

Discussion: “Factors of Safety for Richardson Extrapolation” (Xing, T., and Stern, F., 2010, ASME J. Fluids Eng., 132, p. 061403)

Patrick J. Roache

Visiting Professor (Melchor Chair),
Department of Aerospace and Mechanical Engineering,
University of Notre Dame,
Notre Dame, IN 46556
e-mail: hermosa@sdsc.org

1 Introduction

A recent paper by Xing and Stern [1] provides additional performance data of interest on Verification of Calculations, but is misleading in some important respects. Five methods, all starting from Richardson Extrapolation, are compared: three are based on variants of my Grid Convergence Index method [2–6], and two are based on the authors’ work: the earlier “CF method” and the latest “FS method.”

(1) The method referred to as GCI_2 is successful but its source is cited as a “private communication” from myself (their Ref. [16]). What they refer to as GCI_2 is essentially the GCI method, as I have described and discussed and evaluated at length, Refs. [2–6]. This is the method that, as the authors say [1], “is widely used and recommended by ASME and AIAA” citing [6,7]. I shall refer to it herein as the “real GCI method” for reasons soon to be apparent. The equivalence is obscured because the authors [1] have written their Eq. (12) defining the method using their own ratio term CF, which obscures the simplicity of the original and makes it appear that the factor of safety varies continuously¹ with observed order of convergence p_{RE} ; see Fig. 3 of Ref. [1]. This is merely an artifact of Eq. (12) in which their factor of safety FS is defined² in terms of an error estimate based on observed p_{RE} rather than on the p actually used. As a result, the “factor of safety” FS in Ref. [1] is not always the same concept as the “factor of safety” F_s in Refs. [2–6].

This leaves the question of what are the methods called “GCI” and “ GCI_1 ” by the authors. The answer is that I consider these to be misapplications of the GCI formulas, and therefore misrepresentations of the GCI method.

(2) For specificity, I will change notation and refer to the method of their Eq. (10) as GCI_0 , not as “GCI” since I claim it is not the real GCI method. This Eq. (10) is written in terms of an error estimate based on the “observed” [2–6] or “estimated” [1] order of convergence p_{RE} .

However, the authors correctly indicate in the text following that this is possible only if three grids have been used to calculate a p_{RE} , in which case the recommended Factor of Safety $F_s = 1.25$.

¹In the original GCI method [2], F_s does depend on observed p_{RE} implicitly, in the sense that if p_{RE} is unreasonable one at least reverts to using p_{th} and $F_s = 3$ rather than blindly applying $F_s = 1.25$. The preferred approach would be to continue the grid convergence study with additional grids until reasonable values of p_{RE} were obtained. These rather obvious points have been made more explicit in Ref. [3] which was not available to the authors of Ref. [1].

²There is further unfortunate confusion caused by the authors’ choice of a descriptor for their method and their use of the symbol FS both for their “Factor of Safety” method and for the factors of safety used in all five methods, including the “factor of safety used in the Factor of Safety method.” Likewise for their symbol CF, which denotes both their Correction Factor method and the “correction factor” used in the Correction Factor method as well as in their defining equations for other methods (Eqs. (11) and (12)).

Contributed by the Fluids Engineering Division of ASME for publication in the JOURNAL OF FLUIDS ENGINEERING. Manuscript received October 19, 2010; final manuscript received July 28, 2011; published online October 24, 2011. Assoc. Editor: Dimitris Drikakis.

If only two grids have been used, p_{RE} in Eq. (10) is replaced by a theoretical value p_{th} (e.g., $p_{th} = 2$ for a nominally second-order method) and a more conservative value $F_s = 3$ is recommended [2–6]. This would indeed agree with my GCI method [2–6] in good applications. The trouble is that the authors have applied this regardless of the reasonableness of the estimated p_{RE} . This is contrary to the discussions in Ref. [2], which discussions I claim are part of the real “GCI method” as contrasted with simplistic application of the formula.

Exactly what constitutes a reasonable value was not defined in my work [2–6] and I would not take issue with (say) a 5% difference between p_{RE} and p_{th} , e.g., using an observed $p_{RE} = 2.1$ for a nominally second-order method. (Although it should be obvious that the conservative approach would be to use the minimum of p_{RE} and p_{th} , I would not claim that this was a misapplication of the GCI method.) However, the authors [1] have applied the GCI_0 formula using observed $p_{RE} \gg$ theoretical p_{th} , which is imprudent and not recommended [2–6] no matter what the value of F_s . This point was conveyed to the authors of Ref. [1] in their Ref. [16], the private communication from myself. (I am glad to make this communication, cited here as Ref. [8], available to any interested party.)

Using the authors’ notation $P = p_{RE}/p_{th}$ the authors have claimed rational estimates of uncertainty from studies with P as high as 6.01 for the new ship hydrodynamics problem (see Table 7, second line). Since the numerical method used is nominally second order, this indicates an observed order $p_{RE} = 12.02$, the use of which would be ridiculous. Another instance is the third row entry in Table 6 [1] for the grid triplet (4,5,6) showing $P = 0.08$. These are pretty strong hints that something is wrong. There is no sense in performing a grid convergence study and then ignoring the results. No excuses of the computing difficulties in industrial applications can justify an attempt to obtain a rational estimate of uncertainty from such meaningless corrupted studies. To so attempt is to give merit to meritless results. I can only describe this as a misapplication of the GCI formula, not as a poor result for the real GCI method.

(3) The second method considered in Ref. [1] is GCI_1 based on Eq. (11). The appearance of the CF term in the equation serves to exchange the error estimate based on observed p_{RE} to one based on theoretical p_{th} for the condition $P > 1$, i.e., for $p_{RE} > p_{th}$. This is exactly what is recommended above, and is good conservative practice. However, the Eq. (11) uses $F_s = 1.25$ in either case. This does not make sense. The GCI_1 method as described uses three grid solutions to evaluate an observed p_{RE} , but if this value is too large, the method reverts to using exactly the error estimator that would have been used for only two-grid studies but with the non-conservative value $F_s = 1.25$ instead of the recommended $F_s = 3$. Thus, the authors did use the correct factor of safety $F_s = 3$ as recommended [2–6] for 2 grid studies but they used only $F_s = 1.25$ for 3-grid studies in which the observed order of convergence p_{RE} was not at least approximately consistent with theory.

The authors then concluded from their tests that “These facts suggest that the use of the GCI_1 method is closer to a 68% than a 95% confidence level.” This is a misleading statement since the conclusion follows from using $F_s = 1.25$ and applying it to a dataset from [9] designed with “intentional choice of grid studies with oscillations in both exponent p and output quantity” [9]³ (Ref. [3]). The point of such studies as Ref. [1] should not be to tune large F_s until 95% coverage is obtained no matter how bad the data set is. If such large F_s are required, this should be a flag alerting the investigator that the grids used are inadequate for the problem [3,10,11].

(4) Therefore, the evaluation in Ref. [1] of the real GCI method (GCI_2) is corrupted by confusion with two distorted methods. The authors’ evaluation also does not agree with my own evaluation from some of the same data, notably the wide-ranging study of Cadafalch et al. [12] (Ref. [33] of Ref. [1]) which I presented in

³See also Ref. [3], p. 219. The study [9] also did not use $F_s = 3$, but this would improve the coverage only lightly for this poor data set [10].

Ref. [4] and to which I had alerted the authors by [8]. Cadafalch et al. used an original and reasonable variant of the real GCI in which a mesh-averaged p_{RE} was used for local uncertainty estimates. The literature survey in Ref. [1] omitted my evaluation [4] which itself covered ~ 200 cases. As I had pointed out in Ref. [8], my evaluation of these uncertainty calculations using the GCI with $F_s = 1.25$ showed 92% coverage overall, less for upwind differencing, as expected, and 97.7% for adaptive differencing. There was one case of 176 that was found worrisome; it had $p_{RE} = 1.2$ for an expected first-order method.

(5) When the confusion over the identity of the real GCI is cleared up, the overall evaluation from Ref. [1] of the real GCI method (GCI_2) is excellent. The reporting of the various subsets of data is too convoluted to repeat here completely (and the process is suspect in one aspect — see item (9) below). Briefly, for the largest subsets the real GCI method (GCI_2) produces reliability of about 93% (in some subsets, 94% and higher) and the FS method produces about 96%.

The authors claim that only their latest “FS method provides a reliability larger than 95%.” The FS method certainly works, but I consider it virtually meaningless to distinguish between 93% coverage for the real GCI method (GCI_2) versus 96% for the FS method when the target is roughly 95% [2–6]. Note, we are not talking about the solution itself, or even its error, but about *uncertainty*. It is well recognized that the target of 95%, while traditional and useful, is itself somewhat arbitrary. There is an implicit precision level evidenced by the usual lack of distinction between 95%, or 20:1 odds, or $2-\sigma$ for a Gaussian distribution [4]⁴, and which are only approximately equivalent; this is true also for experimental uncertainties. Note also that greater coverage (say 99%) would not necessarily be considered more successful since the expanded coverage would be considered unnecessarily conservative, giving an unrealistically pessimistic estimate of uncertainty. So the real GCI method (GCI_2) misses the target of roughly 95% by -2% , and the FS method misses it by $+1\%$, for this dataset. Their large dataset used is a good one⁵ and is likely representative of problems of interest, but no one would be surprised if another dataset slightly changed the evaluations.

(6) In this regard, the 93% (and higher) coverage by the real GCI method (GCI_2) represents a confirmation of the method (with a tolerance of -2%) but the results for the authors’ FS method cannot strictly be considered a confirmation because their exercise is actually a calibration. (Here I distinguish *confirmation* versus *calibration* of an uncertainty estimator, analogous to the widely accepted distinction between *validation* and *calibration* of a model to experimental data [2,3,5].) In this study [1], the three free parameters (see Eq. (15)) of their FS method were *tuned* with the objective of attaining 95% coverage, so their claim of *achieving* $> 95\%$ coverage is somewhat circular. One would expect the tuned parameters to hold generally for other datasets, but strictly speaking their FS method could not be considered confirmed until it is applied successfully to another dataset not used

⁴Note that no distribution of any kind has been assumed in the development or testing in Ref. [1] nor in any of the other work cited herein.

⁵I applaud the authors examination on a large number of cases, but note that some of these are old studies, not dependable, with wildly varying results that are perhaps better discarded. On the other hand, the discard of outliers (4 discarded outliers in Sample 21, Table 5 of Ref. [1]) based on Peirce’s criterion may be acceptable but is debatable without a justification based on quality of the studies. See discussion in Ref. [3], Sec. 5.16.

⁶The real GCI method (GCI_2) was published in mature detail in Ref. [2] in 1998. The earlier publications of 1993–1994 [13,14] already considered a likely range of less conservative F_s but recommended $F_s = 3$ pending further extensive studies. “But much experimentation would be required over an ensemble of problems to determine a near-optimum value and to establish the correspondence with statistical measures such as the $2-\sigma$ band.” [15] When those initial studies were completed in 1998, the result was $F_s = 1.25$ for well-behaved problems, as described in Ref. [2]. The method has been stable since, through a wide range of new applications [3]. The important new development has been the extension by Eça and Hoekstra to a Least-Squares GCI (beginning with [16] in 2000 and continuing through the decade) suitable for cases of noisy data; see Ref. [3], Sec. 5.11 for additional references and some details.

in the calibration. By contrast, the real GCI method (GCI_2), by now applied to probably $O(1000)$ cases by many independent modelers, has been stable for over 12 years.⁶ It has one parameter F_s that takes on two possible values, depending on the thoroughness and results of the grid convergence study. $F_s = 1.25$ works well for thorough, well-behaved studies (meaning a minimum of three grids to allow calculation of observed p_{RE} , and p_{RE} in adequate agreement with theory and/or adequately stable over multiple grid triplets, which evaluation is, of course, a judgment call). Presumably, if one tuned it to $F_s = 1.27$, or something equally pointless, the real GCI method also could achieve 96% coverage for the subject dataset, accomplished using only one tuned parameter instead of three as in the latest FS method.

(7) The latest FS method is an improvement over the authors’ previous CF method, in all its versions.⁷ Whether or not the performance is worth the additional tuned empirical parameters (three instead of one) is a matter of opinion. However, it is also true that the authors own results, when examined closely, provide cautionary evidence. In an earlier version of the method [17] (their Ref. [18]) there were problems with performance on the finer grids. (This was communicated to the authors in Ref. [8], their Ref. [16]; see also Ref. [3], p. 229.) The latest version [1] is improved in this regard but is still problematical in the same way, as discussed next.

Less obvious than the previously noted results for values of $P = 6.01$ and 0.08 , but more problematical, is the result for the FS method on the ship hydrodynamics problem. (The finest grid admirably uses 8.1×10^6 points.) Uncertainty estimates for well behaved problems should decrease monotonically as the grid is refined. The uncertainty estimates in Table 6 for the FS method for the three finest grid triplets are 7.62, 8.30, and 5.22, which are not monotonically decreasing. Again, something is wrong. All four other methods show monotonic convergence⁸, e.g., the real GCI method (denoted GCI_2 in the table) gives 4.98, 4.21, 2.10.

In Ref. [17], the authors had noted that the (then new version of the) CF method was “unfortunately invalid” for the finest grid triplet. The authors claimed this was “caused by the contamination of the iterative error on the fine grid.” This is possible, but (a) we do not know how the iterative error was estimated, (b) Eça and Hoekstra [20]⁹ have shown that common methods of estimating iteration error are grossly underpredictive, (c) most importantly, Why is this not a problem for the other methods? They all use the same data (i.e., results from the same computations). At the least, we must conclude that the new FS method is more sensitive to noise than the others. Eça [11] notes that the explanation is simple: the FS method determines the factor of safety from p_{RE} so it is directly affected by the noise in the data. It is intuitively appealing to algorithmically adjust F_s on the basis of “distance from the asymptotic range” as calculated by some measure, but any such uncertainty estimator will be prone to erratic nonmonotone behavior in the same “industrial applications” which motivate its use. As Eça [11] notes, the noisy behavior is not necessarily due to distance from the asymptotic range but can be caused by other factors: lack of iterative convergence, accumulated computer round-off errors, use of block-generated grids or other grid generation methods that compromise strict geometric similarity in a grid sequence, switches embedded in turbulence models, etc. [2,3].

(8) The statement in Ref. [1] that my previous studies did not use statistical techniques apparently refers to my choice to not use the Students t -test or other small sample correction to estimate statistical confidence. Instead, I had just reported straightforward “coverage” or counting of $O(500)$ cases (incrementally accumulated over years) for which the GCI was conservative or not,

⁷See Ref. [3], p. 228 for a spotty history of the confusion of variations in the earlier versions and descriptors.

⁸It is not indicated why the results for the CF method reported in Ref. [1] are different and improved from the results of the CF method in Ref. [17]. See also the critiques in Refs. [18,19].

⁹See also Ref. [3], Sec. 5.10.10.3 for additional references and some details.

compared to the actual error. The evaluation of statistical confidence level is based on a conceptual model in which the examined data set of an individual study is taken to represent the entire population which is inaccessible. If this data set is small, then small sample correction techniques (like Student's t -test) are applicable. The simpler approach just claims (say) "89.2% coverage" for an individual study (perhaps small, or not) with the idea that the results of many studies will eventually be aggregated (probably informally, as I did). If small sample corrections for "confidence interval" are made to each study, later aggregation is confused because the small sample corrections are nonlinear and nondistributive. See Ref. [3], p. 229.

The real GCI (GCI_2) is indeed widely used and recommended, as noted by Ref. [1]. It is impossible to keep track of its many applications. It is surely well over 500, as we claimed in [5], and more likely of $O(1000)$. In Ref. [3], Sections 6.23.3–4 there is a list of 14 types of problems solved by Dr. C. J. Freitas and his group at the Southwest Research Institute. They have applied the real GCI to many realistic problems with far from ideal convergence behaviors with good results. These include confined detonations, vortex cavitation, blast-structure interactions, fluidized beds, shock propagation, structural failure, and turbulent multiphase flows, which problems were solved using FVM, FDM, FEM, ALE, and Riemann solvers. "In all applications the GCI provided a very good estimate of uncertainty [21]."

(9) Although the results in Ref. [1] are consistent with previous studies, their methodology appears to be flawed in one aspect. The authors' define the so-called "actual factor of safety" $FS_{A,i}$ for each case i (each grid triplet) and use average values (averaged over a study) to discriminate performance. The five uncertainty estimators considered are all in the form $U_{est} = Fs \times |E_{est}|$ where E_{est} is some estimate of the discretization error. $FS_{A,i}$ was obtained formally by replacing $E_{est,i}$ by the actual error E_i for each i case and inverting the algebra to solve for $FS_{A,i} = (U_{est,i}/|E_i|)$. But there is no "actual factor of safety" other than the value used. The ratio $(E_{est,i}/E_i)$ has been properly interpreted in Ref. [1] and elsewhere as an "effectivity index" and the "actual factor of safety" is similar to an effectivity index but applied to an uncertainty estimate rather than an error estimate. Values $(E_i/E_{est,i}) \gg 1$ or $FS_{A,i} \gg 1$ would indicate excessive conservatism for a single case. But this would say little or nothing about % coverage attained for an ensemble of cases, which is the agreed-upon target or principal performance metric. Of two uncertainty estimators with the same acceptable coverage, the one with the smaller average or maximum effectivity index would be preferred, but it would not make sense to actually use this "actual factor of safety" in any estimator of uncertainty for a new calculation.

Suspicion of the concept arises when we note that if one case happens to produce the correct solution, the result is an "actual factor of safety" $FS_{A,i} = \infty$, which is difficult to interpret and suggests something is wrong. Indeed, Fig. 4 of Ref. [1] shows three methods producing maximum values ~ 30 and two (including the latest FS method) going off-scale on log plots with $FS_{A,i} \sim 60$.

A single value of $FS_{A,i} < 1$ indicates that the actual single error is outside the uncertainty band, which should happen ~ 1 in 20 times. As the measure of coverage or reliability (Eq. (19) of Ref. [1]) this test is correct. A single value of $FS_{A,i} \gg 1$ would indicate large conservatism for that single case, which, combined with experience and intuition, may suggest an evaluation of excessive conservatism for the method.¹⁰ However, comparing $|E_i|$ to $U_{est,i}$ in whatever formulation cannot be used to faithfully indicate conservatism; it cannot be relied upon as a substitute for examining a large number of cases. Nor can the "average actual $FS_{A,i}$ " of a large study, as used in Ref. [1]. That this can be mis-

leading even for benign data is shown by the following synthetic example.

Consider a data set of 20 cases with the fine-grid solution variables of interest normalized to $S_1 = 1$. For convenience, say all normalized uncertainty estimates are $U_{est} = 0.1$, i.e., an estimated 10% "error band." (This might have been estimated by the real GCI method for well behaved cases using $Fs = 1.25$ from an error estimate of 0.08.) Suppose the true values for the first 19 of 20 cases are given by $T_i = 1 + i/200$ or $T_i = 1.005, 1.01, 1.015, \dots, 1.095$. Suppose $T(20) = 1.11$ so this case is not likely to be discarded as an outlier. It is seen that the GCI uncertainty estimate performs exactly as intended for this synthetic case, with the interval $[S_1 \pm U_{est}] = [0.9, 1.1]$ capturing the first 19 of 20 cases for 95% coverage. However, the "actual $FS_{A,i}$ " ranges from 0.91 to 20, and the "average actual FS_A " = 3.59, which would seem to suggest that the original method using $Fs = 1.25$ is much more conservative than intended. It is not; it had performed perfectly for this example.¹¹ The actual errors are not weirdly distributed; they range from 0.005 to 0.11 with an average error of 0.053, not out of line with the 95% uncertainty estimate of 0.1. If one thinks that the ratio of the original Fs to the "average actual FS_A " can be used to tune the Fs , one is in for a disappointment. The new Fs would be $1.25/3.59 = 0.348$, giving new $U_{est} = 0.028$, producing coverage for the first 5/20 cases or just 25% coverage.

These benign synthetic examples demonstrate that the concept used in Ref. [1] of relying on the so-called "average actual factor of safety" as an indicator of quality is misguided.

(11) Finally, some of the confusion of the authors [1] in not identifying the GCI_2 method as the real (original) GCI method [2–6] is understandable due to the fact that application of the method involves a judgment call as to what constitutes a reasonable value of observed p_{RE} . This makes it difficult for the authors to compare performance of different methods without being subject to unfair criticism. The judgment calls implicit in the earlier presentation of the GCI method in Ref. [2] have been made more explicit in Ref. [3] (which of course was not available to the authors of Ref. [1]) but the method is still not free of judgment calls. It would be preferable if all methods were defined in a recipelike fashion with no room for judgment calls. However, there are other judgment issues possible in a grid convergence study, e.g., how many grid triplets are adequate to claim that p_{RE} is reasonably constant, whether or not to use a least-squares evaluation, which detailed least-squares version to use, etc. I do not expect that such judgment calls will be eliminated for poorly behaved convergence studies.

In their new V&V book, Oberkampf and Roy ([22], p. 326) have proposed a procedure that standardizes the judgment call at least for the question of what constitutes acceptable p_{RE} . Their procedure accepts 10% agreement between p_{RE} and p_{th} , using $p = p_{th}$ and $Fs = 1.25$ in the GCI formula. For disagreement $> 10\%$, the procedure uses $Fs = 3$ and limits the p used in the GCI formula to $p = \min\{\max(0.5, p_{RE}), p_{th}\}$. I consider this a reasonable recipe, although of course other cut-offs could be justified as well. The limits on 10% agreement and the lower limit on $p = 0.5$ (provided that $p_{th} > 0.5$) are obviously judgment calls but are reasonable and unambiguous, and therefore repeatable in a comparative study such as Ref. [1]. A more conservative (perhaps overly conservative) procedure would not set a lower limit on p other than $p > 0$ (allowing for difficult problems such as singularities as in Ref. [3], Sec. 5.10.4.1) and would use $p = \min(p_{RE}, p_{th})$ when these agree within 10% instead of p_{th} .

¹¹Other synthetic examples can be constructed to demonstrate that the concept of "average actual factor of safety" can mislead in the other direction, suggesting adequate conservatism when in fact the method is behaving terribly. Consider true values $T_i = 1.1 + \alpha$ and $\alpha \ll 1$ for $i = 1$ to 19, and $T_{20} = 1.05$, which is not an outlier. Then "average actual FS_A " $\approx 21/20 = 1.05$, falsely indicating mild conservatism when in fact the method has missed 19 of 20 cases, giving 5% coverage rather than the intended 95%.

¹⁰See discussion in Ref. [3], Sec. 5.15 on "Evaluation of Uncertainty Estimators from Small Sample Studies."

These two prescriptive procedures would be additional candidates for evaluation by the authors of Ref. [1] using their extensive dataset.

Acknowledgment

The author gratefully acknowledges discussions with Professor Luis Eça, Dr. C. J. Freitas, Dr. R. W. Logan, Professor C. J. Roy, Maxwell Agnew and Aaron Donahue, and input from an anonymous reviewer.

References

- [1] Xing, T., and Stern, F., 2010, "Factors of Safety for Richardson Extrapolation," *ASME J. Fluids Eng.*, **132**(6), p. 061403.
- [2] Roache, P. J., 1998, *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers, NM.
- [3] Roache, P. J., 2009, *Fundamentals of Verification and Validation*, Hermosa Publishers, NM.
- [4] Roache, P. J., 2003, "Conservatism of the GCI in Finite Volume Computations on Steady State Fluid Flow and Heat Transfer," *ASME J. Fluids Eng.*, **125**(4), pp. 731–732.
- [5] ASME Committee V&V 20, 2009, *ANSI Standard V&V 20. ASME Guide on Verification and Validation in Computational Fluid Dynamics and Heat Transfer*.
- [6] 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), p. 078001.
- [7] 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.
- [8] Roache, P. J., 2009, private communication to F. Stern and others.
- [9] Logan, R. W., and Nitta, C. K., 2006, "Comparing 10 Methods for Solution Verification, and Linking to Model Validation," *AIAA J. Aerosp. Comput. Inf. Commun.*, **3**(7), pp. 354–373.
- [10] Logan, R., 2010, private communication to P. Roache.
- [11] Eça, L., 2010, private communication to P. Roache.
- [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] Roache, P. J., 1993, "A Method for Uniform Reporting of Grid Refinement Studies," *Quantification of Uncertainty in Computational Fluid Dynamics, ASME FED*, **158**, pp. 109–120.
- [14] Roache, P. J., 1993, *Proceedings of the AIAA 11th Computational Fluid Dynamics Conference*, Part 2, Orlando, Florida, 6–9 July 1993, pp. 1057–1058.
- [15] Roache, P. J., 1994, "Perspective: A Method for Uniform Reporting of Grid Refinement Studies," *ASME J. Fluids Eng.*, **116**(9), pp. 405–413.
- [16] Eça, L., and Hoekstra, M., 2000, "On the Application of Verification Procedures in Computational Fluid Dynamics," *2nd MARNET Workshop*, MARIN Maritime Research Institute, the Netherlands.
- [17] Xing, T., and Stern, F., 2009, "Factors of Safety for Richardson Extrapolation for Industrial Applications," IIHR Report No. 469, Univ. of Iowa, Iowa City, IA.
- [18] Roache, P. J., 2003, "Criticisms of the 'Correction Factor' Verification Method," *ASME J. Fluids Eng.*, **125**(4), pp. 732–733.
- [19] Roache, P. J., 2003, "Error Bars for CFD," AIAA 41st Aerospace Sciences Meeting, January 2003, Reno, Nevada, AIAA Paper No. 2003-0408.
- [20] Eça, L., and M. Hoekstra, M., 2006, "On the Influence of the Iterative Error in the Numerical Uncertainty of Ship Viscous Flow Calculations," Proc. 26th Symposium on Naval Hydrodynamics, Rome, Italy, 17–22 Sept.
- [21] Freitas, C. J., 2009, personal communication, 13 May 2009.
- [22] Oberkampf, W. L., and Roy, C. J., 2010, *Verification and Validation in Scientific Computing*, Cambridge University Press, Cambridge, UK.