatever confidence we wish to assess the reliability. This is ***not*** the same as the typical use of a coverage factor, where $2s_i$ =95%, $3s_i$=99.7%, etc. We still want an estimate of σ_i, the standard uncertainty (one standard deviation), but we want it at high confidence. Usually we do not have such high confidence because of a limited sample size. *As a beginning step*, Chi-Square correction factor may be used[17,28,40] to recognize that the true value of the population standard deviation σ_i maybe greater than or less than the sample standard deviation s_i that was computed from the limited test data. In a non-probabilistic treatment, the computation of sample and population standard deviation may not be so simple, and methods such as Evidence Theory[30] or Generalized Information Theory[33] may be used.

Next, we have to examine our confidence in the numerator terms, μ_S and L. Assume for the moment that we know L exactly, as a product requirement, but that μ_S comes from our model, like some of the terms in Eqn. [8] also come from our model. Just like those terms, the confidence in μ_S will depend on validation of the model we used to obtain μ_S. Without model validation, we can take μ_S as the deterministic value of the mean capability, but we can only assume C_1=50% or C_2=0% Confidence. Even if the reliability from Eqn. [4] is assessed at R=0.99 or R=0.9999, the fact that we assessed that at C_1=50% Confidence is not reassuring. If we want higher confidence in our estimate of the Mean Strength μ_S, we will have to use a lower value for it, shifted downward by the width of a Confidence Interval (CI) in the mean, shown in Fig. 3.

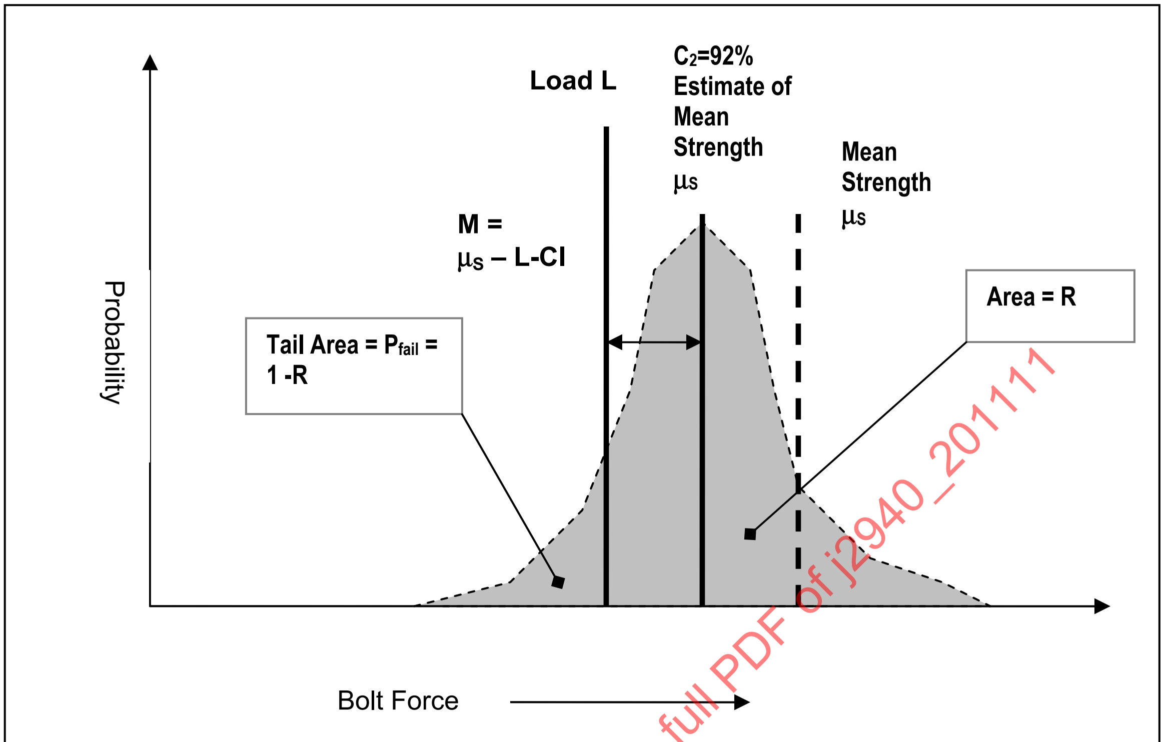

FIGURE 3 - QUANTITIES FOR ASSESSMENT OF RELIABILITY AT CONFIDENCE. A MORE PESSIMISTIC (BUT HIGHER CONFIDENCE) ESTIMATE OF THE MEAN STRENGTH IS USED, FOR EXAMPLE AT C_2=92% CONFIDENCE INSTEAD OF THE MEAN STRENGTH AT C_2=0% CONFIDENCE

A.2.3.2 V&V Links Model Confidence to Assessed Reliability

The reliability at confidence is now assessed as before in Eqn. [3] and [4], except that we will correct the standard deviation σ_T by an epistemic (e.g. small sample) correction factor, denoted "X_c". If we ignore the small sample correction, we would say that $X_c \equiv 1$. This is not the most optimistic assumption, since in reality, there is a chance that the true X_c is $X_c<1$; we do not know the proper value of the small sample correction factor X_c, we can only start with a simple assumption and follow up with more elaborate analyses. One simple assumption is to let the correction factor $X_c = X_\nu^{-1}$, the inverse square root of the reduced Chi-Square term[9]. This gives σ_T at an assessed level of Confidence C:

$$\sigma_T|_C = X_c u_T. \quad [9]$$

The Margin M is also assessed at this level of Confidence C:

$$M|_C = \mu_S|_C - L = \mu_S - L - CI \quad [10a]$$

$$FOS|_C = \mu_S|_C / L = (\mu_S - CI) / L \quad [10b]$$

In the simplest of statistical assessments, we may obtain an estimate of the CI as[21,28,29] :

$$CI = k t_c u_T \sqrt{0 + \frac{1}{N}} \quad [11]$$

In this case N=the number of experimental tests vs. simulations that are compared at the nominal values of input parameters (and therefore the nominal output) of the “Set Point”[6] for Validation. (For the time being, we will only discuss the linkage of the Model and V&V to Reliability when the Model is used at this single “Set Point”.) The quantity “k” is the (large sample) coverage factor[6], intended to provide a CI at a desired Confidence=%C. For normally distributed estimates of the mean, k=1 for a 68% 2-sided CI, k=1.96 for a 95% 2-sided CI, etc. The quantity t_c is the *small-sample correction* factor for estimating the mean, since we may usually assume that the distribution of our estimate of the mean tends toward a normal distribution. Here, Student’s $t = kt_c$. The first term under the radical is the number “0” [zero], to differentiate the CI Eqn. [11] from the PI Eqn. [15] below, which is identical except that a finite term will replace this “0”. Again, for more complex situations, and for the multivariate case, more elaborate formulations[10,31,35] are necessary, if a closed form expression for the Confidence Interval can be obtained at all.

In general, it is not so simple to obtain estimates of either σ_T or the confidence interval CI. However, without an infinite budget for experimental testing, these quantities will always be obtained using Model Validation if we hope to estimate a numerical value of confidence to accompany our estimate of reliability.

This is the key contribution (and the essential contribution) of the formulation herein; linking a validated model to a reliability assessment.

So now we have

$$\beta|_C = (\mu_S - L - CI) / (X_c u_T) \qquad [12]$$

$$R|_C = \Phi(\beta|_C) \qquad [13]$$

Eqn. [12] makes a linear assumption that both a detrimental value of scatter (the population standard deviation in the denominator) and a detrimental value of the mean will coincide. If these two effects are encountered independently, our formulation would look like:

$$\beta|_C = \sqrt{\left((\mu_S - L)^2 - CI^2\right)} / (X_c u_T) \qquad [14]$$

Eqn. [14] is not as easy to visualize in Fig. 3 as the linear subtraction used in Eqn. [12]. The linear subtraction of CI is more conservative, while the RMS version in Eqn. [14] reduces, in the simplest case, to the classic Prediction Interval, PI, if we set R=C and assume small sample corrections as above:

$$PI = k u_T \sqrt{X_c^2 + \frac{t_c^2}{N}} \qquad [15]$$

It has taken us 15 equations to describe the simplest of relations between Model V&V and Reliability. Some of the assumptions (e.g. independence of terms, use of t_c and X_v^{-1}) retain their rigor only for several restrictive assumptions including use of the normal distribution. However, it is often observed that t_c and X_v^{-1}, as used here, compare well to sampling method (Monte Carlo) analysis for the uniform distribution as well. Again, the recipe here is meant as a first pass and for illustrative purposes, to enable us to see each term and to link up Model V&V and Reliability in Design. More elaborate formulations are discussed elsewhere[10,17,33,35,41,44,48]. Furthermore, we have not yet even discussed how to quantify the uncertainty terms in even the simplest form of the Reliability Index denominator, that is, Eqn. [8]:

$$X_c u_T^2 = \sigma_T^2 = \sigma_D^2 + \sigma_N^2 + \sigma_P^2 + \sigma_M^2 \qquad [8c]$$

Model V&V will be essential in quantifying these terms, and therefore σ_T, as well, whether σ_T is obtained by the simple formula in Eqn. [8], or in the more common case where the terms are not independent, and in cases where σ_P may only be assessed via multiple samples of the model, e.g. Monte Carlo methods.

The simple equations shown above contain many assumptions, and in real design situations we are almost certain to violate many of them. However, this formulation is useful for showing in closed form the relation between the terms which must be obtained from Model Validation (e.g. σ_N, σ_P, σ_M, CI, and an assessment of reliability at a specified confidence (e.g. $\beta|_C$ and $R|_C$).

If any of these essential terms are omitted in a Model assessment of Reliability, we can either claim, at best, that R=50%, or that we might have high Reliability but our assessment is limited to Confidence C_2=0% or C_1=50%. The terms discussed here in these 15 equations are in fact, in many cases, less than a minimum set; but they provide an example to see the terms we must assess, the path to assess them, and to illustrate one definitive and complete path from Model V&V to Reliability in Design. This statement itself should prove sufficient incentive to perform quantitative model validation before claims of assessed reliability are obtained from the model.

A.2.4 Summary of Suggested Procedure: Linkage of Model V&V to Assessed Product Reliability

The following steps attempt to summarize the process just described:

1. From Model V&V, obtain and report the u_i -terms (σ_i terms) in Eqn. [8]. Generate *and validate*[5,6,28,38,53] the CI and PI for the quantities of interest as a function of the input variate[s]. The pedigree of any subsequent estimate of reliability and confidence can be no better than the pedigree of the CI and PI generated here.
2. From the design and requirements, obtain and report the mean FOS (at C_2=0% or C_1=50%) and the $FOS|_C$, the desired component confidence.
3. From these quantities, assess and report the mean R (at C_2=0% or C_1=50%) and the $R|_C$, the desired component confidence.
4. From these quantities, assess and report the mean *system reliability* R (at C_2=0% or C_1=50%) and the $R|_C$, the desired system confidence. The relation between component and system confidence must be considered carefully, as it is easy to make assumptions that are often overly conservative, e.g. that the component confidence equals the system confidence. Coupled system reliability analysis may be required if the components are not independent.
5. A starting suggestion is to start with the σ_i terms from Model V&V and equation 8, assume normal distributions, and proceed with the rest of the steps. Then, repeat the steps assuming uniform distributions, calculated from the same σ_i terms. Compare the difference in FOS and R that result. Does the distribution make a difference in your design assessment?
6. The next step, if the assumption of a normal vs. uniform distribution has a large impact on reliability, is to more formally assess the nature of alternate distributions. Is there statistical evidence for the use of a normal or uniform (or other) distribution? This issue can be addressed via:
   a. Statistical tests for distribution assumptions[2], etc.
   b. Interval analysis and Evidence Theory[33]
   c. Bayesian inference of the distributions[18,33]
   d. Etc.
   e. Repeat Steps 1-5 using these other methods to assess the distributions

A.2.4.1 Suggested Sources for Guidance:

There is as yet no *universal* procedure for linking computer models, V&V, and Reliability in Design. The recipe of this narrative is one offering, and can be successful if used as described, and by adding other terms as needed. For additional sources to accomplish this linkage of models, V&V, and Reliability, the following process is suggested:

1. The procedure in this narrative as a starting point.
2. The AIAA[1] and ASME V&V-10[5] Guides for the Model V&V process.
3. The ASME V&V-20[6] Standard for a more complex quantitative procedure for Model V&V. However, the ASME V&V-20 work modestly chooses to stop at the interface between V&V and Reliability in Design.
4. The SAE-AIR5080[42] and SAE-AIR5109[44] documents.
5. The other references listed in this narrative (and many others).
6. Professional judgment and expertise to fill in the missing pieces unique to each situation.

A.2.4.2 Suggested Quantities to be Reported:

For clarity, and for ease of comparison and dialogue between the “Factor Of Safety” (FOS) design culture and the “Reliability Based” design culture, it is suggested that the following in Table 3 be reported:

TABLE 3 - SUGGESTED QUANTITIES TO BE REPORTED IN A COMPLEX SYSTEM DESIGN WITH NDA

For each component: * FOS via the “central margin” i.e. mean (i.e. C_1=50% or C_2=0%). * FOS via the “nominal margin”, i.e. at selected confidence (e.g. C_1=99.86% or C_2=99.73%, etc.) * R at the mean (i.e. C_1=50% or C_2=0%). * R at selected confidence (e.g. C_1=97.7% or C_2=95% or other) For the full system: * Minimum** FOS via the “central margin” i.e. mean (i.e. C_1=50% or C_2=0%). * Minimum** FOS via the “nominal margin”, i.e. at selected confidence (e.g. C_1=99.86% or C_2=99.73%, etc.) * R at the mean (i.e. C_1=50% or C_2=0%). * R at selected confidence (e.g. C_1=84% or C_2=68% or other values you choose) *** For the full system, FOS has no meaning and can only refer to the “weakest link”, i.e. the component with the lowest FOS.*

If the Model’s quantitative validation statement has been carried through the reliability analysis in Eqns. [1-15] and more complex forms, no additional work is needed[41] to report all the quantities in Table 3. Since they are “free”, it is possible that reporting the entire set of quantities in Table 3 will reveal and clarify much about the assessment and role of confidence in the analysis, and allow an intuitive feel for FOS compared to R.

It is not the intent here to be overly prescriptive, even with a suggested procedure. However, our goal should be to generate the information in Table 3 or its equivalent, using an expansion of the procedure above or an equivalent, defensible, documented process. The remainder of the discussion may also provide elaboration and insight on the above “recipe” to link Models, V&V, and Reliability in an NDA.

A.3 ADDITIONAL CONSIDERATIONS

Following are some additional considerations to be addressed when dealing with NDA and linking Model V&V to assessed product reliability.

A.3.1 A Structured Approach and Conceptual Model

In the Conceptual Model phase, before computations are run or new validation tests performed, our approach to modeling the problem at hand is developed. The definition of Conceptual Model may be intuitively obvious, but it is worthwhile to read the AIAA V&V Guide[1], the ASME V&V-10 Guide[5], the ASME V&V-20 Standard[6], and others for further advice. Formal methods like QFD[34] (Quality Function Deployment), PIRT[5] (Phenomena Importance Ranking Table), and FMEA[43] (Failure Modes and Effects Analysis) can be of help in developing the items to be addressed in the Mathematical Model[1,5] (physics equations) and Computational Model[1,5] (e.g. finite element mesh) to follow.

We would strongly urge that the process from the model to V&V to Reliability be performed more than one way. This could be, for example, with a commercial reliability / optimization package (often a front end to a finite element code) backed up by simplified hand calculations. It is helpful to have bounding closed-form calculations to back up a finite element ensemble of calculations. Similarly, it is helpful to use more than one approach to the V&V of that ensemble of calculations. By the same token, it is helpful to back up a complex reliability methodology with other complex and even simple bounding reliability methods.

Next, we enter a discussion regarding identifying sources of possible error. Quantification of these terms will eventually allow us to assess the terms in Eqn. [8], and the possibility that they are correlated inputs, or that they have a correlated effect on the output quantity of interest QOI.

A.3.2 Sources of Non-Determinism

There are several professional societies, in addition to SAE, that are concerned with V&V of numerical simulations, and in the process of writing guidance documents and standards. The ASME V&V-10 Committee recently published its first Guide and ANSI Standard on V&V for Computational Structural Mechanics. In considering a follow-on document at the next level of detail, an ASME V&V-10 Subcommittee has formed and begun to address issues relevant to our SAE Panel charter under "Identify sources of potential error". A few of the relevant points are noted below.

A.3.2.1 Commingled Terms: u_D and u_P

Ideally, we would have a clear separation between the uncertainty terms for parametric inputs and output data, u_P and u_D in the nomenclature of Eqn. [8][6]. Validation tests would only contain u_D, and we would have to test and compute u_P explicitly. However, often the meager test data we can acquire has u_P and u_D commingled, e.g. if the test articles are chosen randomly from a production build and the material properties and geometry, etc. of that specific production unit are not measured. In the all too common case of data with a dubious pedigree, it is most conservative to assume that the variation in output observed on testing experimental assemblies accounts only for u_D, and does not cover the range of u_P. We risk double-counting the impact of u_P in this way, which may lower our subsequent reliability estimate.

A.3.2.2 Coping with Model Form Uncertainty u_M

We are usually faced with limited, imperfect data, especially at the system level. It is at least relatively easy to assess whether or not there is a shortage of data relative to the requirements for reliability and confidence. However, it is not so easy to assess some of the other factors, pertaining to model form, or even the experience of the modeler in using the tools available, which is also a type of model form. Later in the discussion, we will suggest that the surest way to assess model form is to assess the model's Predictive Capability. A "gold standard" test of Predictive Capability is the ability to capture, with a Prediction Interval (PI) built from the validated model, the proper percentage of subsequent new, independent experimental data that falls inside and outside that PI.

It is useful to perform a direct comparison of the integral output quantities from top level tests versus simulations at those same nominal test conditions. It may be that computing the bias error and random error from such a comparison gives us an adequate measure of the error E between the simulation output S and experimental data D (expressed here in ASME V&V-20 nomenclature[6]):

$$E = S - D \qquad [16]$$

Such a comparison may also provide an estimate of σ_T or equivalently u_T in Eqn. [8]. However, the individual terms in Eqn. [8] cannot be discerned from such a top level comparison. Furthermore, when comparing "I" (i=1,I) such pairs of Simulation and Data, if we assume that the standard deviation of these E_i is a good estimate of u_T, we may in fact be underestimating u_T, since we have no assurance that our comparison covers the range of u_N, u_P, and u_M in particular, regardless of the number "I" of such E_i comparisons.

With a sufficient number of experimental replications, we can claim that we have a good assessment of u_D. As our knowledge of u_D improves, an Integral Validation or Top Down Validation beginning with Eqn. [16] is a good first step, but it does not assure we can properly assess each term in u_D, u_N, u_P, u_M.

A.3.3 "How do we evaluate a complex reliability analysis result?"

Reliability assessment at the system level is complex enough when based only on tiers of test data. When the assessment is done using models for part or all of the assessment, the situation is even more complicated. One key part of the solution is that V&V is the link between Model and Reliability. A procedure similar to that depicted with Eqns. [1-15] above is necessary to realize this linkage, and we have already noted that Eqns. [1-15] represent the simplest case. Furthermore, a system reliability analysis will require consideration, whether with closed form approximation or via a sampling method, of how the overall System Confidence and System Reliability will be assessed.

A further related question must also be asked: “How do we relate the Reliability of the real product to the Reliability assessed from the FEA Model?” We might express the Reliability of a real car, for example, as the percent of a given type of car sold in a given year that will provide the customer with 100,000 miles of service with no unscheduled maintenance or breakdowns, for example. So we have now implicitly specified a set of environments that the structure and powertrain of the car must survive, with singular or possibly multiple occurrences. In any case, the requirement specifications for these environments reduce to a multitude of situations not unlike our simple example of the bolt with Strength S and applied Load L. With enough real world test data over the life of the entire fleet of this car type, we will eventually know the fleet “Reliability”, but even then, we will only know its reliability in terms of the ensemble of lifetime loads actually encountered, and these may be more or less severe than product requirement specifications. Even if we could guarantee that each car would be exposed to exactly the environments we specified, it would still do us no good until we learned whether each of the hundreds of thousands of that car model had survived the desired number of miles and years of service. This *post-mortem* reliability assessment is not our goal.

Instead, we want to use FEA (Finite Element Analysis), combined with just a few system tests, past experience and relevant tests, and many component tests, to enable our FEA model to estimate what the reliability of the fleet will have been when its expected lifetime is over. This FEA model-based *estimate* of reliability will be no better than the quantified V&V done on the model for the loading environments that may lead to failure.

The numerical procedure described above for Model V&V can then enable an assessment of design reliability at a given confidence level. The terms are those in Eqn. 8. For simplicity here, we *assume* independent σ terms so that:

$$\sigma_T = \mathrm{RMS}\,(\sigma_{Output.Data} + \sigma_{Solution.Verification} + \sigma_{Model.Parameters} + \sigma_{Model.Form}) \qquad [17]$$

$$\sigma_D = X_{cD}u_D = \sigma_{Output.Data}$$

$$\sigma_N = X_{cN}u_N = \sigma_{Solution.Verification}$$

$$\sigma_P = X_{cP}u_P = \sigma_{Model.Parameters}$$

$$\sigma_M = X_{cM}u_M = \sigma_{Model.Form}$$

These are the uncertainties in the output quantity of interest (QOI) that result from the inputs. In general we will have a correction factor X_c that is unique to each term, i.e. X_{cD}, X_{cN}, X_{cP}, and X_{cM}.

As discussed above when we introduced Eqns. [8a-8d], the RMS notation in Eqn. [17] assumes that each uncertainty contributor listed is independent of the others. It is a good *first step* to do this, but we must remember that we may underestimate *or overestimate* the value of σ_T due to that simplifying assumption. The ASME V&V-20 Standard contains a full mathematical description of how we can treat these terms when they are not independent. Furthermore, the parametric uncertainty term, σ_P, is usually made up of contributing terms from several parameters. Again, we can assume for simplicity that the components of σ_P are independent and so can be summed in quadrature, but we must eventually return and convince ourselves with evidence that it is reasonable to make that assumption.

In design, the choice of the mesh is not random and may well be completely fixed. However, we assess a bias and uncertainty (both systematic and random) likely to result from the choice of this mesh, and the uncertainty term is u_N. Most of the other *input* uncertainty terms are the various terms contributing to the *output* uncertainty u_P, such as geometry, assembly details, and material properties. These may have both a systematic and random component, and various ways of handling these terms are described in the ASME V&V-20 document[6].

The output uncertainty, σ_T or u_T, of Eqn. [17] becomes the distribution used for the *component* reliability assessment as shown in Fig. 1 and Fig. 3. If an integral assessment of the model is being done, then the *component becomes the system*, as described below in the “Hierarchical or Integral” discussion. Otherwise, system reliability must be assessed from a roll-up of component reliabilities[15,20,23,55,31,44,75]. The assessment of component confidence compared to system confidence will complicate this issue.

It is common that the largest uncertainty terms in Eqn. [8] are not the terms u_D or u_N, but the terms in u_P and u_M (often these two terms are difficult to separate as well). As example of this situation, we will consider the steel square section cantilever beam as shown in Figure 4.

The beam is end loaded at the right hand end, and a small length at the left hand end is held fixed. The beam is loaded and end deflection vs. load is measured through the onset and progression of plastic bending. We will model this process with a finite element model, and then validate by comparing the model against test data. We can then use the procedure just described to derive the model assessed reliability of this beam "component" at a given confidence level.

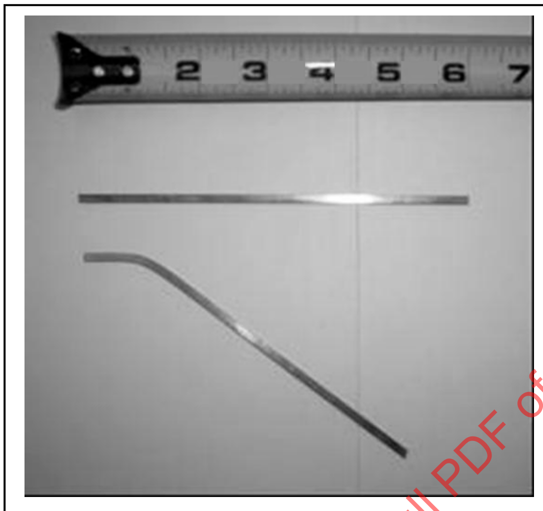

FIGURE 4 - V&V OF RESULTS: SQUARE SECTION BEAM IN PLASTIC BENDING TEST

At each successive load on the beam, contributions of each of the four uncertainty terms in Eqn. [8a-8d] are depicted in the graph in Figure 5. The load for each of the four uncertainty terms is offset slightly for visibility in Figure 5. The bars in the pictorial represent the standard uncertainty term in both directions (+/-), so no distribution form need be assumed at this point. The data term σ_D is depicted by the short "I" shaped bars in the pictorial, 3rd from left at each load level. The numerical term σ_N is depicted by the longer "I" shaped bars 2nd from left in the pictorial, as the standard uncertainty u_N term. The parametric term σ_P is depicted by the bars at the far left for each load step in the pictorial, about the same length as those for the u_N term. These "I" shaped bars at the far left in the pictorial represent the standard uncertainty u_P term. The longest vertical bars at the far right for each load step represent the largest term assessed, which is σ_M, the model form uncertainty term. The length of each "I" shaped bar represents +/- u_M, the standard uncertainty. The terms σ_D and σ_P were assessed using a procedure very similar to those covered in the ASME V&V-20 Standard[6]. Method 10 of (Logan and Nitta, 2006)[28] is used for σ_N; this method is a variant of the Least Squares Solution Verification method described in ASME V&V-20. The final term σ_M was assessed using the procedure described above.

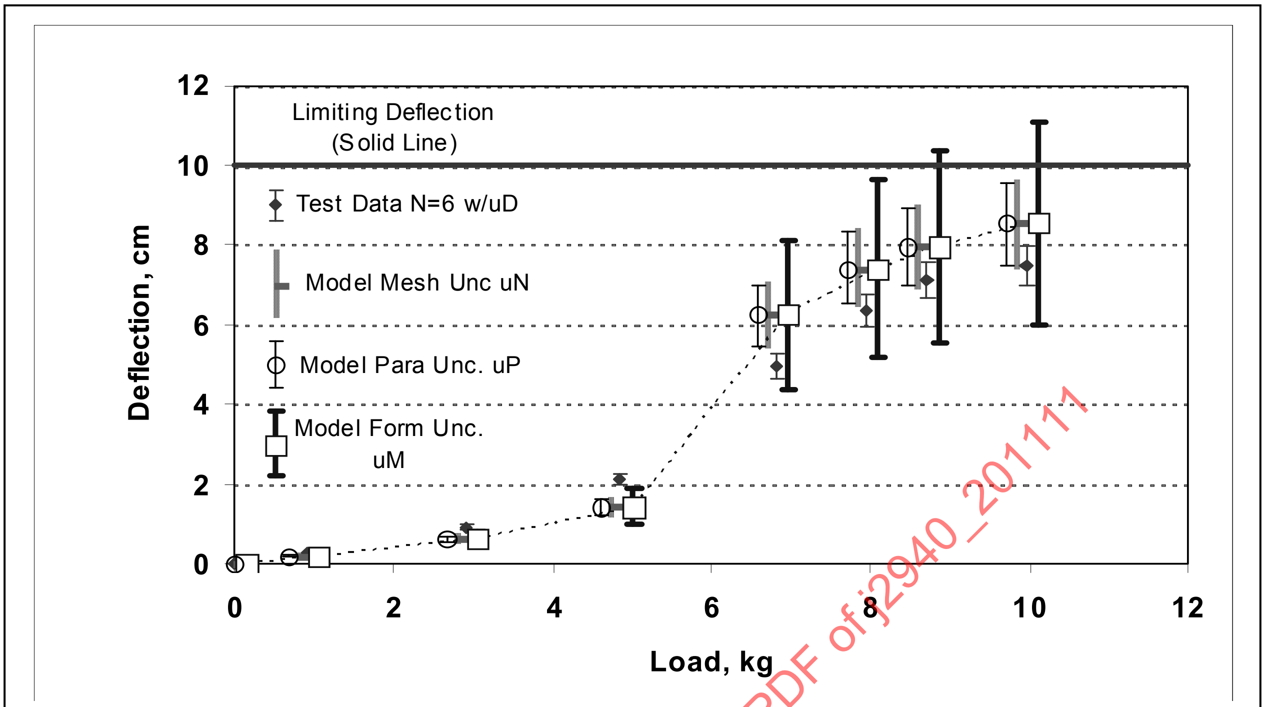

FIGURE 5 - V&V OF RESULTS: MODEL AND VALIDATION TEST DATA FOR PLASTIC DEFLECTION OF END LOADED BEAM SHOWN IN FIGURE 4. EACH STANDARD UNCERTAINTY TERM (U_D, U_N, U_P, U_M) IS SHOWN AT EACH LOAD STEP. MODEL FORM UNCERTAINTY DOMINATES DUE TO THE OVERLY SIMPLE PLASTICITY MODEL THAT WAS DELIBERATELY CHOSEN, BUT THE OTHER MODEL UNCERTAINTY TERMS ALSO DOMINATE THE TEST UNCERTAINTY

Due to the large values of σ_N, σ_P, and σ_M, relative to the data uncertainty σ_D, we would have a very hard time quantifying our *predictive* confidence in a high reliability or low probability of failure for the system in Figure 5, unless the requirement was very benign. However, Figure 5 depicts all the information we would assemble to complete Equations [8-9] above. Given a requirement, for example a maximum allowable beam deflection of 10cm as depicted in Fig. 5, we can then proceed through Equations [10-15] and assess reliability as a function of confidence. Even without a specific design requirement, we can generate confidence and prediction intervals to validate the predictive capability of our model against subsequent data.

To illustrate the process of going from Model V&V to reliability, consider a fixed value of Load=10kg in Fig. 5. A vertical section through Fig. 5 at this load can be depicted as shown in Figure 6, which is similar to the earlier Figure 3, except that the “Requirement Line” is now an upper limit instead of a lower limit. Figure 6 is then a simplified description of the shift of the mean corresponding to the chosen confidence level, and the reliability as the tail of the distribution beyond the requirement line.

Throughout this example, we have implicitly assumed that each of the uncertainty terms σ_D, σ_N, σ_P, σ_M has both an aleatory, or irreducible, component, as well as an epistemic, or reducible, component. We have further avoided separating these two terms, and so we have treated each term as having an epistemic component and possibly an aleatory component as well. For example, if we had enough experimental data at each tier, including the final complete system tier, then we could treat σ_D as a purely aleatory term (the test variability). If we ran each simulation “fully converged” in the spatial, temporal, and iterative domain, we could treat σ_N as an aleatory term. If we had plentiful data for each of the input uncertainties leading to the output σ_P, and we used an infinite number of Monte Carlo iterations to assess σ_P, we could claim that σ_P is aleatory only. We would continue to have to treat σ_M as an epistemic term, by definition. Under these conditions we could separate our validation, reliability, and confidence assessments into a two-step process where the aleatory terms are assessed by inner loop Monte Carlo iterations for example, and then the epistemic terms are assessed separately as an outer loop. It is rare that sufficient data exists to enable a clean separation, but when it does, formulations with an inner aleatory loop (loosely, for reliability) and an outer epistemic loop (loosely, for confidence) may be used[52,30,37,41].

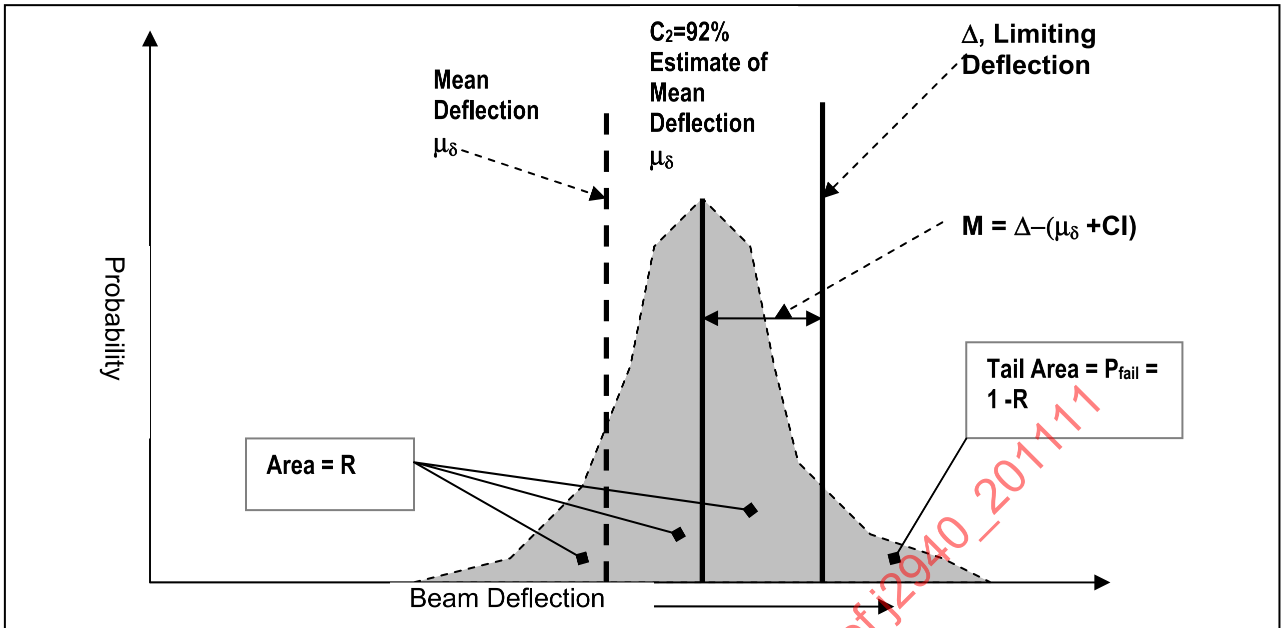

FIGURE 6 - USING VALIDATED MODEL TO ASSESS RELIABILITY; A SIMPLIFIED DEPICTION

In general, there is a limit to how much we can defend, via evidence, about the form of the distribution of the terms σ_D, σ_N, σ_P, σ_M. There is even less defense for our quantification of σ_M if we want to extrapolate out of the domain of our experimental data, or even away from a given experimental point within the domain. The more sparse our data, the less we can say about the form of the distributions, even if we have decent estimates of the values for σ_D, σ_N, σ_P. As a result, it will be very hard for us to estimate very high reliability or very low probability of failure, because in doing so we are making an assumption of the distribution form of these uncertainty terms. We could assume a uniform distribution, with half-width 1.732σ, and *assert* that our model (and real experimental) results will then never fall outside that bound. Our evidence for this assertion would be only a judgment call. We could equally assume a normal distribution, with an infinite "tail", so that we would always predict a non-zero probability of failure. We can *quantify* estimates of reliability and confidence at these extremes of high reliability or low probability of failure, but it is rare that we can avoid *subjectivity*. Sometimes our choices of assumed distributions (usually leading to a different assessed σ_M) will not affect our final design decision, and at other times we must realize that even though our validation and quantification of reliability and confidence might have a good basis inside the domain of our experimental data, we just lack the information to make credible assessments outside the domain of our experimental data.

A.3.4 System Level: Hierarchical or Integral?

As depicted in Figure 7, both the ASME V&V-10 Guide[5] and the AIAA V&V Guide[1] recognize that Model Validation has to account for all the subsystems comprising the product or complete system. If any top level data is available, it is also possible to supplement this tier-style "Hierarchical" or "Bottom-Up" validation with an "Integral"[26] or "Top-Down"[36] validation, using only the top level data. Integral Validation, using whatever top level data is available, is an essential step in the Model-to-V&V-to-Reliability process. However, Integral Validation *alone* is fraught with danger, as it usually requires a greater proportion of empirical "tuning knobs" or "free parameters". This is not bad in itself: For example, the linear regression process on which much of our material level validation is based typically has two "free parameters" when fitting a portion of material behavior to a straight line. A good example is when a Yield Strength S and Tangent Modulus E_T are supplied and used in our model. These quantities are obtained by nothing more than a linear regression fit to experimental data, using two free parameters, slope "m" and intercept "b" in the regression line y=mx+b. Furthermore, classical linear regression provides ranges for "m" and "b" as a function of the confidence level desired. The only difference is that for a material property input, Tangent Modulus E_T is the slope "m", and Yield Strength S is approximately the intercept "b". So we should never have the impression that there are "no free parameters", or "no fitted parameters" anywhere in the realm of V&V. These exist right down to the bottom tier such as the tension test. There is always some danger with "free" or "tunable" parameters floating around during V&V, because they must be accounted for differently than an input parameter such as the distribution of a material property or geometry. To take the extreme, you can always fit a straight line "y=mx+b" perfectly to N=2 data points (x_1,y_1) and (x_2,y_2), but there is zero confidence (or alternately, infinite assessed uncertainty) when this is done. Statistically, terms such as (N-K), where N=2 data points and K=2 free parameters (slope "m" and intercept "b") take care of this by generating N-K=2-2=0 in the denominator of the statistical expressions. However, it is not always obvious as to how we account for the values of "N" and "K" as it is in linear regression. Furthermore, the special danger of Integral Validation is not that there may be free parameters, but that the meaning and procedure to legitimately "tune" these parameters and still have a credible validated model is much more risky.

Therefore, it is recommended, and stressed in the ASME V&V-10 Guide, that Hierarchical Validation should always accompany Integral Validation. This is easy to say, but requires great time and expense. For example, if there is a point in any tier of the system hierarchy (e.g. Figure 7) where we cannot validate that box and assess the terms in Eqn. [8] at a specified confidence, then we are precluded from any assessment of reliability at the top level with any specified confidence. This means that Hierarchical Validation, while in principle more credible and rigorous than Integral Validation, may be impractical to achieve in time to affect the design process.

The tradeoffs between Hierarchical and Integral Validation carry over into the issues regarding reliability assessment for complex systems. In Chapter 15 of reference 33, Thoft-Christensen[33] gives a thorough procedure for system reliability assessment. Considerations for assessment of system reliability at confidence are provided in Blischke and Prabhakar Murthy[10]. These procedures are very complex and require complex computer programs in their own right. Some simplified methods are available, but in using them we must introduce at least as many restrictive assumptions as we already have in the V&V-Reliability-Confidence procedure just described for a single element and failure mode. These extensions will be the topic of future works.

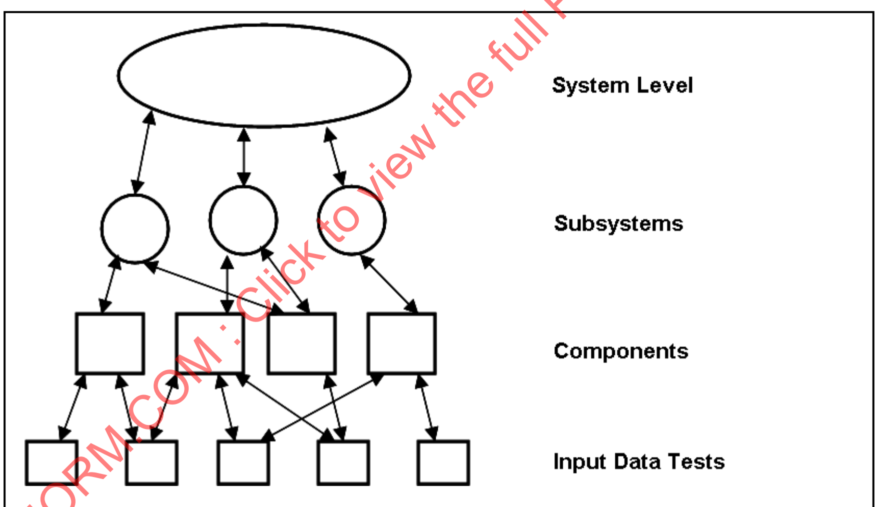

FIGURE 7 - V&V OF RESULTS: EVALUATION OF A COMPLEX RELIABILITY ANALYSIS. THE PROCEDURE HEREIN CAN BE USED AT THE INTEGRAL OR "TOP DOWN" COMPLETE SYSTEM LEVEL, AND ALSO AT THE HIERARCHICAL OR "BOTTOM UP" LEVEL AND IS COMPATIBLE WITH TRADITIONAL SYSTEM RELIABILITY ANALYSIS

A.3.5 Validation and Predictive Capability

We introduced the CI and PI above in the context of generating a process to go from V&V to reliability at confidence, and compared that expression to the simple PI expression that comes from linear regression. However, it is essential that we can produce a CI and PI as the outcome of our quantitative V&V process, for another reason.

Consider the schematic in Figure 8, depicting the idealized results of a univariate linear regression model based on a sample of test data from a large population. Though their many caveats must be remembered, the equations for the CI and PI for *linear* regression assumptions are well known; values for the CI and PI are given above in Eqn. [11] and [15] at the mean, and can even be generalized to approximate extrapolations away from the mean of the data[21,28,29] as:

$$CI = k u_T \sqrt{0 + \frac{t_c^2}{N} + \frac{t_c^2 (x_j - x_m)^2}{\Sigma_i (x_i - x_m)^2}} \quad [18]$$

$$PI = k u_T \sqrt{X_c^2 + \frac{t_c^2}{N} + \frac{t_c^2 (x_j - x_m)^2}{\Sigma_i (x_i - x_m)^2}} \quad [19]$$

In Eqn. [18-19], x_i denote the individual input values for each of the N experimental points, ideally equally spaced along the x-axis, and x_m is the mean of those input values. If we choose an input condition x_j, and conduct additional experiments at x_j, we expect the %C provided by the coverage factor k (e.g. 95% for k=1.96 and normal distribution) of these new experimental points to fall inside the PI, and 5% to fall outside.

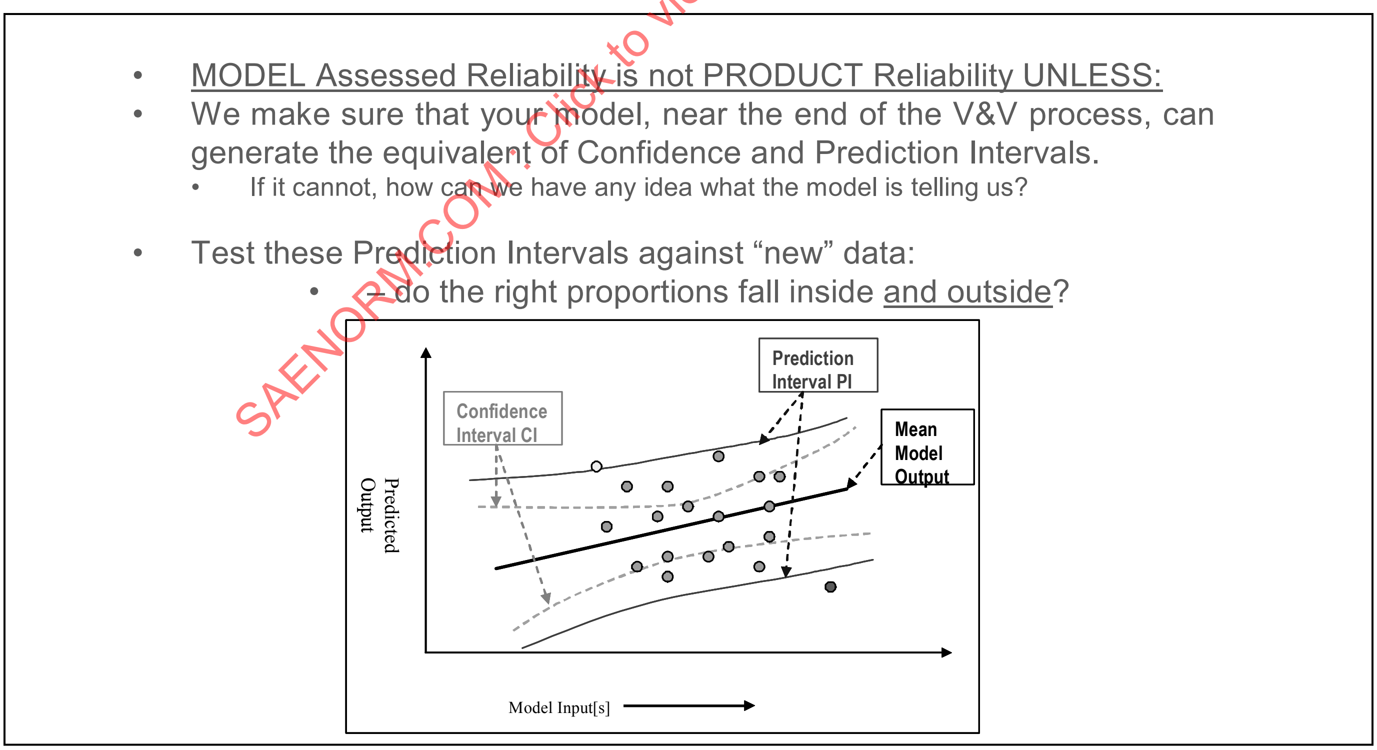

FIGURE 8 - PROCEDURE TO DEMONSTRATE AND ASSESS PREDICTIVE CAPABILITY, A TEST OF BOTH THE MODEL AND THE V&V PROCESS ITSELF. THE RESULT IS A "DATA-CENTRIC" FORM OF PEER REVIEW: THE NEW DATA SERVE AS "PEERS" IN THE REVIEW

It is often a good *start* to consider a model as linearized about the validation set point of most interest, and to use a process nearly as simple as the one shown and described with Eqn. [18] and [19]. These equations for extrapolating the CI and PI within or beyond the referent data used to validate the model contain even more assumptions than the CI and PI equations given just above. They are definitely simplistic, but they can be far more informative, even graphically, than just a line with no extrapolation at all. Eqns. [18] and [19] suggest the use of the Student's t correction t_c for the mean, and the reduced Chi-Square correction X_v^{-1} for the population. Although strictly speaking these correction factors are only appropriate for normal distributions, we find, by comparison to Monte Carlo, that they are good approximations for uniform distribution small sample corrections as well. *The extrapolation term in Eqn. [18] and [19] is outright dangerous, as noted in most statistical texts.* This is because we have no evidence that the assumptions appropriate to these equations hold beyond the referent data, or if there is a step change in the physics themselves just beyond the referent data. *But with proper caveats, even this simple formula is far better than an extrapolated line with no CI or PI at all.* Therefore, plotting out the CI and PI from Eqn. [18] and [19] is still an efficient and informative place to start. Then, think about how many assumptions might have been violated in doing so, and compare the resulting simple CI and PI to the CI and PI generated from a more complex V&V process. Whether we assume a normal distribution, a uniform distribution, or use Bayesian inference or Expert Elicitation to get our distributions, unless our V&V process can generate a CI and PI, e.g. a "95% Prediction Interval", or a "77% Prediction Interval" (or any percent you choose), we do not have a quantitative V&V process.

We hinted above that Eqn. [14] shows why this is necessary: A PI is just a special case of Reliability at Confidence where R≡C and the uncertainty in the population and the mean are assumed to be manifest independently of each other. If our V&V process does not result in the ability to generate a PI, we will not be able to use it to generate R at C either.

The final reason that it is not only essential that our V&V process contain the ability to generate a PI, but that we actually do so, is that this is perhaps the best way to validate the Predictive Capability of our model. As depicted in Figure 8, once we have generated a CI and PI, the next step is to compare "new" experimental data, or at the very least data that was not used or even considered during the V&V process, to the PI that resulted from our V&V process. "Success" does not mean that all the new data falls inside the PI. That would mean our PI is overly conservative, and we will end up wasting money on this portion of the design, and as a result weakening another part of the design due to cost tradeoffs. "Success" means that the correct proportions of the new, independent data fall inside and *outside* the PI. Of course, in addition to the common question of "what is new data", there is also the essential consideration of "what is new *independent* data?"

A.4 NOMENCLATURE

AIAA: American Institute of Aeronautics and Astronautics
ANSI: American National Standards Institute
ASME: American Society of Mechanical Engineers
a: Left-hand edge of a uniform distribution (e.g. Fig. 2)
b: Left-hand edge of a uniform distribution (e.g. Fig. 2)
b: Intercept in the linear regression y=mx+b
B_u: "Reliability Index" for the uniform distribution (see also β)
C: Confidence, 0%≤C≤100%
C_1: One-sided confidence
C_2: Two-sided confidence
CI: Confidence Interval
D: Experimental Data value of QOI
E: E=S-D, difference between Simulation and Data
FMEA: Failure Modes and Effects Analysis
FOS: Factor Of Safety (central or nominal)
k: Coverage factor for desired confidence, e.g. k=1.96 for 95% C_2 with normal distribution
K: Number of free or fitting or tuning parameters
L: Load (in the example for Fig. 1-2)
LRFD: Load Resistance Factor Design
M: Margin (central)
m: Slope in the linear regression y=mx+b
N: Number of "tests" – either experimental (u_D or u_M), grid (u_N), or parametric (u_P)
NDA: Non-Deterministic Analysis
P_{fail}: Probability of Failure, P_{fail} = 1 - R
PI: Prediction Interval
PIRT: Phenomena Importance Ranking Table