FACTORS OF SAFETY FOR RICHARDSON EXTRAPOLATION

by

Tao Xing¹ and Fred Stern

Sponsored by
The Office of Naval Research
Grant N00014-01-1-0073

IIHR Technical Report No. 476
IIHR—Hydroscience & Engineering
College of Engineering
The University of Iowa
Iowa City, Iowa 52242-1585  USA

May 2010

¹ Assistant Professor, Mechanical Engineering Department, Tuskegee University, Tuskegee, AL 36088
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. ERROR AND UNCERTAINTY ESTIMATE USING THE RICHARDSON EXTRAPOLATION METHOD</td>
<td>5</td>
</tr>
<tr>
<td>A. Convergence Studies</td>
<td>5</td>
</tr>
<tr>
<td>B. Error Estimates $\delta$</td>
<td>7</td>
</tr>
<tr>
<td>C. Uncertainty Estimates</td>
<td>10</td>
</tr>
<tr>
<td>D. Confidence Levels</td>
<td>12</td>
</tr>
<tr>
<td>III. STATISTICAL ANALYSIS</td>
<td>12</td>
</tr>
<tr>
<td>A. Reliability</td>
<td>13</td>
</tr>
<tr>
<td>B. Confidence of the Mean Analysis</td>
<td>14</td>
</tr>
<tr>
<td>C. Implementation</td>
<td>14</td>
</tr>
<tr>
<td>IV. ANALYTICAL AND NUMERICAL BENCHMARKS</td>
<td>20</td>
</tr>
<tr>
<td>V. ANALYTICAL AND NUMERICAL BENCHMARK DATA</td>
<td>21</td>
</tr>
<tr>
<td>VI. EXAMPLE FOR SHIP HYDRODYNAMICS APPLICATIONS</td>
<td>25</td>
</tr>
<tr>
<td>VII. CONCLUSIONS</td>
<td>27</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>29</td>
</tr>
<tr>
<td>NOMENCLATURE</td>
<td>29</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>30</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1 Summary of mean effectivity indices for different ranges of $P$ values.........9

Table 2 Analytical and numerical benchmark verification studies........................15

Table 3 Statistics of the (17) case studies and all (98) variables............................16

Table 4 Statistics for different ranges of $P$ values using non-averaged actual factor of safety.................................................................17

Table 5 Statistics excluding outliers at seventeen $P$ values.................................17

Table 6 Verification study for $C_{TX}$ of Athena bare hull with skeg at Froude number 0.48. $U_G$ is $\%S_1$ and $C_{TX}$ is based on the static wetted area........................................27

Table 7 Verification study for motions of Athena bare hull with skeg at Froude number 0.48. $U_G$ is $\%S_1$.................................................................................................27
LIST OF FIGURES

Figure 1 Comparison of error estimates for the one-dimensional wave equation ..............8

Figure 2 Effectivity indices for different error estimates ................................................9

Figure 3 Factor of safety for different verification methods with \( p_{th} = 2 \) and \( r = 2 \) for the CF method ...........................................................................................................11

Figure 4 Actual factor of safety for sample 3, sample 3 averaged using \( \Delta P = 0.01 \), and \( \overline{FS}_x \pm tS_{FS_x} \) for samples 9 to 25: (a) GCI method, (b) GCI1 method, (c) GCI2 method, (d) CF method, (e) FS method .................................................................22

Figure 5 Standard deviation for the mean \( S_{\overline{X}} \) and coefficient of variation \( S_{\overline{X}}(\% \overline{X}) \) for samples 9 to 25: (a) \( S_{\overline{X}} \), (b) \( S_{\overline{X}}(\% \overline{X}) \) .............................................................................................................24
ABSTRACT

A factor of safety method for quantitative estimates of grid-spacing and time-step uncertainties for solution verification is developed. It removes the two deficiencies of the grid convergence index and correction factor methods, namely, unreasonably small uncertainty when the estimated order of accuracy using the Richardson extrapolation method is greater than the theoretical order of accuracy and lack of statistical evidence that the interval of uncertainty at the ninety-five percent confidence level bounds the comparison error. Different error estimates are evaluated using the effectivity index. The uncertainty estimate builds on the correction factor method, but with significant improvements. The ratio of the estimated order of accuracy and theoretical order of accuracy $P$ instead of the correction factor is used as the distance metric to the asymptotic range. The best error estimate is used to construct the uncertainty estimate. The assumption that the factor of safety is symmetric with respect to the asymptotic range was removed through the use of three instead of two factor of safety coefficients. The factor of safety method is validated using statistical analysis of twenty-five samples with different sizes based on seventeen studies covering fluids, thermal, and structure disciplines. Only the factor of safety method, compared to the grid convergence index and correction factor methods, provides a reliability larger than ninety-five percent and a lower confidence limit larger than 1.2 at the ninety-five percent confidence level for the true mean of the parent population of the actual factor of safety. This conclusion is true for different studies, variables, ranges of $P$ values, and single $P$ values where multiple actual factors of safety are available. The number of samples is large and range of $P$ values is wide such that the factor of safety method is also valid for other applications including results not in the asymptotic range, which is typical in industrial and fluid engineering applications. An example for ship hydrodynamics is provided.
I. INTRODUCTION

Current quantitative numerical error/uncertainty estimates for grid-spacing and time-step convergence are based on the Richardson extrapolation method where the error is expanded in a power series with integer powers of grid spacing or time step as a finite sum. It is a common practice to retain only the first term of the series assuming that the solutions are in the asymptotic range, which leads to a grid-triplet study. The grid convergence index (GCI) derived by Roache [1] can be used to estimate the uncertainties due to grid-spacing and time-step errors and is widely used and recommended by ASME [2] and AIAA [3].

Stern et al. [4] derived the correction factor (CF) method with improvements made by Wilson et al. [5]. The CF method uses a variable factor of safety and was validated for correction factor less than one using analytical benchmarks. The factor of safety for correction factor larger than one is obtained by assuming that the factor of safety is symmetric with respect to the asymptotic range.

There are several problems in using the Richardson extrapolation method. As shown by Stern et al. [6], it is difficult to improve the accuracy by retaining more terms in the power series. For instance, use of both first- and second-order terms requires solutions for five grids, which significantly increases the computational cost. Additionally, it requires that all solutions should be sufficiently close to the asymptotic range, i.e., within about six percent of the theoretical order of accuracy of the numerical method $p_{th}$. When solutions are not in the asymptotic range, multiple grid-triplet studies often show non-smooth convergence. In such cases, the estimated order of accuracy $p_{RE}$ approaches $p_{th}$ with oscillations and a wide range of values [2]. The Richardson extrapolation method requires at least three systematic high-quality grids, which may be too expensive for industrial applications. The grid refinement ratio $r$ must be carefully selected. The magnitude of $r$ cannot be too large as the grids may resolve different flow physics. Too small values of $r$ (very close to one) are also undesirable since solution changes will be small and the sensitivity to grid spacing and time step may be difficult to identify compared to iterative errors.
The non-smooth grid convergence problem may be resolved using the least-square [7] or response-surface [8] methods, which requires solutions for more than three grids. There are some issues in using these two methods. The relationship between their estimates for the order of accuracy, error estimate, and numerical benchmark and those for individual grid-triplet studies is not established. They do not discriminate between converging and diverging grid studies and the use of diverging grid studies is not well founded. The requirement of at least four solutions is often too expensive for industrial applications. All the solutions are required to be in the asymptotic range, which is contradictory to the use of solutions that show non-smooth and non-monotonic convergence. They introduce additional uncertainties due to the least-square fit.

The difficulty and computational cost associated with the Richardson extrapolation method may be resolved by the single-grid method. Celik and Hu [9] demonstrated the use of an error-transport equation to quantify the discretization error. The one-dimensional convection-diffusion equation and two-dimensional Poisson equation using uniform grids showed that the method reasonably captured the sign and magnitude of the discretization error. Cavallo and Sinha [10] derived an inviscid error-transport equation for the Euler equation and applied it to a three-dimensional unmanned combat air vehicle 1303 using unstructured grids. The authors concluded that the error-transport equation offers a viable alternative approach to solution verification, particularly for cases where the Richardson extrapolation method cannot be applied. Cavallo et al. [11] extended their method to turbulent flow simulations using Reynolds-averaged Navier-Stokes equations by developing a viscous error-transport equation. Results for wall-bounded and free-shear flows showed marked improvements over the results obtained using an inviscid error-transport equation; however, the sensitivity of the solutions to grid spacing and time step is not provided and control of the spatial discretization error as the simulation progresses needs to be further investigated. Additionally, the applicability of the single-grid method to different discretization schemes and turbulence models needs to be validated.

The GCI and CF methods have two deficiencies. The first is that the uncertainty estimates for \( P_{RE} > P_{th} \) are unreasonably small in comparison to those with the same distance to the asymptotic range for \( P_{RE} < P_{th} \). This is due to the fact that the error
estimate $\delta_{RE}$ for the former is much smaller than that of the latter. The second is that there is no statistical evidence for what confidence level the GCI and CF methods can actually achieve. Roache [1] and the ASME performance test codes committee PTC 61 [12] stated that a ninety-five percent confidence level is achieved for the GCI method with a factor of safety of 1.25 based on over five hundred demonstrated cases by dozens of groups; however, no statistical samples or analyses are reported.

A recent study by Logan and Nitta [8] evaluated ten different verification methods for estimating grid uncertainty using the reliability $R_{sm}$ and reduced Chi-Square $\chi^2_{v}$. The reliability was used to measure the difference between the estimated and expected fractions that the uncertainty estimate will bound the comparison error. The reduced Chi-Square was used to measure the robustness such that a high value indicates that a verification method is too conservative. Methods 1-5 are for solutions that show smooth and monotonic convergence. Method 1 and method 2 implemented the GCI$_1$ method, which is the same as the GCI method except replacing $p_{RE}$ by $p_{th}$ when $p_{RE} > p_{th}$. The expected fractions that the uncertainty estimate will bound the comparison error are ninety-five percent for method 1 and sixty-eight percent for method 2. Methods 3-5 are based on the Richardson extrapolation method with different methods to compute $p_{RE}$. Methods 6-10 use the least-square or response-surface method to account for non-smooth or non-monotonic convergence and use an explicit method to compute the uncertainty due to the curve fit. Method 1 showed sixty percent $R_{sm}$. Method 2 shows ninety-three percent $R_{sm}$. These facts suggest that the use of the GCI$_1$ method is closer to a sixty-eight percent than a ninety-five percent confidence level. The other methods show more than ninety percent $R_{sm}$ except method 5 that shows eighty-two percent. Methods 1-5 predict much higher $\chi^2_{v}$ than methods 6-10, which indicates that there is no correlation between $R_{sm}$ and $\chi^2_{v}$. Since this study has only three structure problems with eighteen individual grid solutions, it was recommended that a sample with the number of grid convergence studies much larger than one hundred is needed to draw general conclusions.

Two other recent studies [13, 14] considered the use of different uncertainty estimates for different ranges of $p_{RE}$ for solutions that show monotonic convergence. Eça
and Hoekstra [13] presented the least-square version of the GCI method for a nominally second-order accurate discretization scheme. Three different uncertainty estimates were provided for $0 < \left( \frac{P - P_{\text{ref}}}{P_{\text{th}}} \right) < 0.475$, $0.475 \leq P < 1.025$, and $P \geq 1.025$. The estimates were based on the experience obtained in a variety of test cases and suggestions and comments of the first Workshop on Computational Fluid Dynamics (CFD) Uncertainty Analysis [15]. Rumsey and Thomas [14] similarly modified the GCI method [2] for a nominally third-order accurate discretization scheme. Three different uncertainty estimates were provided for $0 < P < 0.317$, $0.317 \leq P < 1.017$, and $P \geq 1.017$. These two verification methods were demonstrated for a manufactured solution [13] and the flow over a backward facing step [13, 14] without detailed derivation and validation. Statistical samples or analyses were not reported in either of these studies.

The objective of the present study is to develop a factor of safety (FS) method for solution verification. It removes the two deficiencies previously discussed for the GCI and CF methods. Different error estimates are evaluated using the effectivity index. The uncertainty estimate for the FS method builds on the CF method, but with significant improvements. A better distance metric to the asymptotic range is used. The best error estimate is used to construct the uncertainty. The assumption that the factor of safety is symmetric with respect to the asymptotic range is removed. The FS method is validated using statistical analysis of twenty-five samples with different sizes based on seventeen studies that have analytical or numerical benchmarks and cover fluids, thermal, and structure disciplines. The samples cover a wide range of $P$ values that are within and far from the asymptotic range ($P = 1$). The results of the FS method are compared with the GCI, GCI1, GCI2 [16] and CF methods.

There are two earlier versions of the FS method. The first version [17] resolved the unreasonably small uncertainty estimate for correction factor larger than one by reflecting the uncertainty estimate itself instead of the factor of safety with respect to the asymptotic range. This method was criticized for the lack of validation. The second version [18] extended and improved the first version by modifying the uncertainty for correction factor less than one with the introduction of factors of safety for correction factors equal zero and one, which were validated using statistical analysis as conducted
The second version was criticized by colleagues for the deficiencies of using the correction factor as the distance metric to the asymptotic range, the inapplicability of the method for correction factors larger than two, and the omission of effectivity index for evaluating error estimates. Their comments motivated the improvements over [18] as presented herein. Compared to [18], the present FS method has more general applicability without restriction on the maximum $P$, larger samples with the addition of twenty-five items, and a ninety-five percent reliability for all $P$ ranges.

The number of samples and items are large and range of $P$ values is wide such that the FS method is also valid for other applications including results not in the asymptotic range, which is typical in industrial and fluid engineering applications. An example for ship hydrodynamics is provided.

**II. ERROR AND UNCERTAINTY ESTIMATE USING THE RICHARDSON EXTRAPOLATION METHOD**

It is useful to consider the following four steps in deriving and evaluating solution verification methods: a) convergence studies; b) error estimate $\delta$ with magnitude and sign; c) uncertainty estimate $U$ that indicates the range of likely magnitudes of $\delta$, but no information about its sign; and d) statistical analysis to establish that the interval of $U$ at a ninety-five percent confidence level bounds the comparison error $E$. The comparison error $E$ equals the difference between the true value $T$ and simulation value $S$. For modeling validation, $T$ is the experimental data. For verification method validation, $T$ is either the analytical benchmark solution $S_{AB}$ or numerical benchmark solution $S_{NB}$.

**A. Convergence Studies**

It is assumed that iterative convergence has been achieved such that the iterative uncertainty is at least one order-of-magnitude smaller than the grid-spacing and time-step uncertainty. Grid-spacing and time-step convergence studies are conducted with multiple solutions using systematically refined grid spacing or time steps. First the value of $r$ is
selected. If 3, 2, and 1 represent the coarse, medium, and fine grids with spacing $\Delta x_3$, $\Delta x_2$, and $\Delta x_1$, respectively, then
\[
r = \frac{\Delta x_2}{\Delta x_1} = \frac{\Delta x_3}{\Delta x_2}
\]  
(1)

Constant $r$ is not required [1], but simplifies the analysis. If the solutions for the fine, medium, and coarse grids are $S_1$, $S_2$, and $S_3$, respectively, solution changes $\varepsilon$ for medium-fine and coarse-medium solutions and the convergence ratio $R$ are defined by
\[
\begin{align*}
\varepsilon_{21} &= S_2 - S_1 \\
\varepsilon_{32} &= S_3 - S_2 \\
R &= \frac{\varepsilon_{21}}{\varepsilon_{32}}
\end{align*}
\]  
(2)

When $0 < R < 1$ monotonic convergence is achieved and the Richardson extrapolation method is used to estimate $p_{RE}$, $\delta_{RE}$, and the numerical benchmark $S_C$. The error is expanded in a power series with integer powers of grid spacing or time step as a finite sum. The accuracy of the estimates depends on how many terms are retained in the expansion, the magnitude (importance) of the higher-order terms, and the validity of the assumptions made in the Richardson extrapolation method. With three solutions, only the leading term can be estimated, which provides the one-term estimates
\[
p_{RE} = \ln \left( \frac{\varepsilon_{32}}{\varepsilon_{21}} \right) \ln (r)
\]  
(3)

\[
\delta_{RE} = \frac{\varepsilon_{21}}{r^{p_{RE}} - 1}
\]  
(4)

\[
S_C = S_1 - \delta_{RE}
\]  
(5)

When solutions are in the asymptotic range, then $p_{RE} = p_{th}$; however, in many circumstances, especially for coarse grids and industrial applications, solutions are far from the asymptotic range such that $p_{RE}$ is greater or smaller than $p_{th}$. Stern et al. [4] used the correction factor as a metric for defining the distance of solutions from the asymptotic range based on the fact that $CF \delta_{RE}$ is a better error estimate than $\delta_{RE}$:
\[
CF = \frac{r^{p_{RE}} - 1}{r^{p_{th}} - 1}
\]  
(6)
The deficiency of using the correction factor as the distance metric is that it is also a function of \( r \). Therefore, even for the same \( p_{RE} \) and \( p_{th} \), correction factors could be different. The ratio of \( p_{RE} \) to \( p_{th} \) is used here as the distance metric:

\[
P = \frac{p_{RE}}{p_{th}}
\]  

(7)

The error estimate \( P\delta_{RE} \) is better than \( CF\delta_{RE} \) and useful for statistical analysis, for which analytical and numerical benchmarks can be combined according to the same \( P \) or different ranges of \( P \) values.

### B. Error Estimates \( \delta \)

In the GCI method, \( \delta \) equals \( \delta_{RE} \) for the whole range of \( P \) values. In the GCI\(_1\) and GCI\(_2\) methods, \( \delta \) equal \( \delta_{RE} \) and \( CF\delta_{RE} \) for \( 0 < P \leq 1 \) and \( P > 1 \), respectively. Stern et al. [4] showed that \( CF\delta_{RE} \) is a better error estimate than \( \delta_{RE} \) based on results from the numerical solution of the one-dimensional wave and two-dimensional Laplace equation analytical benchmarks

\[
CF\delta_{RE} = CF \left( \frac{e_{21}}{r_{pRE} - 1} \right)
\]  

(8)

The use of Eq. (8) replaces \( p_{RE} \) by \( p_{th} \) in the error estimate; however, \( p_{RE} \) is not discarded, but included in the correction factor and subsequently in the uncertainty estimate. In the FS method, \( \delta \) equals \( P\delta_{RE} \).

The accuracies of the different error estimates are evaluated using numerical solutions for the one-dimensional wave and two-dimensional Laplace equation analytical benchmarks. Figure 1 shows the comparison error \( E = S_{AB} - S_1 \) along with \( \delta_{RE} \), \( CF\delta_{RE} \), and \( P\delta_{RE} \) for the one-dimensional wave equation as functions of \( P \). The comparison error increases as a second-order polynomial of \( P \) as solutions are farther from the asymptotic range. All error estimates also increase with the distance from the asymptotic range, but with different order polynomials of \( P \). The error estimate \( \delta_{RE} \) is a fifth-order
polynomial and over-predicts the comparison error by up to sixty-one percent. The error estimate $CF\delta_{RE}$ is a first-order polynomial and under-predicts the comparison error by up to twenty-three percent. The error estimate $P\delta_{RE}$ is a second-order polynomial and under-predicts the comparison error by up to nine percent. Differences between different error estimates become smaller when solutions approach the asymptotic range. For the two-dimensional Laplace equation, the $CF\delta_{RE}$ and $P\delta_{RE}$ agree better with the comparison error than $\delta_{RE}$ and the differences between $CF\delta_{RE}$ and $P\delta_{RE}$ are negligible since $P$ and $CF$ were very close to one.

Figure 1 Comparison of error estimates for the one-dimensional wave equation.

The accuracies of the different error estimates can be evaluated using the effectivity index $\theta$, which is defined as the magnitude of the ratio of $\delta$ to $E$

$$\theta = \left| \frac{\delta}{E} \right|$$

where $E$ is the comparison error that is either $S_{AB} - S_1$ or $S_{NB} - S_1$. An ideal error estimate has $1 < \theta < 1 + \tau$, where $\tau$ is a small positive number. As discussed later, the seventeen studies that have analytical and numerical benchmarks provide three hundred and twenty-nine grid-triplet studies that show monotonic convergence, for which the same number of $\delta$, $E$, and $\theta$ are calculated and $\theta$ is shown in Fig. 2. For $P<1$, the
effectivity index for the three GCI methods is the same and much larger than those for the 
CF and FS methods. For \( P > 1 \), the effectivity index for the CF, GCI\(_1\) and GCI\(_2\) methods 
is the same and larger than those for the FS and GCI methods. Table 1 presents the mean 
effectivity index for different ranges of \( P \) values. For the whole range of \( P \) values, the 
comparison error is overestimated by more than 33\%\( E \) for the three GCI methods, 
underestimated by 0.2\%\( E \) for the CF method, and overestimated by 4.2\%\( E \) for the FS 
method. For \( 0 < P \leq 1 \), the comparison error is overestimated by more than 55\%\( E \) for the 
three GCI methods, underestimated by 4.6\%\( E \) for the CF method, and overestimated by 
6.2\%\( E \) for the FS method. For \( 1 < P < 2 \), the GCI method is the only one that 
underestimates the error by 11.2\%\( E \). The overestimations of the comparison error in this 
range for the other methods are 8.3\%\( E \) for the GCI\(_1\), GCI\(_2\), and CF methods and 0.1\%\( E \) 
for the FS method.

![Figure 2 Effectivity indices for different error estimates.](image)

Table 1 Summary of mean effectivity indices for different ranges of \( P \) values

<table>
<thead>
<tr>
<th>( P )</th>
<th>Number of points</th>
<th>( \theta_{GCI} )</th>
<th>( \theta_{GCI,2} )</th>
<th>( \theta_{CF} )</th>
<th>( \theta_{FS} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0 &lt; P &lt; 2 )</td>
<td>329</td>
<td>1.332</td>
<td>1.398</td>
<td>0.998</td>
<td>1.042</td>
</tr>
<tr>
<td>( 0 &lt; P \leq 1 )</td>
<td>218</td>
<td>1.558</td>
<td>1.558</td>
<td>0.954</td>
<td>1.062</td>
</tr>
<tr>
<td>( 1 &lt; P &lt; 2 )</td>
<td>111</td>
<td>0.888</td>
<td>1.083</td>
<td>1.083</td>
<td>1.001</td>
</tr>
</tbody>
</table>
C. Uncertainty Estimates

The uncertainty $U$ is defined as an estimate of an error such that the interval of $U$, $\pm U$, bounds the true value of $\delta$ at a specified level of confidence, which is usually ninety-five percent for experimental fluid dynamics and CFD. For comparison purposes, all the present methods can be written in the form of $U = FS|\delta_{RE}|$, where in the present context $FS$ is the verification method factor of safety over $\delta_{RE}$.

Given an $\varepsilon_{21}$ from a grid convergence study, the GCI is derived by first calculating the error estimate $\delta_{RE}$ using Eqs. (3) and (4), and then calculating an equivalent $\varepsilon_{21}$ that would produce approximately the same $\delta_{RE}$ with $p_{RE} = 2$ and $r = 2$. The absolute value of that equivalent $\varepsilon_{21}$ is the GCI for the fine grid solution, which is expressed as [1]

$$U_{GCI} = FS\frac{|\varepsilon_{21}|}{r^{p_{RE}} - 1} = FS|\delta_{RE}|$$

Roache [1] suggested values of $FS = 3$ for two-grid sensitivity studies using $p_{sh}$ and $FS = 1.25$ for convergence studies with a minimum of three grids using $p_{RE}$. Both approaches use a constant factor of safety for all $P$ values, e.g., $FS = 1.25$ as shown in Fig. 3.

Logan and Nitta [8] used the GCI1 method, for which the factor of safety for the GCI is multiplied by the correction factor for $P > 1$.

$$U_{GCI_1} = FS\left(P, CF\right)|\delta_{RE}| = \begin{cases} 1.25|\delta_{RE}| & P \leq 1 \\ 1.25CF|\delta_{RE}| & P > 1 \end{cases}$$

The GCI2 method [16] is the same as the GCI1 method except $FS = 3$ instead of 1.25 for $P > 1$.

$$U_{GCI_2} = FS\left(P, CF\right)|\delta_{RE}| = \begin{cases} 1.25|\delta_{RE}| & P \leq 1 \\ 3CF|\delta_{RE}| & P > 1 \end{cases}$$
The uncertainty for the CF method is estimated by the sum of the absolute value of the improved error estimate $|\delta_{RE}|$ and the absolute value of the amount of the correction.

$$U_{CF} = FS(CF)|\delta_{RE}| = \begin{cases} 
\left[9.6(1-CF)^2 + 1.1\right]|\delta_{RE}| & 0.875 < CF < 1.125 \\
\left[2|1-CF| + 1\right]|\delta_{RE}| & 0 < CF \leq 0.875 \text{ or } CF \geq 1.125 
\end{cases}$$

(13)

As shown by Wilson et al. [5] and in Fig. 3, the CF method differs from the GCI method since it provides a variable factor of safety. The CF method has the “common-sense” advantage in providing a quantitative metric to determine proximity of the solutions to the asymptotic range and approximately accounts for the effects of the higher-order Richardson extrapolation terms. The CF method has been used in ship hydrodynamics CFD workshops [19].

The procedure of constructing the uncertainty estimate for the FS method builds on the CF method, but with significant improvements: (1) $P$ is used instead of the correction factor as the distance metric to the asymptotic range; (2) an improved error estimate $P\delta_{RE}$ is used; and (3) three factor of safety coefficients at $P = 0$ ($FS_0$), 1 ($FS_1$), and 2 ($FS_2$) are used instead of two factor of safety coefficients. The addition of $FS_2$ enables the removal of the symmetry assumption for the uncertainty estimate used in the
two earlier versions of the FS method; thereby extending the applicability of the FS method for $P > 2$. The uncertainty estimate for the FS method is

$$U_{FS} = FS(P) |\delta_{RE}| = \begin{cases} \left[ FS_{1}P + FS_{0}(1-P) \right] |\delta_{RE}| & 0 < P \leq 1 \\ \left[ FS_{1}P + FS_{2}(P-1) \right] |\delta_{RE}| & P > 1 \end{cases}$$  \hspace{1cm} (14)

The uncertainty estimate for the GCI method is obtained by multiplying the $|\delta_{RE}|$ by a constant factor of safety, either 1.25 or 3. If the same approach is applied for the best error estimate $P|\delta_{RE}|$, then $FS = 1.25P$ or $FS = 3P$. Both approaches are unacceptable for $P < 1$ since $FS = 0$ at $P = 0$ and $FS < 1$ for certain $P$ ranges. For $P > 1$, the factors of safety of both approaches increase linearly with different slopes as $P$ increases, which are not conservative enough.

**D. Confidence Levels**

For the FS method, recommended values of $FS_{0}$, $FS_{1}$, and $FS_{2}$ are determined using statistical analysis for a large number of samples based on analytical or numerical benchmarks. The procedure is to determine the smallest values of the three coefficients until two criteria are met, i.e., reliability is larger than ninety-five percent and the lower confidence limit at the ninety-five percent confidence level is larger than 1.2 for all samples. As a result, $FS_{0} = 2.45$, $FS_{1} = 1.6$, and $FS_{2} = 14.8$ are recommended and the final form of the FS method is:

$$U_{FS} = FS(P) |\delta_{RE}| = \begin{cases} \left( 2.45 - 0.85P \right) |\delta_{RE}| & 0 < P \leq 1 \\ \left( 16.4P - 14.8 \right) |\delta_{RE}| & P > 1 \end{cases}$$  \hspace{1cm} (15)

For comparison purposes, the statistical analyses for the other four methods are also presented. The factors of safety for different methods are shown in Fig. 3.

**III. STATISTICAL ANALYSIS**

Statistical analysis is based on twenty-five samples that consist of actual factor of safety items $FS_{i} (i = 1, N)$ with different sample sizes $N$ ranging from five to three
hundred and twenty-nine. The actual factor of safety for the \( i^{th} \) grid-triplet study of a sample is defined as the ratio of the uncertainty estimate \( U_i \) to the magnitude of \( E_i \):

\[
FS_A = \frac{U_i}{|E_i|}
\]  

(16)

where \( U \) is defined by Eqs. (10), (11), (12), (13), and (15) for the GCI, GCI1, GCI2, CF, and FS methods, respectively. The comparison error \( E_i \) for fine grid solution \( S_{i} \) is

\[
E_i = S_{AB} - S_{i}
\]  

(17)

for an analytical benchmark and

\[
E_i = S_{NB} - S_{i}
\]  

(18)

for a numerical benchmark where \( S_{NB} \) is either the solution on the finest grid or the solution on a very fine grid using a very high order numerical method, which has been considered the “exact” or reference solution. Similar to the use of the effectivity index to evaluate the accuracy of different error estimates, \( FS_A \) can be used as an index to evaluate the conservativeness of \( U \) for different verification methods.

The error and uncertainty estimates are systematic, but \( E_i \) and therefore \( FS_A \) are treated as items drawn from the statistical and random parent population of possible systematic errors, which are similar to the systematic error in experimental fluid dynamics. It is assumed that there are no correlated systematic errors between the different grid-triplet studies. The statistical results suggest that this is a reasonable assumption. Since \( FS_A \) are randomly distributed, the confidence interval for the mean reveals how close the mean value of \( FS_A \), \( \overline{FS_A} \), is to the true mean \( \mu \) of the parent population of \( FS_A \).

**A. Reliability**

Reliability \( R \) is defined as

\[
R = \frac{\sum_{i=1}^{N} \text{number of } FS_A > 1}{N}
\]

(19)
The reliability used in [8], $R_m = 1 - |\text{estimated fraction for } FS_A > 1 - \text{expected fraction for } FS_A > 1|$, is deficient as it does not discriminate between estimated fractions with the same distance below or above the expected fraction.

**B. Confidence of the Mean Analysis**

The confidence of mean analysis is based on the methodology and procedures summarized in [20]. If $X_i (i = 1, N)$ is the $i^{th}$ item of the sample with size $N$, the mean, the standard deviation, and the standard deviation of the mean of the sample are $\overline{X}$, $S_{X_i}$, and $S_{\overline{X}}$ respectively.

The true mean $\mu$ of the parent population at the ninety-five percent confidence level is bounded by $\overline{X} - k$ and $\overline{X} + k$:

$$\text{Pr}(\overline{X} - k \leq \mu \leq \overline{X} + k) \geq 0.95$$  \hspace{1cm} (20)

where $k$ is evaluated using the student t-distribution to account for the effect of a limited number of items,

$$k = tS_{\overline{X}}$$  \hspace{1cm} (21)

The lower confidence limit of the mean is defined by:

$$LCL = \overline{X} - k$$  \hspace{1cm} (22)

**C. Implementation**

The twenty-five samples are based on seventeen studies that have analytical or numerical benchmarks and cover fluids, thermal, and structure disciplines, as summarized in Table 2. The seventeen studies have ninety-eight variables. The largest sample 3 is for the actual factors of safety that are obtained by combining the three hundred and twenty-nine grid-triplet studies that show monotonic convergence from the seventeen studies. The other samples are obtained by combining subsets of sample 3 items for different studies, variables, $P$ ranges, and single $P$ values. For each sample,
the reliability, mean value $\bar{X}$, coefficient of variation $S_{\bar{X}}(\% \bar{X})$, and lower confidence limit are evaluated. It should be noted that the statistical results presented in Section 5 are for the actual factor of safety, i.e., $FS_x$ in Eq. (16) is equivalent to $X_i$ such that $\bar{X} = FS_x$, $S_{\bar{X}} = S_{FS_x}$, $S_X = S_{FS_x}$. For convenience of presentation and discussion, both of the nomenclatures are used interchangeably.

Table 2 Analytical and numerical benchmark verification studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Geometry</th>
<th>Conditions</th>
<th>Verification variables</th>
<th>Number of grids</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1D wave [6]</td>
<td>Wave profile</td>
<td>Maximum/minimum of stream-function, vorticity</td>
<td>10</td>
<td>AB</td>
</tr>
<tr>
<td>2</td>
<td>2D Laplace [22]</td>
<td>-</td>
<td>Arbitrary function</td>
<td>26</td>
<td>AB</td>
</tr>
<tr>
<td>3</td>
<td>2D driven cavity [23]</td>
<td>Re = 1000</td>
<td>Maximum or monitored velocity, location, temperature, and Nu</td>
<td>4</td>
<td>NB [24]</td>
</tr>
<tr>
<td>4</td>
<td>2D natural convection flows in square cavities [25]</td>
<td>Ra = 10^4</td>
<td>Maximum or monitored velocity, location, temperature, and Nu</td>
<td>5</td>
<td>NB (S1)</td>
</tr>
<tr>
<td>5</td>
<td>2D natural convection flows in square cavities [25]</td>
<td>Ra = 10^5</td>
<td>Maximum or monitored velocity, location, temperature, and Nu</td>
<td>6</td>
<td>NB (S1)</td>
</tr>
<tr>
<td>6</td>
<td>2D natural convection flows in square cavities [25]</td>
<td>Ra = 10^6</td>
<td>Maximum or monitored velocity, location, temperature, and Nu</td>
<td>7</td>
<td>NB (S1)</td>
</tr>
<tr>
<td>7</td>
<td>Backward-facing step [26]</td>
<td>Re = 1.5 x 10^5</td>
<td>Reattachment length, velocity</td>
<td>7</td>
<td>NB [27, 28]</td>
</tr>
<tr>
<td>8</td>
<td>2D driven cavity [29]</td>
<td>Re = 100</td>
<td>Velocity</td>
<td>5</td>
<td>NB [30]</td>
</tr>
<tr>
<td>9</td>
<td>2D driven cavity [29]</td>
<td>Re = 1000</td>
<td>Velocity</td>
<td>5</td>
<td>NB [30]</td>
</tr>
<tr>
<td>10</td>
<td>3D cubic cavity [29]</td>
<td>Re = 100</td>
<td>Velocity</td>
<td>4</td>
<td>NB [30]</td>
</tr>
<tr>
<td>11</td>
<td>Axisymmetric turbulent flow through a valve [31, 32]</td>
<td>Re = 10^5</td>
<td>Velocity, TKE, epsilon</td>
<td>5</td>
<td>NB (S1)</td>
</tr>
<tr>
<td>12</td>
<td>1D steady-state convection-diffusion [33]</td>
<td>Pe = 1 and Pe = 10</td>
<td>Arbitrary function</td>
<td>6</td>
<td>AB</td>
</tr>
<tr>
<td>13</td>
<td>Isothermal cylinder enclosed by a square duct [29, 30, 31]</td>
<td>Ra = 10^6, Pr = 10</td>
<td>Velocity, temperature</td>
<td>5</td>
<td>NB (S1)</td>
</tr>
<tr>
<td>14</td>
<td>Premixed methane/air laminar flat flame on a perforated burner [31, 34, 35]</td>
<td>Inlet temperature 298.2K</td>
<td>Velocities, temperature, mass fraction</td>
<td>7</td>
<td>NB (S1)</td>
</tr>
<tr>
<td>15</td>
<td>Data for “exact” grid convergence set [8]</td>
<td>contrived</td>
<td>-</td>
<td>7</td>
<td>AB</td>
</tr>
<tr>
<td>16</td>
<td>Beam bending problem for 2nd series [8]</td>
<td>2nd series</td>
<td>Beam bending stress, Beam end deflection</td>
<td>7 with 5 systematically refined</td>
<td>AB</td>
</tr>
<tr>
<td>17</td>
<td>Beam bending problem for 3rd series [8]</td>
<td>3rd series</td>
<td>Beam bending stress, Beam end deflection</td>
<td>4</td>
<td>AB</td>
</tr>
</tbody>
</table>
Sample 1 is for the seventeen studies, as shown in Table 3. The actual factor of safety for each study is obtained by averaging all actual factors of safety for that study, which may involve different variables and $P$ values. Sample 2 is for the ninety-eight various verification variables, as also shown in Table 3. The verification variables are extracted from the seventeen studies and may involve different numbers of grid-triplet studies. Sample 3 includes all the three hundred and twenty-nine grid-triplet studies covering the largest $P$ range, as shown in Table 4. To evaluate the behavior of the five verification methods at different distances to the asymptotic range, the samples 4 to 8 are for five different $P$ ranges, as also shown in Table 4. The ranges are selected such that a range very close to the asymptotic range (sample 6) and four ranges far from the asymptotic range (samples 4, 5, 7, and 8) are covered with sufficiently large sample sizes.

To evaluate the statistics at individual $P$ values, an averaging process is performed using a tolerance $\Delta P = 0.01$ for sample 3 such that $P$ values are regarded to be the same if the difference between any two $P$ values is less than 0.01. A smaller $\Delta P$ value 0.005 is shown to have limited effects on the statistical results. As a result, there are seventeen individual $P$ values (samples 9 to 25) ranging from 0.705 to 1.205 where at least five items are available after removing the outliers, as shown in Table 5.

### Table 3 Statistics of the (17) case studies and all (98) variables

<table>
<thead>
<tr>
<th>Sample</th>
<th>Factor of safety</th>
<th>Average $\bar{P}$</th>
<th>N</th>
<th>Statistics</th>
<th>GCI</th>
<th>GCI$_1$</th>
<th>GCI$_2$</th>
<th>CF</th>
<th>FS</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Means of studies</td>
<td>0.98</td>
<td>17</td>
<td>R (%N)</td>
<td>94.1</td>
<td>94.1</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>1.746</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\bar{X}$</td>
<td>1.64</td>
<td>1.75</td>
<td>2.42</td>
<td>2.31</td>
<td>3.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$S_{\bar{X}}$</td>
<td>9.1</td>
<td>8.6</td>
<td>8.7</td>
<td>13.0</td>
<td>10.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>1.38</td>
<td>1.49</td>
<td>2.05</td>
<td>1.79</td>
<td>2.60</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Means of all variables of 17 studies</td>
<td>0.96</td>
<td>98</td>
<td>R (%N)</td>
<td>89.8</td>
<td>92.9</td>
<td>96.9</td>
<td>92.9</td>
<td>100</td>
<td>1.663</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\bar{X}$</td>
<td>1.64</td>
<td>1.75</td>
<td>2.45</td>
<td>2.28</td>
<td>3.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$S_{\bar{X}}$</td>
<td>5.5</td>
<td>5.1</td>
<td>4.9</td>
<td>7.9</td>
<td>6.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>1.49</td>
<td>1.60</td>
<td>2.25</td>
<td>1.98</td>
<td>2.82</td>
<td></td>
</tr>
</tbody>
</table>

16
Table 4 Statistics for different ranges of $P$ values using non-averaged actual factor of safety

<table>
<thead>
<tr>
<th>Sample</th>
<th>$P$</th>
<th>N</th>
<th>Statistics</th>
<th>GCI</th>
<th>GCI&lt;sub&gt;1&lt;/sub&gt;</th>
<th>GCI&lt;sub&gt;2&lt;/sub&gt;</th>
<th>CF</th>
<th>FS</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0–2</td>
<td>(0.94)</td>
<td>R (%N)</td>
<td>83.9</td>
<td>90.3</td>
<td>94.2</td>
<td>90.0</td>
<td>96.96</td>
<td>1.645</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\bar{X}$</td>
<td>1.67</td>
<td>1.75</td>
<td>2.39</td>
<td>2.29</td>
<td>3.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$S_\bar{X}$(%$\bar{X}$)</td>
<td>7.8</td>
<td>7.4</td>
<td>5.9</td>
<td>13.1</td>
<td>8.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>1.46</td>
<td>1.54</td>
<td>2.16</td>
<td>1.80</td>
<td>2.63</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0–0.4</td>
<td>(0.24)</td>
<td>R (%N)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>1.796</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\bar{X}$</td>
<td>7.34</td>
<td>7.34</td>
<td>7.34</td>
<td>16.40</td>
<td>13.52</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$S_\bar{X}$(%$\bar{X}$)</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>31.3</td>
<td>30.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>3.51</td>
<td>3.51</td>
<td>3.51</td>
<td>7.19</td>
<td>6.03</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.4–0.9</td>
<td>(0.70)</td>
<td>R (%N)</td>
<td>91.4</td>
<td>91.4</td>
<td>91.4</td>
<td>95.1</td>
<td>95.1</td>
<td>1.664</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\bar{X}$</td>
<td>2.01</td>
<td>2.01</td>
<td>2.01</td>
<td>3.07</td>
<td>3.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$S_\bar{X}$(%$\bar{X}$)</td>
<td>6.5</td>
<td>6.5</td>
<td>6.5</td>
<td>8.8</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>1.79</td>
<td>1.79</td>
<td>1.79</td>
<td>2.62</td>
<td>2.66</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.9–1.1</td>
<td>(0.98)</td>
<td>R (%N)</td>
<td>92.6</td>
<td>93.8</td>
<td>95.5</td>
<td>92.0</td>
<td>97.7</td>
<td>1.646</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\bar{X}$</td>
<td>1.34</td>
<td>1.36</td>
<td>1.88</td>
<td>1.25</td>
<td>1.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$S_\bar{X}$(%$\bar{X}$)</td>
<td>2.2</td>
<td>2.2</td>
<td>3.7</td>
<td>2.4</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>1.29</td>
<td>1.31</td>
<td>1.76</td>
<td>1.20</td>
<td>1.80</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.1–1.5</td>
<td>(1.19)</td>
<td>R (%N)</td>
<td>48</td>
<td>76</td>
<td>94</td>
<td>72</td>
<td>96</td>
<td>1.676</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\bar{X}$</td>
<td>1.04</td>
<td>1.36</td>
<td>3.27</td>
<td>1.35</td>
<td>3.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$S_\bar{X}$(%$\bar{X}$)</td>
<td>7.03</td>
<td>6.45</td>
<td>6.45</td>
<td>6.39</td>
<td>6.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>0.92</td>
<td>1.22</td>
<td>2.92</td>
<td>1.21</td>
<td>3.31</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.5–2.0</td>
<td>(1.63)</td>
<td>R (%N)</td>
<td>30</td>
<td>80</td>
<td>90</td>
<td>90</td>
<td>100</td>
<td>1.833</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\bar{X}$</td>
<td>0.86</td>
<td>1.70</td>
<td>4.08</td>
<td>2.03</td>
<td>7.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$S_\bar{X}$(%$\bar{X}$)</td>
<td>19.8</td>
<td>18.7</td>
<td>18.7</td>
<td>18.4</td>
<td>18.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>0.55</td>
<td>1.12</td>
<td>2.68</td>
<td>1.35</td>
<td>5.08</td>
<td></td>
</tr>
</tbody>
</table>

Table 5 Statistics excluding outliers at seventeen $P$ values

<table>
<thead>
<tr>
<th>Sample</th>
<th>$P$</th>
<th>N</th>
<th>Number of outliers</th>
<th>Statistics</th>
<th>GCI</th>
<th>GCI&lt;sub&gt;1&lt;/sub&gt;</th>
<th>GCI&lt;sub&gt;2&lt;/sub&gt;</th>
<th>CF</th>
<th>FS</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0.705</td>
<td>9</td>
<td>0</td>
<td>R (%N)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>1.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\bar{X}$</td>
<td>2.08</td>
<td>2.08</td>
<td>2.08</td>
<td>3.05</td>
<td>3.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$S_\bar{X}$(%$\bar{X}$)</td>
<td>10.4</td>
<td>10.4</td>
<td>10.4</td>
<td>11.5</td>
<td>10.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>1.68</td>
<td>1.68</td>
<td>1.68</td>
<td>2.40</td>
<td>2.48</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.755</td>
<td>5</td>
<td>0</td>
<td>R (%N)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>2.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\bar{X}$</td>
<td>1.87</td>
<td>1.87</td>
<td>1.87</td>
<td>2.66</td>
<td>2.71</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$S_\bar{X}$(%$\bar{X}$)</td>
<td>11.2</td>
<td>11.2</td>
<td>11.2</td>
<td>11.3</td>
<td>11.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>1.43</td>
<td>1.43</td>
<td>1.43</td>
<td>2.02</td>
<td>2.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R (%N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>11</td>
<td>0.805</td>
<td>15</td>
<td>0</td>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.76</td>
<td>1.76</td>
<td>1.76</td>
<td>2.25</td>
<td>2.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.4</td>
<td>5.4</td>
<td>5.4</td>
<td>6.0</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.59</td>
<td>1.59</td>
<td>2.01</td>
<td>2.26</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.855</td>
<td>6</td>
<td>0</td>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.58</td>
<td>1.58</td>
<td>1.58</td>
<td>1.83</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.6</td>
<td>9.6</td>
<td>10.3</td>
<td>9.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.27</td>
<td>1.27</td>
<td>1.45</td>
<td>1.76</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.905</td>
<td>29</td>
<td>0</td>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.61</td>
<td>1.61</td>
<td>1.61</td>
<td>1.68</td>
<td>2.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.8</td>
<td>6.8</td>
<td>6.7</td>
<td>6.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.42</td>
<td>1.42</td>
<td>1.42</td>
<td>1.49</td>
<td>1.92</td>
</tr>
<tr>
<td>14</td>
<td>0.925</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.42</td>
<td>1.42</td>
<td>1.42</td>
<td>1.38</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.1</td>
<td>3.1</td>
<td>4.8</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.33</td>
<td>1.33</td>
<td>1.24</td>
<td>1.76</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.945</td>
<td>6</td>
<td>0</td>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.37</td>
<td>1.37</td>
<td>1.37</td>
<td>1.27</td>
<td>1.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.32</td>
<td>1.32</td>
<td>1.32</td>
<td>1.20</td>
<td>1.73</td>
</tr>
<tr>
<td>16</td>
<td>0.955</td>
<td>16</td>
<td>0</td>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.42</td>
<td>1.42</td>
<td>1.42</td>
<td>1.31</td>
<td>1.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.8</td>
<td>6.8</td>
<td>6.8</td>
<td>6.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.25</td>
<td>1.25</td>
<td>1.14</td>
<td>1.64</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.965</td>
<td>6</td>
<td>0</td>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.46</td>
<td>1.46</td>
<td>1.46</td>
<td>1.32</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.1</td>
<td>7.1</td>
<td>7.1</td>
<td>7.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.25</td>
<td>1.25</td>
<td>1.25</td>
<td>1.13</td>
<td>1.63</td>
</tr>
<tr>
<td>18</td>
<td>0.975</td>
<td>7</td>
<td>0</td>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.33</td>
<td>1.33</td>
<td>1.33</td>
<td>1.19</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.8</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.24</td>
<td>1.24</td>
<td>1.24</td>
<td>1.10</td>
<td>1.61</td>
</tr>
<tr>
<td>19</td>
<td>0.985</td>
<td>9</td>
<td>0</td>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.27</td>
<td>1.27</td>
<td>1.27</td>
<td>1.13</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.25</td>
<td>1.25</td>
<td>1.25</td>
<td>1.10</td>
<td>1.60</td>
</tr>
<tr>
<td>20</td>
<td>0.995</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>87.5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.24</td>
<td>1.24</td>
<td>1.24</td>
<td>1.09</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.19</td>
<td>1.19</td>
<td>1.19</td>
<td>1.05</td>
<td>1.53</td>
</tr>
</tbody>
</table>

18
<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>1.005</td>
<td>56</td>
<td>4</td>
<td>R (%N)</td>
<td>98.2</td>
<td>98.2</td>
<td>98.2</td>
<td>98.2</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>1.32</td>
<td>1.33</td>
<td>2.08</td>
<td>1.16</td>
<td>1.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S_x (%)X</td>
<td>3.0</td>
<td>3.0</td>
<td>5.6</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>1.26</td>
<td>1.26</td>
<td>1.88</td>
<td>1.11</td>
<td>1.63</td>
</tr>
<tr>
<td>22</td>
<td>1.015</td>
<td>6</td>
<td>0</td>
<td>R (%N)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>83.3</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>1.18</td>
<td>1.21</td>
<td>2.90</td>
<td>1.04</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S_x (%)X</td>
<td>2.9</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>1.11</td>
<td>1.14</td>
<td>2.73</td>
<td>0.98</td>
<td>1.62</td>
</tr>
<tr>
<td>23</td>
<td>1.055</td>
<td>6</td>
<td>0</td>
<td>R (%N)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>1.66</td>
<td>1.83</td>
<td>4.39</td>
<td>1.59</td>
<td>3.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S_x (%)X</td>
<td>9.0</td>
<td>8.9</td>
<td>8.9</td>
<td>8.8</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>1.36</td>
<td>1.50</td>
<td>3.60</td>
<td>1.31</td>
<td>2.68</td>
</tr>
<tr>
<td>24</td>
<td>1.105</td>
<td>13</td>
<td>2</td>
<td>R (%N)</td>
<td>61.5</td>
<td>76.9</td>
<td>100</td>
<td>61.5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>1.32</td>
<td>1.52</td>
<td>3.65</td>
<td>1.38</td>
<td>3.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S_x (%)X</td>
<td>14.4</td>
<td>14.1</td>
<td>14.1</td>
<td>13.9</td>
<td>14.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>0.98</td>
<td>1.14</td>
<td>2.73</td>
<td>1.04</td>
<td>2.56</td>
</tr>
<tr>
<td>25</td>
<td>1.205</td>
<td>12</td>
<td>1</td>
<td>R (%N)</td>
<td>58.3</td>
<td>83.3</td>
<td>100</td>
<td>83.3</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>1.04</td>
<td>1.38</td>
<td>3.32</td>
<td>1.38</td>
<td>4.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S_x (%)X</td>
<td>8.9</td>
<td>8.9</td>
<td>8.9</td>
<td>9.1</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LCL</td>
<td>0.88</td>
<td>1.16</td>
<td>2.79</td>
<td>1.16</td>
<td>3.43</td>
</tr>
</tbody>
</table>

The number of outliers for samples 9 to 25 is summarized in Table 5. The outliers are identified using the Peirce’s Criterion [21]. Compared to the Chauvenet’s criterion, the Peirce’s criterion is more rigorous, does not make an arbitrary assumption concerning the rejection of data, and theoretically accounts for the case where there is more than one suspect data.

The optimal values of $F_{S_0}$, $F_{S_1}$, and $F_{S_2}$ are obtained iteratively by increasing their values from zero until two criteria are satisfied. The first criterion is that for all the twenty-five samples a ninety-five percent reliability is achieved for $F_{S_1} > 1$,

$$R > 95\%$$  \hspace{1cm} (23)

The second criterion is that the lower confidence limit for $F_{S_1}$ at the ninety-five percent confidence level is larger than a specified minimum factor of safety $F_{S_{\min}}$. For practical applications, $F_{S_{\min}}$ values are determined based on risk, reliability, accuracy, and cost, which show a large variation from 1.2 for new bridges and road marks to 4.0 for pressure
vessels. It is reasonable to choose $FS_{\text{min}} = 1.2$ for the present application. The second criterion becomes

$$\Pr(LCL \geq 1.2) \geq 0.95$$

(24)

When the two criteria are met, the smallest values of the three factor of safety coefficients are accepted. When Eq. (24) is satisfied

$$\Pr\left(\bar{X} \geq 1.2\right) \geq 0.95$$

(25)

$$\Pr(\mu \geq 1.2) \geq 0.95$$

(26)

IV. ANALYTICAL AND NUMERICAL BENCHMARKS

The seventeen studies are obtained from published journal or conference proceedings for which either analytical or numerical benchmarks for various verification variables are available. The seventeen studies cover fluid, thermal and structure disciplines. Geometry, conditions, verification variables, number of grids, and indication of analytical or numerical benchmark are summarized in Table 2.

The analytical benchmarks are for the one-dimensional wave, two-dimensional Laplace and one-dimensional steady-state convection-diffusion equations and contrived and beam-bending structures problems. The fluid and thermal studies are for different steady-state flows for two-dimensional and three-dimensional geometries and physical phenomena: laminar flows, turbulent flows using the two-equation ($k$-epsilon) turbulence model and reactive flows. The analytical studies used first- and second-order upwind schemes for the convection terms. The fluid and thermal studies used second-order central-difference schemes for the diffusion terms and first-order upwind or high-order SMART schemes for the convective terms. The types of boundary conditions include constant or non-constant Dirichlet (inlet, wall) and Neumann (axisymmetric, outlet, zero-gradient). The same grid refinement ratio $r = 2$ was used to generate the systematic grids except for the two-dimensional Laplace equation where a 4th root of 2 was also used. All solutions were reported to have achieved iterative convergence. The solutions in total represent three hundred and thirty-nine grid-triplet studies. For each grid-triplet study, a grid convergence study is conducted using Eq. (2), which resulted in a total of three-
hundred and twenty-nine solutions that show monotonic convergence for which \( P_{RE}, \delta_{RE}, \) and \( S_C \) are evaluated using Eqs. (3), (4), and (5), respectively. The remaining ten solutions were either oscillatory or monotonically divergent.

The manner in which the ninety-eight verification variables approach the asymptotic range as the grids are refined is evaluated for each verification variable by the convergence characteristics of \( P \) and \( |E| \) as functions of \( \Delta x_{fine} / \Delta x_{finest} \), where \( \Delta x_{fine} \) is the fine grid spacing for an individual grid-triplet study and \( \Delta x_{finest} \) is the finest grid spacing. Monotonic convergence is defined by \( P \) approaching one monotonically, as the grids are refined. Oscillatory convergence is defined by \( P \) approaching one with oscillations, as the grids are refined. In both cases, \( |E| \) approaches zero monotonically. Determination of the convergence characteristics requires a minimum of three grid-triplet studies. Ten verification variables only have one grid-triplet study with an average \( P \) value (\( \bar{P} \)) for the grid-triplet studies for the finest grids \( \bar{P}_{finest} = 1.026 \). Thirty-three verification variables have two grid-triplet studies with an average \( \bar{P}_{finest} = 0.95 \); however, the errors decrease as the grids are refined. The other fifty-five verification variables have multiple grid-triplet studies ranging from three to twenty-four and show monotonic or oscillatory convergence. Seven verification variables show monotonic convergence with an average \( \bar{P}_{finest} = 0.997 \) covering studies with \( S_{AB} \) (studies 1 and 2) and \( S_{NB} \) (studies 5, 6, and 14). The remaining forty-eight verification variables show oscillatory convergence with an average \( \bar{P}_{finest} = 0.996 \). The complete results are provided in Appendix 1.

**V. ANALYTICAL AND NUMERICAL BENCHMARK DATA**

The actual factor of safety for sample 3, sample 3 averaged using \( \Delta P = 0.01 \), and the upper and lower band of the confidence interval \( \tilde{FS} \pm tS_{FS} \) for samples 9 to 25 are shown in Fig. 4. The three GCI methods are fairly conservative for \( P < 0.9 \) where they show a ninety-three percent reliability. The CF and FS methods are more conservative by
Figure 4 Actual factor of safety for sample 3, sample 3 averaged using $\Delta P = 0.01$, and $\bar{FS}_j \pm t_{FS} S_{FS}$ for samples 9 to 25: (a) GCI method, (b) GCI$_1$ method, (c) GCI$_2$ method, (d) CF method, (e) FS method.
showing a ninety-six percent reliability in this range. Near the asymptotic range when \(0.9 < P < 1.1\), the GCI method has a ninety-three percent reliability, which is a little larger than the CF method but smaller than the other three verification methods for which the FS method is the most conservative with a ninety-eight percent reliability. For \(P > 1.1\), the GCI method is the least conservative with a reliability of forty-five percent. The GCI1 method is almost as conservative as the CF method and both show reliability around seventy-six percent. The GCI2 method is much more conservative than the GCI, GCI1, and CF methods with a reliability of ninety-three percent. Only the FS method shows a reliability larger than ninety-six percent in this range. The GCI, GCI1, and CF methods are not conservative enough near the asymptotic range and far from the asymptotic range when \(P\) is larger than 1.1. This deficiency is partly resolved by the GCI2 method, but with a jump of actual factor of safety at \(P = 1\). The FS method completely resolves the deficiency and has a nearly “symmetric” distribution of actual factor of safety with respect to the asymptotic range. At those \(P\) values where the FS method shows actual factor of safety less than one, all the other verification methods show even smaller actual factor of safety. The FS method estimates uncertainties that bound the largest fraction of the error and provides a minimum lower confidence limit larger than 1.2 for samples 9 to 25. Overall, the magnitudes of the differences of the actual factors of safety between the different verification methods are consistent with the different magnitudes of factor of safety shown in Fig. 3.

Figures 5(a) and 5(b) show the standard deviation for the mean \(S_{\bar{X}}\) and coefficient of variation \(S_{\bar{X}}(\% \bar{X})\), respectively, for samples 9 to 25. The magnitudes of \(S_{\bar{X}}\) for the different verification methods are consistent with the magnitudes of the factors of safety shown in Fig. 3, i.e., larger factor of safety leads to larger \(S_{\bar{X}}\). As a result, larger differences of \(S_{\bar{X}}\) for the different verification methods are shown for \(P > 1\) than for \(0 < P < 1\). Nonetheless, the standard deviations for the mean decrease, as the asymptotic range is approached. The coefficient of variation is a normalized measure of the dispersion for samples 9 to 25. The different verification methods show small differences at each \(P\) value for \(S_{\bar{X}}(\% \bar{X})\) that decrease as the asymptotic range is approached.
Table 3 shows the statistics for samples 1 and 2. Based on the seventeen studies, \( \bar{P} \) is 0.98. The GCI\(_2\), CF, and FS methods achieve one hundred percent reliability. The GCI and GCI\(_1\) methods achieve 94.1% reliability. The mean values of the actual factors of safety for the seventeen studies increase from the minimum value of 1.64 for the GCI method to the maximum value of 3.18 for the FS method. The minimum and maximum coefficients of variation are 8.6 and 13.0 for the GCI\(_1\) and CF methods, respectively. Based on the ninety-eight variables, \( \bar{P} \) is 0.96. The reliability for the GCI method is 89.8%. Only the FS and GCI\(_2\) methods have reliabilities larger than ninety-five percent. Compared to the statistics of sample 1, sample 2 shows that the mean values of the actual factors of safety for all the verification methods are almost the same; however, the coefficient of variation for sample 2 is much lower than that for sample 1 since sample 2 has a much larger sample size. For all the verification methods, the lower confidence limits are larger than 1.2 for both sample 1 and sample 2.

In order to highlight the dependence of the different verification methods on different \( P \) ranges, Table 4 shows the statistics for samples 3 to 8 based on six different \( P \) ranges. For sample 3 (\( 0 < P < 2 \)), only the FS method achieves reliability larger than ninety-five percent. The GCI method achieves the lowest reliability at 83.9%. For sample
4 (0 < P < 0.4), all verification methods show a one hundred percent reliability. For sample 5 (0.4 ≤ P < 0.9), the three GCI methods achieve a 91.4% reliability and the CF and FS methods achieve a reliability larger than ninety-five percent. For sample 6 (0.9 ≤ P < 1.1), only the GCI and FS methods achieve a ninety-five percent reliability. Examination of 18.2% of the data for 1.1 ≤ P < 2.0 which cover samples 7 and 8 shows that only the FS method achieves a ninety-five percent reliability. The reliability for the GCI method is forty-five percent in this range. Based on the means, the most conservative verification methods for samples (3, 7, 8), samples (4, 5), and sample (6) are the FS, CF, and GCI methods, respectively. For sample 3, the lower confidence limit is larger than 1.2 for all the verification methods; however, only the FS method satisfies that the lower confidence limit is larger than 1.2 for all the P ranges.

In order to evaluate the performance of the different verification methods at individual P values, Table 5 shows the statistics at the seventeen P values (samples 9 to 25) ranging from 0.705 to 1.205. For samples 9 to 19 (P < 0.99), all the verification methods achieve one hundred percent reliability except at P = 0.955 where only the FS method achieves such a level. For samples 20 to 23 (P ≈ 1), all the verification methods achieve reliabilities larger than ninety-five percent except the CF method for samples 20 and 22. For samples 24 and 25 (P > 1), only the GCI and FS methods achieve reliabilities larger than ninety-five percent. Based on the means, the most conservative verification method for samples (9 to 20, 25) and samples 21 to 24 are the FS and GCI methods, respectively. Only the FS method satisfies the requirement that the lower confidence limit is larger than 1.2 for samples 9 to 25. The samples for which the lower confidence limits are less than 1.2 are 20, 22, 24, and 25 for the GCI and GCI methods, 20 for the GCI method, and 16 to 25 except 23 for the CF method. The results indicate that the ranges of P values where the verification methods are not conservative enough are P ≥ 1 for the GCI, GCI, and CF methods and P ≈ 1 for the GCI method.

VI. EXAMPLE FOR SHIP HYDRODYNAMICS APPLICATIONS
To evaluate the performance of the different verification methods for practical applications, the verification study by Xing et al. [36] is used. The application is the development of computational towing tank procedures for single run curves of resistance and propulsion for the high-speed transom ship Athena with and without appendages. Verification variables are the total resistance coefficient $C_{TX}$, sinkage, and trim. Seven grids are systematically generated using $r = 2^{0.25}$, i.e., from the coarsest grid 7 with 360,528 points to the finest grid 1 with 8.1 million points. This allows nine grid-triplet studies with five for $r = 2^{0.25}$ (5, 6, 7; 4, 5, 6; 3, 4, 5; 2, 3, 4; and 1, 2, 3), three for $r = 2^{0.5}$ (3, 5, 7; 2, 4, 6; and 1, 3, 5), and one for $r = 2^{0.75}$ (1, 4, 7). The average iterative errors are at least one order of magnitude smaller than the grid uncertainty $U_G$ except for the two finest grids, which indicates that separating the iterative error and $U_G$ for fine grids can be problematic. The convergence studies show that the motions converge more slowly than the resistance. Six grid-triplet studies show monotonic convergence for $C_{TX}$, whereas only two and four grid-triplet studies show monotonic convergence for the sinkage and trim, respectively.

All the grid-triplet studies that show monotonic convergence are presented in Table 6 for the total resistance coefficient $C_{TX}$ and Table 7 for the sinkage and trim. Overall $P$ values vary greatly, which is typical for industrial applications. When $P$ is less than one, the three GCI methods estimate the same $U_G$, which are smaller than the $U_G$ estimated by the CF and FS methods. The CF method is more conservative than the FS method for $P < 0.5$ and less conservative than the FS method when $0.5 < P < 1$. When $P$ is larger than one, but close to the asymptotic range, i.e., $P = 1.07$ for the trim on grids (1, 3, 5), the GCI$_2$ method is the most conservative, followed by the FS, GCI$_1$, GCI, and CF methods. When $P$ is much larger than 1, the FS method is the most conservative, followed by the GCI$_2$, CF, GCI$_1$, and GCI methods. Grid uncertainties estimated by the GCI, GCI$_1$, and CF methods are unreasonably small when $P > 1$ due to the deficiency already discussed, which is resolved using the GCI$_2$ and FS methods.
Table 6 Verification study for $C_{TX}$ of Athena bare hull with skeg at Froude number 0.48. $U_G$ is %$S_i$ and $C_{TX}$ is based on the static wetted area.

<table>
<thead>
<tr>
<th>Grids</th>
<th>Refinement Ratio</th>
<th>$R$</th>
<th>$P$</th>
<th>$U_G$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GCI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GCI1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GCI2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FS</td>
</tr>
<tr>
<td>2, 4, 6</td>
<td>$2^{0.5}$</td>
<td>0.63</td>
<td>0.66</td>
<td>3.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.04</td>
</tr>
<tr>
<td>1, 3, 5</td>
<td>$2^{0.5}$</td>
<td>0.40</td>
<td>1.33</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.02</td>
</tr>
<tr>
<td>4, 5, 6</td>
<td>$2^{0.25}$</td>
<td>0.97</td>
<td>0.08</td>
<td>54.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>54.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>54.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>125.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>104.2</td>
</tr>
<tr>
<td>3, 4, 5</td>
<td>$2^{0.25}$</td>
<td>0.80</td>
<td>0.63</td>
<td>4.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.62</td>
</tr>
<tr>
<td>2, 3, 4</td>
<td>$2^{0.25}$</td>
<td>0.60</td>
<td>1.49</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.30</td>
</tr>
<tr>
<td>1, 2, 3</td>
<td>$2^{0.25}$</td>
<td>0.50</td>
<td>2.00</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.22</td>
</tr>
</tbody>
</table>

Table 7 Verification study for motions of Athena bare hull with skeg at Froude number 0.48. $U_G$ is %$S_i$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Grids</th>
<th>Refinement Ratio</th>
<th>$R$</th>
<th>$P$</th>
<th>$U_G$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GCI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GCI1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GCI2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FS</td>
</tr>
<tr>
<td>Sinkage</td>
<td>1, 3, 5</td>
<td>$2^{0.5}$</td>
<td>0.31</td>
<td>1.70</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.73</td>
</tr>
<tr>
<td>Sinkage</td>
<td>2, 3, 4</td>
<td>$2^{0.25}$</td>
<td>0.13</td>
<td>6.01</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.49</td>
</tr>
<tr>
<td>Trim</td>
<td>1, 3, 5</td>
<td>$2^{0.5}$</td>
<td>0.48</td>
<td>1.06</td>
<td>4.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.74</td>
</tr>
<tr>
<td>Trim</td>
<td>4, 5, 6</td>
<td>$2^{0.25}$</td>
<td>0.86</td>
<td>0.45</td>
<td>24.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>42.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40.48</td>
</tr>
<tr>
<td>Trim</td>
<td>2, 3, 4</td>
<td>$2^{0.25}$</td>
<td>0.53</td>
<td>1.85</td>
<td>3.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>41.48</td>
</tr>
<tr>
<td>Trim</td>
<td>1, 2, 3</td>
<td>$2^{0.25}$</td>
<td>0.53</td>
<td>1.85</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21.62</td>
</tr>
</tbody>
</table>

VII. CONCLUSIONS

A factor of safety method for quantitative estimates of grid-spacing and time-step uncertainties for solution verification is developed. It removes the two deficiencies of the GCI and CF methods, namely, unreasonably small uncertainty when $p_{RE} > p_{ih}$ and lack of statistical evidence that the interval of uncertainty at the ninety-five percent confidence level bounds the comparison error. Different error estimates are evaluated using the effectivity index. The uncertainty estimate builds on the correction factor method, but with significant improvements. The FS method is validated using statistical analysis of
twenty-five samples with different sizes ranging from five to three hundred and twenty-nine based on seventeen studies covering fluids, thermal, and structure disciplines.

The statistical results show that only the FS method, compared to the GCI, GCI$_1$, GCI$_2$, and CF methods, provides a reliability larger than ninety-five percent and a lower confidence limit at the ninety-five percent confidence level larger than 1.2 for the true mean of the parent population of the actual factor of safety. This conclusion is true for different studies, variables, ranges of $P$ values, and single $P$ values where multiple actual factors of safety are available.

The statistical analysis is based on twenty-five samples of grid-triplet studies that show monotonic convergence with $S_{AB}$ or $S_{NB}$ solutions covering a wide range of $P$ values that is within or far from the asymptotic range. The $S_{AB}$ or $S_{NB}$ are required for validation in order to evaluate the comparison error and actual factor of safety for all the grid-triplet studies. The number of samples and items are large and range of $P$ values is wide such that the FS method is also valid for other applications including results not in the asymptotic range, which is typical in industrial and fluid engineering applications.

The confidence level in the results is based on the seventeen studies, ninety-eight variables and three hundred and twenty-nine grid-triplet studies that show monotonic convergence published in journal articles or conference proceedings. Further evaluation and development of the FS method should be performed by adding additional rigorous verification studies with $S_{AB}$ or $S_{NB}$ as they become available, especially those for industrial applications that demonstrate that the asymptotic range is achieved when grids are refined. This has two main benefits: (1) reduced $S_{\bar{X}}$ and thus $k$ for $\bar{X}$ at different $P$ values and (2) Chi-Square analysis for evaluation of the validity of the assumption of the student t-distribution.

The present statistical approach based on many analytical and numerical benchmarks provides a robust framework for developing solution verification methods. There are other unresolved issues and complex factors such as mixed numerical methods, coupled numerical and modeling errors for large-eddy and detached-eddy simulations and single-grid methods. Even though only a small fraction (2.1%) of the data are outliers, it is worthy to investigate the reason.
ACKNOWLEDGEMENT

This study was sponsored by the Office of Naval Research under Grant N00014-01-1-0073, administered by Dr. Patrick Purtell. The authors would like to thank Drs. Logan, Roache, Eça, Roy, Ray and anonymous reviewers for their insightful comments and criticisms.

NOMENCLATURE

AB analytical benchmark
CF correction factor
CFD computational fluid dynamics
$C_{TX}$ total resistance coefficient
$E$ comparison error
FS factor of safety
$FS_A$ actual factor of safety
GCI grid convergence index
$LCL$ lower confidence limit
$N$ sample size
NB numerical benchmark
$P = P_{RE}/P_{th}$ distance metric to the asymptotic range
$ar{P}$ average $P$ value
Pr probability
$P_{RE}$ estimated order of accuracy
$P_{th}$ theoretical order of accuracy
$R$ convergence ratio
$R$ reliability
$R_{im}$ reliability [8]
$r$ refinement ratio

$S_{AB}$ analytical benchmark solution

$S_C$ Richardson extrapolation numerical benchmark

$S_{NB}$ numerical benchmark solution

$S_i$ solution on the $i^{th}$ grid

$S_{X_i}$ standard deviation of the sample

$S_{\bar{X}}$ standard deviation of the mean

$S_{\bar{X}}\left(\% \bar{X}\right)$ coefficient of variation

$T$ true value

$U$ uncertainty

$U_G$ grid uncertainty

$\bar{X}$ mean value

$X^2_{r,*}$ reduced Chi-Square

$\Delta x$ grid spacing

$\Delta x_{fine}$ fine grid spacing for an individual grid-triplet study

$\Delta x_{finest}$ finest grid spacing for a verification variable

$\delta$ error estimate

$\epsilon$ solution change

$\mu$ true mean of the parent population

$\theta$ effectivity index

REFERENCES


Workshop on CFD Uncertainty Analysis, Instituto Superior Tecnico, Lisbon, October.


Appendix A. Analytical and Numerical Benchmarks

Study 1 One-Dimensional Wave Equation

As shown in Figure A.1, with the refinements of the grids, the errors monotonically decrease and solutions approach the asymptotic range as \( P \) changes from \( P = 0.56 \) on the coarsest grid to \( P = 1 \) on the finest grid. \( F_{SA} \) for the three GCI methods are the same as \( P < 1 \) and smaller than that predicted by the FS method. \( F_{SA} \) for the three GCI methods are smaller and larger than those predicted by the CF method for \( P < 0.875 \) and \( 0.875 < P < 1 \), respectively. \( F_{SA} \) of the FS method are smaller and larger than those predicted by the CF method for \( P < 0.6 \) and \( 0.6 < P < 1 \), respectively. These observations are consistent with the FS shown in Figure 3. All methods estimate \( F_{SA} \) larger than one. The average \( F_{SA} \) are 1.44, 1.44, 1.44, 1.61, and 1.99 for the GCI, GCI1, GCI2, CF, and FS methods, respectively. The standard deviations of \( F_{SA} \) are 0.27, 0.27, 0.27, 0.78, and 0.55 for the GCI, GCI1, GCI2, CF, and FS methods, respectively.

Study 2 Two-Dimensional Laplace Equation

As shown in Figure A.2, with the refinements of the grids, the errors monotonically decrease. Solutions approach the asymptotic range (\( P = 1 \)) monotonically for constant Dirichlet boundary conditions and oscillatorially for non-constant value Dirichlet boundary conditions. Comparison of \( F_{SA} \) magnitudes for different methods shows similar trend as discussed in study 1 and shown in Figure 3. All methods estimate \( F_{SA} \) larger than one. The average \( F_{SA} \) are 1.31, 1.32, 1.79, 1.21, and 1.91 for the GCI, GCI1, GCI2, CF, and FS methods, respectively. The standard deviations of \( F_{SA} \) are 0.15, 0.14, 0.71, 0.25, and 0.35 for the GCI, GCI1, GCI2, CF, and FS methods, respectively.

Study 3 Two-Dimensional Driven Cavity (Re=1,000)

As shown in Figure A.3, only two sets of grid study are available and thus it is difficult to draw any conclusions on the convergence characteristics as the grids are
refined. However, errors decrease with the refinements of the grids. Comparison of $FSA$ magnitudes for different methods shows similar trend as shown in Figure 3. All the verification methods except the GCI2 and FS methods estimate $FSA < 1$ for $P > 1$. The average $FSA$ are 1.13, 1.18, 1.57, 1.24, and 1.79 for the GCI, GCI1, GCI2, CF, and FS methods, respectively. The standard deviations of $FSA$ are 0.55, 0.49, 0.61, 0.77, and 0.62 for the GCI, GCI1, GCI2, CF, and FS methods, respectively.

**Study 4 Two-Dimensional Natural Convection Flows in Square Cavities ($Ra=10^4$)**

As shown in Figure A.4, all variables have 3 grid-triplet studies except $U_{mon}$ that has only two grid-triplet studies. When the grids are refined, all errors monotonically decrease and all variable solutions either linearly or oscillatorily approach the asymptotic range. Comparison of $FSA$ magnitudes for different methods shows similar trend as shown in Figure 1. $FSA$ oscillatorily decrease when $P$ is approaching the asymptotic range from $P < 1$. For $P > 1$, GCI2 and FS methods estimate higher $FSA$ than the other methods. For $P > 1.16$, uncertainty estimates using GCI do not bound the true error ($FSA < 1$) and GCI1 and CF have similar magnitudes of $FSA$ around 1.2 whereas the GCI2 and FS methods have much larger $FSA$. The average $FSA$ are 1.47, 1.54, 2.17, 1.69, and 2.42 for the GCI, GCI1, GCI2, CF, and FS methods, respectively. The standard deviations of $FSA$ are 0.41, 0.34, 0.76, 0.79, and 0.61 for the GCI, GCI1, GCI2, CF, and FS methods, respectively.

**Study 5 Two-Dimensional Natural Convection Flows in Square Cavities ($Ra=10^5$)**

As shown in Figure A.5, all variables have 4 grid-triplet studies. When the grids are refined, all errors monotonically decrease and all variable solutions either monotonically or oscillatorily approach the asymptotic range. Similar to the previous studies, uncertainty estimates using GCI, GCI1, and CF methods do not bound the true error for some points at $P > 1$ ($FSA < 1$) whereas the GCI2 and FS methods bound the true error in this range ($FSA > 1$). The average $FSA$ are 1.87, 1.96, 2.76, 2.76, and 3.28 for the GCI, GCI1, GCI2, CF, and FS methods, respectively. The standard deviations of $FSA$ are
1.83, 1.80, 1.82, 4.37, and 3.41 for the GCI, GCI₁, GCI₂, CF, and FS methods, respectively.

**Study 6 Two-Dimensional Natural Convection Flows in Square Cavities (Ra=10⁶)**

As shown in Figure A.6, the number of grid-triplet studies is 5 for all variables except 4 for Nu_{max} and 2 for V_{max}. When the grids are refined, all errors monotonically decrease and all variable solutions either monotonically or oscillatorially approach the asymptotic range. Similar to the previous studies, GCI, GCI₁, and CF methods do not bound the true error for some points for P > 1 whereas the GCI₂ and FS methods bound the true error. The average FS_A are 1.50, 1.56, 2.37, 1.96, and 2.46 for the GCI, GCI₁, GCI₂, CF, and FS methods, respectively. The standard deviations of FS_A are 1.1, 1.07, 1.14, 2.58, and 1.96 for the GCI, GCI₁, GCI₂, CF, and FS methods, respectively.

**Study 7 Two-Dimensional Backward-Facing Step**

As shown in Figure A.7, with the refinement of the grids, the errors monotonically decrease with several orders of magnitude larger than those in other studies and solutions approach the asymptotic range with huge oscillations of P. This suggests that the grids are still far from the asymptotic range for such a high Reynolds number flows. When P > 1.3, the GCI method does not bound the true error while all other methods bounded the true error with FS_A estimated by the FS method is the largest. The average FS_A are 1.16, 1.51, 2.96, 1.53, and 4.88 for the GCI, GCI₁, GCI₂, CF, and FS methods, respectively. The standard deviations of FS_A are 0.54, 0.27, 0.79, 0.20, and 2.32 for the GCI, GCI₁, GCI₂, CF, and FS methods, respectively.

**Study 8 Square Cavity with Moving Top Wall (Re=100)**

As shown in Figure A.8, 4 variables only have two grid-triplet studies and 3 variables have 3 grid-triplet studies, which suggests that further refinement of the grids are necessary. However, all variable errors linearly decrease as the grids are refined.
When $P > 1.4$, all methods except the GCI method estimate sufficient large uncertainty to bound the true errors. The average $F_{SA}$ are 1.45, 1.91, 3.80, 2.15, and 5.32 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively. The standard deviations of $F_{SA}$ are 0.68, 0.56, 1.32, 0.99, and 2.71 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively.

**Study 9 Square Cavity with Moving Top Wall (Re=1,000)**

As shown in Figure A.9, all variables only show monotonic convergence on two grid-triplet studies except 3 grid studies for the x velocity using UDS at 0 and 60 degrees and y velocity using UDS at 60 degrees. When the grids are refined, errors decrease. At $P = 1$ and 1.15, all methods fail to bound the true error except the FS method. The average $F_{SA}$ are 2.32, 2.40, 2.95, 4.07, and 4.37 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively. The standard deviations of $F_{SA}$ are 2.14, 2.08, 1.99, 5.05, and 3.78 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively.

**Study 10 Cubic Cavity with Moving Top Wall (Re=100)**

As shown in Figure A.10, only two grid-triplet studies are available. However, all errors decrease as the grids are refined. Uncertainty estimates estimated by all methods bound the true errors. The average $F_{SA}$ are 2.44, 2.57, 3.47, 3.38, and 4.69 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively. The standard deviations of $F_{SA}$ are 1.13, 1.0, 1.23, 2.14, and 1.60 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively.

**Study 11 Axisymmetric Turbulent Flow through a Valve**

As shown in Figure A.11, similar to study 10, only two grid-triplet studies are available and thus no conclusive behavior of the variables to approach the asymptotic range can be drawn. However, all errors decrease when the grids are refined. Uncertainty estimated by all methods bound the true errors. The average $F_{SA}$ are 3.17, 3.23, 3.87, 4.45, and 5.15 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively. The standard
deviations of $FS_A$ are 2.0, 2.0, 2.82, 4.45, and 3.72 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively.

**Study 12 One-Dimensional Steady-State Convection-Diffusion Process without Source Term**

As shown in Figure A.12, all variables have four grid-triplet studies. When the grids are refined, errors for all variables decrease and solutions are approaching the asymptotic range. Uncertainty estimates of all methods bound the true errors except at $P = 0.95$ where only the FS method bounds the true error. The average $FS_A$ are 1.50, 1.51, 1.63, 1.51, and 2.08 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively. The standard deviations of $FS_A$ are 0.34, 0.33, 0.55, 0.41, and 0.53 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively.

**Study 13 Heat Transfer from Isothermal Cylinder Enclosed by a Square Duct**

As shown in Figure A.13, only two grid-triplet studies are available for all the variables except 3 grid-triplet studies for the y velocity using UDS and temperature using UDS.

When the grids are refined, all errors decrease except the uncertainty estimates using GCI, GCI$_1$, and CF methods for temperature with the SMART scheme increase. At $P = 1.15$, uncertainty estimated by all the methods do not bound the true errors except the GCI$_2$ and FS methods. The average $FS_A$ are 1.39, 1.96, 3.99, 2.17, and 5.77 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively. The standard deviations of $FS_A$ are 0.62, 0.94, 2.54, 1.33, and 5.01 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively.

**Study 14 Premixed Methane/Air Laminar Flat Flame on a Perforated Burner**

As shown in Figure A.14, there are 5 and 4 grid-triplet studies for UDS and SMART methods, respectively. When the grids are refined, errors for most variables continuously decrease, solutions using the UDS method approach the AR either linearly
or oscillatorially, and solutions using the SMART method show oscillations far away from the asymptotic range. For $P$ close to 1, GCI, GCI$_1$ and CF predict $FSA$ around 1. For $P > 1$, $FSA$ estimated by the three methods show large oscillations of which some values are below 1. The GIC$_2$ and FS methods estimate $FSA$ larger than one. However, there is one grid-triplet study at $P = 0.6$ where none of the methods can bound it. The reason behind this is unknown. The average $FSA$ are 1.57, 1.61, 1.84, 2.04, and 2.53 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively. The standard deviations of $FSA$ are 0.81, 0.80, 1.03, 1.86, and 1.58 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively.

**Study 15 Data for “Exact” Grid Convergence Set**

As shown in Figure A.15, solutions for the “exact” grid convergence study are always in the asymptotic range when there is no perturbation or with perturbation #4. Other cases only have two grid-triplet studies. Overall errors decrease when the grids are refined. For $P > 1$ when solutions are far away from the asymptotic range, $FSA$ for the FS method increases while $FSA$ for the other methods decrease. The average $FSA$ are 1.32, 1.51, 2.61, 2.04, and 3.04 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively. The standard deviations of $FSA$ are 1.6, 1.48, 1.31, 2.64, and 2.32 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively.

**Study 16 Beam Bending Problem for 2nd Series on Grid Convergence**

As shown in Figure A.16, there are 3 grid-triplet studies for the three codes investigated. When the grids are refined, both the errors and $P$ show oscillations. Uncertainty estimated by all the methods fail to bound the true errors when $0.7 < P < 1.05$. The average $FSA$ are 0.45, 0.48, 0.68, 0.72, and 0.88 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively. The standard deviations of $FSA$ are 0.36, 0.35, 0.52, 0.72, and 0.64 for the GCI, GCI$_1$, GCI$_2$, CF, and FS methods, respectively.
Study 17 Beam Bending Problem for 3rd Series on Grid Convergence

As shown in Figure A.17, there are only two grid-triplet studies for the three codes investigated. When the grids are refined, the errors decrease. Uncertainty estimated by all the methods fail to bound the true errors when \( P > 1.15 \). The average \( \text{FS}_A \) are 1.17, 1.21, 1.42, 2.15, and 2.13 for the GCI, GCI_1, GCI_2, CF, and FS methods, respectively. The standard deviations of \( \text{FS}_A \) are 0.63, 0.59, 0.54, 1.49, and 0.99 for the GCI, GCI_1, GCI_2, CF, and FS methods, respectively.

![Figure A.1. 1D Wave Equation [6]: (a) error, uncertainties, and correction factor, (b) actual factor of safety](image-url)
Figure A.2. 2D Laplace Equation [22]: (a) non-constant value Dirichlet boundary conditions (Eca and Hoekstra, 2000), (b) non-constant value Dirichlet boundary conditions (Iowa recalculated), (c) constant value Dirichlet boundary conditions (Iowa calculation), (d) actual factor of safety.
Figure A.3. 2D Driven Cavity [23, 24]: (a) maximum stream function using BS scheme, (b) maximum stream function using upwind 3 scheme, (c) maximum stream function Kawamura scheme, (d) maximum vorticity using BS scheme, (e) maximum vorticity using upwind-3 scheme, (f) maximum vorticity using Kawamura scheme, (g) minimum stream function using BS scheme, (h) actual factor of safety.
Figure A.4. 2D Natural Convection Flows in Square Cavities at Ra=10^4 [25]: (a) U_{mon}, (b) V_{mon}, (c) T_{mon}, (d) U_{max}, (e) V_{max}, (f) Nu_{max}, (g) Nu, (h) actual factor of safety.
Figure A.5. 2D Natural Convection Flows in Square Cavities at Ra=10^5 [25]: (a) $U_{mon}$, (b) $V_{mon}$, (c) $T_{mon}$, (d) $U_{max}$, (e) Nu, (f) actual factor of safety.
Figure A.6. 2D Natural Convection Flows in Square Cavities at $Ra=10^6$ [25]: (a) $U_{mon}$, (b) $V_{mon}$, (c) $T_{mon}$, (d) $U_{max}$, (e) $V_{max}$, (f) $Nu_{max}$, (g) $Nu$, (h) actual factor of safety.
Figure A.7. 2D Backward-facing Step at Re=1.5×10^5 [26, 27, 28]: (a) error, uncertainties, and correction factor, (b) actual factor of safety
Figure A.8. Square cavity with moving top wall, Re=100 [29, 30]: (a) UDS-x velocity-0 deg; (b) UDS-x velocity-60 deg; (c) UDS-y velocity-0 deg; (d) SMART-x velocity-0 deg; (e) SMART-x velocity-60 deg; (f) SMART-y velocity-0 deg; (g) SMART-y velocity-60 deg; (h) actual factor of safety.
Figure A.9. Square cavity with moving top wall at Re=1000 [29, 30]: (a) UDS-x velocity-0 deg; (b) UDS-x velocity-60 deg; (c) UDS-y velocity-0 deg; (d) UDS-y velocity-60 deg; (e) SMART-x velocity-0 deg; (f) SMART-x velocity-60 deg; (g) SMART-y velocity-0 deg; (h) SMART-y velocity-60 deg; (i) actual factor of safety.
Figure A.10. Cubic cavity with moving top wall, Re=100 [29, 30]: (a) x-velocity-UDS; (b) y-velocity-UDS; (c) z-velocity-UDS, (d) actual factor of safety.
Figure A.11. Axisymmetric turbulent flow through a valve [31, 32]: (a) TKE-PLDS; (b) dissipation rate of TKE-PLDS; (c) radial-velocity-SMART; (d) axial-velocity-SMART; (e) actual factor of safety.
Figure A.12. One-dimensional steady-state convection-diffusion process without source term, with constant transport properties and with Dirichlet boundary conditions, Pe=1 and Pe=10 [33]: (a) UDS-Pe=1, (b) UDS-Pe=10, (c) SMART-Pe=1, (d) SMART-Pe=10, (e) actual factor of safety.
Figure A.13. Heat transfer from an isothermal cylinder enclosed by a square duct [29, 30, 31]: (a) UDS-x-velocity, (b) UDS-y-velocity, (c) UDS-temperature, (d) SMART-x-velocity, (e) SMART-y-velocity, (f) SMART-temperature, (g) actual factor of safety.
Figure A.14. Premixed methane/air laminar flat flame on a perforated burner [31, 34, 35]:
(a) UDS-radial-velocity, (b) UDS-axial-velocity, (c) UDS-temperature, 
(d) UDS-CH4, (e) UDS-O2, (f) UDS-CO2, (g) UDS-H2O, (h) SMART-radial-velocity, 
(i) SMART-axial-velocity, (j) SMART-temperature, 
(k) SMART-CH4, (l) SMART-O2, (m) SMART-CO2, (n) SMART-H2O, 
(o) actual factor of safety.
Figure A.15. Data for “exact” grid convergence set [8]: (a) exact to F=600, (b) perturbed #3, (c) perturbed #4, (d) all±5, (e) actual factor of safety.
Figure A.16. Beam bending problem for 2nd series on grid convergence [8]: (a) beam bending stress (Code 1), (b) beam bending stress (Code X), (c) beam bending stress (Code W), (d) actual factor of safety.
Figure A.17. Beam bending problem for 3rd series on grid convergence [8]: (a) beam bending stress (Code 1, code X), (b) beam bending stress (Code W), (c) Beam end deflection (Code 1), (d) Beam end deflection (Code X), (e) actual factor of safety.