Closure to "Discussion of 'Factors of Safety for Richardson Extrapolation'" (2011, ASME J. Fluids Eng., 133, p. 115501)

Tao Xing
Assistant Professor
Department of Mechanical Engineering,
College of Engineering,
University of Idaho,
P.O. Box 440902,
Moscow, ID 83844-0902
e-mail: xing@uidaho.edu

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

Introduction

The Technical Brief by Roache [1] presents ten items of discussion of our factor of safety (FS) method for solution verification [2]. Our responses are listed below item-by-item using the same numbering as Roache. The nomenclature mostly follows our own and not Roache's such as p_{RE} for the order of accuracy calculated using the Richardson Extrapolation as opposed to the observed order of convergence and the GCI and GCI_2 methods as opposed to the GCI_0 and the real GCI methods. However, we agree with Roache to use F_S for the factor of safety used in all the verification methods. In response to item (10), we have used our approach to evaluate two new variants of the GCI method and one new variant of the FS method.

Response

(1) The GCI and FS methods can be written in the following general form:

$$U_{GCI/FS} = F_S \left| \frac{\varepsilon_{21}}{r^p - 1} \right| \qquad (1)$$

The FS method is substantially different from and not a variant of the GCI method. In the FS method we use $P = p_{RE}/p_{th}$ to determine F_S and always use $p = p_{RE}$. Only the FS method, compared with different variants of the GCI method, provides a reliability **R** larger than 95% and a lower confidence limit (LCL) 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. F_S is a smooth linear function of P and has no jumps.

There are a few variants of the GCI method. We have used the definition of the GCI method, which arguably is the most common version/interpretation applied in the literature [3–5]. The GCI_1 method was proposed by Logan and Nitta [6]. The guideline for the GCI_2 method was communicated to us by Roache[1] [7] in his criticisms of an earlier version of our FS method [8], which he now

---

[1]The private communication is available to the public upon request.

refers to as the real GCI method. Roache's most recent book [9] does point out that the choice of F_S and p requires user judgment calls; however, no single guideline was provided. The lack of a single guideline clearly has caused considerable confusion. We take no responsibility for this confusion. Statistical analysis showed that none of the GCI variants shows **R** > 95% and LCL > 1.2 for different studies, variables, ranges of P values, and single P values where multiple actual factors of safety are available. As a result, there are high risks to use these GCI variants in certain circumstances, especially for $P > 1$. Except the original GCI method, all variants of the GCI method have jumps of F_S versus P.

Our purpose is not to add to this confusion but rather to evaluate the performance of the outcomes of selecting any of these variants of the GCI method and compare with the FS method using our approach. The correction factor and p_{RE} are used to define the GCI_2 and other verification methods as defined by Eqs. (10)–(15) in Ref. [2] in order to compare their relative conservativeness using the same error estimate δ_{RE}.

(2) We disagree with Roache to refer to the GCI as the GCI_0 method and the GCI_2 as the GCI method for reasons given in item (1). The lack of a single guideline for selecting F_S and p and when to use which variant of the GCI method is highlighted by Roache's current discussion. Roache accepts using p_{RE} when it is within a 5% difference of p_{th} in item (2), whereas later in item (10), Roache considers two other judgment calls as reasonable.

The GCI_2 method discards the "coarse" grid solution in the uncertainty estimate when $P > 1$, which is difficult to justify. For example, four grid solutions from the coarsest grid 4 to the finest grid 1 can build two grid triplet studies, (1, 2, 3) and (2, 3, 4). Grid convergence studies for industrial applications often show the oscillation of p_{RE} such that (1, 2, 3) could estimate $P > 1$ but (2, 3, 4) could estimate $P < 1$. Based on the GCI_2 method, S_3 should be discarded in the uncertainty estimate for (1, 2, 3) but not for (2, 3, 4). Of course, we agree that ideally one would conduct additional grid triplet studies until the solution is at or as close as possible to the asymptotic range; however, clearly this is not always possible especially for industrial applications [10].

We agree that a grid-triplet study with $P = 0.08$ is not desirable. However, it is not uncommon for solution verification studies (e.g., local p_{RE} ranges from 0.012 to 8.47 in Ref. [4]). Additionally, Roache's criticism of using $P = 0.08$ is inconsistent with one of his previous conclusions that there is no necessity to discard results with $p_{RE} < 1$ ($P < 0.5$ for a second order method in Ref. [11]).

(3) The fact that "the use of the GCI_1 method is closer to a 68% than a 95% confidence level" was one of the conclusions by Logan and Nitta [6]. This conclusion was not just based on the dataset with intentional choice of grid studies with oscillations in both exponent p and output quantity. As stated in page 367 in Ref. [6], "However, for our contrived and mechanics example $N_S = 18$ sets (most of which were non-smooth), the use of GCI = 1.25 is much closer to a 68% confidence estimate than 95%." It was also recommended in Refs. [6] and [2] that a sample with the number of grid convergence studies much larger than 100 is needed to draw general conclusions.

We did not recommend the GCI_1 method but rather evaluated it using much larger sample sizes than Ref. [6]. For the largest sample 3 with size $N = 329$, the reliability **R** (Eq. (19) in Ref. [2]) is 90.3% for the GCI_1 method.

(4) We disagree with Roache's evaluation in Ref. [11] where it states that "Briefly, the net result is 14 NC (nonconservative) of 176 entries, or 8.0%." Only 151 of the 176 grid triplet studies have the actual error E. This results in 24 nonconservative of 151 (note there are nine nonconservative grid-triplet studies that estimate $U = E$). So, the reliability for the GCI method [12] is actually 84.1%, which agrees very well with the reliability 83.9% estimated using our 329 grid-triplet studies (sample 3 in Ref. [2]).

Based on our own evaluation above and the fact that Cadafalch et al. [12] used $F_S = 1.25$ for $P > 1$, the method they applied was not the GCI_2 method and more likely the GCI method. The claim of "an original and reasonable variant of the real GCI" [1] again is confusing.

(5) We take 95% coverage as the common uncertainty target for both experiments and computations [5]. Although the GCI_2 method only misses the overall reliability by 0.8% for sample 3, more importantly it fails to provide sufficient conservatism for other samples including the reliabilities of 91.4%, 90%, and 87.5% for samples 5, 8, and 16, respectively [2]. It is possible that another dataset could slightly change our evaluations. Nonetheless, the current sample size is large and the range of P values is wide such that a further increase of the number of samples is not likely to significantly alter the FS method and its results.

(6) The FS method was calibrated/validated against the available dataset. Note that calibration/validation requires that the true error can be evaluated, i.e., the solution numerical benchmark (S_{NB}) or solution analytical benchmark (S_{AB}) is known. We welcome additional validation of the FS method and if necessary recalibration and improvement, but again S_{NB} or S_{AB} must be known. The claim of Roache and others of the 95% reliability for the GCI method is undocumented and based on anecdotal information. We doubt that S_{NB} or S_{AB} is available for many of the cases cited by Roache and others. It should be a simple matter to provide proper documentation.

Note that the FS method is more conservative than the GCI_2 method except for $1 < P < 1.136$ due to the jump of the factor of safety at $P = 1$ for the GCI_2 method. If the FS method is not conservative enough for another dataset, the GCI_2 method will likely be worse.

The claim that the GCI_2 method has been stable for over 12 years is not well founded. Due to the lack of a single guideline on the choice of F_S and p, different variants of the GCI method have been used by different users based on their own judgment calls. For example, Cadafalch et al. [12] did not use the GCI_2 method, and Logan and Nitta [6] used the GCI_1 method. Furthermore, the GCI method may have been applied to $O(1000)$ cases but no statistical evidence for reliability has been documented.

(7) We disagree with Roache's suggestion that the FS method has problems in predicting monotonic convergence for fine grids. The uncertainty estimates in Table 6 for the FS method in Ref. [2] for the three finest grid triplets are not monotonically decreasing since P shows large oscillations, and the factor of safety for the second finest grid triplet (2, 3, 4) at $P = 1.49$ is much larger than that for the other methods evaluated at the same P. However, the larger factor of safety is required to ensure the reliability for $P > 1$. For the three grid triplets discussed, it is interesting to evaluate the convergence ratio R for the fine grid solution S_1 (R_{S_1}), P (R_P), and U_G (R_{U_G}). All the five verification methods have the same R_{S_1} and R_P, which show monotonic convergence. The GCI, GCI_1, and CF methods show monotonic convergence for U_G, whereas the GCI_2 and FS methods show monotonic divergence ($R_{U_G} = 2.74$) and oscillatory divergence ($R_{U_G} = -4.53$), respectively.

The oscillation of P may be caused by many factors. Grid 4 is still too coarse for the solution to be in the asymptotic range. Additionally, reducing the iterative error to machine zero is very difficult for large-scale computations. With the small grid refinement ratio $r = \sqrt[4]{2}$, solution changes ε will be small, and the sensitivity to grid-spacing and time step may be difficult to identify compared with iterative errors U_I. As shown in Fig. 6(b) in Ref. [10], $U_{I,1}/\varepsilon_{12} = 61.6\%$ for the cases in Table 6 [2]. When r increases, U_I/ε will likely decrease. For example, the grid uncertainty decreases from 5.04 for (2, 4, 6) to 4.02 for (1, 3, 5) with $U_{I,1}/\varepsilon_{13} = 20\%$ for $r = \sqrt{2}$. However, it should be noted that a large r may be problematic, too, as different grids may resolve different flow physics.

There are some other cases that the GCI, GCI_1, GCI_2, CF, and FS methods show non-monotonic convergence for multiple grid-triplet studies, including the "well-behaved" problems Cadafalch et al. [12] and Roache [11] used to evaluate the conservativeness of the GCI method. For the radial velocity using the SMART scheme in the study of premixed methane/air laminar flat flame on a perforated burner [13–15], the uncertainty estimates using the

FS and GCI_2 methods monotonically decreased whereas the other three methods did not as the grid is refined. Another example is for the uncertainty estimates for temperature at a monitored location for a two-dimensional natural convection in square cavities at $Ra = 10^6$, which had five grid-triplet studies with $r = 2$ [16]. Uncertainty estimates using the five verification methods discussed in Ref. [2] first monotonically decreased as the grid is refined but suddenly increased for the finest grid-triplet. Thus, it is unreasonable to blame the FS method as the reason for such behavior.

The verification results for our industrial application example are far from the asymptotic range. Although we evaluated the convergence characteristics for the 98 verification variables using P and $|E|$ as functions of $\Delta x_{fine}/\Delta x_{finest}$ [2], a standard criterion for achieving the asymptotic range is still lacking. A possible criterion is that monotonic convergence should be established based on evaluation of the convergence ratio R for fine grid solution S_1 (towards S_C), P (towards 1), and U (monotonically decreasing) for multiple (at least three) grid-triplets with the same grid refinement ratio r and $U_I \ll U$. In some cases, oscillatory convergence may be acceptable; however, this would require many grid triplets [17]. Although R still needs to be evaluated for all the variables in our dataset, 41.5% of the variables that have more than two grid-triplet studies do show that S_1 approaches S_C, P approaches 1, and U monotonically decreases as the grid is refined. For the other 58.5% of the variables, S_1 also approaches S_C as shown by monotonically decreased error magnitude $|E|$, but P and U_G often show mixed convergence conditions as the grid is refined.

(8) As discussed in item (6), without statistical evidence, the claim of the conservativeness of the GCI_2 method is undocumented. Furthermore, we doubt very much how many applications have S_{NB} or S_{AB}. If they do, we will be glad to add them to our dataset. The work by Dr. C. J. Freitas and his group is not publicly available [9]. Therefore, the claim of achieving the 95% reliability is again undocumented and based on anecdotal information.

(9) We agree that the actual factor of safety is undefined when a solution not in the asymptotic range happens to predict the true value. If this happens, it should be excluded from the dataset used to derive the FS method. However, monotonic convergence ensures that the uncertainty estimate is always greater than zero so that a zero error will be automatically bounded by the uncertainty.

The contrived example created by Roache only proves that the average actual factor of safety ($\overline{X}$) cannot be used alone to determine if a solution verification method is conservative enough. But it can be used to determine the relative conservativeness between different verification methods.

It should be noted that we used both the reliability $\mathbf{R}$ and LCL as defined by Eq. (22) in Ref. [2] to develop the FS method and determine if a method is conservative enough. Larger $\overline{X}$ does not necessarily mean larger $\mathbf{R}$ (readers can refer to sample 6 in Ref. [2]).

(10) As requested by Roache, we use our approach to evaluate two new variants of the GCI method proposed by Oberkampf and Roy [18] (GCI_{OR}) and by Roache [1] (GCI_3).

$$U_{GCI_{OR}} = \begin{cases} 1.25 \left| \dfrac{\varepsilon_{21}}{r^{p_{th}} - 1} \right|, & 0.9 \leq P \leq 1.1 \\[2ex] 3 \left| \dfrac{\varepsilon_{21}}{r^{\min(\max(0.5, p_{RE}), p_{th})} - 1} \right|, & 0 < P < 0.9 \ \text{ or } \ P > 1.1 \end{cases}$$

(2)

$$U_{GCI_3} = \begin{cases} 1.25 \left| \dfrac{\varepsilon_{21}}{r^{\min(p_{RE}, p_{th})} - 1} \right|, & 0.9 \leq P \leq 1.1 \\[2ex] 3 \left| \dfrac{\varepsilon_{21}}{r^{\min(p_{RE}, p_{th})} - 1} \right|, & 0 < P < 0.9 \ \text{ or } \ P > 1.1 \end{cases}$$

(3)

To address Roache's concern of using p_{RE} when $p_{RE} >> p_{th}$, we also evaluate an alternative form of the FS method (FS_1 method). The FS_1 method is the same as the FS method for $P < 1$ but uses p_{th} instead of p_{RE} in the error estimate for $P > 1$. Thus, Eq. (14) in Ref. [2] becomes
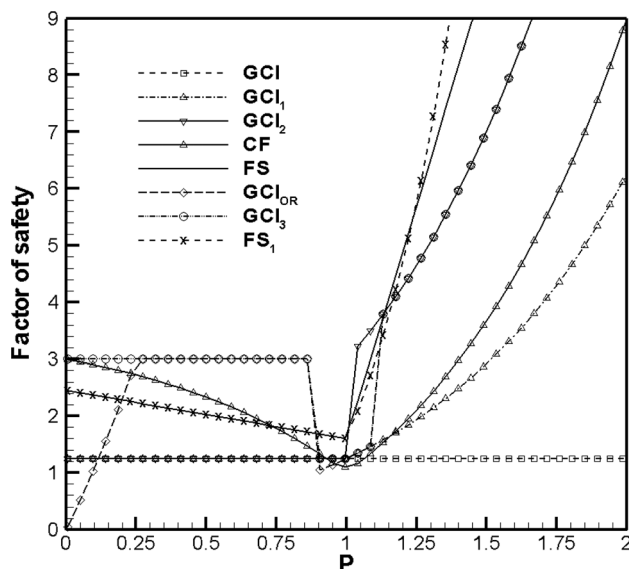
Fig. 1 Factor of safety for different verification methods with $p_{th} = 2$ and $r = 2$

$$U_{FS_1} = \begin{cases} [FS_1 P + FS_0(1-P)] \left| \dfrac{\varepsilon_{21}}{r^{p_{RE}} - 1} \right|, & 0 < P \leq 1 \\ [FS_1 P + FS_2(P-1)] \left| \dfrac{\varepsilon_{21}}{r^{p_{th}} - 1} \right|, & P > 1 \end{cases} \quad (4)$$

Following the same procedure described in Sec. 2.4 of Ref. [2], $FS_0 = 2.45$, $FS_1 = 1.6$, and $FS_2 = 6.9$ are recommended, and the final form of the FS_1 method is

$$U_{FS_1} = \begin{cases} (2.45 - 0.85P) \left| \dfrac{\varepsilon_{21}}{r^{p_{RE}} - 1} \right|, & 0 < P \leq 1 \\ (8.5P - 6.9) \left| \dfrac{\varepsilon_{21}}{r^{p_{th}} - 1} \right|, & P > 1 \end{cases} \quad (5)$$

To compare the relative conservativeness between different verification methods, the three new methods are rewritten in terms of the same error estimate δ_{RE}.

$$U_{GCI_{OR}} = \begin{cases} 1.25 CF |\delta_{RE}|, & 0.9 \leq P \leq 1.1 \\ \dfrac{3(r^{p_{RE}} - 1)}{r^{\min(\max(0.5, p_{RE}), p_{th})} - 1} |\delta_{RE}|, & 0 < P < 0.9 \ \text{ or } \ P > 1.1 \end{cases} \quad (6)$$

$$U_{GCI_3} = \begin{cases} \dfrac{1.25(r^{p_{RE}} - 1)}{r^{\min(p_{RE}, p_{th})} - 1} |\delta_{RE}|, & 0.9 \leq P \leq 1.1 \\ \dfrac{3(r^{p_{RE}} - 1)}{r^{\min(p_{RE}, p_{th})} - 1} |\delta_{RE}|, & 0 < P < 0.9 \ \text{ or } \ P > 1.1 \end{cases} \quad (7)$$

$$U_{FS_1} = \begin{cases} (2.45 - 0.85P)|\delta_{RE}|, & 0 < P \leq 1 \\ (8.5P - 6.9)CF|\delta_{RE}|, & P > 1 \end{cases} \quad (8)$$

The factors of safety for all the verification methods discussed so far are shown in Fig. 1. One problem of the GCI_2 method is the jump of factor of safety across the asymptotic range at $P = 1$. For two grid-triplet studies with one at $P = 0.999$ and the other at $P = 1.001$, the factor of safety suddenly increases from 1.25 to 3 even though P only varies by less than 0.2%. Eça et al. [19] gave similar comments on this issue: "However, it is not easy 'to accept' a jump of a factor of 2.4 in the uncertainty when the observed order of accuracy may vary by only 0.1." Similar problems exist for the GCI_{OR} and GCI_3 methods when p_{RE} differs from p_{th} by 10%. It should be noted that the GCI_{OR} method set the lower limit of p_{RE} to be larger than 0.5, which corresponds to $P \geq 0.25$ for a nominal second order method. Thus, the factor of safety for $P < 0.25$ for the GCI_{OR} method shown in Fig. 1 is only a result of the mathematical reformulation. Figure 1 also shows that the GCI_{OR} and GCI_3 methods are much more conservative than the other methods for $0.25 < P < 0.9$ and coincide with the GCI_2 method for $P > 1.1$. The FS_1 method is less and more conservative than the FS method for $1 < P \leq 1.235$ and $P > 1.235$, respectively.

Table 1 Statistics for different ranges of P values using non-averaged actual factor of safety

Sample	P	N	Statistics	GCI_2	GCI_{OR}	GCI_3	FS	FS_1	t
3	0–2 ($\overline{P} = 0.94$)	329	**R** (%N)	94.2	93.62	93.92	96.96	96.96	1.645
			$\overline{X}$	2.39	2.78	3.10	3.04	3.02	
			$S_{\overline{X}}(\%\overline{X})$	5.9	5.4	8.6	8.2	7.1	
			LCL	2.16	2.53	2.67	2.63	2.67	
4	0–0.4 ($\overline{P} = 0.24$)	12 (3.65%)	**R** (%N)	100	100	100	100	100	1.796
			$\overline{X}$	7.34	9.9	17.63	13.52	13.52	
			$S_{\overline{X}}(\%\overline{X})$	29	15	29	30.8	30.8	
			LCL	3.51	7.23	8.44	6.03	6.03	
5	0.4–0.9 ($\overline{P} = 0.70$)	81 (24.62%)	**R** (%N)	91.4	96.3	96.3	95.1	95.1	1.664
			$\overline{X}$	2.01	4.81	4.83	3.03	3.03	
			$S_{\overline{X}}(\%\overline{X})$	6.5	6.5	6.5	7.3	7.3	
			LCL	1.79	4.28	4.30	2.66	2.66	
6	0.9–1.1 ($\overline{P} = 0.98$)	176 (53.50%)	**R** (%N)	95.5	93.18	93.75	97.7	97.7	1.646
			$\overline{X}$	1.88	1.29	1.36	1.87	1.83	
			$S_{\overline{X}}(\%\overline{X})$	3.7	2.1	2.2	2.1	2.3	
			LCL	1.76	1.25	1.31	1.80	1.76	
7	1.1–1.5 ($\overline{P} = 1.19$)	50 (15.20%)	**R** (%N)	94	90	90	96	96	1.676
			$\overline{X}$	3.27	2.77	2.77	3.74	3.48	
			$S_{\overline{X}}(\%\overline{X})$	6.45	6.99	6.99	6.93	7.23	
			LCL	2.92	2.44	2.44	3.31	3.06	
8	1.5–2.0 ($\overline{P} = 1.63$)	10 (3.04%)	**R** (%N)	90	90	90	100	100	1.833
			$\overline{X}$	4.08	4.08	4.08	7.67	9.06	
			$S_{\overline{X}}(\%\overline{X})$	18.7	18.7	18.7	18.4	17.7	
			LCL	2.68	2.68	2.68	5.08	6.12	

Table 2 Statistics excluding outliers at seventeen *P* values

| Sample	P	N	Number of outliers	Statistics	GCI_2	GCI_{OR}	GCI_3	FS	FS_1	t
9	0.705	9	0	**R** (%N)	100	100	100	100	100	1.86
				$\overline{X}$	2.08	4.99	4.99	3.08	3.08	
				$S_{\overline{X}}(\%\overline{X})$	10.4	10.4	10.4	10.4	10.4	
				LCL	1.68	4.02	4.02	2.48	2.48	
10	0.755	5	0	**R** (%N)	100	100	100	100	100	2.13
				$\overline{X}$	1.87	4.49	4.49	2.71	2.71	
				$S_{\overline{X}}(\%\overline{X})$	11.2	11.2	11.2	11.2	11.2	
				LCL	1.43	3.42	3.42	2.07	2.07	
11	0.805	15	0	**R** (%N)	100	100	100	100	100	1.76
				$\overline{X}$	1.76	4.22	4.22	2.49	2.49	
				$S_{\overline{X}}(\%\overline{X})$	5.4	5.4	5.4	5.4	5.4	
				LCL	1.59	3.82	3.82	2.26	2.26	
12	0.855	6	0	**R** (%N)	100	100	100	100	100	2.02
				$\overline{X}$	1.58	3.79	3.79	2.18	2.18	
				$S_{\overline{X}}(\%\overline{X})$	9.6	9.6	9.6	9.6	9.6	
				LCL	1.27	3.06	3.06	1.76	1.76	
13	0.905	29	0	**R** (%N)	100	93.1	100	100	100	1.70
				$\overline{X}$	1.61	1.36	1.61	2.17	2.17	
				$S_{\overline{X}}(\%\overline{X})$	6.8	6.9	6.8	6.8	6.8	
				LCL	1.42	1.20	1.42	1.92	1.92	
14	0.925	5	0	**R** (%N)	100	100	100	100	100	2.13
				$\overline{X}$	1.42	1.27	1.42	1.89	1.89	
				$S_{\overline{X}}(\%\overline{X})$	3.1	2.2	3.1	3.1	3.1	
				LCL	1.33	1.21	1.33	1.76	1.76	
15	0.945	6	0	**R** (%N)	100	100	100	100	100	2.02
				$\overline{X}$	1.37	1.27	1.37	1.81	1.81	
				$S_{\overline{X}}(\%\overline{X})$	2.0	1.5	2.0	2.0	2.0	
				LCL	1.32	1.23	1.32	1.73	1.73	
16	0.955	16	0	**R** (%N)	87.5	87.5	87.5	100	100	1.75
				$\overline{X}$	1.42	1.32	1.42	1.87	1.87	
				$S_{\overline{X}}(\%\overline{X})$	6.8	6.6	6.8	6.8	6.8	
				LCL	1.25	1.16	1.25	1.64	1.64	
17	0.965	6	0	**R** (%N)	100	100	100	100	100	2.02
				$\overline{X}$	1.46	1.39	1.46	1.91	1.91	
				$S_{\overline{X}}(\%\overline{X})$	7.1	7.0	7.1	7.1	7.1	
				LCL	1.25	1.19	1.25	1.63	1.63	
18	0.975	7	0	**R** (%N)	100	100	100	100	100	1.94
				$\overline{X}$	1.33	1.28	1.33	1.73	1.73	
				$S_{\overline{X}}(\%\overline{X})$	3.6	3.3	3.6	3.6	3.6	
				LCL	1.24	1.20	1.24	1.61	1.61	
19	0.985	9	0	**R** (%N)	100	100	100	100	100	1.86
				$\overline{X}$	1.27	1.24	1.27	1.64	1.64	
				$S_{\overline{X}}(\%\overline{X})$	1.1	1.0	1.1	1.1	1.1	
				LCL	1.25	1.22	1.25	1.60	1.60	
20	0.995	8	0	**R** (%N)	100	100	100	100	100	1.90
				$\overline{X}$	1.24	1.22	1.24	1.59	1.59	
				$S_{\overline{X}}(\%\overline{X})$	2.0	2.1	2.0	2.0	2.0	
				LCL	1.19	1.18	1.19	1.53	1.53	
21	1.005	56	4	**R** (%N)	98.2	98.2	98.2	100	100	1.68
				$\overline{X}$	2.08	1.33	1.33	1.71	1.71	
				$S_{\overline{X}}(\%\overline{X})$	5.6	3.0	3.0	3.0	3.0	
				LCL	1.88	1.26	1.26	1.63	1.62	
22	1.015	6	0	**R** (%N)	100	100	100	100	100	2.02
				$\overline{X}$	2.90	1.21	1.21	1.74	1.67	
				$S_{\overline{X}}(\%\overline{X})$	3.0	3.0	3.0	3.6	3.4	
				LCL	2.73	1.14	1.14	1.62	1.56	
23	1.055	6	0	**R** (%N)	100	100	100	100	100	2.02
				$\overline{X}$	4.39	1.83	1.83	3.25	2.98	
				$S_{\overline{X}}(\%\overline{X})$	8.9	8.9	8.9	8.6	8.6	
				LCL	3.60	1.50	1.50	2.68	2.46	
24	1.105	13	2	**R** (%N)	100	84.6	84.6	100	100	1.78
				$\overline{X}$	3.65	1.71	1.71	3.43	2.99	
				$S_{\overline{X}}(\%\overline{X})$	14.1	12.2	12.2	14.3	14.1	
				LCL	2.73	1.33	1.33	2.56	2.24	
25	1.205	12	1	**R** (%N)	100	100	100	100	100	1.80
				$\overline{X}$	3.32	3.32	3.32	4.08	3.66	
				$S_{\overline{X}}(\%\overline{X})$	8.9	8.9	8.9	8.9	8.9	
				LCL	2.79	2.79	2.79	3.43	3.07	
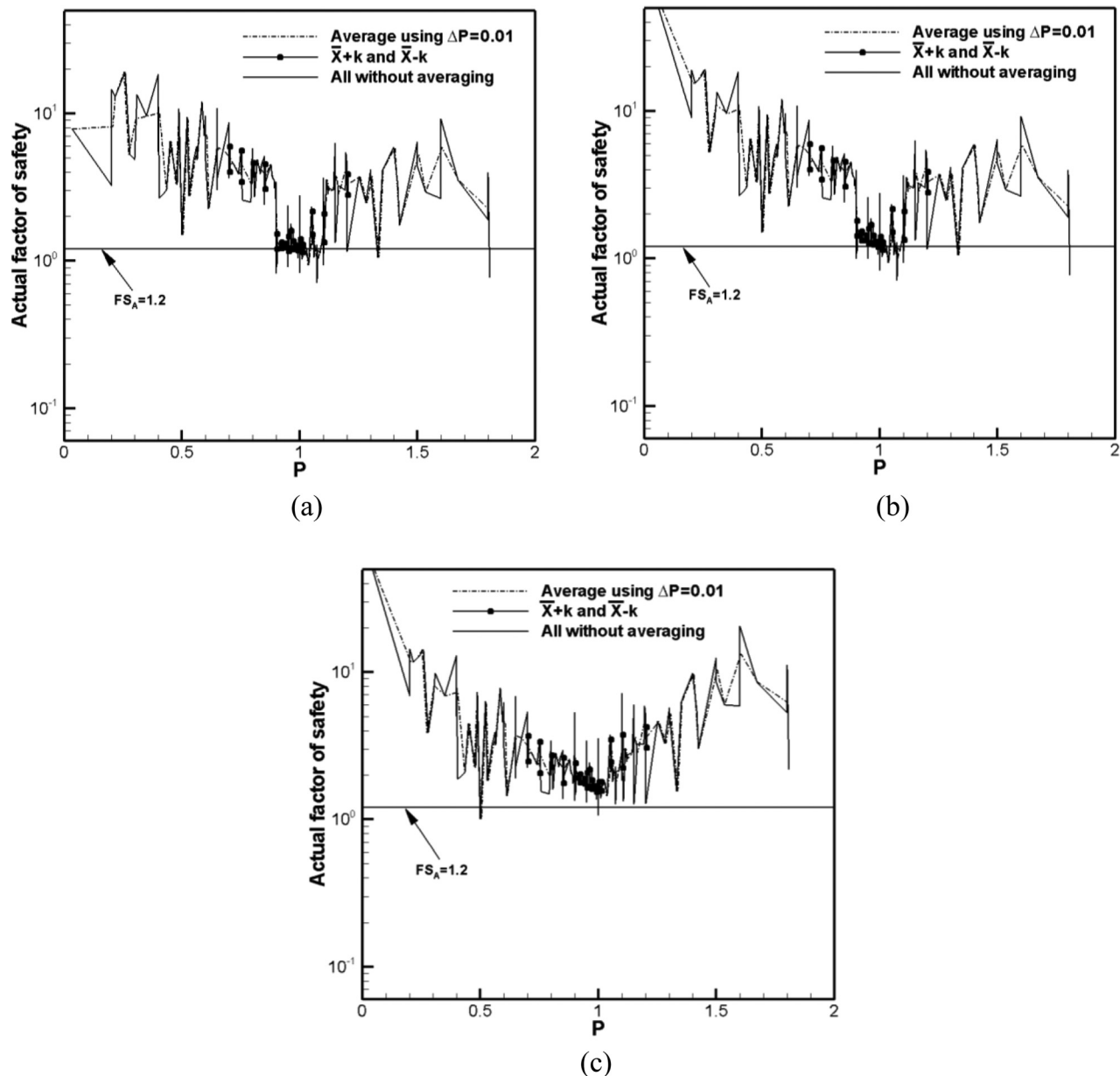
Fig. 2 Actual factor of safety for sample 3, sample 3 averaged using $\Delta P = 0.01$, and $\overline{X} \pm t S_{\overline{X}}$ for samples 9 to 25: (*a*) GCI_{OR} method, (*b*) GCI_3 method, and (*c*) FS_1 method

The GCI_{OR}, GCI_3, and FS_1 methods are evaluated using statistical analysis of the 25 samples following Ref. [2], with focus on samples 3 to 25. Table 1 shows the statistics for samples 3 to 8 [2] based on six different P ranges for the three new methods. The FS_1 method has the same reliability as the FS method for samples 3 to 8. The GCI_{OR} and GCI_3 methods almost have the same reliability, but the GCI_3 method is a little more conservative. Compared to the GCI_2 method, the GCI_{OR} and GCI_3 methods improve the reliability for $P < 1$ to be larger than 95% but are not conservative enough for $P \geq 1$, especially near the asymptotic range. 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 and FS_1 methods achieve 95% reliability, but the GCI_{OR} and GCI_3 methods achieve only 90%. The largest $\overline{X}$ for samples 3-5, sample 6, sample 7, and sample 8 are the GCI_3, GCI_2, FS, and FS_1 methods, respectively. For all the verification methods, the LCLs are larger than 1.2 for all the P ranges.

Table 2 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 reliabilities larger than 95% except 93.1% for the GCI_{OR} method at $P = 0.905$, 87.5% for the three GCI methods at $P = 0.955$, and 84.6% for the GCI_{OR} and GCI_3 methods at $P = 1.105$. The largest $\overline{X}$ for samples 9-12, samples 13-20, samples 21-24, and sample 25 are the GCI_{OR} and GCI_3 methods, FS and FS_1, GCI_2, and FS_1 methods, respectively. Only the FS and FS_1 methods satisfy the requirement that $LCL > 1.2$ for samples 9 to 25. The GCI_2 method has $LCL < 1.2$ for sample 20; the GCI_{OR} method has $LCL < 1.2$ for samples 13, 16, 17, 18, 20, and 22; and the GCI_3 method has $LCL < 1.2$ for samples 20 and 22.

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 $\overline{X} \pm t S_{\overline{X}}$ for samples 9 to 25 are shown in Fig. 2. t is the factor for the student-t distribution and $S_{\overline{X}}$ is the standard deviation of the mean of the sample, as defined in Ref. [2]. The GCI_{OR} and GCI_3 methods do not satisfy $LCL > 1.2$ near the asymptotic range. Compared to the FS method (Fig. 4(*e*) in Ref. [2]), the FS_1 method shows a larger actual factor of safety when solutions are farther from the asymptotic range for $P > 1$.

Concluding Remarks

The choice of F_S and p in the GCI method requires user judgment calls, for which no single guideline is currently available. We recommend that a single guideline be provided.

The GCI_{OR} and GCI_3 methods have almost the same reliability. But the GCI_3 method is a little more conservative. Compared to the GCI_2 method, the GCI_{OR} and GCI_3 methods improve the reliability for $P < 1$. However, they are too conservative for $P < 0.9$ using a factor of safety 3 and not conservative enough for $P \geq 1.1$.

The FS_1 and FS methods are the same for $P \leq 1$. For $p_{th} = 2$ and $r = 2$, the FS_1 method is less and more conservative than the FS method for $1 < P \leq 1.235$ and $P > 1.235$, respectively. As a result, the FS_1 method may have an advantage for uncertainty estimates when $P > 2$ where the FS and other verification methods likely predict unreasonably small uncertainties due to small error estimates. However, since the current dataset is restricted to $P < 2$, the pros/cons of using the FS or FS_1 method cannot be validated. Thus, until additional data is available for $P > 2$, all verification methods should be used with caution for such conditions and, if possible, additional grid-triplet studies conducted to obtain $P < 2$.

The authors' statistical approach based on many analytical and numerical benchmarks provides a robust framework for developing solution verification methods. The authors welcome additional validation of the FS method and, if necessary, re-calibration and improvement using additional rigorous verification studies with S_{AB} or S_{NB} available. More research is needed to establish the criterion for achieving the asymptotic range along with its use in providing high quality numerical benchmarks.

Acknowledgment

References

[1] Roache, P. J., 2011, "Discussion: 'Factors of Safety for Richardson Extrapolation'," ASME J. Fluids Eng., **133**(11), p. 115501.
[2] Xing, T., and Stern, F., 2010, "Factors of Safety for Richardson Extrapolation," J. Fluids Eng., **132**(6), p. 061403.
[3] Roache, P. J., 1998, *Verification and Validation in Computational Science and Engineering*, Hermosa, Albuquerque, NM.
[4] Celik, I. B., Ghia, U., Roache, P. J., Freitas, C. J., Coleman, H., and Raad, P. E., 2008, "Procedure for Estimation and Reporting of Uncertainty Due to Discretization in CFD Applications," ASME J. Fluids Eng., **130**(7), p. 078001.
[5] ASME Committee PTC-61, 2008, "ANSI Standard V&V20," *ASME Guide on Verification and Validation in Computational Fluid Dynamics and Heat Transfer*, Nov. 30, 2009.
[6] 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.
[7] Roache, P. J., 2009, private communication.
[8] Xing, T., and Stern, F., 2009, "Factors of Safety for Richardson Extrapolation for Industrial Applications," IIHR Report No. 469.
[9] Roache, P. J., 2009, *Fundamentals of Verification and Validation*, Hermosa, Albuquerque, NM.
[10] 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), p. 101102.
[11] Roache, P. J., 2003, "Conservatism of the Grid Convergence Index in Finite Volume Computations on Steady-State Fluid Flow and Heat Transfer," ASME J. Fluids Eng., **125**(4), pp. 731–732.
[12] 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.
[13] 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.
[14] 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.
[15] 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.
[16] Hortmann, M., Perić, M., and Scheuerer, G., 1990, "Multigrid Benchmark Solutions for Laminar Natural Convection Flows in Square Cavities," *Benchmark Test Cases for Computational Fluid Dynamics*, I. Celik and C. J. Freitas, eds., ASME, New York, pp. 1–6.
[17] Coleman, H. W., Stern, F., Di Mascio, A., and Campana, E., 2001, "The Problem With Oscillatory Behavior in Grid Convergence Studies," ASME J. Fluids Eng., **123**(2), pp. 438–439.
[18] Oberkampf, W. L., and Roy, C. J., 2010, *Verification and Validation in Scientific Computing*, Cambridge University Press, New York.
[19] Eça, L., Vaz, G., and Hoekstra, M., 2010, "Code Verification, Solution Verification and Validation in RANS Solvers," *Proceedings of the ASME 2010 29th International Conference on Ocean, Offshore, and Arctic Engineering (OMAE2010)*, June 6–11, 2010, Shanghai, China.