

# Factors of Safety for Richardson Extrapolation

## Tao Xing

Assistant Professor  
Department of Mechanical Engineering,  
College of Engineering, Architecture and Physical  
Sciences,  
Tuskegee University,  
Tuskegee, AL 36088  
e-mail: taox@tuskegee.edu

## Frederick Stern

Professor  
Research Engineer  
IHHR-Hydroscience and Engineering,  
C. Maxwell Stanley Hydraulics Laboratory,  
University of Iowa,  
Iowa City, IA 52242-1585  
e-mail: frederick-stern@uiowa.edu

*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 95% 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 25 samples with different sizes based on 17 studies covering fluids, thermal, and structure disciplines. Only the factor of safety method, compared with the grid convergence index and correction factor methods, provides a reliability larger than 95% and a lower confidence limit greater than or equal to 1.2 at the 95% 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 the 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. [DOI: 10.1115/1.4001771]*

*Keywords:* factors of safety, Richardson extrapolation, solution verification

## 1 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 ( $FS$ ) and was validated for a correction factor less than 1 using analytical benchmarks. The factor of safety for correction factor larger than 1 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 6% 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 nonsmooth 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 with iterative errors.

The nonsmooth grid convergence problem may be resolved using the least-squares [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 nonsmooth and nonmonotonic convergence. They introduce additional uncertainties due to the least-squares 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

Contributed by the Fluids Engineering Division of ASME for publication in the JOURNAL OF FLUIDS ENGINEERING. Manuscript received February 11, 2009; final manuscript received April 19, 2010; published online June 23, 2010. Assoc. Editor: Dimitris Drikakis.

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 95% confidence level is achieved for the *GCI* method with a factor of safety of 1.25 based on over 500 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  $X_{v^*}^2$ . 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. Methods 1 and 2 implemented the *GCI*<sub>1</sub> 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 95% for method 1 and 68% 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-squares or response-surface method to account for nonsmooth or nonmonotonic convergence and use an explicit method to compute the uncertainty due to the curve fit. Method 1 showed 60%  $R_{sm}$ . Method 2 shows 93%  $R_{sm}$ . These facts suggest that the use of the *GCI*<sub>1</sub> method is closer to a 68% than a 95% confidence level. The other methods show more than 90%  $R_{sm}$  except method 5 that shows 82%. Methods 1–5 predict much higher  $X_{v^*}^2$  than methods 6–10, which indicates that there is no correlation between  $R_{sm}$  and  $X_{v^*}^2$ . Since this study has only three structure problems with 18 individual grid solutions, it was recommended that a sample with the number of grid convergence studies much larger than 100 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-squares version of the *GCI* method for a nominally second-order accurate discretization scheme. Three different uncertainty estimates were provided for  $0 < (P = p_{RE}/p_{th}) < 0.475$ ,  $0.475 \leq P < 1.025$ , 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 *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 25 samples with different sizes based on 17 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*, *GCI*<sub>1</sub>, *GCI*<sub>2</sub> [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 1 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 a correction factor less than 1 with the introduction of factors of safety for correction factors equal 0 and 1, which were validated using statistical analysis as conducted herein. 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 2, and the omission of effectivity index for evaluating error estimates. Their comments motivated the improvements over [18] as presented herein. Compared with Ref. [18], the present *FS* method has more general applicability without restriction on the maximum  $P$ , larger samples with the addition of 25 items, and a 95% reliability for all  $P$  ranges.

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

## 2 Error and Uncertainty Estimates 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 95% 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  $S_{AB}$  or numerical benchmark  $S_{NB}$ .

**2.1 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 spacings  $\Delta x_3$ ,  $\Delta x_2$ , and  $\Delta x_1$ , respectively, then

$$r = \frac{\Delta x_2}{\Delta x_1} = \frac{\Delta x_3}{\Delta x_2} \quad (1)$$

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

$$\begin{aligned} \varepsilon_{21} &= S_2 - S_1 \\ \varepsilon_{32} &= S_3 - S_2 \\ R &= \varepsilon_{21}/\varepsilon_{32} \end{aligned} \quad (2)$$

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,

$$p_{RE} = \frac{\ln(\varepsilon_{32}/\varepsilon_{21})}{\ln(r)} \quad (3)$$

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

$$S_C = S_1 - \delta_{RE} \quad (5)$$

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  $\delta_{RE}$ :

$$CF = \frac{r^{p_{RE}} - 1}{r^{p_{th}} - 1} \quad (6)$$

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:

$$P = \frac{p_{RE}}{p_{th}} \quad (7)$$

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.

**2.2 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 the results from the numerical solution of the one-dimensional wave and two-dimensional Laplace equation analytical benchmarks

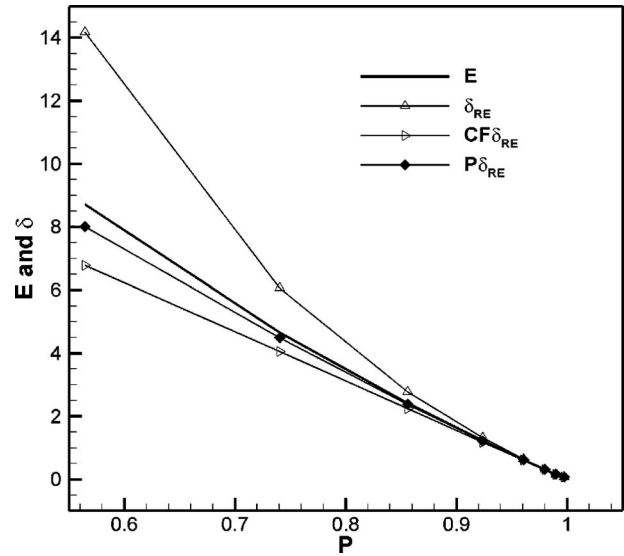


Fig. 1 Comparison of error estimates for the one-dimensional wave equation

$$CF\delta_{RE} = CF \left( \frac{\varepsilon_{21}}{r^{p_{RE}} - 1} \right) \quad (8)$$

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 overpredicts the comparison error by up to 61%. The error estimate  $CF\delta_{RE}$  is a first-order polynomial and underpredicts the comparison error by up to 23%. The error estimate  $P\delta_{RE}$  is a second-order polynomial and underpredicts the comparison error by up to 9%. 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 1.

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  $E$

$$\theta = \left| \frac{\delta}{E} \right| \quad (9)$$

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 17 studies that have analytical and numerical benchmarks provide 329 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.

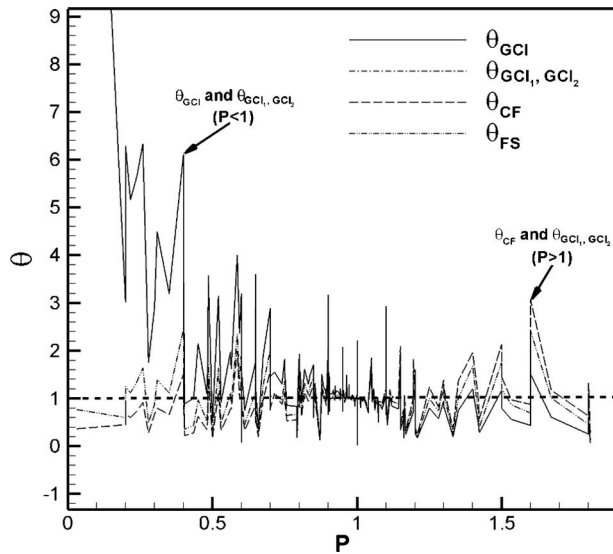


Fig. 2 Effectivity indices for different error estimates

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.

**2.3 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 95% 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]

$$U_{GCI} = FS \frac{|\varepsilon_{21}|}{r^{p_{RE}-1}} = FS|\delta_{RE}| \quad (10)$$

Roache [1] suggested values of  $FS=3$  for two-grid sensitivity studies using  $p_{th}$  and  $FS=1.25$  for convergence studies with a minimum of three grids using  $p_{RE}$ . Both approaches use a constant

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

$P$	No. of points	$\theta_{GCI}$	$\theta_{GCI_{1,2}}$	$\theta_{CF}$	$\theta_{FS}$
$0 < P < 2$	329	1.332	1.398	0.998	1.042
$0 < P \leq 1$	218	1.558	1.558	0.954	1.062
$1 < P < 2$	111	0.888	1.083	1.083	1.001

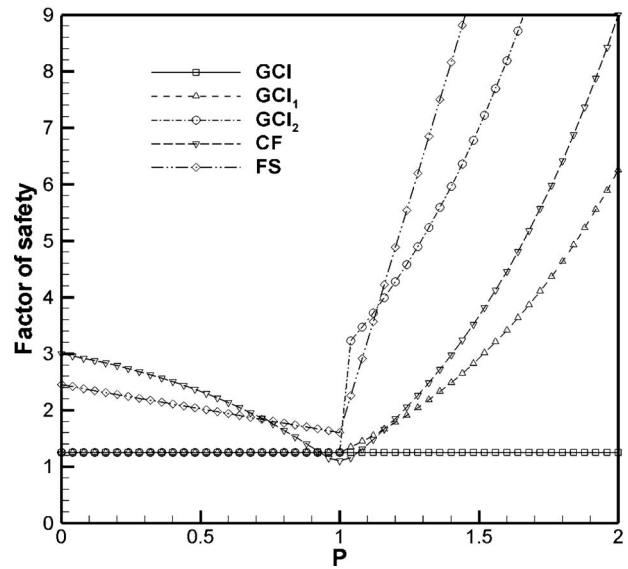


Fig. 3 Factor of safety for different verification methods with  $p_{th}=2$  and  $r=2$  for the  $CF$  method

factor of safety for all  $P$  values, e.g.,  $FS=1.25$ , as shown in Fig. 3.

Logan and Nitta [8] used the  $GCI_1$  method, for which the factor of safety for the  $GCI$  is multiplied by the correction factor for  $P > 1$ .

$$U_{GCI_1} = FS(P, CF)|\delta_{RE}| = \begin{cases} 1.25|\delta_{RE}|, & P \leq 1 \\ 1.25CF|\delta_{RE}|, & P > 1 \end{cases} \quad (11)$$

The  $GCI_2$  method [16] is the same as the  $GCI_1$  method except  $FS=3$  instead of 1.25 for  $P > 1$ .

$$U_{GCI_2} = FS(P, CF)|\delta_{RE}| = \begin{cases} 1.25|\delta_{RE}|, & P \leq 1 \\ 3CF|\delta_{RE}|, & P > 1 \end{cases} \quad (12)$$

The uncertainty for the  $CF$  method is estimated by the sum of the absolute value of the improved error estimate  $CF|\delta_{RE}|$  and the absolute value of the amount of the correction.

$$U_{CF} = FS(CF)|\delta_{RE}| = \begin{cases} [9.6(1 - CF)^2 + 1.1]|\delta_{RE}|, & 0.875 < CF < 1.125 \\ [2|1 - CF| + 1]|\delta_{RE}|, & 0 < CF \leq 0.875 \text{ or } CF \geq 1.125 \end{cases} \quad (13)$$

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 the 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



$$U_{FS} = FS(P)|\delta_{RE}| = \begin{cases} [FS_1 P + FS_0(1 - P)]|\delta_{RE}|, & 0 < P \leq 1 \\ [FS_1 P + FS_2(P - 1)]|\delta_{RE}|, & P > 1 \end{cases} \quad (14)$$

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.

**2.4 Confidence Levels.** For the *FS* method, the recommended values of  $FS_0$ ,  $FS_1$ , and  $FS_2$  are determined using a 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 95% and the lower confidence limit at the 95% confidence level is greater than or equal to 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

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

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.

### 3 Statistical Analysis

Statistical analysis is based on 25 samples that consist of actual factor of safety items  $FS_{A_i}$  ( $i = 1, N$ ) with different sample sizes  $N$  ranging from 5 to 329. 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  $E_i$ :

$$FS_{A_i} = \frac{U_i}{|E_i|} \quad (16)$$

where  $U$  is defined by Eqs. (10)–(13) and (15) for the *GCI*, *GCI*<sub>1</sub>, *GCI*<sub>2</sub>, *CF*, and *FS* methods, respectively. The comparison error  $E_i$  for fine grid solution  $S_{1_i}$  is

$$E_i = S_{AB} - S_{1_i} \quad (17)$$

for an analytical benchmark and

$$E_i = S_{NB} - S_{1_i} \quad (18)$$

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,  $FS_{A_i}$  can be used as an index to evaluate the conservativeness of  $U$  for different verification methods.

The error estimate is systematic, but  $E_i$  and therefore  $FS_{A_i}$  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  $FS_{A_i}$  are randomly distributed, the confidence interval for the mean reveals how close the mean value of  $FS_{A_i}$ ,  $\overline{FS_A}$ , is to the true mean  $\mu$  of the parent population of  $FS_{A_i}$ .

**3.1 Reliability.** Reliability  $R$  is defined as

$$R = \frac{\sum_{i=1}^N \text{number of } FS_{A_i} > 1}{N} \quad (19)$$

The reliability used in Ref. [8],  $R_{sm} = 1 - |\text{estimated fraction for } FS_{A_i} > 1 - \text{expected fraction for } FS_{A_i} > 1|$ , is deficient as it does not discriminate between estimated fractions with the same distance below or above the expected fraction.

**3.2 Confidence of the Mean Analysis.** The confidence of mean analysis is based on the methodology and procedures summarized in Ref. [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  $\bar{X}$ ,  $S_{X_i}$ , and  $S_{\bar{X}}$ , respectively.

The true mean  $\mu$  of the parent population at the 95% confidence level is bounded by  $\bar{X} - k$  and  $\bar{X} + k$ :

$$\Pr(\bar{X} - k \leq \mu \leq \bar{X} + k) \geq 0.95 \quad (20)$$

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

$$k = tS_{\bar{X}} \quad (21)$$

The lower confidence limit of the mean is defined by

$$LCL = \bar{X} - k \quad (22)$$

**3.3 Implementation.** The 25 samples are based on 17 studies that have analytical or numerical benchmarks and cover fluids, thermal, and structure disciplines, as summarized in Table 2. The 17 studies have 98 variables. The largest sample 3 is for the actual factors of safety that are obtained by combining the 329 grid-triplet studies that show monotonic convergence from the 17 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 Sec. 5 are for the actual factor of safety, i.e.,  $FS_{A_i}$  in Eq. (16) is equivalent to  $X_i$  such that  $\bar{X} = \overline{FS_A}$ , and  $S_{X_i} = S_{FS_{A_i}}$ ,  $S_{\bar{X}} = S_{\overline{FS_A}}$ . For convenience of presentation and discussion, both of the nomenclatures are used interchangeably.

Sample 1 is for the 17 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 98 various verification variables, as also shown in Table 3. The verification variables are extracted from the 17 studies and may involve different numbers of grid-triplet studies. Sample 3 includes all the 329 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, samples 4–8 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 17 individual  $P$  values (samples 9–25) ranging from 0.705 to 1.205 where at least five items are available after removing the outliers, as shown in Table 5.

The number of outliers for samples 9–25 is summarized in

**Table 2 Analytical and numerical benchmark verification studies**

Study	Geometry	Conditions	Verification variables	No. of grids	Benchmark
1	1D wave [6]	-	Wave profile	10	AB
2	2D Laplace [22]	-	Arbitrary function	26	AB
3	2D driven cavity [23]	Re=1000	Maximum/minimum of stream-function, vorticity	4	NB [24]
4	2D natural convection flows in square cavities [25]	Ra=10 <sup>4</sup>	Maximum or monitored velocity, location, temperature, and Nu	5	NB (S <sub>1</sub> )
5	2D natural convection flows in square cavities [25]	Ra=10 <sup>5</sup>	Maximum or monitored velocity, location, temperature, and Nu	6	NB (S <sub>1</sub> )
6	2D natural convection flows in square cavities [25]	Ra=10 <sup>6</sup>	Maximum or monitored velocity, location, temperature, and Nu	7	NB (S <sub>1</sub> )
7	Backward-facing step [26]	Re=1.5 × 10 <sup>5</sup>	Reattachment length, velocity	7	NB [27,28]
8	2D driven cavity [29]	Re=100	Velocity	5	NB [30]
9	2D driven cavity [29]	Re=1000	Velocity	5	NB [30]
10	3D cubic cavity [29]	Re=100	Velocity	4	NB [30]
11	Axisymmetric turbulent flow through a valve [31,32]	Re=10 <sup>5</sup>	Velocity, TKE, epsilon	5	NB (S <sub>1</sub> )
12	1D steady-state convection-diffusion [33]	Pe=1 and Pe=10	Arbitrary function	6	AB
13	Isothermal cylinder enclosed by a square duct [29–31]	Ra=10 <sup>6</sup> , Pr=10	Velocity, temperature	5	NB (S <sub>1</sub> )
14	Premixed methane/air laminar flat flame on a perforated burner [31,34,35]	Inlet temperature 298.2 K	Velocities, temperature, mass fraction	7	NB (S <sub>1</sub> )
15	Data for “exact” grid convergence set [8]	Contrived	-	7	AB
16	Beam bending problem for second series [8]	Second series	Beam bending stress, beam end deflection	7 with 5 systematically refined	AB
17	Beam bending problem for third series [8]	Third series	Beam bending stress, beam end deflection	4	AB

Table 5. The outliers are identified using Peirce’s criterion [21]. Compared with Chauvenet’s criterion, 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  $FS_0$ ,  $FS_1$ , and  $FS_2$  are obtained iteratively by increasing their values from zero until two criteria are satisfied. The first criterion is that for all the 25 samples a 95% reliability is achieved for  $FS_{A_i} > 1$ ,

$$R > 95\% \tag{23}$$

The second criterion is that the lower confidence limit for  $\overline{FS_A}$  at the 95% confidence level is greater than or equal to a specified minimum factor of safety  $FS_{min}$ . For practical applications,  $FS_{min}$  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  $FS_{min}=1.2$  for the present application. The second criterion becomes

$$\Pr(LCL \geq 1.2) \geq 0.95 \tag{24}$$

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

$$\Pr(\bar{X} \geq 1.2) \geq 0.95 \tag{25}$$

$$\Pr(\mu \geq 1.2) \geq 0.95 \tag{26}$$

#### 4 Analytical and Numerical Benchmarks

The 17 studies are obtained from published journal or conference proceedings for which either analytical or numerical benchmarks for various verification variables are available. The 17 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

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

Sample	Factor of safety	Average $P$ ( $\bar{P}$ )	$N$	Statistics	$GCI$	$GCI_1$	$GCI_2$	$CF$	$FS$	$t$
1	Means of studies	0.98	17	<b>R</b> (%N)	94.1	94.1	100	100	100	1.746
				$\bar{X}$	1.64	1.75	2.42	2.31	3.18	
				$S_{\bar{X}}(\% \bar{X})$	9.1	8.6	8.7	13.0	10.4	
				$LCL$	1.38	1.49	2.05	1.79	2.60	
2	Means of all variables of 17 studies	0.96	98	<b>R</b> (%N)	89.8	92.9	96.9	92.9	100	1.663
				$\bar{X}$	1.64	1.75	2.45	2.28	3.14	
				$S_{\bar{X}}(\% \bar{X})$	5.5	5.1	4.9	7.9	6.1	
				$LCL$	1.49	1.60	2.25	1.98	2.82	

**Table 4 Statistics for different ranges of  $P$  values using nonaveraged actual factor of safety**

Sample	$P$	$N$	Statistics	$GCI$	$GCI_1$	$GCI_2$	$CF$	$FS$	$t$
3	0–2 ( $\bar{P}=0.94$ )	329	$\mathbf{R} (\%N)$	83.9	90.3	94.2	90.0	96.96	1.645
			$\bar{X}$	1.67	1.75	2.39	2.29	3.04	
			$S_{\bar{X}}(\%\bar{X})$	7.8	7.4	5.9	13.1	8.2	
			$LCL$	1.46	1.54	2.16	1.80	2.63	
4	0–0.4 ( $\bar{P}=0.24$ )	12 (3.65%)	$\mathbf{R} (\%N)$	100	100	100	100	100	1.796
			$\bar{X}$	7.34	7.34	7.34	16.40	13.52	
			$S_{\bar{X}}(\%\bar{X})$	29	29	29	31.3	30.8	
			$LCL$	3.51	3.51	3.51	7.19	6.03	
5	0.4–0.9 ( $\bar{P}=0.70$ )	81 (24.62%)	$\mathbf{R} (\%N)$	91.4	91.4	91.4	95.1	95.1	1.664
			$\bar{X}$	2.01	2.01	2.01	3.07	3.03	
			$S_{\bar{X}}(\%\bar{X})$	6.5	6.5	6.5	8.8	7.3	
			$LCL$	1.79	1.79	1.79	2.62	2.66	
6	0.9–1.1 ( $\bar{P}=0.98$ )	176 (53.50%)	$\mathbf{R} (\%N)$	92.6	93.8	95.5	92.0	97.7	1.646
			$\bar{X}$	1.34	1.36	1.88	1.25	1.87	
			$S_{\bar{X}}(\%\bar{X})$	2.2	2.2	3.7	2.4	2.1	
			$LCL$	1.29	1.31	1.76	1.20	1.80	
7	1.1–1.5 ( $\bar{P}=1.19$ )	50 (15.20%)	$\mathbf{R} (\%N)$	48	76	94	72	96	1.676
			$\bar{X}$	1.04	1.36	3.27	1.35	3.74	
			$S_{\bar{X}}(\%\bar{X})$	7.03	6.45	6.45	6.39	6.93	
			$LCL$	0.92	1.22	2.92	1.21	3.31	
8	1.5–2.0 ( $\bar{P}=1.63$ )	10 (3.04%)	$\mathbf{R} (\%N)$	30	80	90	90	100	1.833
			$\bar{X}$	0.86	1.70	4.08	2.03	7.67	
			$S_{\bar{X}}(\%\bar{X})$	19.8	18.7	18.7	18.4	18.4	
			$LCL$	0.55	1.12	2.68	1.35	5.08	

convection-diffusion equations and contrived and beam-bending structure 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 nonconstant 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 fourth root of 2 was also used. All solutions were reported to have achieved iterative convergence. The solutions in total represent 339 grid-triplet studies. For each grid-triplet study, a grid convergence study is conducted using Eq. (2), which resulted in a total of 329 solutions that show monotonic convergence for which  $p_{RE}$ ,  $\delta_{RE}$ , and  $S_C$  are evaluated using Eqs. (3)–(5), respectively. The remaining ten solutions were either oscillatory or monotonically divergent.

The manner in which the 98 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  $\Delta x_{\text{fine}}/\Delta x_{\text{finest}}$ , where  $\Delta x_{\text{fine}}$  is the fine grid-spacing for an individual grid-triplet study and  $\Delta x_{\text{finest}}$  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  $\bar{P}_{\text{finest}}=1.026$ . Thirty-three verification variables have two grid-triplet studies with an average  $\bar{P}_{\text{finest}}=0.95$ ; however, the errors decrease as the grids are refined. The other 55 verification variables have multiple grid-triplet studies ranging from 3 to 24 and show monotonic or oscillatory convergence. Seven verification variables show monotonic convergence with an average  $\bar{P}_{\text{finest}}=0.997$  covering studies with  $S_{AB}$  (studies 1 and 2) and  $S_{NB}$  (studies 5, 6, and 14). The remaining 48 verification variables show oscillatory convergence with an average  $\bar{P}_{\text{finest}}=0.996$ . The complete results are provided in Ref. [36].

## 5 Statistical Results

The actual factor of safety for sample 3, sample 3 averaged using  $\Delta P=0.01$ , and the upper and lower bands of the confidence interval  $FS_A \pm tS_{FS_A}$  for samples 9–25 are shown in Fig. 4. The three  $GCI$  methods are fairly conservative for  $P < 0.9$  where they show a 93% reliability. The  $CF$  and  $FS$  methods are more conservative by showing a 96% reliability in this range. Near the asymptotic range when  $0.9 < P < 1.1$ , the  $GCI$  method has a 93% 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 98% reliability. For  $P > 1.1$ , the  $GCI$  method is the least conservative with a reliability of 45%. The  $GCI_1$  method is almost as conservative as the  $CF$  method and both show reliability around 76%. The  $GCI_2$  method is much more conservative than the  $GCI$ ,  $GCI_1$ , and  $CF$  methods with a reliability of 93%. Only the  $FS$  method shows a reliability larger

**Table 5 Statistics excluding outliers at 17 *P* values**

Sample	<i>P</i>	<i>N</i>	No. of outliers	Statistics	<i>GCI</i>	<i>GCI</i> <sub>1</sub>	<i>GCI</i> <sub>2</sub>	<i>CF</i>	<i>FS</i>	<i>t</i>
9	0.705	9	0	<b>R</b> (% <i>N</i> )	100	100	100	100	100	1.86
				$\bar{X}$	2.08	2.08	2.08	3.05	3.08	
				$S_{\bar{X}}(\% \bar{X})$	10.4	10.4	10.4	11.5	10.4	
				<i>LCL</i>	1.68	1.68	1.68	2.40	2.48	
10	0.755	5	0	<b>R</b> (% <i>N</i> )	100	100	100	100	100	2.13
				$\bar{X}$	1.87	1.87	1.87	2.66	2.71	
				$S_{\bar{X}}(\% \bar{X})$	11.2	11.2	11.2	11.3	11.2	
				<i>LCL</i>	1.43	1.43	1.43	2.02	2.07	
11	0.805	15	0	<b>R</b> (% <i>N</i> )	100	100	100	100	100	1.76
				$\bar{X}$	1.76	1.76	1.76	2.25	2.49	
				$S_{\bar{X}}(\% \bar{X})$	5.4	5.4	5.4	6.0	5.4	
				<i>LCL</i>	1.59	1.59	1.59	2.01	2.26	
12	0.855	6	0	<b>R</b> (% <i>N</i> )	100	100	100	100	100	2.02
				$\bar{X}$	1.58	1.58	1.58	1.83	2.18	
				$S_{\bar{X}}(\% \bar{X})$	9.6	9.6	9.6	10.3	9.6	
				<i>LCL</i>	1.27	1.27	1.27	1.45	1.76	
13	0.905	29	0	<b>R</b> (% <i>N</i> )	100	100	100	100	100	1.70
				$\bar{X}$	1.61	1.61	1.61	1.68	2.17	
				$S_{\bar{X}}(\% \bar{X})$	6.8	6.8	6.8	6.7	6.8	
				<i>LCL</i>	1.42	1.42	1.42	1.49	1.92	
14	0.925	5	0	<b>R</b> (% <i>N</i> )	100	100	100	100	100	2.13
				$\bar{X}$	1.42	1.42	1.42	1.38	1.89	
				$S_{\bar{X}}(\% \bar{X})$	3.1	3.1	3.1	4.8	3.1	
				<i>LCL</i>	1.33	1.33	1.33	1.24	1.76	
15	0.945	6	0	<b>R</b> (% <i>N</i> )	100	100	100	100	100	2.02
				$\bar{X}$	1.37	1.37	1.37	1.27	1.81	
				$S_{\bar{X}}(\% \bar{X})$	2.0	2.0	2.0	3.0	2.0	
				<i>LCL</i>	1.32	1.32	1.32	1.20	1.73	
16	0.955	16	0	<b>R</b> (% <i>N</i> )	87.5	87.5	87.5	87.5	100	1.75
				$\bar{X}$	1.42	1.42	1.42	1.31	1.87	
				$S_{\bar{X}}(\% \bar{X})$	6.8	6.8	6.8	7.1	6.8	
				<i>LCL</i>	1.25	1.25	1.25	1.14	1.64	
17	0.965	6	0	<b>R</b> (% <i>N</i> )	100	100	100	100	100	2.02
				$\bar{X}$	1.46	1.46	1.46	1.32	1.91	
				$S_{\bar{X}}(\% \bar{X})$	7.1	7.1	7.1	7.3	7.1	
				<i>LCL</i>	1.25	1.25	1.25	1.13	1.63	
18	0.975	7	0	<b>R</b> (% <i>N</i> )	100	100	100	100	100	1.94
				$\bar{X}$	1.33	1.33	1.33	1.19	1.73	
				$S_{\bar{X}}(\% \bar{X})$	3.6	3.6	3.6	3.8	3.6	
				<i>LCL</i>	1.24	1.24	1.24	1.10	1.61	
19	0.985	9	0	<b>R</b> (% <i>N</i> )	100	100	100	100	100	1.86
				$\bar{X}$	1.27	1.27	1.27	1.13	1.64	
				$S_{\bar{X}}(\% \bar{X})$	1.1	1.1	1.1	1.1	1.1	
				<i>LCL</i>	1.25	1.25	1.25	1.10	1.60	
20	0.995	8	0	<b>R</b> (% <i>N</i> )	100	100	100	87.5	100	1.90
				$\bar{X}$	1.24	1.24	1.24	1.09	1.59	
				$S_{\bar{X}}(\% \bar{X})$	2.0	2.0	2.0	2.0	2.0	
				<i>LCL</i>	1.19	1.19	1.19	1.05	1.53	
21	1.005	56	4	<b>R</b> (% <i>N</i> )	98.2	98.2	98.2	98.2	100	1.68
				$\bar{X}$	1.32	1.33	2.08	1.16	1.71	
				$S_{\bar{X}}(\% \bar{X})$	3.0	3.0	5.6	3.0	3.0	
				<i>LCL</i>	1.26	1.26	1.88	1.11	1.63	
22	1.015	6	0	<b>R</b> (% <i>N</i> )	100	100	100	83.3	100	2.02
				$\bar{X}$	1.18	1.21	2.90	1.04	1.74	
				$S_{\bar{X}}(\% \bar{X})$	2.9	3.0	3.0	3.0	3.6	
				<i>LCL</i>	1.11	1.14	2.73	0.98	1.62	



**Table 5 (Continued.)**

Sample	$P$	$N$	No. of outliers	Statistics	$GCI$	$GCI_1$	$GCI_2$	$CF$	$FS$	$t$
23	1.055	6	0	$\mathbf{R} (\%N)$	100	100	100	100	100	2.02
				$\bar{X}$	1.66	1.83	4.39	1.59	3.25	
				$S_{\bar{X}}(\% \bar{X})$	9.0	8.9	8.9	8.8	8.6	
				$LCL$	1.36	1.50	3.60	1.31	2.68	
24	1.105	13	2	$\mathbf{R} (\%N)$	61.5	76.9	100	61.5	100	1.78
				$\bar{X}$	1.32	1.52	3.65	1.38	3.43	
				$S_{\bar{X}}(\% \bar{X})$	14.4	14.1	14.1	13.9	14.3	
				$LCL$	0.98	1.14	2.73	1.04	2.56	
25	1.205	12	1	$\mathbf{R} (\%N)$	58.3	83.3	100	83.3	100	1.80
				$\bar{X}$	1.04	1.38	3.32	1.38	4.08	
				$S_{\bar{X}}(\% \bar{X})$	8.9	8.9	8.9	9.1	8.9	
				$LCL$	0.88	1.16	2.79	1.16	3.43	

than 96% in this range. The  $GCI$ ,  $GCI_1$ , 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  $GCI_2$  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 an actual factor of safety less than 1, 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 9–25. 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 9–25. 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 9–25. 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 17 studies,  $\bar{P}$  is 0.98. The  $GCI_2$ ,  $CF$ , and  $FS$  methods achieve 100% reliability. The  $GCI$  and  $GCI_1$  methods achieve 94.1% reliability. The mean values of the actual factors of safety for the 17 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 98 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 95%. Compared with 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 samples 1 and 2.

In order to highlight the dependence of the different verification methods on different  $P$  ranges, Table 4 shows the statistics for samples 3–8 based on six different  $P$  ranges. For sample 3 (0

$< P < 2$ ), only the  $FS$  method achieves reliability larger than 95%. The  $GCI$  method achieves the lowest reliability at 83.9%. For sample 4 ( $0 < P < 0.4$ ), all verification methods show a 100% reliability. For sample 5 ( $0.4 \leq P < 0.9$ ), the three  $GCI$  methods achieve a 91.4% reliability and the  $CF$  and  $FS$  methods achieve a reliability larger than 95%. For sample 6 ( $0.9 \leq P < 1.1$ ), only the  $GCI_2$  and  $FS$  methods achieve a 95% reliability. Examination of 18.2% of the data for  $1.1 \leq P < 2.0$ , which cover samples 7 and 8, shows that only the  $FS$  method achieves a 95% reliability. The reliability for the  $GCI$  method is 45% in this range. Based on the means, the most conservative verification methods for samples 3, 7, and 8, samples 4 and 5, and sample 6 are the  $FS$ ,  $CF$ , and  $GCI_2$  methods, respectively. For sample 3, the lower confidence limit is larger than 1.2 for all the verification methods; however, only the  $GCI_2$ ,  $CF$ , and  $FS$  methods satisfy that the lower confidence limit is greater than or equal to 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 17  $P$  values (samples 9–25) ranging from 0.705 to 1.205. For samples 9–19 ( $P < 0.99$ ), all the verification methods achieve 100% reliability except at  $P=0.955$  where only the  $FS$  method achieves such a level. For samples 20–23 ( $P \approx 1$ ), all the verification methods achieve reliabilities larger than 95% except the  $CF$  method for samples 20 and 22. For samples 24 and 25 ( $P > 1$ ), only the  $GCI_2$  and  $FS$  methods achieve reliabilities larger than 95%. Based on the means, the most conservative verification method for samples (9–20 and 25) and samples 21–24 are the  $FS$  and  $GCI_2$  methods, respectively. Only the  $FS$  method satisfies the requirement that the lower confidence limit is larger than 1.2 for samples 9–25. The samples for which the lower confidence limits are less than 1.2 are 20, 22, 24, and 25 for the  $GCI$  and  $GCI_1$  methods, 20 for the  $GCI_2$  method, and 16–25 except 23 for the  $CF$  method. The results indicate that the ranges of  $P$  values, where the verification methods are not conservative enough, are  $P \geq 1$  for the  $GCI$ ,  $GCI_1$ , and  $CF$  methods and  $P \approx 1$  for the  $GCI_2$  method.

## 6 Example for Ship Hydrodynamics Applications

To evaluate the performance of the different verification methods for practical applications, the verification study by Xing et al. [37] 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  $8.1 \times 10^6$  points. This

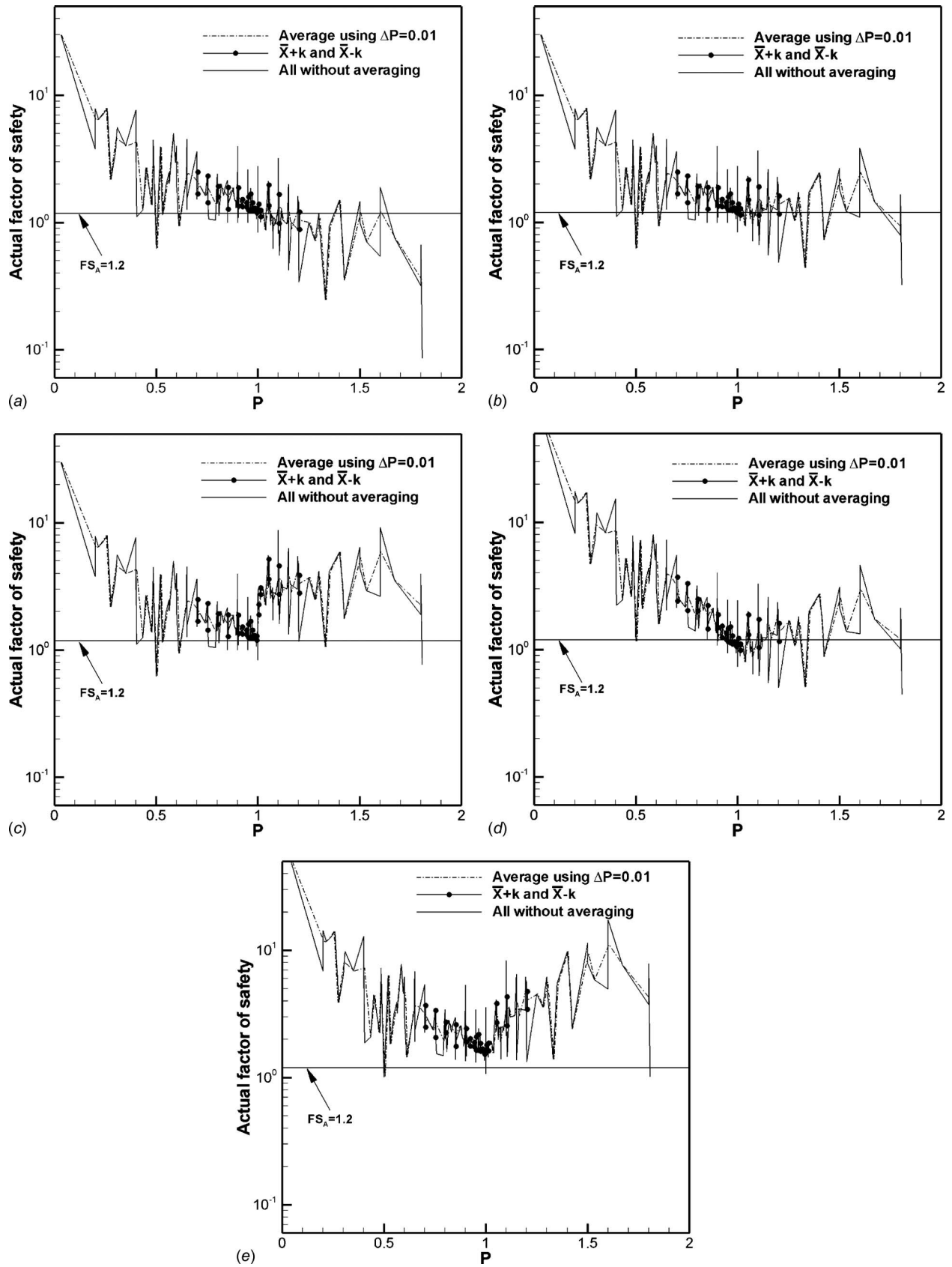


Fig. 4 Actual factor of safety for sample 3, sample 3 averaged using  $\Delta P=0.01$ , and  $\overline{FS}_A \pm t_{\overline{FS}_A}$  for samples 9–25: (a)  $GCI$  method, (b)  $GCI_1$  method, (c)  $GCI_2$  method, (d)  $CF$  method, and (e)  $FS$  method

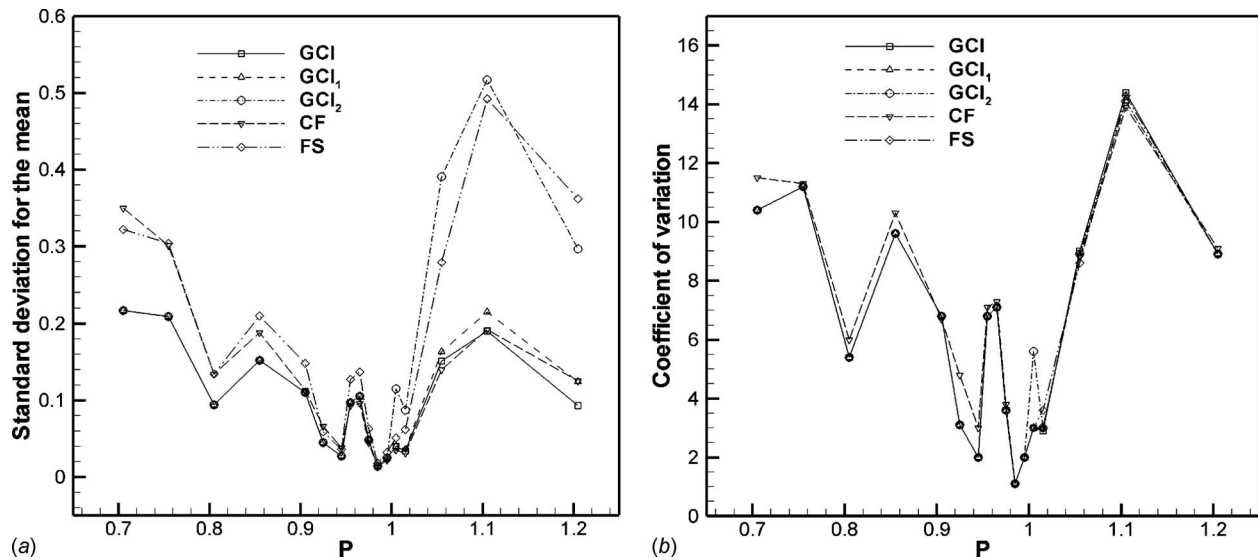


Fig. 5 Standard deviation for the mean  $S_{\bar{x}}$  and coefficient of variation  $S_{\bar{x}}(\% \bar{X})$  for samples 9–25: (a)  $S_{\bar{x}}$  and (b)  $S_{\bar{x}}(\% \bar{X})$

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 the  $P$  values vary

greatly, which is typical for industrial applications. When  $P$  is less than 1, 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 1, but close to the asymptotic range, i.e.,  $P = 1.06$  for the trim on grids (1, 3, and 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 a Froude number of 0.48.  $U_G$  is  $\%S_1$  and  $C_{TX}$  is based on the static wetted area.

Grids	Refinement ratio	$R$	$P$	$U_G$ (%)				
				$GCI$	$GCI_1$	$GCI_2$	$CF$	$FS$
2, 4, 6	$2^{0.5}$	0.63	0.66	3.34	3.34	3.34	4.90	5.04
1, 3, 5	$2^{0.5}$	0.40	1.33	0.72	1.09	2.61	1.16	4.02
4, 5, 6	$2^{0.25}$	0.97	0.08	54.7	54.7	54.7	125.2	104.2
3, 4, 5	$2^{0.25}$	0.80	0.63	4.98	4.98	4.98	7.23	7.62
2, 3, 4	$2^{0.25}$	0.60	1.49	1.07	1.75	4.21	1.95	8.30
1, 2, 3	$2^{0.25}$	0.50	2.00	0.36	0.87	2.10	1.11	5.22

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

Parameter	Grids	Refinement ratio	$R$	$P$	$U_G$ (%)				
					$GCI$	$GCI_1$	$GCI_2$	$CF$	$FS$
Sinkage	1, 3, 5	$2^{0.5}$	0.31	1.70	0.64	1.45	3.47	1.80	6.73
Sinkage	2, 3, 4	$2^{0.25}$	0.13	6.01	0.05	0.88	2.11	1.37	3.49
Trim	1, 3, 5	$2^{0.5}$	0.48	1.06	4.12	4.49	10.78	3.88	8.74
Trim	4, 5, 6	$2^{0.25}$	0.86	0.45	24.42	24.42	24.42	42.87	40.48
Trim	2, 3, 4	$2^{0.25}$	0.53	1.85	3.35	7.25	17.40	8.92	41.48
Trim	1, 2, 3	$2^{0.25}$	0.53	1.85	1.73	3.77	9.04	4.64	21.62

## 7 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  $p_{RE} > p_{th}$  and lack of statistical evidence that the interval of uncertainty at the 95% 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 25 samples with different sizes ranging from 5 to 329 based on 17 studies covering fluids, thermal, and structure disciplines.

The statistical results show that only the *FS* method, compared with the *GCI*,  $GCI_1$ ,  $GCI_2$ , and *CF* methods, provides a reliability larger than 95% and a lower confidence limit at the 95% confidence level greater than or equal to 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 25 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 the 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 17 studies, 98 variables, and 329 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.

## Acknowledgment

This study was sponsored by the Office of Naval Research under Grant No. N00014-01-1-0073, administered by Dr. Patrick Purtell. The authors would like to thank Dr. Logan, Dr. Roache, Dr. Eça, Dr. Roy, Dr. Ray and anonymous reviewers for their insightful comments and criticisms.

## Nomenclature

$C_{TX}$	= total resistance coefficient
$E$	= comparison error
$FS_A$	= actual factor of safety
$N$	= sample size
$P = p_{RE} / p_{th}$	= distance metric to the asymptotic range
$\bar{P}$	= average $P$ value
$Pr$	= probability
$p_{RE}$	= estimated order of accuracy
$p_{th}$	= theoretical order of accuracy

$R$	= convergence ratio
$R$	= reliability
$R_{sm}$	= reliability [8]
$r$	= refinement ratio
$S_{AB}$	= analytical benchmark
$S_C$	= Richardson extrapolation numerical benchmark
$S_{NB}$	= numerical benchmark
$S_i$	= solution on the $i$ th grid
$S_{X_i}$	= standard deviation of the sample
$S_{\bar{X}}$	= standard deviation of the mean
$S_{\bar{X}}(\% \bar{X})$	= coefficient of variation
$T$	= true value
$U$	= uncertainty
$U_G$	= grid uncertainty
$\bar{X}$	= mean value
$X_{\nu}^2$	= 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
$\varepsilon$	= solution change
$\mu$	= true mean of the parent population
$\theta$	= effectivity index

## References

- [1] Roache, P. J., 1998, *Verification and Validation in Computational Science and Engineering*, Hermosa, Albuquerque, NM.
- [2] Celik, I. B., Ghia, U., Roache, P. J., Freitas, C. J., Coleman, H. W., and Raad, P. E., 2008, "Procedure for Estimation and Reporting of Uncertainty Due to Discretization in CFD Applications," *ASME J. Fluids Eng.*, **130**(7), pp. 078001.
- [3] Cosner, R. R., Oberkampf, W. L., Rumsey, C. L., Rahaim, C. P., and Shih, T. I.-P., 2006, "AIAA Committee on Standards for Computational Fluid Dynamics: Status and Plans," 44th Aerospace Sciences Meeting and Exhibit, Reno, NV, Jan. 9–12, AIAA Paper No. 2006-889.
- [4] Stern, F., Wilson, R. V., Coleman, H. W., and Paterson, E. G., 2001, "Comprehensive Approach to Verification and Validation of CFD Simulations—Part 1: Methodology and Procedures," *ASME J. Fluids Eng.*, **123**(4), pp. 793–802.
- [5] Wilson, R. V., Shao, J., and Stern, F., 2004, "Discussion: Criticisms of the 'Correction Factor' Verification Method (Roache, P. J., 2003, *ASME J. Fluids Eng.*, **125**(4), pp. 732–733)," *ASME J. Fluids Eng.*, **126**(4), pp. 704–706.
- [6] Stern, F., Wilson, R. V., Coleman, H. W., and Paterson, E. G., 1999, "Verification and Validation of CFD Simulations," IHR Report No. 407.
- [7] Eça, L., and Hoekstra, M., 2002, "An Evaluation of Verification Procedures for CFD Applications," 24th Symposium on Naval Hydrodynamics, Fukuoka, Japan, Jul. 8–13.
- [8] Logan, R. W., and Nitta, C. K., 2006, "Comparing 10 Methods for Solution Verification, and Linking to Model Validation," *J. Aerosp. Comput. Inf. Commun.*, **3**(7), pp. 354–373.
- [9] Celik, I., and Hu, G. S., 2004, "Single Grid Error Estimation Using Error Transport Equation," *ASME J. Fluids Eng.*, **126**(5), pp. 778–790.
- [10] Cavallo, P. A., and Sinha, N., 2007, "Error Quantification for Computational Aerodynamics Using an Error Transport Equation," *J. Aircr.*, **44**(6), pp. 1954–1963.
- [11] Cavallo, P. A., Sinha, N., and O'Gara, M. R., 2008, "Viscous Error Transport Equation for Error Quantification of Turbulent Flows," 38th Fluid Dynamics Conference and Exhibition, Seattle, WA, Jun. 23–26, AIAA Paper No. 2008-3851, pp. 1–16.
- [12] ASME Performance Test Codes Committee PTC 61, 2008, "V&V 20: Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer."
- [13] Eça, L., and Hoekstra, M., 2006, "Discretization Uncertainty Estimation Based on a Least Squares Version of the Grid Convergence Index," Proceedings of the Second Workshop on CFD Uncertainty Analysis, Instituto Superior Tecnico, Lisbon, Oct.
- [14] Rumsey, C. L., and Thomas, J. L., 2008, "Application of FUN3D and CFL3D to the Third Workshop on CFD Uncertainty Analysis," NASA Report No. TM-2008-215537.
- [15] 2004, Proceedings of the Workshop on CFD Uncertainty Analysis, L. Eça and M. Hoekstra, eds., Lisbon, Portugal, Oct.
- [16] Roache, P. J., 2009, private communication.
- [17] Xing, T., and Stern, F., 2008, "Factors of Safety for Richardson Extrapolation for Industrial Applications," IHR Report No. 466.
- [18] Xing, T., and Stern, F., 2009, "Factors of Safety for Richardson Extrapolation for Industrial Applications," IHR Report No. 469.
- [19] Larsson, L., Stern, F., and Bertram, V., 2003, "Benchmarking of Computa-



- tional Fluid Dynamics for Ship Flows: The Gothenburg 2000 Workshop," *J. Ship Res.*, **47**(1), pp. 63–81.
- [20] Kreyszig, E., 1993, *Advanced Engineering Mathematics*, 7th ed., Wiley, New York, Chaps. 23 and 24, pp. 1148–1271.
- [21] Ross, S. M., 2003, "Peirce's Criterion for the Elimination of Suspect Experimental Data," *J. Eng. Technol.*, **20**(2), pp. 38–41.
- [22] Eça, L., and Hoekstra, M., 2000, "An Evaluation of Verification Procedures for Computational Fluids Dynamics," Instituto Superior Tecnico Report No. D72-7.
- [23] Bruneau, C. H., and Saad, M., 2006, "The 2D Lid-Driven Cavity Problem Revisited," *Comput. Fluids*, **35**(3), pp. 326–348.
- [24] Botella, O., and Peyret, R., 1998, "Benchmark Spectral Results on the Lid-Driven Cavity Flow," *Comput. Fluids*, **27**(4), pp. 421–433.
- [25] Hortmann, M., Perić, M., and Scheuerer, G., 1990, "Multigrid Benchmark Solutions for Laminar Natural Convection Flows in Square Cavities," in *Benchmark Test Cases for Computational Fluid Dynamics*, I. Celik and C. J. Freitas, eds., ASME, New York, pp. 1–6.
- [26] Celik, I., and Karatekin, O., 1997, "Numerical Experiments on Application of Richardson Extrapolation With Nonuniform Grids," *ASME J. Fluids Eng.*, **119**(3), pp. 584–590.
- [27] Thangam, S., and Speziale, C. G., 1992, "Turbulent Flow Past a Backward-Facing Step: A Critical Evaluation of Two-Equation Models," *AIAA J.*, **30**(5), pp. 1314–1320.
- [28] Avva, R. K., Kline, S. J., and Ferziger, J. H., 1998, "Computation of Turbulent Flow Over a Backward-Facing Step-Zonal Approach," AIAA 26th Aerospace Sciences Meeting, Reno, NV, Jan.
- [29] Pérez-Segarra, C. D., Oliva, A., and Cònsul, R., 1996, "Analysis of Some Numerical Aspects in the Solution of the Navier-Stokes Equations Using Non-Orthogonal Collocated Finite-Volume Methods," *Proceedings of the Third EC-COMAS Computational Fluid Dynamics Conference*, Paris, France, Wiley, New York, pp. 505–511.
- [30] Demirdžić, I., Lilek, Ž., and Perić, M., 1992, "Fluid Flow and Heat Transfer Test Problems for Non-Orthogonal Grids: Bench-Mark Solutions," *Int. J. Numer. Methods Fluids*, **15**(3), pp. 329–354.
- [31] Pérez-Segarra, C. D., Oliva, A., Costa, M., and Escanes, F., 1995, "Numerical Experiments in Turbulent Natural and Mixed Convection in Internal Flows," *Int. J. Numer. Methods Heat Fluid Flow*, **5**(1), pp. 13–33.
- [32] Pérez-Segarra, C. D., Cadafalch, J., Rigola, J., and Oliva, A., 1999, "Numerical Study of Turbulent Fluid Flow Through Valves," *Proceedings of the International Conference on Compressors and Their Systems*, City University, London, pp. 13–14.
- [33] Cadafalch, J., Pérez-Segarra, C. D., Cònsul, R., and Oliva, A., 2002, "Verification of Finite Volume Computations on Steady-State Fluid Flow and Heat Transfer," *ASME J. Fluids Eng.*, **124**(1), pp. 11–21.
- [34] Sommers, L. M. T., 1994, "Simulation of Flat Flames With Detailed and Reduced Chemical Models," Ph.D. thesis, Technical University of Eindhoven, Eindhoven, The Netherlands.
- [35] Soria, M., Cadafalch, J., Cònsul, R., and Oliva, A., 2000, "A Parallel Algorithm for the Detailed Numerical Simulation of Reactive Flows," *Proceedings of the 1999 Parallel Computational Fluid Dynamics Conference*, Williamsburg, VA, pp. 389–396.
- [36] Xing, T., and Stern, F., 2010, "Factors of Safety for Richardson Extrapolation," IIHR Report No. 476.
- [37] Xing, T., Carrica, P., and Stern, F., 2008, "Computational Towing Tank Procedures for Single Run Curves of Resistance and Propulsion," *ASME J. Fluids Eng.*, **130**(10), pp. 101102.