PSFC/JA-03-20

Beyond Benchmarking — How Experiments and Simulations Can Work Together in Plasma Physics

Martin Greenwald

September 12, 2003

Plasma Science and Fusion Center
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

This work was supported by the U.S. Department of Energy, Cooperative Grant No. DE-FC02-99ER54512. Reproduction, translation, publication, use and disposal, in whole or in part, by or for the United States government is permitted.

Submitted for publication to *Computer Physics Communications*.

Beyond Benchmarking - How Experiments and Simulations Can Work Together in Plasma Physics

Martin Greenwald
MIT – Plasma Science & Fusion Center

Abstract

There has been dramatic progress in the scope and power of plasma simulations in recent years; and because codes are generally cheaper to write, to run and to diagnose than experiments, they have a well-recognized potential to extend our understanding of complex phenomena like plasma turbulence. However, simulations are imperfect models for physical reality and can be trusted only so far as they demonstrate agreement, without bias, with experimental results. This "validation" process tests the correctness and completeness of the physical model along with the assumptions and simplifications required for solution. At the same time, it must be understood that experimental measurements are almost always incomplete and subject to significant uncertainties and errors. For optimum scientific progress, simulations and experiments must be seen as complementary not competitive approaches. We need experiments dedicated to answering critical questions raised by the simulations and which examine the validity of models and which explicitly test their assumptions. A premium should be placed on ongoing collaborations which are open and candid about the sources of error and the strengths and weaknesses of each approach. Ultimately both experiments and simulation have much to gain by adopting a an approach of co-development, where simulations are continuously and carefully compared to experimental data and where experiments are guided by the results of simulations.

1. Introduction

Over the past several decades, computation has joined analytic theory and experimentation as a powerful tool in the advancement of plasma science. The interaction and interrelationship between these approaches has had a profound impact on the quality of research and the rate of its progress. This paper discusses the benefits of a close and ongoing collaboration between numerical simulation and experiments and outlines principles and practicalities that are involved. The critical question we must address is, with a full analytic theory intractable, with numerical simulations complex and constrained by available hardware and with experimental measurements incomplete and imperfect, how do we work together to make the maximum progress?

The ability to predict is related intimately to the scientific process. Predicting the results of experiments that have not yet been carried out is viewed as the strongest demonstration that some significant level of understanding of a physical phenomena has been achieved. Prediction has practical importance as well, since it allows science to be applied to practical problems – for example the design of an aircraft, the forecast of severe weather or the construction of a fusion power plant. Because of the wide range in temporal and spatial scales, strong isotropy, complex geometry and essential non-linearities, plasma physics presents a particularly difficult problem for both computational and standard theoretical approaches. Researchers must choose between two less than ideal alternatives, to obtain "exact solutions to approximate equations" or "approximate solutions to exact equations". Experiments are the arena where these non-ideal results are ultimately tested. Experiments have other important roles of course. Discovery of new and unexpected phenomena through extension of a parameter range or application of new diagnostic tools has always been a goal of experimental or observational science. For those areas with practical applications, extending the performance or efficiency of devices or processes can be a significant goal as well. For fusion systems this has concrete implications, "achieving ignition with pencil or paper or in a computer doesn't count [1]". None the less, a significant fraction of experimental effort needs to be devoted to detailed comparisons with theory and simulation. To cope with these difficult issues, a rigorous approach to simulation/experiment comparisons needs to be adopted. The aim is to improve models through qualitative and quantitative tests and to determine the regimes and parameters over which they can be useful.

Theory and computation on the one hand and experimentation on the other should be seen not as competitive activities but as complementary ones. Simulations, though dealing with imperfect models or solutions, can have near perfect diagnostics and are often cheaper and faster then corresponding experiments. They offer a high degree of flexibility; numerical experiments can be performed by turning particular terms on or off, isolating particular physical effects, or by varying input parameters, boundary or initial conditions. On the other hand, while experiments are performed on a "perfect" model, measurements of that reality are highly incomplete and imperfect. Parameters can be varied only over a limited range and particular physical effects are difficult to isolate. The interaction of computation and experiments should not be characterized as simply one of benchmarking a particular code or calculation, but rather should permeate the scientific process and include the mutual identification of interesting or important phenomena, testing of basic physical models and validation of codes and calculations. The process might best be described as "co-development" where each approach contributes its own strengths and recognizes its own limitations. Experimentalists gain by participating more fully in the development of the science. The advantage to theory is equally clear, experiments represent their only means of contact with the physical world. It is worth noting that the concept of theory and experiments as distinguishable but mutually dependent activities has a long and important place in the history in science. These can be traced through two schools of philosophy, rationalism, the logical development of a model based on indisputable axioms and empiricism which requires that every axiom, deduction, assumption or outcome be empirically confirmed [2].

How does one make meaningful comparisons when measurements and calculations are both incomplete and limited? Falsification is not necessarily a useful or easily applied concept as some degree of non-agreement is to be expected. Rather we should ask, "what degree of non-agreement falsifies a theory/simulation?" For engineering problems it is often sufficient to take an approach of continuous improvement as a practical measure for products or processes. Science however holds out for a deeper level of understanding. It is important to remember that when comparing with simulations it is not simply a matter of blindly trusting experiments. Typically systems that are harder and less reliable to model are harder to measure as well and it is certainly naïve to accept a single experiment as the final word. The complementary strength and weakness of the two approaches can be illustrated by a pair of quotes taken from the aerodynamics community.

“The greatest disaster one can encounter in computation is not instability or lack of convergence but results that are simultaneously good enough to be believable but bad enough to cause trouble [3].

“No one believes the CFD [simulation] results except the one who performed the calculation, and everyone believes the experimental results except the one who performed the experiment.”

There are several ways in which simulations and experiments can aid one another. Experiments can identify important problems and motivate development of particular physical models. They can confirm success through quantitative comparison or by searching for phenomena predicted but not yet observed. They can help elucidate numerical problems; apparent non-convergence or non-uniqueness may actually reflect real system intermittency. Experiments can be used to ‘calibrate’ models by providing parameters that cannot yet be calculated. Finally, they can be run in progression, accessing increasingly complex physics and geometry to guide codes in their development. Codes can help identify critical physics and critical measurements for a given experiment. At times they may be able to identify measurable quantities which can stand in for more fundamental but unmeasurable ones. They can be used to interpret experimental observations. They can explain longstanding observations which are not consistent with simpler theories. They can be used to define experiments, important measurements, critical regimes or parameter ranges and needed levels of accuracy for validation.

2. Case Studies

Numerous case studies exist which illustrate the synergy between experimental and computational/theoretical research. Consider the course of research into the role of ITG (Ion Thermal Gradient) turbulence and transport in tokamak plasmas. Earlier work on the Alcator experiments showed a clear decrease in transport as the plasma density was raised [4]. However, subsequent studies on Alcator-C found that this effect saturated at a relatively low density. During the same period, theoretical and computation studies suggested that important instabilities were excited when η_i, the ratio of density profile scale length, to temperature profile scale length exceeded a critical value on the order of 1 [5]. It was predicted that plasmas with steeper density profiles would be immune to this instability and thus might have lower levels of transport. Experiments to test this prediction were carried out using injection of high-speed deuterium pellets to fuel the plasma core and peak the density. The result was a dramatic drop in energy and particle transport consistent with predictions[6]. These experiments, among other results, spurred interest and activity in a class of instabilities and turbulence which are now believed to be the principle cause of anomalous transport in tokamaks. Experimental observation of transport “barriers” in the core and edge [7-10] motivated theoretical research into stabilization mechanisms. Out of this work, a new paradigm arose in which sheared plasma flows were seen as the principle agent of ITG turbulence regulation and suppression [11]. While far from complete, this theory is now the “standard model” for anomalous ion transport.

In a second example, consider the role of turbulent cross field transport in the physics and modeling of the plasma edge. In recent years elaborate models of edge plasmas have been written and tested[12,13]. These codes do a reasonable job of modeling parallel physics, neutrals and radiation in complex geometries but tend to use only a simple parameterization for perpendicular transport. On the other hand, extensive experimental studies of edge plasmas had documented high levels of fluctuations as a nearly universal result [14]. Finally, and over the same period, sophisticated non-linear turbulence codes were being developed which could treat the special conditions that exist in the edge[15-18]. These three elements proceeded almost independently until very recently and it was only when the threads were pulled together that a coherent picture began to emerge. It now appears that cross-field transport can play a dominant role in the dynamics of divertor and edge physics and that phenomena as disparate as the distribution of neutrals, impurity sources, the L/H transition, and the density limit may be manifestations of a common turbulence mechanism [19,20].

As a final example, in a field with close analogs to plasmas, we consider the Atmospheric Radiation Measurement program (ARM)[21]. Almost uniquely in science, this experiment was devised in direct response to the needs of computer simulations. During the 1970’s and 80’s a number of powerful climate modeling codes were developed and tested. Disconcertingly, the predictions of these codes varied dramatically. After years of close study and comparison between models, it was found that the principle uncertainty and the main difference between the models was the treatment of clouds in atmospheric radiation transfer [22]. To remedy this situation, arrays of sophisticated diagnostics were deployed at three sites (Oklahoma, northern Alaska, and the southwest Pacific) deemed typical for different meteorological regimes. The goal of the program was to learn as much as possible about the interaction between clouds and radiative heat flux and to embody that knowledge in the simulation codes.

These examples all illustrate the need to go beyond mere benchmarking in experiment/simulation interactions. Progress has been accelerated by close and ongoing interactions and slowed in the absence of that interaction. Even in their present incomplete and imperfect state, simulations have been extremely useful in focusing and interpreting experimental results. At the same time, experimental data, even when sparse compared to the output of simulations, can guide code development along the most productive channels.

3. Code Testing and Validation

The growing importance of simulations and recognition of their limitations and complexity led the computational fluid dynamics (CFD) community to a strenuous debate over the role that codes play in their vast and varied field [23]. Simulations are used to model flows in aerospace ship, turbo-machinery and automobile design; for wind loads and in architecture and bridge design; for weather forecasting; for cooling systems in buildings, nuclear reactors and on circuit boards; and for analysis of dispersal of pollutants in air, water and land. In these applications, accuracy and reliability of codes can have important economic ramifications and consequences for human safety. The close relation between the physics of fluids and plasmas suggest that our own field has much to learn from their experiences.

Early (and unwarranted) optimism in the mid-1970’s led to predictions that wind tunnels would soon be superfluous [24]. However, many practitioners understood the difficulties that lay ahead and produced cogent rebuttals [25]. A large number of codes were tested in the Stanford “turbulence Olympics” in 1981 with a result summarized by “the numerical quality of the solutions was so poor that it was impossible to draw meaningful conclusions about the relative merits of various turbulence models" [26]. By 1988, despite advances in computing power and numerical techniques beyond what had been previously assumed, talk of replacing wind tunnel experiments had vanished [23]. Even now, direct numerical simulation for high Reynolds number flows, with real boundary conditions complex geometry is far in the future. Instead, practical applications still rely on replacing all (Reynolds averaged Navier-Stokes) or some (Large eddy simulations) details of the flow dynamics with simplified transport expressions in their governing equations.

Assessment of predictive models has been divided into two distinct activities, verification and validation, with formal definitions and recommended practice for their application. Verification and validation are essentially confidence building activities, aimed at improving the quality of predictions from simulations. The first definition of these terms came from Schlesinger [27].

"Model verification: substantiation that a computerized model represents a conceptual model within specified limits of accuracy."

"Model validation: substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model."

It is important to note the highly conditional nature of these definitions. Codes are validated in the context of a particular problem or set of nearby problems, for a particular set of parameters, in a particular range and to a particular level of accuracy. Formally a code is not validated, but rather a particular calculation. There is no unambiguous way to define 'nearby', transitions or boundaries between regimes may be crucial and confounding. The emphasis on accuracy implies quantitative measures and attention to errors and uncertainties.

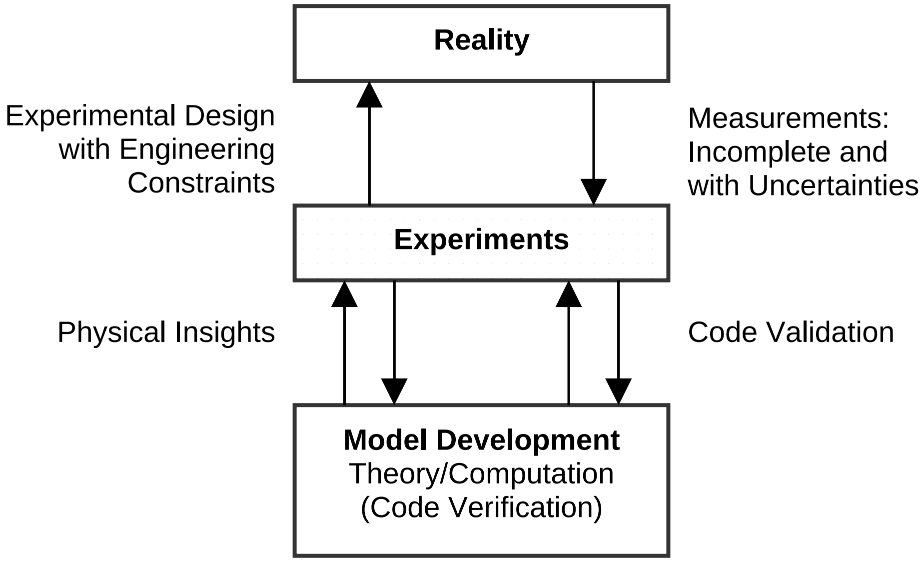

Figure 1. The interrelationship between simulations, experiments and physical reality are illustrated along with the processes that connect them.

Verification assesses the degree to which a code correctly implements the chosen physical model and is essentially a mathematical problem. Sources of error include algorithms, numerics, spatial or temporal gridding, coding errors, language or compiler bugs, convergence difficulties and so forth. It will not be the subject of further discussion in this paper; readers are referred to several excellent reviews [23,28,29]. Validation assesses the degree to which a code describes the real world. It is a physical problem and one without a clearly defined endpoint. The relation between the various processes can be illustrated in figure 1. Verification should, in principle, precede validation. Comparison between experiments and incompletely converged or otherwise inaccurate solutions are at best useless and at worst misleading. Agreement with experiments might be fortuitous, causing researchers to neglect critical tests. At this point, it is worth discussing "calibration", a related process involving the adjustment of modeling parameters in a computational model in order to improve agreement with data. Calibration may be justified to account for physics beyond the scope of the model, however it should not be used to obscure essential errors in a model or its implementation. Calibration may add or detract from the predictive ability of a code depending on how it has been applied.

4. Code Validation: Practical Issues

Meaningful comparisons between experiments and simulation require some hard thinking about what constitutes rigorous tests of a model in a particular area of physics. Differences between measurements and code results can arise through several sources.

- Model formulation errors – missing or incorrect physics
- Numerical solution errors due to discretization, boundary conditions or implementation
- Measurement errors and scarcity

The goal of validation is to test for errors in the physical model for verified codes, taking into account limitations of experimental observations. Sources of error in measurements can include conceptual errors with measurement technique; differences arising from temporal or spatial averaging; resolution issues relating to aperiodic data requiring long sample times; statistical or counting errors; calibration errors; electronic noise and data acquisition errors and data reduction errors. Some errors reduce the absolute accuracy of a measurement without affecting its relative precision or repeatability. Thus it is often more meaningful to compare trends than absolute values. Researchers must be alert however, to cases where this approach fails [23]. Difficulties can arise because of "parameter transition boundaries" where behavior changes rapidly with small changes in parameters.

Ideally, validation is carried out throughout the code development process. Comparisons should proceed in phases proceeding from the simplest systems to the most complex (see figure 2) [23, 28].

1. Unit problems
2. Subsystem cases
3. Complete systems

Each level represents a different degree of physics coupling and geometric complexity. The early tests attempt to isolate basic physical phenomena in the simplest geometry. Dedicated experiments, with specialized diagnostics are required at this stage and straightforward comparisons with analytic solutions may be possible. As experience and confidence is gained, researchers move to more demanding, more "realistic" cases. Note that as this hierarchy is traversed, the number of code runs and experiments tend to decrease. The quality and quantity (especially spatial coverage) tend to decrease as well while experimental errors and uncertainties increase. Information on boundary and initial conditions decrease as well. These trends suggest that at least as much attention should be paid to the lower levels of the hierarchy as to the top.

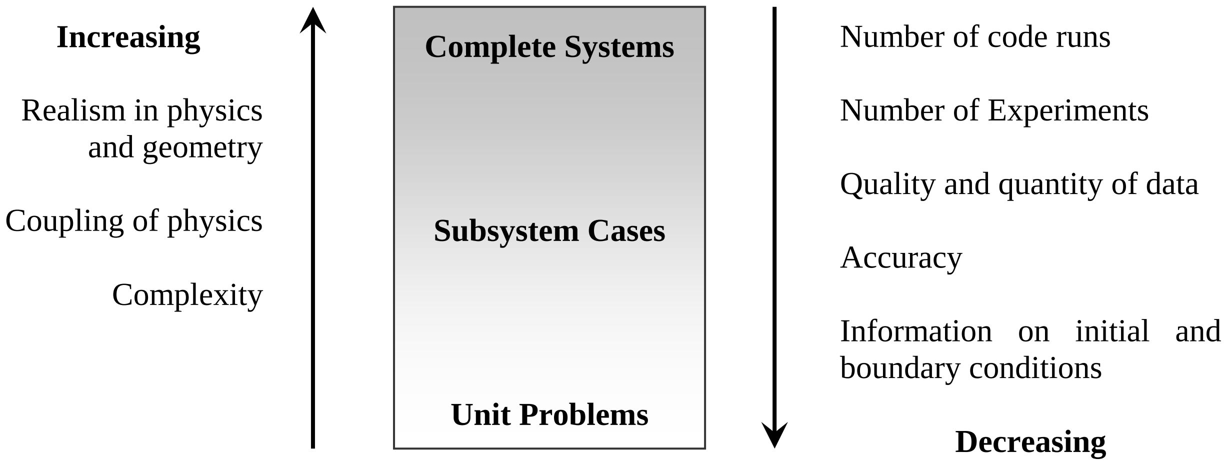

Figure 2. The hierarchy of validation begins with experiments designed to isolate particular physical phenomena and proceeds to more complex and realistic cases. As the hierarchy is traversed, the quantity and quality of data decreases, making comparisons less definitive.

While much can be learned from mining databases of previous experiments, the most useful comparisons are carried out in experiments designed and dedicated to the purpose. Older data is often not well enough documented or characterized and in any event direct interaction with experimentalists is essential to the process. Principles for the design and execution of these experiments have been thoroughly discussed [23,28-31]. These would include:

1. Verify codes first.

2. Plan a hierarchy of experiments beginning with the simplest physics and geometry.

3. Design experiments jointly by experimentalists and computationalists.

4. Experiments should test crucial features of the model, especially its assumptions or important simplifications. Perturbing effects should be minimized. Geometry, boundary and initial conditions must be well characterized and documented. Critical measurements

should be defined and limitations, uncertainties, and sources of error discussed with openness and candor.

5. Document code predictions in advance.

6. While jointly designed, carry out experiments and code runs independently.

7. Make as complete measurements as possible when carrying out experiments. Multiple diagnostics to measure the same quantities are desirable. Statistically sufficient data sets should be collected, repeating runs as required. It can be valuable to conduct experiments at more than one facility if this is practical.

8. Pay special attention to analysis of errors and uncertainties. Use modern statistical techniques to design experiments and to identify random and bias errors.

9. When analyzing results, don't paper over differences. The goal is not to prove that a code is correct, but to assess it's reliability and point the way towards improvement.

10. Document process and results including data reduction techniques and error analysis.

5. Summary and Discussion

Progress on most problems of interest in plasma physics involve both simulation and experiments and require more closely coordinated and collaborative work than is typical today. Comparisons need to be carried out over a hierarchy of experiments, ranging from the simplest, testing isolated aspects of physics in simple geometries, to full scale systems with the full range of physics integration and complex geometry. It is worth noting that while the latter attracts the most attention and often the most funding, the difficulties of integrated simulation and the difficulty of making measurements for the most complex systems suggest that important opportunities are being lost by neglect of the former.

In general, experiments do not measure quantities of interest directly. Typically there is a significant amount of physics involved in interpretation and analysis. In many cases, it is preferable to make comparisons through "synthetic" diagnostics – that is by post-processing simulation data in a manner which is as analogous as possible to the physical diagnostic. Such comparisons can be more direct and allow simpler and more quantitative assessment of differences. The underlying dynamics of non-linear systems may be revealed and compared by applying advanced techniques for analysis of time series or imaging data to both sets of data. These approaches can be facilitated through development of infrastructure and tools that aid in data sharing, visualization and analysis. Common data structures or common interfaces will be especially useful. Assembling databases dedicated for comparisons between simulations and experiments can be a useful activity. These data bases should have the following characteristics:

- Dynamic and interactive – able to be updated, annotated, appended
- Include metadata (data about the data) for every data item. This would document, for example, where the data came from, when it was written, who was responsible for it as well

as basic information on the data type, size, structure and so forth creating a complete, coherent self-descriptive structure
- Include both experimental and modeling data
- Contain all auxiliary data, assumptions, geometry, boundary and initial conditions
- Contain estimations of error
- Regimes well defined
- Able to be queried - searchable by content or by address
- Able to be browsed
- Linked to publications

Dedicated workshops, which bring researchers doing simulations and experiments together, can be especially useful if they focus on well defined and relatively narrow subjects. These are places to discuss the nuts and bolts of comparisons in addition to recent results and plans. The can provide a forum for discussions of analysis techniques, data sharing protocols and statistical methodology. Such workshops foster collaboration and aid in planning of coordinated activities.

References

[1] W. Nevins in presentation to US-Transport Task Force 2001

[2] U. Mehta, AIAA Journal of Propulsion and Power **12**, 1996, p. 940-948.

[3] J.H. Ferziger, Proceedings of ASME Fluids Engineering Division Summer Meeting, Washington DC, 1993, ASME FED-Vol. 158, p. 18.

[4] R. R. Parker, M. Greenwald, S. C. Luckhardt, E. S. Marmar, M. Porkolab, S. M. Wolfe, Nucl. Fusion **25**, 1985 p. 1127.

[5] T. Antonsen, B. Coppi, R. Englade, Nucl. Fusion **19**, 1979, p. 641.

[6] M. Greenwald, D. Gwinn, S. Milora, et al., Phys. Rev. Lett. **53** 4, 1984, p. 352.

[7] K.H. Burrell, M.E. Austin, T.N. Carlstrom, et al., Plasma Physics and Controlled Fusion, Vol. 1, 1995, p. 221.

[8] H. Kimura and JT-60 Team, Phys. Plasmas **3**, 1996, p. 1943.

[9] E.J. Synakowski, S. H. Batha, M.A. Beer, et al., Phys. Plasmas **4**, 1997, p. 1736.

[10] C.L. Rettig, K.H. Burrell, R.J. LaHaye, T.H. Osborne, W.A. Peebles, and R.J. Groebner, Plasma Physics and Controlled Fusion **36**, 1994, p. A207.

[11] H. Biglari, P.H. Diamond, P.W. Terry, Phys. Fluids B **2**, 1990, p. 1.

[12] G.D. Porter, S.L. Allen, M. Brown, et al., Phys. Plasmas **3**, 1996, p. 1967.

[13] R. Schneider, R. Reiter, D. Coster, et al., J. Nucl. Mat. **196-198**, 1992, p. 810.

[14] Wootton, A.J. J. of Nucl. Mat. **176-177**, 1990, p. 77-88

[15] B.N. Rogers and J.F. Drake, Phys. Rev. Letters **79** (1997) p.229-232

[16] B.D. Scott, Plasma Physics & Controlled Fusion **39** (1997) p.471-504.

[17] X.Q. Xu, et al., Phys. Plasmas **7**, 2000, p 1951.

[18] K. Hallatschek, Phys. Rev. Lett. **84**, 2000, p. 5145.

[19] B. LaBombard, et al., Phys. Plasmas **8**, 2001. p. 2107-2117.

[20] W. Suttrop, et al., J. Nucl. Mat. **266-269**, 1999, p. 118-123.

[21] US Department of Energy (DOE), *Atmospheric Radiation Measurement Program Plan,* Washington, D.C. DOE/ER-0442, 1990

[22] R.D. Cess, et al., J. Geophys. Res. 95, 1990, p. 16601

[23] P.J. Roache, "Verification and Validation in Computational Science and Engineering", Hermosa Publishers, Albuquerque, 1998.

[24] D.R. Chapman, H. Mark, M.W. Pirtle, Aerospace America, April 1975, p. 22-35.

[25] P. Bradshaw, Letters Astronautics and Aeronautics, September 1975, p. 6.

[26] S.J. Kline, B.J. Cantwell, G.M. Lilley, Eds., *"Objectives, Evaluation of Data, Specifications of Test Cases, Discussion, and Position Papers"*, Proceedings of the 1980-81 AFOSR/HTTM-Standford Conference on Complex Turbulent Flows, Vol. 1, ThermoSciences Division, Mechanical Engineering Dept. Stanford, University, 1981.

[27] S. Schlesinger, Simulation **32,** pg. 103, 1979.

[28] W. Oberkampf in "Verification and Validation of Computational Fluid Dynamics", F. Grasso, J. Perlaux, H. Deconinck editors, von Karman Institute for Fluid Dynamics publishers, Belgium, 2000.

[29] P.J. Roache, "Fundamentals of Computational Fluid Dynamics", Hermosa Publishers, Albuquerque, 1998.

[30] D.P. Aeschliman and W.L. Oberkamf, AIAA Journal **36**, 1998, p. 733.

[31] A. Rizzi, J. Vos, AIAA Journal **36**, 1998, p. 668.