Recent Developments in Verification and Validation

Patrick J. Roache
Consultant, Socorro, New Mexico, USA
hermosa@swcp.com
2nd Workshop on CFD Uncertainty Analysis
19-20 October 2006
Instituto Superior Técnico
Lisbon, Portugal

Abstract

A brief review will be given of several recent developments in CFD that bear on Calculation Verification and more generally on Verification and Validation. Topics include the following.

1. Effect of Incomplete Iteration Error on Accuracy and Calculation Verification
2. V&V Committee Progress: AIAA, ASME CSM, ASME CFD, ASCE
3. Journal Policy Statements on V&V
4. An Invalid Approach to Verification and Validation
5. Observed Grid Convergence Rate *p* Affected by Asymmetrical Grid Refinement
6. Uncertainty Calculations for Large Eddy Simulation
7. Grid Convergence Index for Unstructured Grids
8. Singularity Treatment in Calculation Verification
9. Non-Uniqueness in Turbulence Modeling
10. New K-ω Turbulence Model
11. Commercial Code Use
12. Challenges to the Concept of Validation

1. Effect of Incomplete Iteration Error on Accuracy and Calculation Verification

Eça and Hoekstra [1] recently presented a paper on the influence of the incomplete iteration error in the numerical uncertainty estimates in CFD. At the risk of embarrassing our hosts at Lisbon II, honesty compels me to state that I consider it a major contribution to a largely overlooked area. The paper continues the tradition of the authors for performing work notable for its thoroughness. The methodology of the study is well chosen; it takes the "machine accuracy" solution as the benchmark, analogous to the exact manufactured solution used here at Lisbon II in evaluations of uncertainty estimators.

The common approaches are hardly rational; the arbitrary and too lenient 10^{-3} reduction in initial residuals (widely used and set as default in some commercial codes) on the one extreme, and on the other, machine zero, equally arbitrary but too strict. [1] quantifies a more rational result, a usable rule of thumb that iteration error 2-3 orders smaller than discretization error is safe.

There are several results that will prove to be unpopular - another case of "inconvenient truth." Iteration error estimators based upon the last performed iteration systematically underestimate the iterative error. This is serious, since this practice is the most common type of iteration criterion used. The L_∞ norm is more dependable than L_1 or L_2, which will make analyses difficult. And iterative convergence rate depends on the turbulence model, therefore iterative error estimates are model dependent. All these results belie the casualness of the 10^{-3} reduction in residual criterion commonly recommended as a rule of thumb.

2. V&V Committee Progress: AIAA, ASME CSM, ASME CFD, ASCE

At least four current committees of professional societies are devoted to V&V issues. The AIAA has a standing committee on the AIAA Guide for V&V in CFD [2]. There is no publication forthcoming, but several of the members were recently successful in getting approval of an expanison of the earlier 1993 V&V publication policy [3] with a new one [4]. I had chaired the AIAA Fluid Dynamics Subcommittee that produced the first [3] but have been involved only as an outside commentator since. By contrast, I have first-hand experience on the other three committees.

The ASME Committee PTC-60 (recently re-designated V&V10) has completed its Guide for V&V in Computational Solid Mechanics which is now in the process of external reviews, with official publication expected in 2007 [5]. The project was inspired by the AIAA Guide on V&V in Fluid Dynamics [2] and like that publication, it is valuable for overview, definitions, philosophy, and motivation but it is not intended to provide detailed procedures. (One equation crept into the penultimate draft.) It is even broader and more management oriented than its AIAA predecessor. The committee intends to produce a procedural document in the future. I was unable to convince the committee to include distinctions between "model" in the strong and weak senses [6]. This lack causes some internal contradictions in a close reading.

The younger ASME Committee PTC-61 (recently re-designated V&V10) is working on its Guide for V&V in CFD [7], which might appear in 2007 but 2008 is more likely. Despite the similarity in title to [2] and [5], this document is targeted to engineering practice rather than management. It includes overview and definitions, but also is being constituted as a detailed guide to methodology.

A similar approach is being taken by the ASCE committee developing a monograph [8] on V&V for the more limited class of fluid dynamics problems of free surface problems. This project is more ambitious and noteworthy for including detailed case studies of both laboratory experiments and field studies. The field studies are both real and extensive, e.g. simulations of San Francisco Bay, Victoria Bend of the Mississippi River, and others. Clearly the document will be valuable if it is ever published. Unfortunately, the committee has hit a surprising impasse at a very late date. After ten years of work (my involvement has only covered eight years) the committee finds itself sharply divided on a few very fundamental issues. One of these is sure to surprise attendees at Lisbon II. Should Calculation Verification be performed before or after Calibration and Validation? (!) This audience can well imagine the tone of committee discussions.

3. Journal Policy Statements on V&V

The expanded AIAA V&V publication policy [4] has already been mentioned. Another has been published in *Clinical Biomechanics* [9]. Although this policy confuses some V&V terminology, it is a significant contribution to setting quality standards for CSM (Computational Solid Mechanics) in difficult problems. For example, Validation is complicated by the aging of animal tissues during the experiments.

4. An Invalid Approach to Verification and Validation

In November 2004 a Workshop on V&V was organized by the U.S. National Institute of Standards and Technology (NIST) at their offices. The Workshop was called [10, Ch. 1] to "advance the research on a framework of methodologies for the verification and validation (V&V) of computer models of complex engineering systems *with or without experimental data*." (Emphasis added.) Also ostensibly requested was input from V&V practitioners (31 of the 50 participants were from outside NIST) on the NIST approach, notably to "assess and improve the resulting metrology-based approach to V&V" [11]. I did not attend, but my contacts within the non-NIST participants were critical of the manner in which the Workshop was conducted and of MV&V. Apparently the Workshop was a complete contrast to Lisbon I [12] in every way, e.g. stifled discussion, lack of experience, naiveté over definitions, lack of understanding of past and current work in the V&V community. Of course, Validation "*with or without experimental data*" is an oxymoron, so credibility was lost in the first sentence.

The basic approach and flaw of metrology-based V&V is to treat the results of simulations like one treats statistical variations in a production manufacturing run. MV&V would replace the inconvenience of experiments with a kind of democratic approach called a "consensus mean" in which the results of various simulations are weighted and combined with statistics to arrive at the benchmark values. One might hope that at least a 4th order solution on 10^7 cells would be weighted more heavily than a 1st order solution on 10^2 cells, rather than a pure democracy of "one code, one vote." But the coarse grid, 1st order solution would be counted as part of the Validation replacement for experimental data. We in the V&V community must be clear and unequivocal:

$$\textit{No Experiments} \Rightarrow \textit{No Validation}$$

When it comes to systems, I am not a purist. I would be open to the possibility of considering a claim to Validation of a system based on Validation of components and their interactions. There will be doubts about unanticipated coupling, exceeding parameter ranges, etc. Many have made a good case for the more demanding claim that only full systems experiments should deserve the claim of "Validation." Likewise, I can defer to engineering judgment of experts in regard to interpolation and even extrapolation in the input parameter space, determining the limits of the "domain of Validation." And of course the criteria for acceptable level of validation may necessarily be compromised by one's inability to perform good experiments (e.g. limitations on weapons testing) or the impossibility of controlled experiments (e.g. true astrophysical experiments, climate modeling, etc.). Again, I have no problem with claiming "Validation" for weak agreements, provided that some quantitative correspondence exists (e.g. parameter trends) and that the coarse level of agreement has some engineering or scientific utility.

But the referent must be physical measurements, not simply an ensemble of un-validated computations.

The suggested MV&V approach also covers Calculation Verification, in the same democratic fashion. This approach is not just questionable; I believe that it can proven definitively to be wrong.

If MV&V for Calculation Verification works at all, it should work for the best-behaved computational cases:

(a) correct codes (i.e. no coding errors),
(b) simple well-behaved linear problems (no singularities, no advection terms, e.g. a Poisson problem with a smooth source term),
(c) regular mesh generation (e.g. uniform quadrilaterals or triangles),
(d) rigorous iterative convergence criteria or use of direct solvers with negligible round-off accumulation, and
(e) high enough resolution on all meshes used to provide monotonic mesh convergence behavior (very easy to achieve for the example Poisson problem).

For these best-behaved cases, consider the MV&V approach of processing the results of N1 hypothetical codes on N2 meshes. The hypothetical N1 codes all use the same continuum equations, including boundary conditions, so there exists an unambiguous correct mathematical answer. The hypothetical codes cover a range of orders of formal accuracy; to be specific, let us consider orders of convergence (theoretical and observed) p = 2,4,6,8. Consider meshes starting from coarse resolution (say 10 elements in each direction, giving accuracy of ~20%) but still fine enough (for this simple, well-behaved problem) to be within the asymptotic range so that mesh convergence is monotonic. At the high end, we consider a very large number, say 10^4 elements in each direction.

The MV&V method will always give a mean solution that is worse than the best solution (finest mesh and highest order method), because it weights the best with the worst solutions (coarsest mesh and lowest order method). Furthermore, the variance of the solutions will always give error bars (i.e. mean $\pm$ δ) that are inside the extremes of the results; yet the true answer is in fact always outside the set of all results. This is because the convergence in these best-behaved cases is monotonic, i.e. the convergence is one-sided, so the correct answer is approached asymptotically, and it will never be obtained by any interpolation or averaging of the individual computations. In fact, the best estimate is obtained by *extrapolation*, whereas the MV&V approach is always some kind of interpolation.

If we looked only at codes using 2nd order methods, and if the mesh resolution sampling were well covered, the MV&V estimate might loosely be expected to be close to the solution on a 10^3 mesh, whereas the true answer lies beyond that of the 10^4 mesh, and outside the error bars provided by the variance of the results.

If we looked at only one mesh, say the 10^2 mesh, the MV&V estimate might loosely be expected to be close to the solution obtained by the p = 2 or 4 method, whereas the true answer lies beyond the p = 8 solution. However, for this situation the variance might provide a conservative error band (probably excessively conservative, based on our experiences).

Not surprisingly, the example MV&V exercise in [11] showed no significant mesh convergence tests (except for case CPS4). The essential point to bear in mind is that

all the (correct) codes will give the correct answer asymptotically as the mesh is refined. So why not do it? And why give any weight at all to the CPS4 solutions in 1×2×10 and 1×4×20 meshes, when we have a 1×4×80 solution that must be better than the first two? (The information from the first two solutions can of course be well used to extrapolate to a better estimate and provide error estimates, but this is not considered in [11].)

Also, since all the (correct) codes will give the correct answer asymptotically as the mesh is refined, there is no need to run the case "in as many FEA codes as possible."

In summary, the MV&V has nothing to do with Validation (experimental agreement) and demonstrably cannot correctly do the Verification. The basic choice of a "consensus mean" is incorrect. To continue the analogy used in [11] with experimental results from different laboratories, the MV&V approach would be like statistically analyzing 100 laboratory results for the speed of light, one of which was provided by the Stanford Physics Labs, one by CERN, and the other 98 by high school physics projects.

5. Observed Grid Convergence Rate *p* Affected by Asymmetrical Grid Refinement

Eça and Hoekstra [13, and elsewhere] demonstrated thoroughly that lack of strict geometric similarity in the grid sequence is a major contributor to noisy values of observed rate of convergence p. It is obvious that geometric similarity requires the same grid refinement factor in each coordinate. A recent paper by Salas [14] investigates this effect in a systematic way, and discloses a widespread mistake associated with it. By citing error estimation results from workshops, he points out the prevalent current practice of using the power series form (suggested by the well-known Taylor series analysis for truncation error) for a 1-D problem

$$f_e \sim f_c + c\, h^p \qquad \text{(1.2 of [14])}$$

and using it in multidimensional problems. Users calculate a grid refinement factor r as the ratio of representative grid spacings defined as $h = (hx\ hy)^{1/2}$ or possibly others, such as the diagonal $h = (h_x^2 + h_y^2)^{1/2}$ .The form does not really matter, since the grid refinement ratio r scales out. However, as most of us recognize, that practice only makes sense if the same r applies in each coordinate. Otherwise, another coefficient is introduced for each coordinate, and it would require four grid solutions (rather than three) to determine observed p using the correct 2-D form,

$$f_e \sim f_c + a\, h_x^p + b\, h_y^p \qquad \text{(1.4 of [14])}$$

Salas claims there is widespread misunderstanding of this effect, and that current practices are flawed, particularly in external aerodynamics.

For precise language, let us refer to the observed p calculated using the 1-D form applied to a grid sequence with $r_x \neq r_y$ as observed pseudo-1-D p, denoted p_{1s}.

Using an exact solution for transonic flow (the Ringleb solution) Salas demonstrates that his theoretically second-order code exhibits observed $p = 2.2$ when $r_x = r_y$, but exhibits an observed pseudo-1-D $p_{1s} = 9.94$ when $r_x \neq r_y$. (See Eça and Hoekstra [13] for sometimes comparable anomalous results.)

Salas includes a simple synthetic problem to illustrate his point. Upon examination, this problem raises another point that, I think, should be further investigated. It bears on the question of how one might use observed pseudo-1-D $p_{1s} > p$ theoretical in uncertainty estimations.

As already stated, in order for the analysis for order of convergence p to apply in a multidimensional problem, the grid refinement ratio r must be the same in all directions, unless one solves for the coordinate coefficients separately. It is also true that *convergence* itself does not depend on this condition. At the risk of being too long-winded, I arrange computational solutions on all possible 2-D (I×J) grids, starting with a 2x2 cell grid, as follows.

2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9
3,2	*3,3*	**3,4**	3,5	3,6	3,7	3,8	3,9
4,2	4,3	*4,4*	4,5	**4,6**	4,7	4,8	4,9
5,2	5,3	5,4	*5,5*	5,6	5,7	**5,8**	5,9
6,2	6,3	6,4	6,5	*6,6*	6,7	6,8	6,9
7,2	7,3	7,4	7,5	7,6	*7,7*	7,8	7,9
8,2	8,3	8,4	8,5	8,6	8,7	*8,8*	8,9
9,2	9,3	9,4	9,5	9,6	9,7	9,8	*9,9* ...

The computational solutions f_g (I,J) would form a (discrete) single-valued solution surface above this discrete 2-D domain of definition. The exact solution f_e is approached down and to the right (but not just down, and not just to the right, which are only one-coordinate refinements). The grid doubling sequence is underlined; it is not necessary to follow this path. The *preferred* paths are anything along the diagonal (*italics*), for which r is not necessarily constant in the sequence, but it is the same in each direction. But if we took another path down and to the right, e.g. the **bold font** path, we would still be heading towards the exact solution f_e as (I,J) $\rightarrow (\infty,\infty)$.

The evaluation of an observed p enables us to extrapolate along the path, analogous to a directional derivative. (I suppose the analogy could be made precise by somehow generalizing the discrete surface to a continuum surface, but such precision seems unwarranted.) The extrapolated value can be used as a better estimate of the exact (converged) value f_e and so gives an error estimate. Any path along the diagonal corresponds to the same r in each direction. This diagonal direction does not uniquely determine the observed value of p because different values of r give different p. The grid doubling sequence gives a different value of the observed p than the unit sequence (2,2), (3,3), (4,4), (5,5), etc. {The latter would produce a better approximation of a secant evaluation of a directional derivative, if we were envisioning a continuum surface of solutions. But in actual calculations, it would contribute to noise in p because of incomplete iteration error and round-off error, especially as (I,J) $\rightarrow (\infty,\infty)$.}

The extrapolation slope is clearly path dependent, since the solution surface is not flat. Although any path along the diagonal corresponds to the same r in each direction, others are possible. If the observed p or the p_{1s} is not real, presumably it cannot be used, and we are far outside the asymptotic region. But if it is real, can it be used? If we are following the bold font path, we clearly do ***not*** want to use theoretical p, even if it is correct for the diagonal path. We want to use the slope appropriate for our path.

The question is: Will this extrapolation be accurate? Might it be that the somewhat arbitrary definition of h and therefore r, and the evaluation of a path-dependent p_{1s}, produces an accurate extrapolation when used with a consistent evaluation of the coefficient c of Eq. (1.2)?

It would be of interest to try it for the exact (Ringleb) solution of [14], but solution values are not presented (only L_1 errors in velocity). But when applied to the synthetic second-order accurate problem results given in Tables 1 and 2 of [14] it works very well indeed. (The following are results of Salas[14], not mine.)

The synthetic problem is devised with $f_e = 1$, $p = 2$, $a = 1$, $b = 5$ in Eq. (1.4). Using the same r in both directions ($r_x = r_y = 2.0$ for the first refinement, = 1.5 for the second) reproduces the "true" (diagonal path) value of $p = 2$, along with $c = 12.9$ and the exact solution value $f_e = 1$, as expected. If the same r is not used in each direction ($r_x = 1.6$ and $r_y = 2$ for both refinements), the three grid solutions produce $p_{1s} = 2.36$ and $c = 34.75$, very different from the diagonal path, as is to be expected; however, used consistently with Eq. (1.2) they produce $f_e = 0.999 \sim 1$. Coincidence?

The conditions for use of p_{1s} to estimate error remain to be determined. In a real problem with other contributors to noise, I would still be reluctant to base an uncertainty estimate on some observed $p = 9.94$, and the consensus at Lisbon I was to enforce an upper limit on p to avoid too-optimistic uncertainty calculations. My suggestion has been to limit $p \sim$ theoretical, but perhaps this is unnecessarily conservative.

6. Uncertainty Calculations for Large Eddy Simulation

Our fellow participant at Lisbon I, Prof. I. Celik, and his colleagues from Darmstadt have presented [15] several new approaches for calculation uncertainty (generally, assessment measures) for Large Eddy Simulations (LES). Along with other contributions, they notably make an effort to segregate numerical errors from modeling errors, which are often confused in LES and other simulations in which h is part of the model. Of special interest is an approach using 3 solutions on 2 grids. The "standard" (unperturbed) LES model is applied on two grids to estimate the purely numerical error using an assumed theoretical rate of numerical convergence. Then a modified (perturbed) LES model is used on the finer grid to estimate model error. The numerical and modeling errors are combined using absolute values to attain some conservatism, in the spirit of an uncertainty estimate. The paper is highly recommended for the presentation of the background as well as the new approaches.

7. Grid Convergence Index for Unstructured Grids

I still wait for an unstructured grid, structured refinement application of the GCI, as suggested in [6]. Perhaps one will be presented here at Lisbon II. I would expect larger Fs to be required, since true grid similarity will be compromised. Perhaps Fs = 1.5? = 3?

8. Singularity Treatment in Calculation Verification

In a paper on Finite Element Analysis (FEA) of Computational Solid Mechanics (CSM) stress problems, Sinclair et al. [16] presented a thorough work on the detection

and treatment of singularities during grid convergence studies. The work should be equally applicable to CFD, although it remains to be tested on CFD problems. The methods can automatically detect and distinguish between cases of power singularities, logarithmic singularities, or simply grids not yet in the asymptotic range. Of course, there is a gray area and account is taken of ambiguous results; the most difficult cases to distinguish are weak singularities (e.g. a terribly difficult behavior of Hertzian contact stress converging at ~ 0.1.)

Interestingly, the authors used the approach we all agreed upon for Lisbon II (also [17]), evaluating the performance of the singularity treatments by using realistic exact solutions (with prescribed singular behavior [17]) produced by MMS. However, their development was completely independent, and they refer to it as a Tuned Test Model (TTM) in an earlier paper with both CSM and FD examples [18]. (Like many of our CFD colleagues, the CSM audience often finds the MMS concept elusive [19].) The evaluation exercises are thorough, with 21 numerical experiments on 14 trial problems with power singularities, 21 on 5 problems with log singularities, and 103 on 18 problems with nonsingular stresses. Alternative methods from the previous literature are found to be lacking: the two-mesh check (i.e. no evaluation of observed p) and the "linearly-increasing mesh sequence."

{The latter uses a mesh sequence such as 100, 110, 120 total elements *irrespective of the dimensionality*, and convergence is judged (effectively extrapolating) by a log-log plot. However common it may be in FEA CSM, it has no basis in analysis. Nevertheless, I have seen it done in CFD by some notables in aerodynamics. The foolishness is easily demonstrated by a synthetic problem that diverges in 1-D, with computational solution $f = C\times(\text{I-1}) = C/\Delta x$. This divergence would be evident in a plot of f vs. I. To disguise this divergence, extend the dimensionality to 3-D with no variation in y or z, and plot f vs. N = (I-1)(J-1)(K-1). The result will appear to be asymptotically approaching a converged solution with linear convergence rate. Go to 4-D for quadratic convergence.}

Rather than targeting our 95% certainty as in the GCI approach [6,12,13], Sinclair et al. adopt a less ambitious approach to evaluation of accuracy and prediction, aiming to achieve and predict (in the error estimation) one of the four accuracy levels: excellent or < 1% error in stress prediction, good or <5%, satisfactory or <10%, and unsatisfactory or ≥ 10%. They evaluate the error estimate as an unambiguous success if the correct *level* is predicted, and acceptable if they miss by only one level, without worrying about being inside or outside error bars, which is the GCI approach. I am married to the GCI uncertainty approach, and it is becoming well established, but their approach is reasonable and in step with engineering practice "in the trenches."

9. Non-Uniqueness in Turbulence Modeling

One can imagine how confusing evaluation of the GCI would be if the computational solutions were non-unique, dependent on extraneous factors such as initial conditions and iterative algorithms and paths. In two very interesting papers, Rumsey et al. [20,21] have shown this indeed is the case for some popular low-Re K-ε models.

The non-uniqueness problem is exposed with real CFD problems and elegantly described by a nullcline analysis. The result is of course exceedingly important to the diminishing group of K-ε users, being applicable to the most commonly used low-Re K-ε

models by Jones-Launder and Launder-Sharma (although not to the Chen model). It ought to drive the final nail in the coffin of low-Re K-ε models, but it probably will not. The fundamental failings of K-ε in regard to the incorrect limit behavior in the defect layer, and the resulting inaccuracy for transitional flows, for adverse pressure gradients flows and for compressible flows, is well documented (e.g. see Wilcox book [22]). Yet, as one expert told me, the K-ε approach remains "surprisingly resilient."

The non-uniqueness is limited to some K-ε models, and the authors present an easy fix. Fortunately, the non-uniqueness phenomenon is not applicable to the more important and more accurate K-ω and Spalart-Allmaras models [22]. The K-ω simply does not allow pseudo-laminar regions. This is shown explicitly by the same authors [21] and for the "standard" K-ε, i.e. without the low-Re bludgeon of the sublayer.

It is unlikely that this non-uniqueness would have been discovered routinely during Code Verification by MMS. At most, the non-uniqueness could result in a false-negative Code Verification, i.e. the coding is correct, but the code converged to a different solution than the one built into the MMS. (I addressed this possibility in [6].) My guess is that in most cases, the source term would dominate and the Manufactured Solution would be attained with no non-uniqueness problem, demonstrating coding correctness but not revealing any non-uniqueness problems. Likewise for Calculation Verification; the error estimation from grid convergence tests would be polluted by the non-uniqueness but might escape attention. Even if it were not, the result obtained would be just the estimation of the discretization error, again not revealing anything about non-uniqueness. [6, p. 263]. Further, I doubt that usual parameter sensitivity analysis would have disclosed the problem. Many codes set default initial conditions from the free-stream boundary conditions, as done in [20,21]. If one is testing for solution sensitivity to boundary conditions, each run would have produced a plausible and apparently unique solution. Tests would not expose the problem unless one tested for the unexpected result of dependence on initial conditions independent of boundary conditions, which is not what people usually mean by parameter sensitivity testing. Note also that the non-unique results displayed in Rumsey et al. [20] (Fig. 3) result from a 4-order-of-magnitude variation in free stream turbulence ε (rate of energy dissipation). This is perhaps the reason that the non-uniqueness has not been experienced enough in practice to alert the CFD community earlier.

However, non-uniqueness is also disclosed by different mesh-sequencing in multigrid procedures, as shown in Fig. 6 of [20]. I find it incredible (literally) that none of the thousands of low-Re K-ε users have ever experienced this. Why has it not been reported before? However [20], "many CFD practitioners have noticed that the K-ε equations often fail to go fully turbulent." I find it easier to excuse this oversight, because it might seem that the non-uniqueness here was mimicking a physically plausible non-uniqueness (transition hysteresis) [6, p. 27]. They just did not pursue it far enough to discover, as Rumsey et al. did, that it is not like physical non-uniqueness; instead they pursued the heavy-handed fixes described in [20, p. 1588]. (Also, most studies fix boundary layer transition at the leading edge or at some small fixed percentage of chord distance from the leading edge, and start the K-ε calculations with turbulent flow, which disguises the problem.)

The analysis in Rumsey et al. also sheds light on the numerical fragility of K-ε. See the discussion on degenerate critical point [20, p. 1589]. Many practitioners have

experienced this, failing to achieve iterative convergence on some problems depending on initial conditions, even after resorting to severe damping. Again, the K-ω equations are more robust.

Even when the parameters are right and a stable turbulence solution is attained, the phase-plane trajectories (Fig. 9 of [20] and discussion following) are circuitous and intuitively (to me) non-physical. Again, this is a criticism specific to low-Re K-ε, not to other turbulence models or to computational physics in general. Compare the phase-plane trajectories in the second paper [21] of Figs. 4 and 5 for low-Re K-ε with the intuitively correct progression shown in Fig. 6 for K-ω.

10. New K-ω Turbulence Model

Wilcox has re-visited his Standard K-ω turbulence model and made some embellishments that are simple to incorporate but significant in improving accuracy. He has revisited a suite of 100 problems and claims significant improvement in accuracy, and that all 100 are of acceptable engineering accuracy. Notable improvements are in the sensitivity to free-stream values of ω and more accurate (reduced) spreading rate for free shear layers. The results will be presented at the January 2007 AIAA Annual Meeting in Reno [23] and in the third edition of his book to be released in early 2007 [22]. This book also expands his examples of Calculation Verification using the GCI.

11. Commercial Code Use and Comparisons

Hutton [24] cited estimates of the worldwide number of commercial CFD code users of 25-30,000, including 2/3 of the Fortune 500 companies, and licensing revenues growing at 15-20% annually. From a V&V viewpoint, this is intimidating. Industry colleagues tell me that the Code Verification exercises provided to users by vendors are not as convincing as we here at Lisbon II would like. With consolidation of commercial CFD vendors, the Code Verification quality is not likely to improve. All the more reason to do your own Calculation Verification using MMS (or to just become very trusting).

Hutton [24] also provided an example of a code comparison exercise on turbulence modeling. As is often the case, a wide range of results were obtained, even between groups using the same "code." As expected, the difficulty was one with which we are all familiar: "code" ≠ "model", and the problem specification was not complete because of unspecified radial velocity at inflow. Unlike our own exercises here and at Lisbon I, "very few" of the contributors performed grid convergence studies. He referred to one university group who "actually" refined the grid! This presented an excellent example of a K-ε calculation that agreed better with experiment for a coarse grid than a fine grid, not just with a single-valued solution functional (e.g. see [6]) but over a distribution. The systematic cancellation of discretization error and modeling error occurs because K-ε overproduces turbulent kinetic energy while under-resolution reduces it. This is yet another warning against rushing into Validation without first ascertaining grid convergence in Calculation Verification. With his experience, Hutton also affirmed the fact, widely acknowledged in the V&V community but always requiring reiteration, that old experimental data is generally inadequate for good Validation exercises. He presented

some examples of a Best Practices Guideline and a QA system for such code exercises, accessible at www.QNET-CFD-KB.com

12. Challenges to the Concept of Validation

The challenges to the very concept of Validation of Models addressed in Appendix C of [6] continue, e.g. [25]. All relate to the science philosophy of Karl Popper [26] and to a rarefied view of Validation that (a) has nothing to do with practical engineering or science and (b) is contradictory with widely accepted and pragmatic definitions [2-9,12,24]. It is interesting that Popper is quoted to support the impossibility of Validation of codes and computer models, but his 1935 views, whatever value they may have, are much more general and would apply, for example, to "models" due to Newton or Einstein. I am writing on a modest paper [27] to present a down-to-earth rebuttal, not in order to change anyone's mind, which would be highly unlikely, but to give a little support to anyone who needs to address and dismiss the issue (because of reviewers, funders, regulatory agencies, stakeholders, etc.) to be able to get on with their work.

Acknowledgements

I gratefully acknowledge discussions with L. Eça, M. Hoekstra, G. B. Sinclair, M. Salas, D. C. Wilcox, and C. L. Rumsey.

References

1. 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. 2006.
2. AIAA (1998), *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, AIAA G-077-1998.
3. AIAA (1993), Editorial Policy Statement, January 2006, all AIAA Journals.
4. AIAA (2006), Editorial Policy Statement, January 2006, all AIAA Journals.
5. ASME (2007), *Guide for the Verification and Validation in Computational Solid Mechanics*, ASME V&V10, to appear.
6. Roache, P. J. (1998), *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers, Albuquerque, August 1998.
7. ASME (2008), *Guide for the Verification and Validation of Computational Fluid Dynamics*, ASME V&V20, to appear.
8. ASCE/EWRI (2007), 3D Free Surface Flow Model Verification and Validation, to appear.
9. Burton, K., Viceconti, M., Olsen, B., Nolte, L.-P. (2005), "Extracting clinically relevant data from finite element simulations," Editorial, *Clinical Biomechanics*, Vol. 20, p. 451-454.
10. Fong, J. T. and de Wit, R., (Eds.), (2004), *Proc. Workshop on Verification and Validation of Computer Models for Design and Performance Evaluation of High-Consequence Engineering Systems*, U.S. National Institute of Standards and Technology, Gaithersburg, MD, 8-9 Nov. 2004.

11. Fong, J. T., Filliben, J. J., Fields, R. J., deWit, R., and Bernstein, B. (2004), "A Reference Benchmark Approach to V&V of Computer Models of High-Consequence Engineering Systems," Chapter 4, Preprints of 41 Invited Papers of [10].
12. Eça, L., Hoekstra, M., and Roache, P. J. (2005), "Verification of Calculations: an Overview of the Lisbon Workshop," AIAA Paper No. 4728, AIAA Computational Fluid Dynamics Conference, Toronto, June 2005.
13. Eça , L. and Hoekstra, M., "An Evaluation of Verification Procedures for CFD Algorithms," 24th Symposium on Naval Hydrodynamics, Fukuoka, Japan, 8-13 July 2002.
14. Salas, M. D. (2006), "Some Observations on Grid Convergence," *Computers and Fluids*, Jan. 2006.
15. Celik, I, Klein, M. and Janicka, J. (2006), "Assessment Measures for LES Applications," FEDSM2006-98015, Proc. of FEDSM2006, ASME Joint U.S. European Fluids Engineering Summer Meeting, 17-20 July 2006, Miami, Florida, U.S.A.
16. Sinclair, G. B., Beisheim, J. R., and Sezer, S. (2006), "Practical Convergence-Divergence Checks for Stresses from FEA," Proc. 2006 International ANSYS Users Conference and Exposition, 2-4 May 2006, Pittsburgh, PA, U.S.A.
17. Roache, P. J. (2002), "Code Verification by the Method of Manufactured Solutions", *ASME Journal of Fluids Engineering*, Vol. 114, No. 1, March 2002, pp. 4-10.
18. Sinclair, G. B., Anaya-Dufresne, M., Meda, G, and Okajima, M. (1997), "Tuned Test Problems for Numerical Methods in Engineering," *International Journal for Numerical Methods in Engineering*, Vol. 40, pp. 4183-4209.
19. Sinclair, G. B. (2006), pers. comm.
20. Rumsey, C. L., Pettersson-Reif, B. A., and Gatski, T. B. (2006), "Arbitrary Steady-State Solutions with the K-ε Model," AIAA Journal, Vol. 44, No. 7, July 2006, pp. 1586-1592.
21. Pettersson-Reif, B. A., Gatski, T. B., and Rumsey, C. L. (2006), "On the behavior of two-equation models in homogeneous turbulence," Physics of Fluids, Vol. 18, 065109.
22. Wilcox, D. C. (2007), *Turbulence Modeling for CFD*, Third Edition, DCW Industries, La Canada, CA, to appear.
23. Wilcox, D. C. (2007), "Formulation of the K-ω Turbulence Model Revisited," AIAA Paper 2007-1408, Jan. 2007, to appear.
24. Hutton, A. G. (2006), "Quality and Reliability in CFD," Plenary Presentation, FEDSM2006, ASME Joint U.S. European Fluids Engineering Summer Meeting, 17-20 July 2006, Miami, Florida, U.S.A.
25. Hazelrigg, G. A. (2003), "Thoughts on Model Validation for Engineering Design," DETC2003/DTM-48632, Proc. ASME 2003 Design Engineering Technical Conferences and Computers and Information in Engineering Conferences, Chicago, IL, U.S.A., 2-6 September 2003.
26. Popper, Karl (1992), *The Logic of Scientific Discovery*, Routledge Classics, New York. Translation of *Logic der Forschung* (1935), Verlag von Julius Springer, Vienna.
27. Roache, P. J. (2007), "A Defense of Computational Physics," to appear.