JOURNAL OF AEROSPACE COMPUTING, INFORMATION, AND COMMUNICATION
Vol. 3, July 2006

Comparing 10 Methods for Solution Verification, and Linking to Model Validation

Roger W. Logan* and Cynthia K. Nitta†
Lawrence Livermore National Laboratory, Livermore, CA 94551

Grid convergence is often assumed as a given during computational analyses involving finite difference and finite element analyses. In practice, perfect grid convergence is rarely achieved or assured, and this fact must be addressed to make statements about model verification and validation (V&V) or the use of models in risk analysis. When modeling a problem that is either discontinuous spatially (e.g. contact and impact) or discontinuous in physics (e.g. shocks, melting, etc) we may observe non-smooth or non-monotonic behavior of the output quantity as the grid is refined. This can lead to erroneous conclusions about the rate of convergence, and a lack of means to estimate residual uncertainty in the output quantity due to the use of non-converged grids. We compare ten techniques for grid convergence assessment, each formulated to enable a quantification of uncertainty estimates due to the use of non-converged grids, and rate of convergence for monotonic and non-monotonic grid convergence studies. The more rigorous of these methods require a minimum of four grids in a grid convergence study to quantify the grid convergence uncertainty. The methods supply the quantitative terms for solution verification error and uncertainty estimates needed for inclusion into subsequent model validation, confidence, and reliability analyses.

Nomenclature

AIAA	American Institute of Aeronautics and Astronautics
ANSI	American National Standards Institute
ASME	American Society of Mechanical Engineers
Δ\$B	Benefit, usually in \$\$\$
B	Bias error estimate (ASME/ANSI treatment as an uncertainty)
BCR	Benefit/Cost Ratio (Δ\$B-Δ\$C)/Δ\$C
Δ\$C	Cost, usually in \$\$\$
C	Confidence, a numerical value
CFD	Computational Fluid Dynamics
E	Bias error as percent of fine grid solution F_{fg}
F_{fg}	Quantity of Interest, discretized (computational) model solution, for finest grid
F_i	Quantity of Interest, discretized (computational) model solution, for i^{th} grid
F_i^*	Quantity of Interest, response surface estimate solution, for i^{th} grid

Received 28 October 2005; revision received 18 April 2006; accepted for publication 20 April 2006.

* Senior Staff Engineer, Engineering Associate Director's Office, L-151 PO Box 808, AIAA Member.
† Group Leader, B-Division, L-096 PO Box 808.

F_o	Quantity of Interest, exact solution
F_o^*	Quantity of Interest, response surface estimate of exact solution
FEM	Finite Element Method
FPI	Fast Probability Integration
GCI	Grid Convergence Index
h_i	Grid spacing for the ith grid
h_m	Grid spacing, mean of $I = 1$, N_g grids
h_o	Grid "spacing" at infinite refinement; $h_o = 0$.
h_1	Grid spacing for the finest grid
I	Beam moment of inertia in the beam deflection examples
K	Tuning Dials or "Knobs"; free parameters
L	Beam length in the beam examples
LLNL	Lawrence Livermore National Laboratory
LSQ	Least SQuares method
M	Margin, where Factor of Safety $= M + 1$
MV	Mean Value
N	Number of trials as in coin-flipping
N_g	Number of grids in a grid convergence study
N_s	Number of sets of grids
p	Convergence *Rate* Exponent in REE Equation $F_i = F_o + \alpha h_i^p$
p^*	Exponent in response surface approximation of REE Equation $F_{i^*} = F_{o^*} + \alpha^* h_i^{p^*}$
PDF	Probability Distribution Function
r^2	Pearson product moment correlation coefficient
r_h	Ratio of grid cell sizes along a given dimension
R	Reliability
R_{sm}	Reliability of a given Solution Verification Method
REE	Richardson Extrapolation Estimation
RMS	Root Mean Square
RSM	Response Surface Model
s	Sample standard deviation
S	Random Uncertainty Estimate (ASME/ANSI treatment)
SQA	Software Quality Assurance
U	Uncertainty, General or "System" [always at a confidence C]
U_s	Solution Verification RSM uncertainty of fit as percent of F_{fg}
u_s	Solution Verification RSM *standard* uncertainty of fit (at 1σ) as percent of F_{fg}
U_{sver}	Solution Verification uncertainty (combination of E and U_s), as percent of F_{fg}
$U_{sver}\|C$	Solution Verification uncertainty assessed at confidence percent or level "C"
V&V	Verification & Validation
w	uniform distributed load on the beam in the beam examples
x	Distance along the beam in the beam examples
y	deflection axis (normal to x axis) in the beam examples
Z	Standard Normal Distribution Variable for variable X, $Z = (X - \mu)/\sigma$
α	Constant in REE Equation $F_i = F_o + \alpha h_i^p$
α^*	Constant in response surface approximation of REE Equation $F_{i^*} = F_{o^*} + \alpha^* h_i^{p^*}$
$X^2_{\nu^*}$	Modified (biased) reduced Chi-Squared function
δ_{RE}	Richardson Extrapolation Error
ε	Richardson Extrapolation Error as a percent of F_{fg}
E	Young's Modulus in the beam examples
μ	population mean (or estimate)
σ	population standard deviation (or estimate)

I. Introduction

THE use of computational models is increasingly prevalent in product and process design, qualification, and certification, because computing costs continue to go down while testing costs go up. With this increasing reliance on computational models, a rigorous Verification and Validation (V&V) process is essential to assess and make clear the accuracy and predictive capability of such models. Since complex models are often involve a discretization of the fundamental equations of physics, the V&V process must include a means to assess errors and uncertainties due to insufficiently fine discretization. One such implementation of Verification and Validation (V&V) of a computational model of a physical system can be described simply as a 4-step "A-B-C-D' process.[1,2]

The first step [A] is the planning and requirements phase. This includes description of the physical system (product) and its requirements, and a plan accounting for available test data, what codes are available, and their status regarding software quality assurance (SQA) and Code Verification. A simple interpretation of Code Verification is that "(code) verification means solving the equations right", i.e. if one intends to give an answer to the equation "$2 + 2 =$", then one must run the resulting code to assure that the answer "4" results. In other words, the features and physics of the code that are relevant to the problem at hand should be tested and documented to assess the ability of the code to provide mathematically correct answers. These tests should be as comprehensive as possible. Obviously this becomes at least a quantitative value judgment, and often a qualitative value judgment, as to how much code verification and SQA is actually cost effective. Step [A] is the time to address deficiencies if Code Verification or SQA are deemed insufficient.

Second [B] is Solution Verification (the focus of this paper). Solution Verification is the process of assuring that a computational model[7] using discretization of a physical reality converges in each discretized spatial, temporal, and iterative domain to a converged answer on the quantity of subsequent validation interest. Ideally, assurance of a converged answer would also be assurance of a mathematically correct answer. We must assure ourselves that, with fine enough discretization and tight enough tolerances, our code is capable of obtaining the correct (i.e. $2 + 2 = 4$) answer to problems that are similar to our real-world problem but are amenable to closed-form solution; this is the realm of code verification. If we modeled with finer and finer meshes and converged to $2 + 2 = 5$, with a rate of convergence consistent with our numerical technique, we would recognize this as a *code* verification problem, not necessarily a *solution* verification problem. So the process for solution verification would be to build a model, using a discretization that would provide, in our best judgment, a high quality answer for the quantity of interest in the regime of application of the model. The discretization should be made bearing in mind that we will want to refine or coarsen the grid to assess grid convergence and quantify the estimates of solution bias error and uncertainty for any chosen grid size. We can then run our discretized continuum model over a wide range of coarse grids to fine grids in the asymptotic range,[8] and observe both a converged solution and, ideally, a constant *order* or *rate* of convergence.

We are often modeling a reality of interest that is *neither* continuous spatially (e.g. contact and impact) nor smooth in relevant physics (e.g. shocks, melting, etc). The typical result is a non-smooth or even non-monotonic convergence plot, typically a plot of output quantity versus grid spacing (an example is shown in Fig. 1). Non-smooth or non-monotonic convergence can lead to erroneous conclusions about the rate of convergence, and a lack of means to estimate residual error or uncertainty. We offer one emerging technique that provides a quantification of solution verification uncertainty at confidence ($U|C$, defined in Section II below) and rate of convergence for monotonic and non-monotonic mesh convergence studies. The observed rate of convergence is useful in at least two ways. First, it can sometimes be compared to the theoretical rate of convergence of the mathematical discretization and physics description. Second, the observed rate of convergence can be used to extrapolate to the solution at zero grid size. In the ideal solution verification case, monotonic behavior of the output quantity during refinement to an infinitely fine grid leads only to an error estimate term; a bias error. This bias error may be compensated during the parameter estimation process inherent to some V&V methods. However, even if the bias error is adjusted or fit to be zero using a free model parameter somewhere along the line, the value of the original bias should still be carried through as an additional[15] uncertainty term in validation and reliability assessment, and this is the method we describe in Section II below. Non-smooth or non-monotonic grid convergence adds an uncertainty of fit term (U_s, described in Section II below), in addition to our treatment of the bias error estimate as an uncertainty. We will assess the solution verification uncertainty term, $U = U_{sver}$, at a given level of statistical confidence, i.e. $C = 68\%$ or $C = 95\%$, etc. as described just below in this section. Unless otherwise stated, we will define $U|C$ where $C =$ 1-sigma or 68% confidence

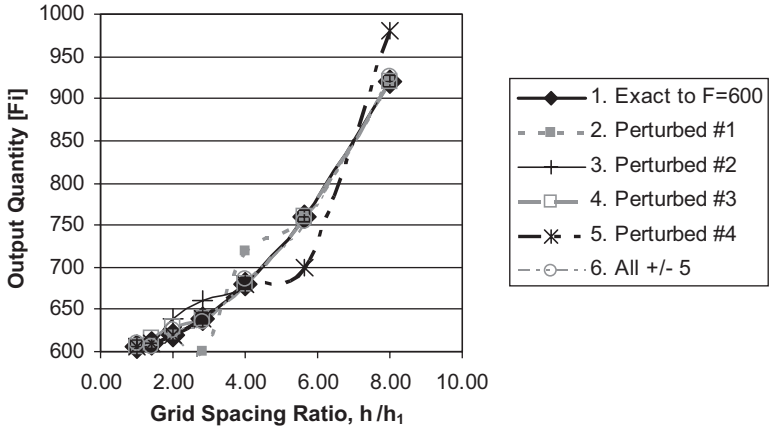

Fig. 1 Plots that begin a grid convergence study. Quantity of interest F_i vs. grid spacing, h_i, on linear scale. Exact second order "p = 2.00" convergence rate results with *contrived* perturbations to test assessment of solution verification uncertainty.

(2-tailed) or 84% confidence (1-tailed) as assessed in this work with the assumption of a normal distribution. It is also essential that the term U_{sver} represent the uncertainty in the model output *quantity of interest*, rather than some intermediate model output whose relation to the quantity of interest is not clear. This will enable the U_{sver} term of Step [B] of our "ABCD" V&V to be incorporated into the subsequent model validation, Step [C]. We refer to U_{sver}, the "solution verification uncertainty", as one of the standard (or expanded) uncertainty[15] terms to be added in quadrature or otherwise and then used to generate confidence and prediction intervals[20] on the simulation output quantity of interest.

The third step [C] of our V&V process is Model Validation over the domain of the available validation data referent. This requires the generation of both a mean and a confidence interval for model predictions, in light of comparison to the available referent data, and with inclusion of other uncertainty terms not directly accounted for in model-vs.-data comparisons. The error and uncertainty estimate from Solution Verification is one of these uncertainty terms; hence it must be quantified in a manner consistent with the other terms in this third step; that is, a bias or systematic error and an uncertainty at a specified confidence level.

The fourth step [D] of our V&V process is the extrapolation of the third step outside the domain of model validation into the domain of application; that is, the quantification of a predictive capability.[3] The resulting uncertainty terms at given confidence levels, from V&V, may be used in combination with reliability methods to establish risk-based inputs into an adequacy or acceptance assessment.[4] Even if the domain of application is "inside" the validation space of the referent data of Step "C", Step "D" must still justify the use of the physics, confidence intervals, and uncertainties for interpolation within the space of referent data in the validation domain. Step "D" should provide the quantitative information about the model and its predictive capability such that, given a requirement, an adequacy assessment can be made to determine if more computing capacity, code development, validation analyses or referent data are needed.

Oberkampf et al. state that V&V must address tradeoffs for a "balance of sufficiency and efficiency",[5] while Logan and Nitta assert that V&V must acknowledge (and ideally, quantify) the point when "better has become the enemy of good enough".[4] Such tradeoffs involve timing and funding for many issues including compute platforms, code development, analyses, and certification issues and schedules. In this work, we note that Solution Verification forms an important part of quantification of the tradeoff between speed of the computational analysis, and improvement in confidence. The improvement in confidence obtained due to a smaller assessed Solution Verification error and uncertainty estimate must be balanced against the cost of obtaining the additional compute capability to run the computational model at finer meshing. We have discussed such benefit/cost ratio tradeoffs previously, in terms of the benefit/cost of additional code development or obtaining more experimental data.[6] In this paper, we will not extend our analysis to the benefit/cost of purchasing a larger compute platform based on our quantification of Solution

Verification; but we will demonstrate and compare procedures to develop the terms to do just that. Naturally, most such methodologies are still evolving, and this work represents the views of the authors and not necessarily the views of Lawrence Livermore National Laboratory. Regarding definitions of terms, we have tried to adhere as best we can to the existing AIAA Guide for the Verification and Validation of Computational Fluid Dynamics Simulations.[7]

II. Grid Convergence in Solution Verification: Principles and Goals

Step "B" in the 4-step validation process is that of Solution Verification. Solution Verification is the process of assuring that the output quantity of interest in a model approximating a physical reality with a discretized continuum (e.g. finite element) code converges in each discretized domain (spatial, temporal, and iterative). This is accomplished in the spatial domain by subdividing the elements or cells on the entire grid or portions of the grid. Grid convergence is only part of what is necessary to fully address Solution Verification, but we will limit the scope of our discussion to grid convergence to describe and demonstrate various methods.

It is important to emphasize that while solution verification is an assessment of the rate and proximity of our model to a converged solution, we can only assess convergence to a correct solution in cases where a closed form solution is available. Lacking a closed form solution, many authors use model validation to assess the ability of a model to duplicate the quantity of interest compared to experiments approximating the real world application. In this all too common case, we do not want the model validation process to be contaminated by the ambiguity of not knowing whether our grid is converged or how much error and uncertainty might result from the grid we choose for our model representation of experiment. It is the job of solution verification to provide and quantify this information via quantitative grid convergence bias and uncertainty terms that can then work in conjunction with the model validation process.

The first essential step in a grid convergence study is to plot the quantity of interest, F_i, vs. the mesh spacing along a given direction, h_i, for a series of $i = 1, N_g$ grids. We will assume in this work that a best practice is to refine the grid simultaneously in each direction, although selective directional refinement (i.e. refining the grid in only one direction at a time) can be useful in some instances. To generate inputs to the subsequent validation, reliability, and risk processes, we will need outputs from solution verification in the form of (error + uncertainty) at a given confidence level. Our demand for a confidence level means that, since we will in general have $K_g = 3$ free parameters in our grid convergence models as described below, we will find that in general, we need a minimum of $N_g = 4$ grids, with $N_g > 4$ being highly desirable. Table 1 shows an example of a *contrived* grid convergence study, comparing the grid convergence behavior of 6 different algorithms, all solving the same physical problem.

Each of the 6 algorithms is used with $N_g = 7$ grids (7 steps of grid refinement), to represent the results that might be obtained from comparing 6 different codes or algorithms for solution. The exact solution is $F_o = 600$. Each of the 6 algorithms and each of their 7 grids provide a solution F_i, where $F_i \neq F_o$ due to discretization errors. The first of the 6 algorithms is contrived so that during the 7 grids used for refinement, the solution will converge at $h_o = 0$ to exactly $F_o = 600$, with a rate of convergence of $p = 2.0$ in Eqn. [1a]. The remaining 5 algorithms are perturbed slightly. In this sense, the latter 5 algorithms will yield a non-constant value for the apparent grid convergence exponent p, since the output quantity F_i is non-smooth. In the extreme, such as in the second algorithm, even the output quantity F_i may be non-monotonic, and at this time some traditional grid convergence approaches will fail entirely. Typically such results are plotted first on a linear scale as in Fig. 1, then on a log-log scale as log assessed error $(F_i - F_o)$ vs. log grid spacing.

Table 1 Data for "exact" grid convergence set: $F_o = 600$ with p = 2.0.

h_i/h_1	1. Exact to $F = 600$	2. Perturbed #1	3. Perturbed #2	4. Perturbed #3	5. Perturbed #4	6. All ±5
8.000	920.00	920.00	920.00	920.00	980.00	925.00
5.657	760.00	760.00	760.00	760.00	700.00	755.00
4.000	680.00	720.00	680.00	680.00	680.00	685.00
2.828	640.00	600.00	660.00	640.00	640.00	635.00
2.000	620.00	620.00	640.00	630.00	620.00	625.00
1.414	610.00	610.00	610.00	615.00	610.00	605.00
1.000	605.00	605.00	605.00	605.00	605.00	610.00

Figure 1 has features typical of a grid convergence study. It looks like our quantity of interest F_i converges to $F_o \sim 600$, but how can we be sure? What bias error and uncertainty can we assess given that we may only have F_{fg}, our finest grid solution? We want to know how far we are from the exact solution $F_o = 600$, which is known in this case but unknown in general. There are many assumptions made during grid convergence studies,[8] leading to the frequent use of the Richardson Extrapolation Estimation (REE) technique.[9] The REE method *assumes* that the grid convergence results are in the asymptotic convergence regime, neglects higher order terms in its approximation assumptions, and assumes that the discretized model solutions F_i are of the form[10] $F_i = F_o + \alpha h_i^p$ as given in Eqn. [1a] just below. To accommodate non-smooth and even non-monotonic behavior of the quantity of interest during grid convergence, we will build a Least-Squares (LSQ) or Response Surface Method (RSM[11,12,13,14]) model to provide monotonic solutions approximating the discretized solutions. As is usually the case in the use of the RSM technique, we are building a *model of a model*, so we must quantify the *additional* uncertainty term that results from doing so.

We construct the RSM solutions F_i^* of the same *form* for the assumed behavior, Eqn. [1a], and the RSM, Eqn. [1b]:

$$F_i = F_o + \alpha h_i^p \tag{1a}$$

$$F_i^* = F_o^* + \alpha^* h_i^{p^*} \tag{1b}$$

In Eqn. [1b], F_i^* = the solution from the i^{th} grid (either the code result F_i, or a Least-Squares (LSQ) or Response Surface Method (RSM) approximation F_i^*). If the grid convergence is perfectly monotonic with constant convergence rate exponent 'p = p^*', we will have $F_i = F_i^*$ for all grids. We will see that in practice this is rarely the case except in contrived examples. We will assess the uncertainty term resulting from our RSM approximation to the model's grid convergence results using the standard deviation of fit between F_i and F_i^* In addition, we desire to obtain F_o = (the exact known solution), or F_o^* = (the estimate of the exact solution), where h_i = the i^{th} grid spacing, p = the convergence rate (either theoretical or RSM fitted p^*), and α = a fitting constant (α^* for the RSM fit).

Eqn. [2] and subsequent equations below follow from Eqn. [1a] and [1b] and hence carry the same caveats and assumptions such as a constant grid refinement ratio "r_h". This restriction could be lifted, as there are more general methods as discussed by Roache[8] that do not require a constant r_h. If the model results are consistent with a constant observed convergence rate exponent, we can use any combination of grid results for the quantity of interest F_i in Eqn. [2] and obtain the same exponent p.

We first obtain an estimate of the exponent p of convergence rate to the fully converged solution F_o. For any grid triplet or series of three grids, with a constant p and grid refinement ratio "r_h", we can calculate the convergence rate (exponent p), as in Roache,[8] as:

$$\text{p} = \log[(F_1 - F_3)/(F_1 - F_2)]/\log[r_h] \tag{2}$$

If we fit the grid convergence output quantity results to an RSM with an assumed constant exponent p, we can then calculate an estimate of the exact solution at zero grid size as in Roache[8] as:

$$F_o^* = F_o = F_1 + (F_1 - F_2)/(r_h^p - 1) \tag{3}$$

If this is true, we will obtain a correlation coefficient $r^2 = 1$ to the regression of the line fitting $\log(F_i^* - F_o^*)$ versus $\log(h_i)$. We will also obtain a sample standard deviation $s = 0$ (and bias $B = 0$) when comparing the set of computational model results and regression fit results. In practice however, even with smooth mesh convergence results, it seldom happens that all the F_i will yield the exact same exponent p. For example, to obtain exactly p = 2 at all points in a grid convergence study, we would need to be modeling a problem whose exact solution was a quadratic, with a numerical formulation that is second order convergent. Therefore to obtain an estimate of p over the entire domain of grid sizes, we suggest a regression fit "response surface model" (RSM) to the linearization of (Eqn. [1b]) to obtain a constant regression slope p^*.

We can obtain a regression value of p^*, with a correlation coefficient (hopefully $r^2 \sim 1$), and standard deviation comparing the regression fit to the computational model grid refinement ratios. We can estimate the error E_1 in our

finest grid solution (F_1) using Eqn. [4a] if p = constant, or Eqn. [4b] for the RSM where we define p^* = constant:

$$E_1[\text{fine grid}] = (F_2 - F_1)/(r_h^p - 1) \tag{4a}$$

$$E_1^*[\text{fine grid}] = (F_2^* - F_1^*)/(r_h^{p^*} - 1) \tag{4b}$$

This value may also be taken as an uncertainty, in addition to a systematic error or bias estimate. The proper choice here is not universally established in the community, since in application of the model, we cannot always assure ourselves of the direction of the bias error; only that we have a finite bias error at any grid that does not produce the exact, but often unknown, F_o. Therefore we might take the bias error and also sum in quadrature[15] (i.e. root mean squared (RMS) as in Eqn. [5a]) the *uncertainty* taken from the Richardson Extrapolation Estimation (REE), with the uncertainty taken from the (small sample corrected) standard deviation of the computational model result minus the regression fit (response surface model) at constant p^*. Or, we might simply take a linear sum of the REE estimate, (E) and the response surface model (RSM) random uncertainty estimate, (U_s). When the bias error term B, or E as we denote it for the Richardson Extrapolation term, is taken as an additional uncertainty, we choose in this work to scale E to E_C, scaled from E by the statistically assessed percent confidence used for the rest of the analysis. We now briefly contrast our treatment of the systematic uncertainty or observed bias B with the ASME suggested treatment regarding test uncertainty.[15] For this discussion we will neglect small sample corrections and so assume that sample standard deviation equals population standard deviation, or $s = \sigma$. The standard quadrature process we adopt[15] essentially defines use of the bias as a systematic uncertainty term (B), in quadrature with a normal 2-sigma or 95% confidence random uncertainty 2S, or $2u_s$ in our notation. Our procedure uses a normal 1-tail 2-sigma fraction 97.7% of the bias error B (or $E_C = 0.977$E here) in quadrature with $U_s = 2u_s$ (at 2-sigma). For a 1-sigma analysis, we combine the 1-tailed 84% ($E_c = 0.841$E) with $U_s = 1u_s$ at 1-sigma. The general expression for combining the systematic component and random (uncertainty of fit) component is given in our notation as:

$$U|C = U = (E_c^2 + U_s^2)^{(1/2)} \tag{5a}$$

In Eqn. [5a], both the bias term E_c and the random term U_s are given at a specified confidence level. Specific treatment using the ASME procedure[15] would give a combined uncertainty at 95% assessed confidence (U_{95}) or 68% assessed confidence (U_{68}), with $E_c = 1.000$E or $E_c = 0.500$E, respectively, of:

$$U|C = U_{95} = ((1.000\text{E})^2 + (2u_s)^2)^{(1/2)} \tag{5b}$$

$$U|C = U_{68} = ((0.500\text{E})^2 + (1u_s)^2)^{(1/2)} \tag{5c}$$

In contrast, the treatment of systematic uncertainty combined with random uncertainty used in the current work would provide, with $E_c = 0.977$E or $E_c = 0.841$E, respectively:

$$U|C = U_{95} = ((0.977\text{E})^2 + (2u_s)^2)^{(1/2)} \tag{5d}$$

$$U|C = U_{68} = ((0.841\text{E})^2 + (1u_s)^2)^{(1/2)} \tag{5e}$$

We are not aware of a well accepted standard for combining model assessed bias (B or E) with model fit uncertainty (u_s here) and would welcome such a standard. Meanwhile, we are exploring more rigorous treatments of the bias terms when treated as added uncertainties. The procedure we suggest and use here does not greatly influence our results or conclusions, but this may not always be the case. We will use the linear combination ($E_c + U_s$) in part of this work as noted, but unless otherwise stated, this work will use the root-mean-squares (RMS) combination to obtain an estimate of the total Uncertainty (U) from our grid convergence study:

$$U|C = U = (E_c^2 + U_s^2)^{(1/2)} \tag{5a}$$

This quantity (U) is hereafter expressed as a percent of the estimated exact solution F_o, so that:

$$U = (100\%) \times U/F_{fg}^* \tag{6}$$

For a smooth, monotonic example, $U_s \sim 0$ since the regression fit response surface is essentially an exact duplicate ($r^2 = 1$) of the computational model, so the entire error "$E_c + U_s$" is just "E_c", the REE estimate taken

as a confidence scaled bias uncertainty for the fine grid, after Roache:[8]

$$E_1[\text{fine grid}] = (F_2 - F_1)/(r_h^p - 1) \tag{4a}$$

Eça and Hoekstra[10] define the REE as

$$\delta_{RE} = E_1 = F_i^* - F_o^* = \alpha^* h_i^{p^*} \tag{7a}$$

For the finest of the $i = 1, N_g$ grids ($i = fg$) we have:

$$\delta_{RE} = E_1 = F_{fg}^* - F_o^* = \alpha^* h_{fg}^{p^*} \tag{7b}$$

We report percent error estimate δ_{RE} or E as a percent of the finest grid solution or its RSM estimate F_{fg}^*:

$$E = (100\%) \times E_1/F_{fg}^* \tag{8a}$$

$$|\varepsilon| = E = \%\delta_{RE} = (100\%) \times |(F_{fg} - F_{(fg+1)})/F_{fg}| \tag{8b}$$

If we neglect the uncertainty of fit component U_s in Eqn. [5], our process is complete since now $U_{sver} = U|C = E_c$ in Eqn. [5]. However, as noted above, it is rare that a grid convergence study of $i = 1, N_g$ grids will exhibit an exact fit to Eqn. [1a], with p = constant, α = constant, and no oscillations. To address the reality of non-smooth grid convergence observations, two basic methods have been used. One approach sets uncertainty of fit $U_s = 0$, and depends on a Factor-of Safety F_s approach (see Method #1 and Method #2 below) to modify the value of δ_{RE} or E as follows:

$$U_{sver} = \%GCI(\text{fine grid}) = F_s|\varepsilon|/(r^p - 1) \tag{9}$$

The other approach (see Method #6 to Method #10 below) uses a least-squares or response surface method to account for non smooth or non monotonic grid convergence, and uses an explicit uncertainty of fit term U_s instead of a factor of safety F_s. In this approach, the three free parameters $K_g = 3$, (F_o^*, α^*, and p^*) are best-fit to the grid convergence study of $I = 1, N_g$ grids. To account for the fitting process, Eça and Hoekstra[10] also include a Least–Squares (LSQ) uncertainty of fit term, which can also be called, as we shall do here, the Response Surface Method (RSM) uncertainty term:[11–14]

$$U_s = \sqrt{\sum_{i=1}^{N_g} (F_i - (F_o^* + \alpha^* h_i^{p^*}))^2/(N_g - K_g)} \tag{10}$$

Our estimate of solution verification (SVER) grid convergence uncertainty $U = U_{sver}$ (at a given confidence level) given by Eqn. [5] is the term that we can use as a Mean Value (MV) or other Fast Probability Integration (FPI) method[16] term in subsequent model validation, reliability, risk, and benefit/cost quantification. Therefore, the body of this work will describe various ways, from simple to elaborate, to obtain U_{sver} estimates, and how these methods behave on several challenging (oscillatory or non-monotonic) grid convergence examples.

III. Grid Convergence in Solution Verification: Methods

Using the notation above, we present a range of methods from simple to elaborate, to obtain a quantified estimate of $U = U_{sver}$, the solution verification grid convergence uncertainty estimate of Eqn. [5a]. There are 10 such methods, each building on the others, which we now describe.

A. Method #1–#5: Grid Convergence Index and Richardson Extrapolation

The first 5 methods are based on Grid Convergence Index (GCI) and Richardson Extrapolation Estimates (REE) with some fairly simple, empirical suggestions. These methods are very simple and easy to key into a spreadsheet program and we would encourage the reader to try them, both as an introduction to GCI and REE, and as a starting point to the more advanced methods.

1. Method #1: Grid Convergence Index as 2σ (GCI, 2σ)

For a series of grids $i = 1, N_g$ that shows monotonic convergence, use the Grid Convergence Index (GCI) method as described by Roache[8] and others[17] *as a first estimate*. For Computational Fluid Dynamics (CFD) problems, there are good and suggested correlations whereby the basic REE estimate is multiplied by a GCI factor of safety, $F_s = 1.25$ for $N_g = 3$ or more, and $F_s = 3.00$ for $N_g = 2$. With these F_s in the GCI, one can obtain an estimate of $U = U_{sver}$ at what is contended to be an expanded uncertainty at 2-sigma or 95% confidence[8] as given in Eqn. [9].

The use of these F_s with the contention of 95% confidence is *completely empirical*, but is based on correlation with an extensive database of agreement for CFD.[8] It is not clear that the same 95% confidence level correlates with the use of these same F_s for computational mechanics or other finite element, finite difference, or finite volume grid studies.

2. Method #2: Grid Convergence Index as 1σ (GCI, 1σ)

Use the GCI process described as Method #1, with $F_s = 1.25$ and $N_g = 4$ or more. Take the resulting $U_{sver} = U$ as a 68% confidence estimate. We suggest this procedure because, based on our small but growing database of non-smooth grid convergence studies with known analytical solutions as discussed below, a claim of 68% confidence fits much better than a claim of 95% confidence. The section below on Results and Reliability will help support this suggestion. However, like the original values of $F_s = 1.25$ or $F_s = 3.00$ for the GCI, our suggestion for taking the U_{sver} from Method #2 as a 68% confidence (1-sigma standard uncertainty) estimate is purely empirical. Nevertheless, since Method #1 and Method #2 can be used together with no extra work, and since they are a good motivation to pursue more complex methods, we suggest that they be used as first steps in a solution verification assessment. We will show examples of this below.

Method #1 and #2 will give generally conservative estimates of U_{sver} for smooth, monotonic grid convergence results, where the physics and mathematics of the problem actually yield a constant, or even approximately constant,[18] rate of convergence p for the problem at hand. However, Method #1 and #2 will fail for non-monotonic grid convergence results. This is because these methods require terms of Log(error), such as $\text{Log}(F_i - F_{i-1})$ or $\text{Log}(F_i - F_o^*)$ to obtain estimates of the slope p in Eqn. [1a]. For non-monotonic grid convergence, the sign of $(F_i - F_{i-1})$ changes sign as "i" is incremented to compare the next two grid results, so $\text{Log}(F_i - F_{i-1})$ becomes undefined during the assessment. Because of this, we suggest the next three methods. For each of these Methods, #3, #4, and #5, the first step is to guess a value for p in Eqn. [1a] and then plot, on a linear scale, the observed grid solutions F_i vs. h_i^p. The intercept is then F_o^*, and the slope is α in Eqn. [1a].

3. Method #3: Fit Output F vs. Grid Spacing, with Factor of Safety = 3 (Fit, $F_s = 3$)

For this method, we suggest iterating to find p that maximizes the correlation coefficient r^2 of the linear fit of F_i versus h_i^p. In other words, since we have assumed that Eqn. [1a] ($F_i = F_o + \alpha h_i^p$) describes our grid convergence behavior, so that h_i^p should explain all the variation in F_i, ideally we would obtain $r^2 = 1.000$ if we select the right p. In reality this will not be generally true, partly because Eqn. [1a] may not hold, and partly because our grid convergence results are not smooth. However, since we have assumed that Eqn. [1a] does describe our grid convergence behavior, maximizing r^2 is a way to attempt to enforce our assumption that h_i^p should explain all the variation in F_i. Then, after obtaining α and p to maximize r^2, use a safety factor $F_s = 3$ and take the resulting GCI (Eqn. [9]) as the 1-sigma estimate of $U_{sver} = U$. Choosing a value of $F_s = 3$ or any other value is just as empirical a process as choosing F_s for the GCI procedure, and so far, our suggestion of $F_s = 3$ is based on an even smaller database than the suggested values for the F_s in GCI. The value of $F_s = 3$ with 1-sigma may seem extreme, but with our limited data base of non-smooth grid studies with exact solutions, it is the best suggestion we have so far.

4. Method #4: Richardson Extrapolation, Maximum Error (δ_{RE}, MAX)

Method #4 is one step simpler than Method #3, which is because Method #3 is just a simplified Richardson Extrapolation without extension to GCI or any further modifications. For Method #4, the steps are simply to plot F_i versus h_i^p, choosing first p = 1 and then p = 2. Take the *maximum* of the two uncorrected $\delta_{RE} = E_1$ (E as a percent) as the 1-sigma estimate of $U_{sver} = U$.

5. Method #5: Richardson Extrapolation, Average Error (δ_{RE}, AVG)

This method is a variant on Method #4. Obtain $\delta_{RE} = E_1$ as in Method #4 for both p = 1 and p = 2, and take the *average* of these same two uncorrected $\delta_{RE} = E_1$ as the 1-sigma estimate of $U_{sver} = U$.

B. Method #6–#10: Use of Response Surface Methods (RSM)

Methods #1–#5 are either well described in the literature (Method #1 and #2) or they are robust and simple to implement (#3, #4, and #5). They have the advantage that Method #1 and #2 are obtained together with no additional work, and have been used with as few as $N_g = 2$ grids, although we know of no one in the community advocating $N_g = 2$. In defense of $N_g = 2$ or even $N_g = 3$ for simulations of low risk scenarios,[19] we recognize that generating the meshes and grids appropriate for solution verification is one of the most challenging and labor-intensive tasks in the finite element community today. (We also recognize that all too often, we see the use of $N_g = 2$ or $N_g = 3$, or even $N_g = 1$ for simulations that could hardly be regarded as applied to "low risk" scenarios, e.g. where the consequences involve minimal dollar amounts and negligible health or safety considerations.) Method #3, #4 and #5 can be done very efficiently at the same time simply by selecting p = 1, then p = 2, and then iterating the value of p to complete Method #3. These first 5 methods are so simple that we suggest looking at them as a first simple check on the number being used for $U_{sver} = U$.

Method #6 through Method #10 all involve generation of an uncertainty term U_s explicitly. To do this, there is a minimum requirement that $N_g > 3$, since we have $K = 3$ free parameters (F_o^*, α, p) to fit. Four grids $N_g = 4$ is a minimum needed, and even $N_g = 4$ will lead to use of fairly large small sample corrections with the many assumptions inherent to small sample statistics. Our examples contain studies with $N_g = 6$ and $N_g = 7$. However, application specific grid convergence studies with $N_g = 6$ or more are rare, especially for complex problems with complex geometry.

6. Method #6: Eça and Hoekstra 2004, Minimize U *(EH04u)*

Method #6 is the method exactly as described in Eça and Hoekstra[10] and hence denoted the EH04u method here. Of note, the estimate of E is obtained by multiplying δ_{RE} by a GCI $F_s = 1.25$, which is often taken as 95% confidence, based on previous work and empirical correlations regarding the GCI.[8] This estimate of systematic error E is added to U_s at 1-sigma to obtain $U_{sver} = U$. We will call this a 1-sigma estimate of $U_{sver} = U$ in the examples below. However, as discussed above, there is no standard or consensus on combining bias error style uncertainty terms (δ_{RE} or E or B in general) with model fit uncertainty or variability terms (U_s in this case). Of note is in the Eça and Hoekstra (EH04u) method, the fit is obtained by choosing the $K = 3$ free parameters to minimize the U_s of Eqn. [10].

7. Method #7: Nitta and Logan 2004, Minimize U (NL04u)

Method #7 was developed and implemented independently by Nitta and Logan[11] (denoted NL04u) and in about the same time frame as the EH04u method. Both methods are processes built with logical elements and they differ only in the details. In Method #7 (NL04u), we do not currently use the GCI style $F_s = 1.25$ but rather $F_s = 1.00$; in other words, we use U_{sver} to construct confidence intervals and prediction intervals[20] as given by Eqn. [11] below, rather than applying any safety factor such as F_s. However, we do use a small sample correction (Student's t or other depending on conservatism) based on the number of grids = N_g, the number of free parameters $K = 3$, and the assumption of a normally distributed LSQ uncertainty of fit U_s. As in Method #6 (EH04u), we perform a least squares fit to minimize the term U_s in Eqn. [10].

8. Method #8: Nitta and Logan 2004, Minimize $E +$ U *(NL04eu)*

Method #8 is an extension of Method #7, also from Nitta and Logan,[11] and is denoted NL04eu. The procedure is the same as Method #7 (NL04u), except that we use a least squares procedure to generate a response surface model of the code results on the $i = 1$, N_g grids, this time to minimize the total term $U_{sver} = U$ in Eqn. [5]. That is, instead of minimizing U_s of Eqn. [10], we minimize Eqn. [5a], the combination of E with U_s. This does not seem as faithful to the principles of Richardson Extrapolation or even Least Squares Richardson Extrapolation with RSM as does Method #7 (NL04u) or Method #6 (EH04u).

However, as we noted on some non-monotonic examples, minimizing U_s in Eqn. [10] can give spurious results for the free parameters $K = (F_o^*, \alpha^*, p^*)$. Method #8 (NL04eu) is far more stable in these situations. The (NL04eu)

procedure will give slight errors in the case of an exactly correct grid convergence study, compared to Method #7 (NL04u) or Method #6 (EH04u) which will, if the uncertainty of RSM fit $U_s = 0$, converge to the exact analytical solution $F_o^* = F_o$. For example, consider the grid convergence in the left column of Table 1. This contrived set of solutions gives an exact convergence rate (p = 2.000) as the grid is refined, with perfect convergence predicted to the exact solution of $F_o^* = F_o = 600$. Both Method #7 (NL04u) and Method #6 (EH04u) will converge to $F_o^* = F_o = 600$, and provide correct estimates of discretization error E for the finest grid used (bottom row of Table 1, e.g. $F_{fg} = 605$ vs. $F_o^* = F_o = 600$). To show that even in this situation of perfect grid results for perfect $p = 2.000$ convergence rate, the approximation of Method #8 (NL04eu) is a good one, we used both Method #7 (NL04u) and Method #8 (NL04eu) on this exact grid study. We used progressively coarser values for h_i at $i = N_g$, the finest grid, and compared the error in the estimate of $F_o = 600$ or $F_o = 100\%$ (known exactly), in light of the known ratio of the fine grid solution F_{fg}/F_o. We also compared the overall $U_{sver} = U$, which is the quadrature combination of E and U_s (Eqn. [5a]). Figure 2 shows that at fine grids, Method #7 (NL04u) and Method #8 (NL04eu) are indistinguishable. At coarser values of $h_i, i = N_g$, we see that Method #7 (NL04u) continues to converge to the exact $F_o^*/F_o = 100\%$, while Method #8 (NL04eu) drifts away from the exact solution, giving $F_o^*/F_o = 104\%$ for very coarse grids. However, at this same grid, we have $F_{fg}/F_o = 153\%$. In other words, by the time the difference in F_o^* estimate between Method #7 (NL04u) and Method #8 (NL04eu) is 4% different, the grid is so coarse that we are far from the true solution anyway (53% too high) so the difference between Method #7 and #8 would be the least of our concerns. We feel that given the robustness of Method #8 (NL04eu), it is worth tolerating this small difference. Similarly, Fig. 3 shows a comparison of $U_{sver} = U$, the quadrature of E and U_s, for both Method #7 (NL04u) and Method #8 (NL04eu) methods. On the coarse grid, the U_{sver} estimate is $U = 109\%$ for Method #7 (NL04u), and only $U = 105\%$ for Method #8 (NL04eu). Both estimates of U (at 1-sigma confidence) are so large as to make this grid choice practically useless. And still, Method #7 (NL04u) and Method #8 (NL04eu) are very close in their estimate of $F_o^* \sim 600$ and the assessed U_{sver}.

Therefore, given the small discrepancy resulting from the use of Method #8 (NL04eu), we strongly recommend using jointly Method #7 and #8. While Method #7 (NL04u) has sounder foundations in the limit of vanishing mesh size "h", Method #8 (NL04eu) is an essential partner *because it is more robust*, and it will alert us to spurious results that may result from Method #7 (NL04u) in non-smooth grid convergence studies. As shown in Fig. 2 and Fig. 3, there should not be much difference between the U_{sver} obtained from Method #7 (NL04u) and Method #8 (NL04eu). If there is, we suggest using the larger of the two U_{sver} so obtained while the reason for the large discrepancy is investigated.

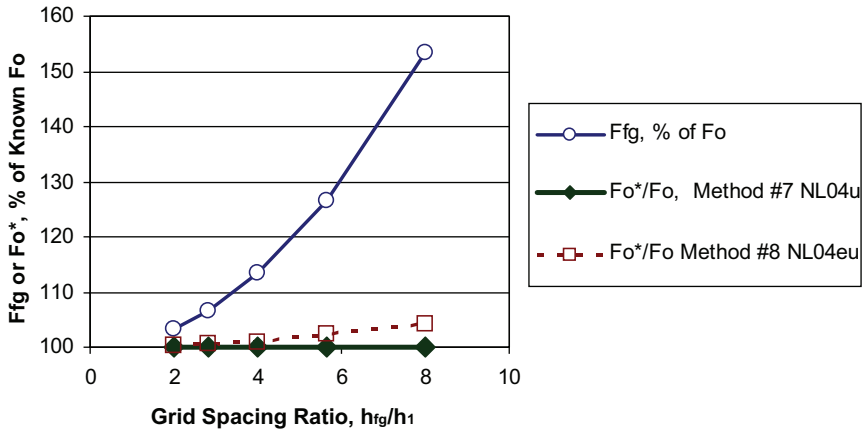

Fig. 2 Comparison of the estimate of F_o^*/F_o for Method #7 (NL04u) and Method #8 (NL04eu, more robust but approximate). Both estimates are excellent compared to the error in F_{fg} vs F_o for the very "coarse" choices of finest grid spacing h_{fg}.

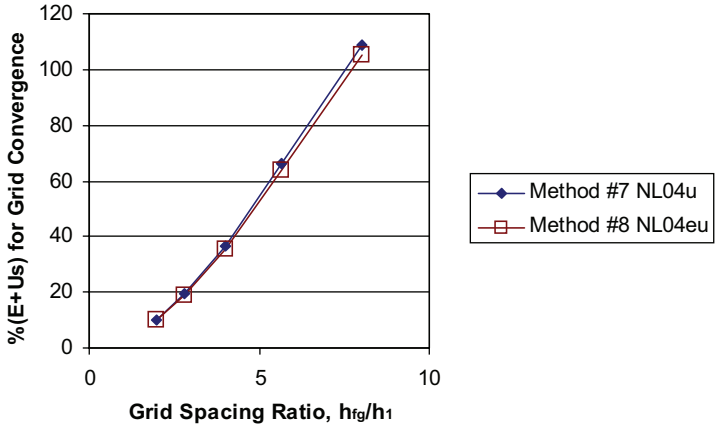

Fig. 3 Comparison of the estimate of $U_{sver} = U$ for Method #7 (NL04u) and Method #8 (NL04eu, more robust but approximate). Both estimates are very close even at high values of U_{sver} for the "coarse" choices of finest grid spacing h_{fg}.

9. Method #9: Logan and Nitta 2005, Minimize U *(LN05u)*

Method #9 is the same as Method #7 (NL04u) but with one important addition. In Method #9 (denoted LN05u), we note that, consistent with the original intent of the GCI, we are extrapolating our grid convergence study, with $h(i = N_g)$ the coarsest grid, and $h(i = 1)$ the finest grid, to $h_o = 0$, the estimate F_o^* of the fully converged solution F_o. Beginning with Method #6 (EH04u) and Method #7 (NL04), we are now mixing in the statistics of a least squares uncertainty term U_s. In Method #9 (NL05u), we note this, and as part of our extrapolation to $h_o = 0$, we correct with a statistical multiplier for the prediction interval extrapolated to $h_o = 0$ as in Eqn. [12] below. This is the simplest of standard parabolic corrections[20] that is a function of the interval from $h_{fg}(i = 1)$ to $h_{cg}(i = N_g)$, with mean h_m. The prediction interval correction is a function of distance from $h_o = 0$ to the mean of the grids used, $h_i = h_m$, compared to the span of the coarsest $h_i = h_{cg}$ and finest $h_i = h_{fg}$ grids used. This extrapolation prediction interval estimate will be small if $h_i = h_{fg}$ is close to $h_o = 0$ already, and large if $h_i = h_{fg}$ is far from $h_o = 0$.

10. Method #10: Logan and Nitta 2005, Minimize $E + $ U *(LN05eu)*

Method #10 (LN05eu) is the same as Method #9 (LN05u), except that we minimize Eqn. [5a], the combination of $E + U_s$, so we denote this as Method #10 (LN05eu). Method #9 and #10 (LN05u and LN05eu) form a pair in that both use the prediction interval correction extrapolation to $h_o = 0$ in Eqn. [12]. This is analogous to Method #7 and #8 (NL04u and NL04eu) which form a similar pair to Method #9 and #10 (LN05u and LN05eu), except that Method #7 and #8 (NL04u and NL04eu) do not use the prediction interval correction extrapolation, and so Method #7 and #8 obtain a prediction interval (P.I.) equation as:

$$P.I. = U_s\sqrt{1 + \frac{1}{N_g}} \tag{11}$$

Method #9 and #10 obtain a prediction interval (P.I.) that contains an explicit term for extrapolation to $h_o = 0$ by multiplying the term U_s by:

$$P.I. = U_s\sqrt{1 + \frac{1}{N_g} + \frac{(0 - h_m)^2}{\Sigma_i (h_i - h_m)^2}} \tag{12}$$

IV. Grid Convergence in Solution Verification: Results and Reliability

We will now demonstrate one quantitative example to evaluate the usefulness of each of our 10 methods for estimating $U_{sver} = U$, our solution verification contribution to uncertainty at (1-sigma in this case) confidence. We compared each of the 10 methods on grid convergence studies that had known solutions. The first of these was the

set of grid convergence results given in Table 1. This study has $N_s = 6$ sets of results for comparison, and each set of results has $N_g = 7$ grids with successive refinement as shown in Table 1 to study grid convergence.

The second series was taken from the 2nd Verification Suite, a comparison by Harrison and Conway[21] at LLNL. This annual Verification Suite was begun as part of the first edition of our V&V Methodology[22] several years ago, with results first reported in the 1st Verification Suite by Sam et al.[23] This second series contains a grid convergence study with $N_g = 7$ grids, for a beam bending problem. We compare to the exact solutions for bending stress and end deflection for a beam represented by *shell* elements loaded on the shell surface. The bending behavior through the shell thickness should approximate that of a beam element. The beam bending model is a discretized model of a beam assumed to follow the governing equation:

$$EI(d^2y/dx^2) = -(w/2)(xL - x^2) \tag{13}$$

In this case, a uniform distributed load w is applied along the whole length of the upper surface of a beam ($x = 0$ to $x = L$). The beam has elastic modulus E and moment of inertia I. Deflection is measured in the direction y at the free end of the beam, $x = L$. We use a series of three different finite element codes, for a total of $N_s = 6$ sets of results. The results of each of the $N_s = 6$ finite element code grid refinement sets (stress for each of 3 codes and deflection for each of 3 codes) are given in Table 2, for the 7 grid refinement steps h_i/h_1. The exact solution is given in the bottom row for comparison to the direct finite element outputs shown.

The third series is also taken from the 2nd Verification Suite, and is a comparison on $N_g = 4$ grids (the minimum number of grids for Methods #6 to #10). Again the problem is that of beam bending, with a known exact solution. This time the same three finite element codes are compared, on stress and deflection (giving an additional $N_s = 6$ sets of grids), but with a *brick* element mesh. The results of each of the $N_s = 6$ finite element code grid refinement sets (stress from each of 3 codes and deflection from each of 3 codes) are given in Table 3, for the 4 grid refinement steps h_i/h_1.

Altogether, we have $N_s = 18$ sets of grid convergence studies, with $N_g = 7$ or $N_g = 4$ grids for each. For each of our $N_s = 18$ sets of grid convergence studies, we will, in the end, generate an estimate (Eqn. [5a]) of $U_{sver} = U$; our estimate of the uncertainty at 1-sigma of our fine-grid solution F^*_{fg}, compared to our estimate F^*_o of the exact solution. Since all of the $N_s = 18$ sets have known analytical solutions, we can compare this U_{sver} to the actual difference E_{actual} between the code-produced fine grid solution F_{fg} and the known solution F_o. If we claim 1-sigma confidence assuming normal distributions of U_{sver}, and we are neither too conservative nor too optimistic, we should find that about 68% of the time (12/18) the U_{sver} estimate should enclose E_{actual}. In about 6 of 18 cases, we should find that U_{sver} does not enclose E_{actual}. Similarly, if we use the 2-sigma estimate of $U_{sver}|2\sigma$, we should find that in 95% of cases (about 17/18) U_{sver} should enclose E_{actual}. In about 1 of 18 cases, E_{actual} should be larger than

Table 2 Finite element code results and exact solution data for 2nd series on grid convergence: 7 grids for refinement.

Relative grid size	Beam bending stress (MPa)			Beam end deflectgion (mm)		
h_i/h_1	Code 1	Code X	Code W	Code 1	Code X	Code W
Coarse						
16	98.4	98.4	98.2	1.2718	1.2718	1.2718
12	100.7	100.9	100.4	1.2717	1.2717	1.2720
8	102.1	102.1	101.7	1.2718	1.2714	1.2720
6	102.5	102.6	102.2	1.2716	1.2714	1.2719
4	102.8	102.9	102.4	1.2716	1.2713	1.2719
2	103.1	103.1	102.8	1.2713	1.2713	1.2719
Fine						
1	103.2	103.2	102.9	1.2713	1.2713	1.2719
Exact solution:	103.4	103.4	103.4	1.2700	1.2700	1.2700

Table 3 Finite element code results and exact solution data for 3rd series on grid convergence: 4 grids for refinement.

Relative grid size	Beam bending stress (MPa)			Beam end deflection (mm)		
h_i/h_1	Code 1	Code X	Code W	Code 1	Code X	Code W
Coarse						
8	38.7	38.7	29.7	1.7060	1.5621	1.2488
4	46.5	46.5	43.9	1.3674	1.3674	1.2700
2	51.8	51.8	51.2	1.2996	1.3039	1.2742
Fine						
1	54.7	54.7	54.7	1.2869	1.2827	1.2827
Exact solution:	58.2	58.2	58.2	1.2700	1.2700	1.2700

$U_{sver}|2\sigma$. We define our Solution Verification Method Reliability R_{sm} as

$$R_{sm} = 1 - |\text{Fraction inside, Method Estimate} - \text{Fraction inside, expected}| \tag{14}$$

That is, at 1-sigma for $N_s = 18$, we expect 12 of the series to have $U_{sver} > E_{actual}$, or Fraction Inside, Expected $= 12/18 = 0.667$. If a given Method #X (where #X is #1 thru #10 in turn) assesses all 18 $U_{sver} > E_{actual}$, then [*Fraction Inside, Method #X*] $= 1.000$ and $R_{sm} = 1 - |1.000 - 0.667| = 67\%$. If a given Method assesses 12 $U_{sver} > E_{actual}$, then $R_{sm} = 1 - |0.667 - 0.667| = 100\%$.

If a given Method assesses only 9 $U_{sver} > E_{actual}$, then $R_{sm} = 1 - |0.500 - 0.667| = 83.3\%$. For a method that is neither too optimistic (risk induced due to non-conservatism) nor too pessimistic (too risk-averse) we want R_{sm} to be as close as possible to 100%. As just stated, we would expect at 2σ (95% confidence) that about 5% of the time, or in about $1/18$ cases, E_{actual} would be larger than $U_{sver}|2\sigma$. It is worth noting that our definition of Method Reliability means that even if we find, at 2σ, that E_{actual} is larger than $U_{sver}|2\sigma$ zero out of eighteen $(0/18)$ or two out of eighteen $(2/18)$ times instead of our expected one out of eighteen $(1/18)$ times, the Method Reliability will only be lowered by 6% (one in 18). Naturally, a sample size of $N_s \gg 18$ would be better, but our sample size of $N_s = 18$ will make our points without excessive truncation artifacts due to small integer sample size. This definition of "Method Reliability" is similar, though not identical, to the development and usage given by Urbina, Paez, et al.[24] Figure 4 shows the results on our $N_s = 18$ grid sets, for the 10 methods. All 10 of the Methods show fairly high Method Reliability R_{sm}, except for the use of Method #1 (GCI, 2σ), with GCI $= 1.25$ as a "95% confidence" estimate. We realize that there is a large database of CFD solutions (perhaps mostly smooth and monotonic) that supports[8] the use of Method #1 (GCI, 2σ). However, for our contrived and mechanics example $N_s = 18$ sets (most of which were

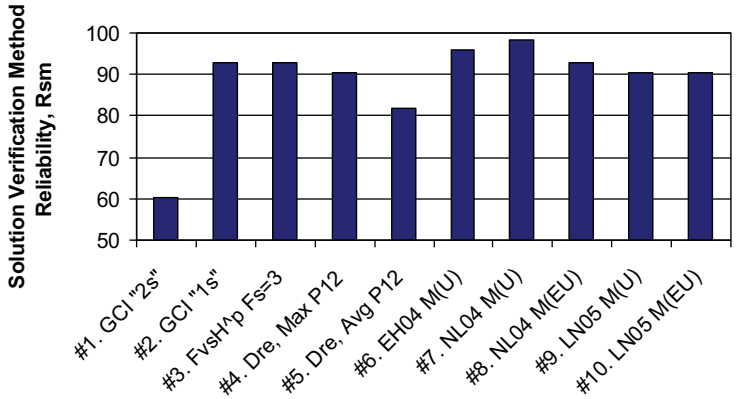

Fig. 4 Solution verification method reliability R_{sm}, for the 10 Methods #1 to #10 on the 18 grid set series discussed.

non-smooth), the use of GCI = 1.25 is much closer to a 68% confidence estimate than 95%. A larger sample set (so that the number of grid convergence sets $N_s \gg 100$) is needed to draw any general conclusions in this regard, but we suspect that this observation is due perhaps to the characteristics of solid mechanics applications, but most likely to our intentional selection of non-smooth grid convergence results.

It is important to note that we did not allow values of exponent p > 2 in our fitting procedures, since none of our examples had numerical formulations that would enable convergence of order or rate p > 2. We note that methods in solution verification can give spurious values[11] of p ≫ 2, or p larger than the theoretical value of the formulation, in the presence of non-smooth grid convergence results. We attempted to avoid this situation by not allowing any fits with p > 2. It has been suggested to us that in the GCI procedures, it is best to use the lesser of the observed p and the theoretical p of the numerical algorithm being used in assessments of U_{sver}. We agree and would extend the suggestion to the entirety of Method #1–#10.

While encouraging, the "Method Reliability" shown in Fig. 4, cannot be used to identify any one of Method #1–#10 as superior to all the others. Perhaps an extension of our data set beyond $N_s = 18$ will reveal more about the method reliabilities.

A measure of robustness of each of the Methods #1 to #10 is given by evaluating a variant of the reduced Chi-Square ($X^2_{\mathbf{v}}$) statistic[25] as:

$$X^2_{\mathbf{v}^*} = \sum_{I=1}^{N_s=18} [U_{sver}/(F_{fg} - F_o)]^2/(N_s = 18) \tag{15}$$

We stress that our $X^2_{\mathbf{v}^*}$ is a modified form of $X^2_{\mathbf{v}}$, in that we are not attempting to use model free parameters to remove bias as is commonly done which normally leads to terms such as (N-1) in the denominator of standard deviation. A high value for this $X^2_{\mathbf{v}^*}$ indicates a solution verification method that is too conservative; U_{sver} is in general much larger than E_{actual}. We computed this value of $X^2_{\mathbf{v}^*}$ for each method, but also added the standard deviation of the $X^2_{\mathbf{v}^*}$ ratio for each of the $N_s = 18$ sets. Ideally, if $U_{sver} = E_{actual}$ in each of the $N_s = 18$ series, we would have $X^2_{\mathbf{v}^*} = 1.00$ and the standard deviation $s(X^2_{\mathbf{v}^*}) = 0.00$ so $X^2_{\mathbf{v}^*} + s(X^2_{\mathbf{v}^*}) = 1.0 + 0.0 = 1.0$. The actual values are shown in Fig. 5 for each of Solution Verification Methods #1–#10. This figure measures the combination of over-conservatism (bias) and uncertainty (standard deviation) of the individual and composite $X^2_{\mathbf{v}^*}$ assessment of each method. Considering that the Method Reliability R_{sm} numbers were quite good, we would expect these $X^2_{\mathbf{v}^*}$ values to be lower than those shown in Fig. 5. We believe that our intentional choice of grid studies with oscillations in both exponent p and output quantity F_i might explain why we have good R_{sm} but $X^2_{\mathbf{v}^*} \gg 1$ in general, and why there is no apparent correlation between R_{sm} and $X^2_{\mathbf{v}^*}$.

We observed overall that while most of the Methods #1 to #10 gave high Method Reliability R_{sm}, Method #1–#5 give unpredictable results for any given individual series, as reflected in their higher $X^2_{\mathbf{v}^*}$ numbers in Fig. 5. *Hence we*

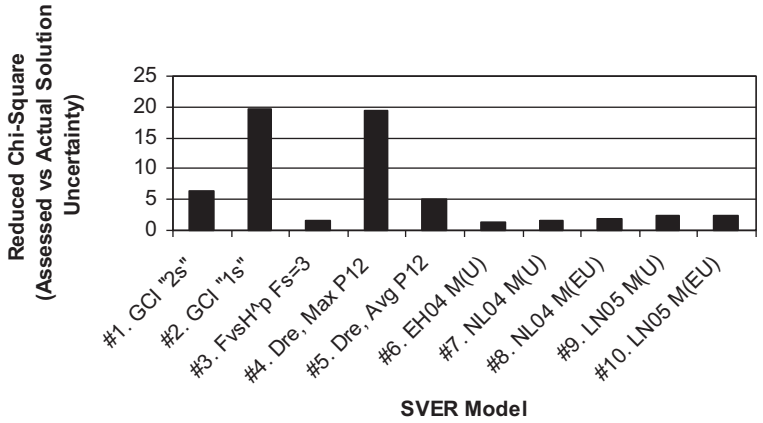

Fig. 5 Reduced Chi-Square measure of solution verification method overconservatism added to scatter in conservatism. Values closer to one are best; high values are either systematically or erratically overconservative.

Table 4 Summary of features of solution verification method #1 to method #10.

Method	Symbol	Basic reference	Basis	Uses factor of safety F_s?	Key equations	Statistical confidence and prediction intervals?	Prediction interval extrapolation to $h_o = 0$?
Method #1	GCI,2σ	Roache[8]	GCI	YES	1–9	–	–
Method #2	GCI,1σ	Roache[8]	GCI	YES	1–9	–	–
Method #3	Fit, $F_s = 3$	Richardson[9]	REE	YES	1–9	–	–
Method #4	δ_{RE}, MAX	Richardson[9]	REE	–	1–8	–	–
Method #5	δ_{RE}, AVG	Richardson[9]	REE	–	1–8	–	–
Method #6	EH04u	Eça and Hoekstra[10]	RSM	YES	1–10	YES	–
Method #7	NL04u	Nitta and Logan[11]	RSM	–	1–11	YES	–
Method #8	NL04eu	Nitta and Logan[11]	RSM	–	1–11	YES	–
Method #9	LN05u	(This work)	RSM	–	1–12	YES	YES
Method #10	LN05eu	(This work)	RSM	–	1–12	YES	YES

highly recommend the use of one or more of Method #6–#10 to obtain the most reliable and robust estimates of U_{sver}. We stress that the for the most part, use of Method #1–#10 for a solution verification uncertainty at confidence estimate (i.e., Eqn. [5a]) show high method reliability R_{sm}, and therefore we urge the use of at least a few of these methods as opposed to no estimate at all. We intend to add more data sets to our study and compare smooth convergence studies with oscillatory ones.

A brief summary of solution verification Method #1–#10 is given in Table 4. The simplicity of the table may not capture all of the subtleties of the different methods, but should give a quick overall view of each.

V. Linking Solution Verification to Model Validation

We used Method #1–#10 on the grid convergence study data of Nitta and Logan,[11] and on a few of the examples from Eça and Hoekstra.[10] These grid convergence studies gave either non-smooth or even non-monotonic results, making them ideal for testing the different solution verification methods. In general, these tests confirm that it is best to use several of the Methods during a grid convergence study, and in particular we recommend running several variants of Method #1–#10. We observed several cases where the inclusion of the prediction extrapolation parabola (Method #9 and #10, LN05u and LN05eu) provided added stability by avoiding a spurious fit. We also observed instances where minimizing $(E + U_s)$ as in Method #8 and #10 (NL04eu and LN05eu), avoided optimizing to a spurious solution[11] with apparent super-convergence. That is, we have seen instances where Method #7–#9 (NL04u or LN05u) would indicate a spurious best-fit p = 3, whereas Method #8 or Method #10 (NL04eu or LN05eu) would show a best-fit with $p \ll 2$. However, all of Methods #1–#10 are partly empirical in nature, and none has proven totally robust, so far. Hence, we recommend several different combinations. Eça and Hoekstra[10] offered one additional simple, and perhaps close to bounding, estimate of U_{sver} for extremely difficult non-monotonic grid sets, and we provide an illustrative example. Consider the grid convergence results shown in Fig. 6. These are contrived numbers, but very close to situations we have seen in nonlinear large deformation mechanics problems. Note that only $N_g = 3$ grids are available. However, in this case, we make the reluctant assumption that for similar codes, physics, algorithms, and problem application, the same convergence rate "p^*" applies to all four grid series. We now have only $K_g = 2$ free parameters (F_o^* and α^*), so we need a minimum of $N_g = 3$ grids. Method #10 (LN05eu), was able to generate solution verification uncertainty at confidence estimates that are both stable and compare well with those of "Emergency Method #E" as proposed by Eça and Hoekstra,[10] as shown in the comparison in Fig. 7.

For what we call Method #E, Eça and Hoekstra propose to take the largest magnitude of observed difference in two adjacent grids, and triple that value to obtain an estimate of the U_{sver} uncertainty. The choice of tripling the largest observed difference is at least as arbitrary as the choice of $F_s = 3.00$ in the GCI or other empirical methods, and is simply a suggestion[10] for use where grid convergence information already borders on being inadequate. For the examples considered, we find that Method #10 (LN05eu) bounds their number with 1.00 sigma and 1.65 sigma analyses. All methods tell us one very clear thing about this analysis: Three grids $N_g = 3$ are rarely enough for a

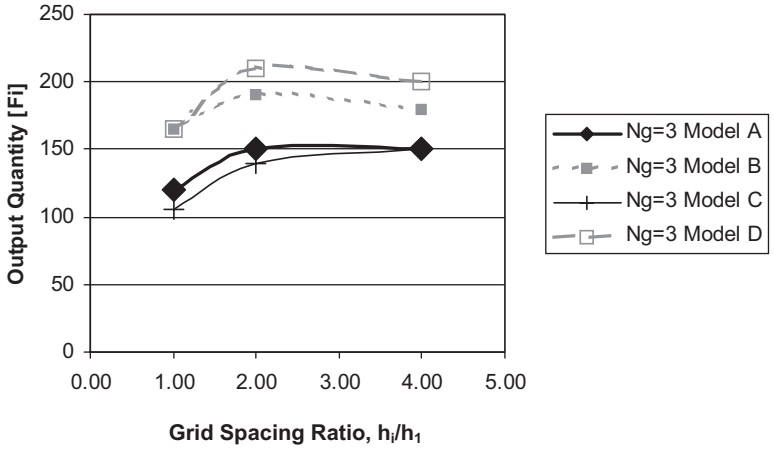

Fig. 6 Special case example for $N_g = 3$; assume same p for all four grid series leaving $K_g = 2$. Convergence for the sparse $N_g = 3$ ranges from oscillatory to ambiguous.

credible assessment of U_{sver}, even when assuming the same p-value to make the analysis possible. Furthermore, with grid results shown here, the message is simply that more work is needed. If a number must be had with only the model results shown in Figs. 6–7, the value of U_{sver} shown in Fig. 7 should be sufficiently large to avoid any misinterpretation. In this case, an assessment that the solution verification uncertainty is 50 to 100% of the quantity of interest would convey the proper message; that this model would only give us a converged answer "within a factor of two or so" as measured on the output quantity of interest. The inadequacy of such grid convergence behavior can now be reflected quantitatively even if only the "Emergency Method" assessment of U_{sver} is carried forward into a risk analysis and systems engineering construct.

Having described 10 methods for quantification of a term U_{sver} as $U|C$, an uncertainty at an assessed statistical confidence level e.g. $C = 95\%$, we now proceed to link this result into a quantitative validation statement that results from step [C] of our V&V process.

First of all, we offer a suggestion on how to proceed for the case where more than one of Method #1–#10 is used. If more than one of Method #1–#5 is used, and large differences between the U_{sver} estimates are observed,

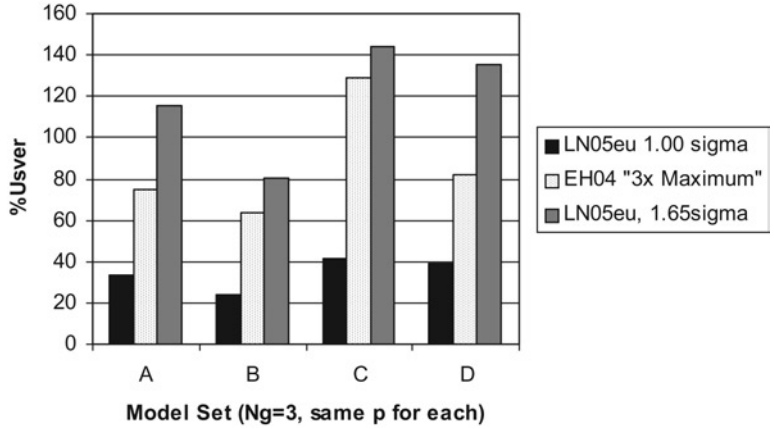

Fig. 7 Special case example for $N_g = 3$; assume same p for all four grid series leaving $K = 2$. Uncertainty estimates at 1-sigma are 24% to 40% using Method #10 at 1-sigma. Eça and Hoekstra "3x max difference" gives uncertainties of 60% to 130%, about equal to Method #10 at 1.65 sigma. Either way, the message has been sent; this situation heeds help.

we suggest using one or more of Method #6–#10, which should give very similar estimates of U_{sver}. If they do not, we suggest using the largest of the U_{sver} from Method #6–#10, while investigating the causes for the observed large discrepancies between the methods. If all else fails and time constraints intervene, use the "Emergency Method #E" discussed and illustrated in Figs. 6–7.

We have now assessed an uncertainty at confidence $(U_{sver}|C)$ for the grid convergence portion of solution verification. We have previously provided examples of how this $(U_{sver}|C)$ term is used in proceeding from solution verification to validation, reliability, risk, and finally the generation of Benefit/Cost Ratios (BCRs) for future actions.[11] The essential steps linking these processes are:

- Plot the quantity of interest versus grid spacing
- Use a mixture of Method #1–#10, or Method #E as a temporary method if needed, to generate the solution verification uncertainty at confidence, $U_{sver}|C$. This is "Step B" of the "ABCD" V&V process outlined earlier in the paper.
- If the analysis is assessed to be low risk, or if one is confident that the conditions for the use of the GCI are valid for the analysis, then the GCI (Method #1) with $N_g = 3$ or even $N_g = 2$ offers a fast assessment of $U_{sver}|C$ that may indicate whether more advanced methods (Method #6–#10) and more grids $N_g \geq 4$ are needed.
- In general however, and as illustrated in our last example, three grids $N_g = 3$ are rarely enough for a credible assessment of U_{sver}, even if we assume the same p-value to make the analysis even possible.
- Use this $U_{sver}|C$ value as a mean value (MV) contribution carried forward into the Model Validation Process "step C and D" of V&V. (There are methods more elaborate than including $U_{sver}|C$ as a mean value contribution, and if $U_{sver}|C$ is large enough to have an impact on the validation statement about the model, the inclusion of $U_{sver}|C$ as a mean value term may not be adequate. Methods more advanced than MV are sometimes considered as integral with the type of reliability analysis[4,26] to follow).
- Combine the total and components of uncertainty at confidence from validation with system requirements, to generate a reliability %R, $(0\% < \%\text{R} < 100\%)$ for the model of the system (at a given assessed percent confidence level, $0\% < \%\text{C} < 100\%$).
- Combine these measures of model + system reliability at confidence with a consequence assessment of success and failure of the system, leading to quantitative risk and potential for risk mitigation or reduction.
- Assess the Benefit/Cost Ratio BCR of the opportunities for risk mitigation or reduction.

The measures for risk mitigation or reduction might include taking more experimental data or improving model physics,[6] or perhaps buying more compute capability[11] to allow the chance to lower solution verification uncertainty.

VI. Conclusions: Linking Solution Verification to Validation, Reliability, Risk, and Benefit/Cost

Solution verification is a difficult and tedious, but essential part of the model V&V process. We must acknowledge that solution verification (grid convergence studies in particular) can take a lot of time and computer resources, and so we must provide ways to show the value of the process with quantitative measures.

We have taken our discussion from the first step in a grid convergence study (plotting the quantity of interest versus grid spacing) to the point where an input term for validation is generated. This term is an uncertainty at confidence, $U_{sver}|C$. We described 5 relatively simple methods and 5 more complex methods to assess U_{sver}, with 1 emergency method as a backup. We recommend that more than one be used in an analysis. The choice, and subsequent action, may depend on the discrepancy between the value of U_{sver} obtained from each method, as well as the risk involved in using the simulation being considered. Our (limited) experience indicates that the values of U_{sver} range, at 1-sigma confidence, anywhere from less than 5% of the output quantity to as much as 100% of the output quantity. It is not unusual to see values of U_{sver} at 1-sigma of 30% of the output quantity when grid convergence studies are non-monotonic in the output quantity. We do not know of any method for assessing grid convergence uncertainty that will circumvent the fact that non-smooth grid convergence results lead to large values for the solution verification uncertainty term.

There appears to be no unique method to assess solution verification uncertainty, especially for non-smooth grid convergence results. We have presented 10 methods, and recommend using the largest value of U_{sver} so obtained while further investigation is pursued if warranted. We have structured each method with the goal of a U_{sver} term that

links directly into a quantitative validation process, i.e. quantification of assessed risk and the potential for reduction in assessed risk at the cost of more work in solution verification. This linkage is important, because if we can show the value, in terms of risk and benefit/cost ratio (BCR), of solution verification, perhaps it will be easier to justify the effort spent on the process. Since there is a balance between the amount of effort (cost) spent on V&V of a model, and the amount of value (reduced assessed risk) that can be gained, we have presented a method to quantify the contribution of solution verification to this balance.

Developing our computational models in a systems engineering context will enable us to balance these two aspects, and defend our determination of this balance. However, as we show in the analysis, the balancing point depends on the details and fidelity of the quality of the codes, solution verification issues, referent data and model validation status, consequences leading to risk assessments, and the important role of the BCR in determining when "better has become the enemy of good enough". The procedures just outlined will enable us to (hopefully) show, given sufficient computational resources, a reduced value of assessed solution verification uncertainty via Eqn. [5a]. This will enable us to credibly assess increased product reliability[26] and reduced risk, leading to a positive dollar benefit[4] $\Delta\$B$. If this benefit outweighs the cost $\Delta\$C$ of the computational power needed for solution verification, then we will have a favorable BCR and will have made the business case for solution verification and the computational power needed.

The dangers of making decisions based on computational models with inadequate grid refinement are present, whether we quantify and acknowledge the situation or not. Since there is no completely universal or theoretical way to quantify the uncertainty and risk associated with grid convergence issues, we present several methods, from very simple to very complex. Our message is to try at least one quantitative solution verification method, and carry its quantitative impact through as an input to a risk-informed decision process. When decision makers realize how important solution verification can be in a risk analysis context, support for solution verification will increase.

Acknowledgments

This work was performed under the auspices of the U.S. Dept of Energy by Lawrence Livermore National Laboratory under contract W-7405-Eng-48. The authors are grateful for their interaction with a large community in these fields. In particular, we wish especially to thank Luis Eça, Patrick Roache, Tim Trucano, Henry Hsieh, Ben Thacker, David Sam, and Tom Paez, for the discussions that motivated this work on solution verification and its ultimate linkage to the decision process. We also wish to acknowledge the reviewers' time and their many helpful comments and suggestions.

Disclaimer

References

[1]Logan, R. W., and Nitta, C. K., "Verification & Validation: Process and Levels Leading to Qualitative or Quantitative Validation Statements", LLNL UCRL-TR-200131 Oct. 2003.

[2]Trucano, T. G., Pilch, M., and Oberkampf, W. L., "General Concepts for Experimental Validation of ASCI Code Applications", Sandia National Laboratories SAND-2002-0341, Mar. 2002.

[3]Easterling, R. C., "Statistical Foundations for Model Validation: Two Papers", SAND2003-0287, Sandia National Laboratories, Feb. 2003.

[4]Nitta, C. K., and Logan, R. W., "Qualitative and Quantitative Linkages from V&V to Adequacy, Certification, Risk, and Benefit/Cost Ratio", UCRL-TR-205809, Proc. Foundations 04, by Society for Modeling & Simulation International (SMSI), at Arizona State U. Tempe, AZ, Oct. 2004.

[5]Oberkampf, W. L., Trucano, T. G., and Hirsch, C., "Verification, Validation, and Predictive Capability in Computational Engineering and Physics", Proc. Foundations 02, Johns Hopkins Univ., Laurel, MD, Oct. 2002.

[6]Logan, R. W., Nitta, C. K., and Chidester, S. K., "Risk Reduction as the Product of Model Assessed Reliability, Confidence, and Consequence", AIAA-2004-1175, 42nd AIAA Aerospace Sciences Meeting, Jan. 2004.

[7]AIAA, *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, American Institute of Aeronautics and Astronautics, Reston, VA, AIAA-G-077-1998, 1998.

[8]Roache, P. J., *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers, Albuquerque, NM, 1998.

[9]Richardson, L. F., "The Deferred Approach to the Limit", Trans. Royal Society of London Series A, Vol. 226, p. 299.

[10]Eça, L., and Hoekstra, M., "An Uncertainty Estimation Exercise with the Finite-Difference and Finite-Volume Versions of PARNASSOS", Proc. Workshop on CFD Uncertainty Analysis, 21-22 Oct. 2004, Instituto Superior Technico, Lisbon, Portugal.

[11]Nitta, C. K., and Logan, R. W., "Solution Verification Linked to Model Validation, Reliability, and Confidence", UCRL-CONF-206544, Sep. 2004.

[12]Unal, R., Lepsch, R. A., and McMillin, M. L., "Response Surface Model Building and Multidisciplinary Optimization using D-Optimal Designs", AIAA-98-4759, American Institute of Aeronautics and Astronautics, 1998.

[13]Myers, R. H., *Response Surface Methodology*, Virginia Commonwealth University, Allyn and Bacon Inc., Boston, MA, 1971.

[14]Box, G. E., and Draper, N. R., *Empirical Model Building and Response Surfaces*, John Wiley, New York, NY, 1987.

[15]ASME, *Test Uncertainty*, Performance Test Codes PTC 19.1, an American National Standard, ASME, New York, NY, Dec. 1998.

[16]Ang, A., and Tang, W., *Probability Concepts in Engineering Planning and Design*, Wiley & Sons, 1984.

[17]Coleman, H. W., and Stern, F., "Uncertainties in CFD Code Validation", ASME J. Fluids Eng., Vol. 119, pp. 795–803, Dec. 1997.

[18]Roache, P. J., "Conservatism of the GCI in Finite Volume Computations on Steady State Flow and Heat Transfer", ASME Journal of Fluids Engineering, Vol. 125, No. 4, July 2003, pp. 731–732.

[19]ASME, "The Role of Risk Analysis in Decision Making", ASME Board of Governors Position Paper ID #02-15, available at http://www.asme.org/gric/ps/2002/02-15.html, American Society of Mechanical Engineers, April 2002.

[20]Moore, D. S., and McCabe, G. P., *Introduction to the Practice of Statistics*, W.H. Freeman & Co., 1989.

[21]Harrison, J. L., and Conway, R., "Second Verification Test Suite for ASCI Codes for the Complex", LLNL report CODTU-2003-1225, Rev. 1, Oct. 2003.

[22]Logan, R. W., and Nitta, C. K., "Verification & Validation (V&V) Methodology and Quantitative Reliability at Confidence (QRC): Basis for an Investment Strategy", UCRL-ID-150874, Nov. 2002.

[23]Sam, D. D., Murty, S. S., Perfect, S. A., Castano, D., and Pott, J., "First Verification Test Suite for ASCI Codes for the Complex", LLNL report CODTU-2002-0113, Jan. 2002.

[24]Urbina, A., Paez, T. L., Gregory, D. L., and Resor, B. R., "Response Comparisons for Mechanical Joint Model Validation", Proc. 9th ITEA Modeling & Simulation Conference, Las Cruces, NM, Dec. 2003.

[25]Bevington, P. R., and Robinson, D. K., *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, 1969.

[26]Hsieh, S. R., and Sam, D. D., "Virtual Proving Ground for Assessing Reliability and Uncertainty", UCRL-CONF-205448, Proc. Foundations 04, by Society for Modeling & Simulation International (SMSI), at Arizona State U. Tempe, AZ, Oct. 2004.

Christopher Rumsey
Associate Editor