

11

Design and execution of validation experiments

Chapter 10, Model validation fundamentals, discussed the philosophy of validation experiments and how they differ from traditional experiments as well as calibration experiments. A validation experiment is conducted for the primary purpose of determining the predictive accuracy of a mathematical model. In other words, a validation experiment is designed, executed, and analyzed for the purpose of quantitatively determining the ability of a mathematical model and its embodiment in a computer code to simulate a well-characterized physical process. In this chapter we describe six primary guidelines for the design and execution of validation experiments. Our discussion will also deal with how these guidelines can be carried out and why they are sometimes difficult to execute in practice.

We will then discuss a high-quality validation experiment that the authors, as well as several others, designed and executed. The experiment is referred to as the Joint Computational/Experimental Aerodynamics Program (JCEAP). From the beginning of the project, it was designed to synergistically couple computational fluid dynamics development and an experimental research in a hypersonic wind tunnel. The program was initiated in 1990 at Sandia National Laboratories and came to a successful close in 1997. The program helped develop the six fundamental guidelines for the design and execution of high-quality validation experiments. We use JCEAP to exemplify these six guidelines and make recommendations for how they can be applied to validation experiments in general.

11.1 Guidelines for validation experiments

These guidelines were developed over a period of several years and were originally reported in Aeschliman and Oberkampf (1998); Oberkampf and Blottner (1998); and Oberkampf and Trucano (2002). Although JCEAP greatly assisted in refining these ideas, most of these ideas had been presented individually by computational and experimental researchers in the past. Some of the key contributors to developing these ideas were Cosner (1995); Marvin (1995); Porter (1996); Barber (1998); Benek *et al.* (1998); and Roache (1998).

In each of the guidelines we refer to a *validation experiment*. We want to stress that this term is simply for convenience. A high quality validation experiment is always a *suite* or *ensemble* of many experiments, sometimes numbering into the hundreds, that all share a common goal: critical evaluation of the predictive accuracy of a model.

11.1.1 Joint effort between analysts and experimentalists

Guideline 1: A validation experiment should be jointly designed by experimentalists, model developers, code developers, and code users working closely together throughout the program, from inception to documentation, with complete candor about the strengths and weaknesses of each approach.

By *model developers* we mean applied researchers who build the mathematical models that are incorporated into computer codes. By *code developers* we mean those who write the computer software or implement the mathematical models into the software. By *code users* we mean those who use the code for analysis of engineering systems or processes. For our purposes here, we will group these various types of people into the term *analysts*.

To some, Guideline 1 may sound relatively easy to accomplish. Our experience has shown, however, that it is extraordinarily difficult to accomplish in practice. There are several reasons why this is true: some practical and some human nature. The first requirement to form a team of experimentalists and analysts for the design and execution of a validation experiment is that funding is available for both at the same time. Experimentalists and analysts have traditionally sought funding from separate sources. If these two groups wanted to design and conduct a validation experiment, they would each go their own way to seek funding. It would be rare for both to obtain support for their work during the same time frame, if at all. Contributing to this is the long held tradition, both in industry and in government, that one source is not responsible for funding both experiments and computations. It is certainly understandable why these traditions are in place, but if they continue, high quality validation experiments will remain rare.

If funding does become available for both experimentalists and analysts at the same time, then one must deal with the technical synchronization of both activities. The experimental facility and diagnostic instrumentation needed for the validation experiment must be available and ready for use; not in the construction or development phase. Similarly, the code and the options needed for validation must be operational and have been adequately tested through code verification activities. Some analysts will question this last requirement. They feel that code debugging and testing can be conducted in parallel with validation. This is a gross misjudgment that was discussed in Chapter 2, Fundamental concepts and terminology, and Chapter 10, Model validation fundamentals. Having both the experimental and computational capabilities fully prepared and the needed personnel available at the same time requires careful planning.

Another practical reason why it is very difficult to form a unified team of experimentalists and analysts is that team members are commonly from different organizations. For example, suppose a group of analysts is from a company that does not have the needed experimental facilities. If the analysts' company obtains funding for validation activities related to their models and codes, then the analysts' company would seek bids for the work and subcontract with a company that could conduct the experiments. Although technical qualifications of potential bidders for the experimental work are important, there are two business issues that come into play. First, when the analysts' company is seeking a facility in which to conduct

the experiment, this is clearly a competitive situation for potential bidders. The potential bidders will be reluctant to expose any weakness, limitations, or deficiencies within their facility. Not only would they risk losing the potential contract, they would also fear that the information could be used against them in the future by their competitors. Second, the analysts' company, once the contract is placed, is in control of funding and the direction of the project. From the perspective of the experimentalists' company, they are the supplier and the analysts' company is the customer. As a result, the experimentalists are the subordinate team members.

Suppose the experimentalists and the analysts are from two different organizations in the same corporation, government organization, or university. As is very commonly the case, the experimentalists and the analysts may be familiar with each other, but they have not developed a close working relationship. In fact, their history may be that they have been competitors for funding or recognition of achievement in the past. For example, the "computers versus wind tunnels" mentality (Chapman *et al.*, 1975) that has existed in the past in some areas of the computational fluid dynamics (CFD) community is a perfect example of competition, not collaboration. Also, because of the nature of their technical training and professional work, there is typically a gulf separating computationalists' and experimentalists' technical interests and possibly backgrounds. These factors, both obvious and insidious, will deter the candor and openness needed to forthrightly discuss the strengths and weaknesses in the computational and experimental approaches, during both the design and the execution of the validation experiment. These can be overcome in time, but it requires patience, respect, and understanding. In the fast-paced modern, competitive environment, these can be in short supply.

Some changes are obvious concerning how to overcome these substantive difficulties, but some are not. Funding organizations need to recognize the importance of funding both experimental and computational activities together. Even if it were recognized, there would probably be structural changes needed within the funding organization to actually accomplish funding of joint experimental and computational activities.

Organizations that contract for experimental services to provide validation data need to understand the new requirements in high quality validation experiments. Valuing candor and forthrightness in subcontracting for these types of services will be a difficult change to make for both sides of the business enterprise.

Finally, within organizations that conduct validation experiments and simulations, a change in management perspective is needed. Funding groups within an organization must understand that experimental and computational activities need to be funded at the same time. For applied research groups, it is reasonable to imagine this change. For project groups that are focused on building and delivering hardware and staying on budget and schedule, this is extremely difficult to achieve. Some farsighted project managers may devote project funds to core competency development such as validation, but our experience is that these are rare individuals. For organizations that *are* able to fund both experimental and computational groups in a validation experiment project, management must understand that it is the success or failure of the joint activity that should be openly recognized and rewarded.

A common viewpoint of scientific computing managers is that if the joint project shows that the model performed poorly, then the project is considered a failure. Stated differently, many scientific computing managers feel that the goal of a validation activity is to make the model look good. It is difficult for more enlightened managers to change this viewpoint.

11.1.2 Measurement of all needed input data

Guideline 2: A validation experiment should be designed to capture the essential physics of interest, and measure all relevant physical modeling data, initial and boundary conditions, and system excitation information required by the model.

The validation experiment should be designed to address the physics of interest with regard to its spatial dimensionality, temporal nature, and geometric complexity. For example, one may be interested in conducting a two-dimensional (2-D), either planar or axisymmetric, experiment. The question is: are the three-dimensional (3-D) effects in the experiment negligibly small so that a 2-D simulation can be appropriately compared with the experimental data? Since the ability of the experimentalist to detect 3-D effects is very limited, the analyst should attempt to answer this question. Both 2-D and 3-D simulations of the experiment should be computed to predict the spatial dimensionality effect on the system response quantities (SRQs) of interest. If the change in the SRQs between the 2-D and 3-D simulations is comparable to the expected experimental measurement uncertainty, and the 3-D simulations are credible, then a 3-D simulation should be used in model validation. Note that both solutions must be mesh resolved and iteratively converged to the same level so that a proper comparison of results can be made. If the 3-D simulations are not computed, then the difficulty is what to conclude if poor agreement is obtained between the experimental measurements and the 2-D simulations. Is the disagreement due to the 2-D assumption in the model, or other weaknesses in the model? Recent experience by several researchers in fluid dynamics has shown that some experiments that were thought to be essentially planar 2-D were found to have significant 3-D effects. An analogous question arises with regard to temporal assumptions in the model. Can a steady-state assumption be made in the model, or is an unsteady simulation required?

Guideline 2 points out that the goal-directed activities of the validation experiment must be understood by both the computational and experimental members of the team. From the computational side, the analysts must understand what spatial dimensionality, temporal nature, and geometric simplifications are appropriate in modeling a specific experiment. From the experimental side, the experimentalists must decide what design and operational features should exist in the experimentally tested system, in the experimental facility, and in the instrumentation so that the specified modeling assumptions are properly tested. If the parameters that are initially requested by the analyst for the simulation cannot be satisfied in the proposed experimental facility, it may be feasible to alter the model inputs and still satisfy the primary validation goals. Or it may be necessary to look for another facility. For example, in a fluid dynamics experiment, can a laminar boundary layer on a surface be ensured and

can it be accurately characterized? Is the type and quantity of instrumentation appropriate to provide the required data in sufficient quantity and at the required accuracy and spatial resolution? Conversely, analysts must understand, or be informed by the experimentalists, concerning the limitations of the experimental equipment, the instrumentation, and the facility.

As part of the requirement for Guideline 2, the experimentalists should measure all important physical modeling data, initial conditions (ICs), boundary conditions (BCs), and system excitation information needed as input for the simulation. Any important experimental conditions needed by the code, but not measured, will significantly undermine the value of the validation experiment. If key input data are not measured, as discussed in Chapter 10, the best result that can be obtained is that the simulation will yield imprecise results, e.g., either SRQ intervals or probability-boxes (p-boxes), which must be compared with the measurements. (p-boxes were introduced in Chapter 3, Modeling and computational simulation, and will be discussed in more detail in Chapter 12, Model accuracy assessment.) The much more likely result is that the analyst will use the unmeasured quantities as adjustable parameters to calibrate their models to obtain the best agreement with the measured system responses. This, of course, destroys the fundamental goal of validation: assessment of model predictive accuracy.

The input quantities that should be used in the computational simulation are those that are *actually measured* in the validation experiment. Some of these input quantities from the experiment may not be known or even knowable for various reasons. Two different reasons for not precisely knowing input quantities can arise. First, a quantity can be epistemically uncertain, for example: (a) a quantity was not measured because of poor communication between the analyst and the experimentalist, but was characterized by the experimentalist as an interval; (b) a quantity was not measured and the experimentalist simply estimated the quantity based on experience, or took the value from an engineering handbook, as opposed to actual measurement; or (c) the geometric description of the test article is taken from the fabrication drawings of the hardware, as opposed to determining the deflection of the article due to mechanical and thermal loading in the actual experiment.

Second, a quantity is known to be a random variable, *and* it is not independently measurable before, during, or after the validation experiment. Some examples are (a) a quantity cannot be measured independently from the validation experiment because the measurement alters the quantity, such as a destructive test; (b) a quantity cannot be measured independently from the experiment because its characterization changes during the experiment; and (c) a quantity in the experiment is not controllable from one experiment to the next, and it is not independently measurable. If the input quantity is uncertain because of both a lack of knowledge (epistemic uncertainty) and a random process (aleatory uncertainty), then the quantity should be characterized as a p-box.

Input quantities that are aleatory uncertainties and, in addition, are *not* independently measurable as part of the validation experiment, should be characterized in terms of a probability density function or a cumulative distribution function. For example, suppose a material property is an input quantity in a validation experiment, but it has inherent

variability due to manufacturing. The property can be measured, but the measurement alters the specimen, rendering it useless for the validation experiment. The proper procedure is to characterize the variability of the property by drawing samples from the parent population, e.g., a given production lot. After a number of samples are obtained, a probability distribution is constructed for the population and it is used as the input to the model. This probability distribution is propagated through the model to obtain a probability distribution for the SRQ of interest. A quantitative comparison can then be made between a probability distribution of experimental measurements and the probability distribution from the simulation. This procedure is discussed in more detail in Chapter 12.

To give an indication of the level of detail that is needed in characterizing a validation experiment, various fluid dynamics experiments will be discussed. The following are examples of the level of detail that may be needed for physical modeling and boundary conditions for a vehicle tested in a wind tunnel:

- accurate measurements of the actual model dimensions, as opposed to nominal or as requested specifications;
- surface roughness condition, including imperfections or mismatches in body components or attachments;
- location of boundary layer transition over the surface of the vehicle for all free-stream conditions, angles of attack, and control surface deflections;
- measurement of free-stream turbulence quantities in the wind tunnel;
- accurate measurement of locations and geometric detail of all instrumentation, as opposed to requested locations or installations stated in the fabrication drawings;
- location of where the free-stream conditions were measured in the wind tunnel, especially for a subsonic free-stream;
- for a subsonic free-stream, pressure measurements on the tunnel walls inside the test section near the beginning and end of the computational domain;
- detailed dimensions and geometric detail of all hardware for mounting the model in the wind tunnel.

For aircraft configurations, wings, and deformable bodies, an additional important detail is the measurement of the actual deformed geometry under the experimental load. A much less desirable approach is to calculate of the shape of the deformed structure. The measurements or calculations of the deformed structure may involve time-dependent histories. In long duration hypersonic wind tunnels, the deformation of the vehicle due to aerodynamic heating should be measured or, if necessary, estimated. For hypersonic and supersonic tunnels, the model's surface temperature should be measured at an appropriate number of locations over the surface of the model during the run-time of the wind tunnel.

In wind tunnels, the free-stream flow provides the inflow boundary conditions for the simulation. This information comes from measurements of flowfield quantities in the test section of the tunnel, such as the free-stream Mach number; total pressure and static pressure; and total temperature and static temperature. These data are measured at a number of points in the test section during the calibration of the wind tunnel. Commonly, these data are spatially averaged over the test section and these averages are used in simulations. For

boundary layer transition experiments, the calibration measurements should also include free-stream turbulence intensity and Reynolds stresses. Some facility managers may be reluctant to share such detailed flow quality data with users (and competitors). However, for experimental data that will be compared with CFD simulations that use sophisticated turbulence models or transition models, these data are critical.

For supersonic wind tunnels, the flowfield calibration measurements at the beginning of the test section could be used to set the flow properties as location dependent boundary conditions upstream of the bow shock wave of the vehicle. Such an approach, although conceptually feasible, is just now being attempted in simulations for validation experiments. Some might consider this approach excessive and unnecessary for wind tunnels with high quality, i.e., very uniform, flowfields. However, in Section 11.4.2 we show that flowfield nonuniformity is the largest contributor to experimental measurement uncertainty in three high quality supersonic wind tunnels that have been used for validation experiments. We contend this is true for most wind tunnels.

For subsonic wind tunnels, the question of boundary conditions becomes much more complex because of the elliptic character of the PDEs describing the flowfield. For low-speed wind tunnels, even with low levels of model blockage, one of the first issues that must be addressed by the CFD analyst is: should I model the flow in the entire test section of the tunnel, or assume an infinite-size tunnel? This question could be restated as: for the system response quantities that will be measured in the tunnel and compared with the CFD simulations, what are the changes in these quantities if I compute the simulation of the wind tunnel used in the experiment, versus an infinite-sized tunnel? Although, to our knowledge, no detailed analyses addressing this question have been published, we believe that the sensitivity to tunnel blockage will be significant even at low to moderate blockage. Wind tunnel experimentalists use correction factors to try to eliminate blockage effects on various measured quantities, but these can be of questionable accuracy. For transonic flow tunnels these correction factors are much more in doubt.

11.1.3 Synergism between computation and experiment

Guideline 3: A validation experiment should strive to emphasize the inherent synergism that is attainable between computational and experimental approaches.

By *synergism* we mean a joint computational and experimental activity that generates improvements in the capability, understanding, or accuracy in both approaches. The improvements in each approach can be immediate, for example, during the present activity, or they can be improvements that will benefit future projects. Those discovering the benefits of joint computational/experimental efforts sometimes claim that synergism is the primary benefit of the effort. Discovering the strong positive reinforcement of analysts and experimentalists working closely together can be pleasantly surprising, but validation experiments contribute much more than this. We give two examples from fluid dynamics of how this synergism can be exemplified.

First, the strength of one approach can be used to offset a weakness of the other approach. Consider the example of perfect-gas, laminar flow in a supersonic wind tunnel. Assume that a wind tunnel model is designed so that it can be easily reconfigured from simple to complex geometries. For the simple geometry at a low angle of attack with no separated flow, one should be able to compute flowfield solutions with very high confidence, exclusive of the separated flow in the base region of the vehicle. This may require independent CFD codes and analysis teams, but it is certainly possible with present CFD technology. High-accuracy solutions can then be compared with wind tunnel measurements to detect a wide variety of shortcomings and weaknesses in the facility, the instrumentation, and the data recording system. An example of this synergism will be discussed in Sections 11.2 and 11.4. If the high accuracy solution is for the flowfield of the probe geometry used in the calibration of the test section, then the calibration measurements can be more accurately interpreted to improve the calibration of the test section flowfield. If the wind tunnel model is reconfigured into a complex vehicle geometry, then the situation reverses from the simple geometry case because this geometry would have strongly 3-D flows, separated flows, and shock wave/boundary layer separation. For this complex flow case, highly accurate CFD models do not yet exist. The experimental measurements would be expected to be more accurate than the CFD simulation, and the complex geometry case would then be viewed as a validation experiment to test the physics models in the code.

Second, one can use CFD simulations in the planning stages of a validation experiment to dramatically improve the design, instrumentation, and execution of the experiment. For example, one can compute shock wave locations and their impingement on a surface, separated flow and reattachment locations, regions of high heat flux, and vortical flows near a surface. Such computations allow the experimentalist to improve the design of the experiment, especially the type, sensitivity, and location of the instrumentation. This strategy can also be taken a step further by optimizing the design of the experiment to most directly stress the models in the code, i.e., design the experiment to *break the model*. Optimizing the experimental design to break the model can be done by (a) optimizing the physical modeling parameters, such as the Reynolds number and the Mach number; (b) modifying the boundary conditions, such as model geometry and wall surface conditions; and (c) changing the initial conditions for an initial value problem, such as a blast wave problem. We should point out that analysts commonly do not find it appealing to pursue a strategy of optimizing the experiment to break their models.

11.1.4 Independence and dependence between computation and experiment

Guideline 4: Although the experimental design should be developed cooperatively, independence must be maintained in obtaining the computational and experimental system response results.

The goal of this guideline is to achieve a blind test prediction for comparison with the experimentally measured results. There are varying opinions as to the value added of blind test predictions contrasted to comparisons made with measurements known to the

analysts. Guideline 4 makes clear our view of the importance of blind test predictions. Most experimentalists and experienced project managers who have relied on computational predictions in the past share this view. Guideline 4 can be viewed as the essence of what predictive accuracy should be and how it should be experimentally evaluated.

Analysts often view Guideline 4 as a personal affront to their integrity. This is clearly *not* the intention. The intention is to learn from the experiences of many individuals and projects as to the impact of knowing the correct (experimental) result when a computation–experiment comparison is to be made. It should also be noted that very rarely are blind test predictions openly conducted and published in the literature. If a blind test prediction exercise is conducted within an industrial organization, then the results are sometimes considered proprietary, whether the comparisons are good or bad. Analysts, or more precisely their organizations, see very little benefit to be gained in a blind test prediction, and a great deal of risk if poor comparison results are to become public knowledge. It is not unheard of for organizations that have produced poor results to resort to devious means to conceal or limit the distribution of the results of a blind prediction. There must be an extraordinary incentive for organizations to participate in these blind test prediction activities. The most common motivator is that if the organization does not participate, then they will be barred from certain future bidding opportunities or government certification of their codes.

It is difficult to accomplish the close cooperation of the analysts and the experimentalists promoted in Guideline 1 and, at the same time, retain the appropriate independence of computational and experimental results. However, this challenge can be met by careful attention to the procedural and management details, outlined in the following. When the experimental measurements are processed and analyzed, the analysts should *not* be given the system response results initially. The analysts should be given the complete details of all of the input data needed for their modeling approach, e.g., the physical modeling parameters, ICs, BCs, and system excitation, *exactly* as it was conducted. That is, everything that is needed for the analysts to compute solutions must be provided – but no more. The analysts should not be shown the experimental measurements of the SRQs. In addition, the analysts must be required to quantify the numerical solution errors and propagate any aleatory or epistemic uncertainties in the input so as to obtain uncertainties in the quantities that will be compared with experimental measurements. Then the comparisons between the computational and experimental results are made, preferably by the experimentalists or organizers of the effort. Even if the agreement is poorer than hoped for, investigating the causes of differences invariably leads to a deeper understanding of the experiments and the simulations.

As a final procedural comment, we recommend that management *not* be involved in the initial comparisons and discussions. The discussions should just involve analysts and experimental staff. Often it is found in the initial discussions that miscommunications or misinterpretations between team members will be exposed and appropriate changes are made in the results of one or both team members. If management is involved before these necessary discussions or corrections are made, over zealousness on one side of the team can sometimes occur, possibly leading to behavior that is destructive of teamwork.

Guideline 4 stresses the independence of the computational and experimental validation activities for an engineering environment, not a research code or a mathematical model-building environment. For example, one may have the situation where experiments are being conducted for the purpose of building mathematical models of physical processes, e.g., turbulence models. As was pointed out in Chapter 10, experiments to better understand physical processes are not validation experiments, but are model building experiments. Model building experiments require very close cooperation and communication between the model builder and the experimentalist; sometimes they are the same person.

It is also recommended that validation experiments should not produce data that fundamentally depend on closely related model calculations for critical data reduction tasks. By *closely related model calculations* we mean calculations from a data reduction code that shares many of the same physical assumptions as the model that is being validated. Data from such validation experiments do not properly address the need for independence of experimental data and simulations. Experimental data that require closely related model calculations for data processing can never be a desirable outcome for a validation experiment, although this exact situation may arise in other kinds of experiments. For an example of what we mean, consider the problem of determining the material temperature from experimentally acquired shock hydrodynamic data (density, pressure, and velocity fields) using simulations of shock wave physics rather than some type of diagnostic that directly measures the temperature. The only possible validation data that will emerge from shock hydrodynamics experiments without temperature diagnostics are the shock hydrodynamic data. This problem is relevant, since it arises in investigations of temperature dependence in high-pressure, shock-driven, material response. Such experiments often need simulations to estimate the associated thermal conditions under shock loading. For purposes of scientific discovery, this is permissible though speculative. Such experiments, however, cannot be claimed to provide validation data for the high-pressure thermomechanical response of the material because of the lack of independence of calculations and experiment.

11.1.5 Hierarchy of experimental measurements

Guideline 5: Experimental measurements should be made of a hierarchy of system response quantities, for example, from globally integrated quantities to local quantities.

A discussion was given in Chapter 10 concerning the concept of a hierarchy of difficulty to predict and measure various SRQs. The spectrum of difficulty was ordered in terms of integrals and derivatives of the dependent variables appearing in the mathematical model. Here we repeat the figure discussed in Chapter 10 as Figure 11.1. Measurements in a validation experiment should be made, if possible, over multiple levels of the hierarchy. That is, design the experiment with the strategy such that measurements of the SRQs are made over at least two or three levels of the hierarchy. Experience has shown that a strategy of concentrating on measuring one or two SRQs at the same level can produce inconclusive

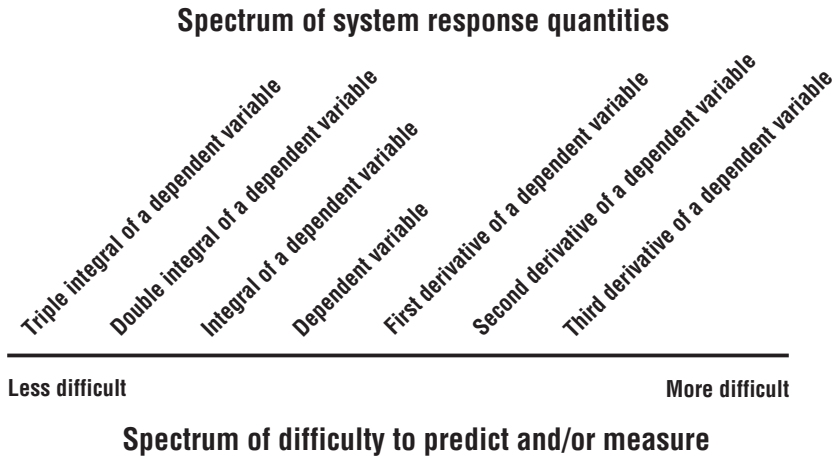


Figure 11.1 Spectrum of SRQs and the difficulty to predict and measure each.

and misleading validation results. A model with substantive predictive capability should yield robust predictions over at least two, if not more, levels of the hierarchy.

An important practical advantage also comes into play by taking measurements over a range of difficulty. There are occasions where the model performs much more poorly than expected. If measurements are only taken at one level of difficulty, or only a few SRQs are measured at the same level, then it can be very difficult for the model builders to discover the root causes of the weaknesses in the model. In addition, multiple levels of measurement could also help determine if there are inconsistencies, flaws, or mistakes in the experimental results. If, on the other hand, measurements were made over a range of difficulty, and a range of different types of SRQs, then this is much more helpful to the model builders in uncovering the weaknesses. This is especially important if there is a wide range of spatial and temporal scales occurring in the physics, such as multi-scale modeling, or there is strong coupling of different types of physics.

There is another advantage in conducting measurements over multiple levels of difficulty. Suppose that measurements were only made at one level of difficulty, but it was a very demanding level for predictive capability. Also, suppose that the model agreed well, maybe surprisingly well, with the measurements. Sometime afterwards, however, a similar validation experiment is conducted but it is found that there is a surprisingly large *disagreement* between the model and the measurements. Further investigation shows that the good agreement on the first set of experiments was entirely fortuitous. It was discovered that cancellation of modeling errors, or the fortuitous benefit of numerical solution errors, was the reason for the good agreement; not the high fidelity physics of the model. If measurements are made at different levels of difficulty and different types of SRQs, fortuitous good agreement can usually be spotted. For example, suppose one observes that very good agreement is obtained for fine spatial or temporal detail in the physics, but the model is not able to predict larger scale features of the physics. Something is seriously amiss in either

the model or the experiment. Our recommendation is that whether the agreement between computation and experiment is good or bad, computation and experiment each should be subjected to critical examination.

Two pragmatic realities that must be factored into the strategy of measurements at multiple levels of difficulty are the following. First, a wider range of experimental diagnostic capability must be available. This is an expense not only for the additional diagnostics and possibly modifications to the experimental facility, but also in the wider range of technical staff capabilities that must be supported. Second, the experimental measurement uncertainty typically increases as one moves to higher levels of difficulty in the spectrum shown in Figure 11.1. The rate of increase of measurement uncertainty can be substantial as one moves to higher levels in the spectrum, particularly if new diagnostic measurement techniques must be developed and perfected.

One diagnostic method to cover two, and possibly three, levels on the measurement hierarchy is the use of modern optical techniques. With the wide range in framing speeds and pixel resolutions available in digital recording cameras, one can record many time-dependent features of a system's response over a large two- or three-dimensional field. This breadth of spatial and temporal scales allows quantitative data to be obtained over multiple levels of the measurement hierarchy. For example, in certain types of experiment it is very difficult, or impossible, to place sensors in locations of interest. Some examples are (a) flight motion of small objects such as bullets, birds, or insects; (b) impact and penetration of high-speed objects; (c) interaction of a fluid and a very thin flexible structure; (d) in situ shock wave dynamics in both fluids and solids; and (e) various types of free surface flow. In some situations where the sensors influence the physical phenomena of interest one can include the physical and geometric characteristics of the sensors directly in the simulations.

In fluid dynamics, for example, modern techniques, such as particle image velocimetry (PIV) and planar laser-induced fluorescence (PLIF), can measure local fluid velocity, mass concentration, temperature, and pressure over a plane of data. This is an extraordinary increase in the quantity of data compared to local or point measurements in a field. Experimental data from video recording systems can commonly be compared directly with computed SRQs. Alternatively, computational solutions could be post-processed by solving an additional set of physics equations for modeling optical responses through the domain of interest. For example, in fluid dynamics the computed flowfield solution, including the mass density field, can be used to compute a Schlieren or interferometer photograph that can then be compared with the experimental image. Obtaining quantitative experimental data over a plane and comparing it with computational results is much more demanding of a model's predictive capability than local measurements in a field.

Video and audio recording of an experiment or setup of the experiment has proven to be very helpful in detecting experimental setup problems, improving the understanding of the physics, and documenting the experiment. Depending on the time scales in the experiment, one could use inexpensive standard framing rate digital AV equipment or use high-speed digital video recordings. Some examples where standard AV equipment can prove very beneficial are (a) unexpected interactions of sensors with the phenomenon that

is being sensed, (b) unexpected flowfield interactions either within the facility in which the experiment is conducted or between various elements of the geometry being tested, (c) discovering that modeling assumptions concerning BCs or ICs were inappropriate, and (d) discovering that surface characteristics or assembly of test articles were not what was intended or what was modeled. High-speed video recordings can now be coupled with automatic feature recognition algorithms to produce quantitative data for position, velocity, and acceleration. In addition, with the large amount of AV records that can now be stored, new technologies are becoming available that can search AV records for certain classes of objects or words.

11.1.6 Estimation of experimental uncertainty

Guideline 6: The experimental design should be constructed to analyze and estimate the components of random (precision) and systematic (bias) experimental uncertainties.

This guideline is, of course, recommended in any type of experimental measurement, but in validation experiments its importance is crucial. One could ask: why is it so important? Since the primary goal of a validation experiment is to assess the accuracy of a model, the *core issue is uncertainty estimation*. By contrast, in the field of metrology the focus is on estimating the *true value* of a quantity along with the estimated uncertainty. Granted, these two perspectives are closely related, but the emphasis is different. In validation experiments, it is more important to quantitatively and rigorously estimate the uncertainty of all needed quantities than to infer what the true value might be based on those measurements. We are *not* suggesting that high accuracy measurements are unimportant. We are stressing that accurate estimation of uncertainty of (a) experimental conditions needed for a simulation, and (b) measured SRQs are of the utmost importance in validation experiments.

The standard technique for estimating experimental measurement uncertainty is the International Organization of Standardization (ISO) technique described in the *Guide to the Expression of Uncertainty in Measurement* (GUM) (ISO, 1995; ISO, 2008). The same procedure is codified in the *US Guide to the Expression of Uncertainty in Measurement* published by the American National Standards Institute (ANSI, 1997). The ISO/ANSI technique has also been described in detail in certain disciplines, such as wind tunnel testing (AIAA, 1999, 2003). This technique is the minimum level of effort that should be required for uncertainty estimation in validation experiments. The technique is primarily concerned with the expression of uncertainty in the measurement of a fixed quantity that is referred to as the true value. Uncertainties are not categorized according to the usual terms random and systematic, but according to the *method* which is used to characterize an uncertainty. The ISO/ANSI technique divides uncertainties into

Type A: uncertainties that are evaluated using a statistical analysis of a collection or ensemble of observations;

Type B: uncertainties that are evaluated by means other than a statistical analysis of a collection or ensemble of observations.

The combined standard uncertainty of a quantity (the result of a number of other measured quantities) is the positive square root of a sum of terms. The terms are the variances or covariances of the other measured quantities weighted according to how the measurement result varies with changes in these quantities. The fact that Type B uncertainties are supposed to come from “non-statistical means” does not deter the ISO/ANSI approach from using statistical concepts anyway.

A different approach has been used by the statistical sampling community for at least the last four decades (Montgomery, 2000; Box *et al.*, 2005; Hinkelmann and Kempthorne, 2008). This is an entirely statistical approach that is based on analyzing comparisons of multiple measurements of quantities of interest. It is usually referred to as statistical design of experiments (DOE) because it analyzes the final measured result of the quantities of interest, based on specially designed sampling techniques. In the DOE approach one identifies various classes of source that are believed to contribute to the combined uncertainty in the final quantities of interest. Then experimental strategies are designed to quantify the statistical contribution that each of these sources contributes to the final quantities of interest. The strategy of the DOE is quite different from the ISO/ANSI approach mentioned above. The approach uses replication, randomization, and blocking techniques in the design of the sample data collected in an experiment. This approach has been widely used in analyzing data from many fields, for example, production process control, system and component reliability, environmental statistics, biostatistics, medication testing, epidemiology, etc. This approach, however, has seen limited use in experiments in engineering and physics. Oberkampf, Aeschliman, and colleagues have used this technique extensively in wind tunnel validation experiments and compared it with the results obtained from the ISO/ANSI approach (Oberkampf *et al.*, 1985; Oberkampf and Aeschliman, 1992; Oberkampf *et al.*, 1993, 1995; Aeschliman and Oberkampf, 1998). Both the ISO/ANSI and DOE approaches will be discussed in detail in Section 11.3.

11.2 Validation experiment example: Joint Computational/Experimental Aerodynamics Program (JCEAP)

11.2.1 Basic goals and description of JCEAP

In 1990, Sandia National Laboratories initiated a long-term, coupled CFD/experimental research effort, referred to as the Joint Computational/Experimental Aerodynamics Program (JCEAP). The goal of the program was twofold: (a) improve Sandia’s hypersonic wind tunnel experimentation and CFD simulation capabilities, and (b) improve the understanding of how CFD and experiments could synergistically work together for a more rapid development of each. The research program came to a successful close in 1997 and a number of reports, conference papers, and journal articles were produced describing the effort (Oberkampf and Aeschliman, 1992; Walker and Oberkampf, 1992; Oberkampf *et al.*, 1993; Aeschliman *et al.*, 1994, 1995; Oberkampf *et al.*, 1995, 1996; Aeschliman and Oberkampf, 1998; Oberkampf and Blottner, 1998). The program produced an extensive database of force and moment, and surface pressure measurements on a hypersonic vehicle

configuration. More importantly, the program developed the fundamental concepts for the design and execution of high quality validation experiments. The six guidelines discussed above evolved from the experiences of JCEAP, and not all were learned easily. The discussion that follows excludes many details that would be needed to use the experimental data as a validation database for assessing CFD codes in hypersonic flow. See the references just given for the details. Here we will concentrate on using JCEAP to exemplify the six guidelines discussed.

The first phase of the program involved the computational and experimental study of the aerodynamic forces and moments on a flight vehicle configuration in a hypersonic flowfield. During this phase of the program, the geometry of the vehicle was determined, the force and moment (F&M) wind tunnel model was fabricated, the experimental conditions were determined, and the F&M data were obtained. The second phase of the program addressed the next level of difficulty in prediction and experimentation; surface pressures on an identically sized and shaped geometry at the same flow conditions. As part of both the F&M and pressure experiments, an experimental uncertainty estimation procedure was developed. The procedure was based on a statistical DOE procedure that had been used in a previous wind tunnel experiment, unrelated to JCEAP (Oberkampf *et al.*, 1985).

All of the experimental measurements were obtained in the Sandia National Laboratories long-duration, blowdown, hypersonic wind tunnel (HWT). The tunnel is a blowdown-to-vacuum facility that can operate up to a minute, depending on operating conditions. The nozzle and test section consist of three contoured axisymmetric nozzles, arranged as a large fixture that can revolve around a common hub. Figure 11.2 shows an overall view of the wind tunnel with the test section open. Depending on which nozzle is used, the test section Mach number is nominally 5, 8, or 14. The diameter of the circular test section of each nozzle is roughly the same, about 0.35 m. Each nozzle is provided with its own electric resistance heater to prevent flow condensation in the test section. The total pressure and temperature of the flow is controlled during a run so that a chosen Reynolds number can be obtained within the operating envelope of the facility. During a run the total pressure is manually controlled, but all other operating parameters are computer controlled. Usable run times are typically 30–60 s, depending on the flow Reynolds number, and the turnaround time between runs is one hour or less. Model pitch angle (angle of attack) is varied during a run using a computer-controlled, arc sector drive mounted far aft of the model. Model roll angle and configuration changes are made between runs by retracting the aft portion of the wind tunnel and manually making the needed changes to the mounting of the model on the sting support.

11.2.2 Joint planning and design of the experiment

11.2.2.1 Wind tunnel conditions

From the very beginning of the effort, JCEAP was designed to be a validation experiment. What that meant, however, was not fully recognized. The first design decision that needed to be made was: what Mach number condition should be used in the HWT? The primary

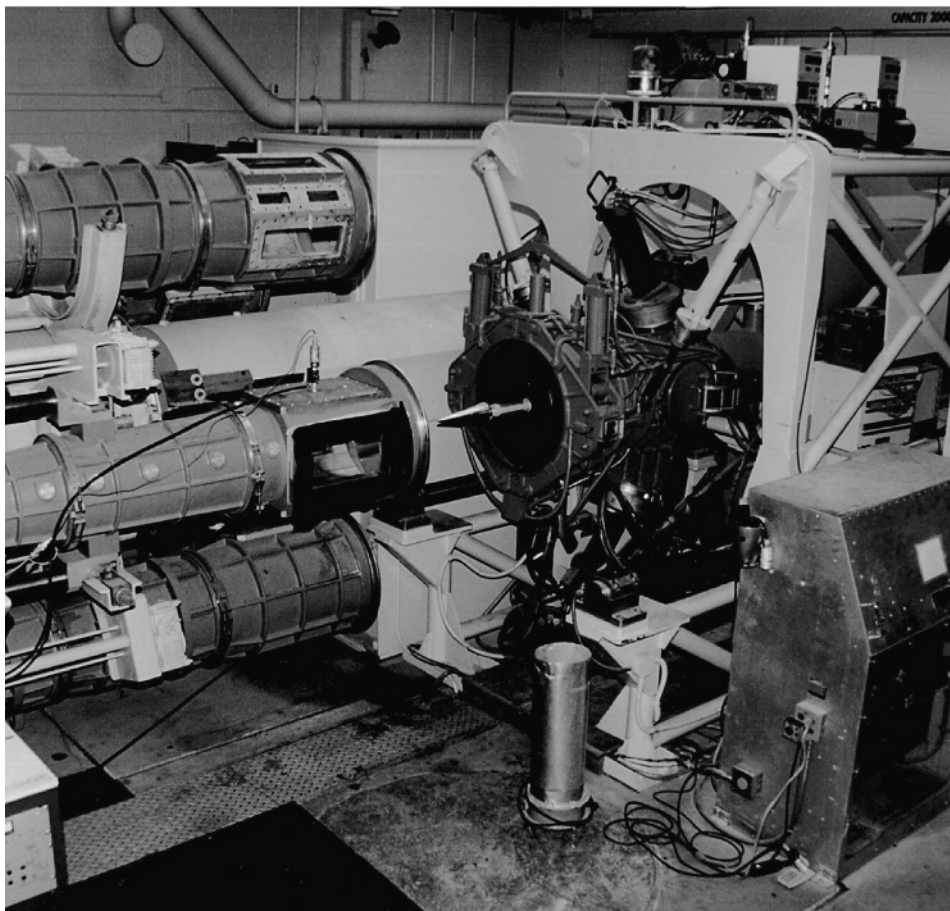


Figure 11.2 Sandia hypersonic wind tunnel with the aft section retracted from the test section (flow is from left to right). The model is visible in the open section and the yellow support is part of the rear-mounting fixture to hold the model (Oberkampf *et al.*, 1993). (See color plate section.)

CFD requirement was that the flow over the model should be entirely laminar flow. This was a very unusual request in wind tunnel testing because the normal desire is to have turbulent flow over the vehicle that matches the flowfield of the vehicle in flight. The reason laminar flow was the critical driver was that only with laminar flow was it conceivable that the CFD simulations could *possibly* compute some flow conditions more accurately than experimental measurements could be made. With turbulent flow, no such claim could be made; and this is still true two decades after the initiation of JCEAP.

The Reynolds number of the flow over the vehicle is the primary factor in determining if the flow is laminar or turbulent over a vehicle. The Reynolds number is a dimensionless ratio defined as $R_L = \rho V L / \mu$, where ρ is the free-stream density, V is the free-stream velocity, L is a characteristic length scale (usually the total length of the vehicle), and

μ is the absolute fluid viscosity in the free-stream. The lower the Reynolds number, the more likely the flow will remain laminar over the entire vehicle. As the Reynolds number increases, the flow will transition to turbulent flow over different portions of the vehicle (Wilcox, 2006).

The operating conditions of the HWT facility for the Mach 5 nozzle were examined, given a reasonable estimate on the length of the model that could be used in the test section. It was found that at the lowest Reynolds number conditions that were producible for Mach 5, it was unavoidable that a mixture of laminar and turbulent flow would occur over the geometry. For the Mach 14 nozzle, the range of Reynolds numbers attainable is much lower because of the larger expansion of the flow into the test section. As a result, laminar flow could be assured over the model at any operating condition of the facility. The primary argument against using the Mach 14 nozzle was that the pressures in the test section were so low that very accurate measurement of surface pressure would become questionable.

For the range of operating conditions available for the Mach 8 nozzle, Reynolds numbers could be produced that balanced the needs of the experiment. Operating at the higher-pressure conditions of the operating envelope would produce a mixture of laminar and turbulent flow. Operating at the lower end of the envelope would produce the all-laminar flow desired, and the pressure in the test section is sufficiently large to measure surface pressures accurately. The Mach 8 test section has a diameter of 0.350 m and is provided with 0.2×0.38 m Schlieren-grade windows on the top, bottom, and sides. The high-grade windows provide excellent access for optical diagnostics.

11.2.2.2 Model geometry

We were interested in designing a vehicle geometry that could produce a range of flow features. It was required that the simplest flowfield would have the possibility of being computed more accurately than it could be measured. This could be attained with a simple geometry and using low angles of attack. Complex flowfields could be attained by modifying the geometry and using higher angles of attack. Also factored into our thinking was the desire to produce a geometry that could be of interest to the Sandia project group that flight-tested re-entry vehicles. They were interested in maneuvering re-entry vehicles, that is, vehicles that could maneuver during hypersonic flight by deflecting control surfaces on the vehicle.

After a number of design iterations for the geometry using CFD, we settled on a spherically blunted cone with a half-angle of 10° with a sliced region on the aft portion of the cone. We did not choose a sharp-nosed vehicle for three reasons. First, no matter how sharp a machinist fabricates a nose tip, there is always finite nose bluntness. If a sharp tip were used, then a very precise inspection would be required to accurately characterize exactly what the geometry was for the CFD simulations. Second, a very small radius nose tip would change the flow regime from continuum fluid dynamics that exists over the rest of the geometry to rarified flow over the minute nose tip. Stated differently, one of the key assumptions in the mathematical model would be violated, damaging the high confidence

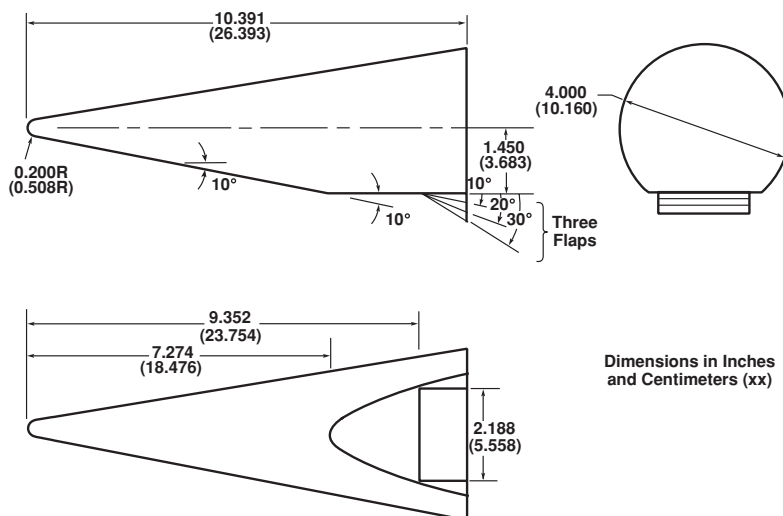


Figure 11.3 JCEAP geometry (Aeschliman and Oberkampf, 1998).

that is desired for certain flowfield solutions. Third, a sharp tip can be easily damaged over a long period of usage in a wind tunnel test. The damage could be from slight erosion due to impact from minute particles in the free-stream, or it could be due to accidental damage due to handling the model. As a result, the nose tip would have to be re-inspected regularly during the course of the experiment to determine if any changes occurred. Based on CFD simulations, 10% nose bluntness was chosen. That is, the nose had a radius equal to 0.1 of the base radius of the cone. This nose radius provided a good balance between the having an entropy gradient approaching the slice on the vehicle, but not an excessively large region of subsonic flow near the nose of the vehicle.

Based on CFD simulations, it was decided to begin the slice, which was parallel to the axis of the cone, at 0.7 of the length of the vehicle. See Figure 11.3 for the final vehicle design. Given that the test section diameter of the Mach 8 nozzle was 0.355 m, a base diameter of 0.1016 m was chosen. With this base diameter and the 10° cone half-angle, the resulting length of the model was 0.2639 m. To increase the difficulty of the CFD simulation for certain geometries, and to make the geometry more appealing to the flight test project group, the model was designed so that different flaps could be attached to the aft portion of the slice. Flap deflection angles of 0, 10, 20, and 30° were chosen, and all of the flaps extended to the baseplane. By designing the flaps to extend to the model baseplane for all flap deflections, a substantial simplification became possible in constructing the computational mesh for the body geometry and for the base flow. This also simplifies setting the outflow boundary conditions across the baseplane in the numerical simulation. Note that, in an actual flight vehicle, this would not be the case. In a flight vehicle, the flap would be hinged near the leading edge of the flap, and as the flap was deflected, the trailing edge of the flap would not extend to the baseplane of the vehicle.

After the vehicle geometry was finalized, we discussed the geometry of JCEAP with the flight test project group. They were interested in the maneuvering re-entry vehicle geometry and its potential advantages from their perspective. However, when we explained to them that we were only going to test in laminar flow, their interest flagged considerably. As a result, they were not interested in providing funding for the project. Fortunately, there was strong support for the project from internal research and development (IR&D) funding. As was briefly discussed in Section 10.1.2, there is commonly conflict between project needs for a validation experiment and the needs for a high quality validation experiment. If funding for the experiment comes from the project group, their requirements usually take precedent.

11.2.2.3 Model fabrication and instrumentation

Two different physical models were fabricated, but both had the same geometry (Figure 11.3). The F&M model was used in conjunction with a precision six-component strain gauge balance. The balance was installed internal to the model geometry, as is typically done to improve the accuracy of the measurements of the aerodynamically induced forces and moments. The pressure model was significantly more complicated in design and fabrication because of the additional instrumentation required. A total of 96 pressure ports were machined in the model surface. The orifices were connected to either the 2482 N/m² or the 6895 N/m² electronically scanned pressure module, both of which were located inside the model. The choice of which ports were attached to which module was based on CFD predictions of the pressure level expected. The model also incorporated nine semiconductor bridge Kulite gauges to detect high-frequency surface pressure fluctuations at certain locations along the surface. The model incorporated four thermocouples mounted in the model wall at two different axial locations along the model. The thermocouples provided model surface temperature for input boundary conditions required for the final CFD calculations.

The pressure model was constructed in four sections (plus detachable nose tips) to permit machining operations and the installation of steel pressure port inserts during fabrication. Figure 11.4 shows a longitudinal cross section of the model, the adjacent portion of the base cover and sting, and the sting cover. This design approach also greatly facilitated tubing connections between ports and internally mounted pressure modules during subsequent model assembly. A base cover extension, slightly smaller in diameter than the base diameter of the model, was fabricated as part of the sting and provided the attachment point to the model. It also provided a hollow recess with additional volume for the pressure modules, pressure tubing, and electrical wiring. The base cover extension had a flat section on one side to conform to the slice on the model. The model was supported in the tunnel by a stiff, thick-walled sting to reduce the deflection of the model due to aerodynamic loading during a run. The sting connects the model to the arc sector drive located downstream of the test section that pitches the model to angle of attack. The sting length is designed so that the center of rotation of the arc sector is at about the mid-length of the model. In this way,

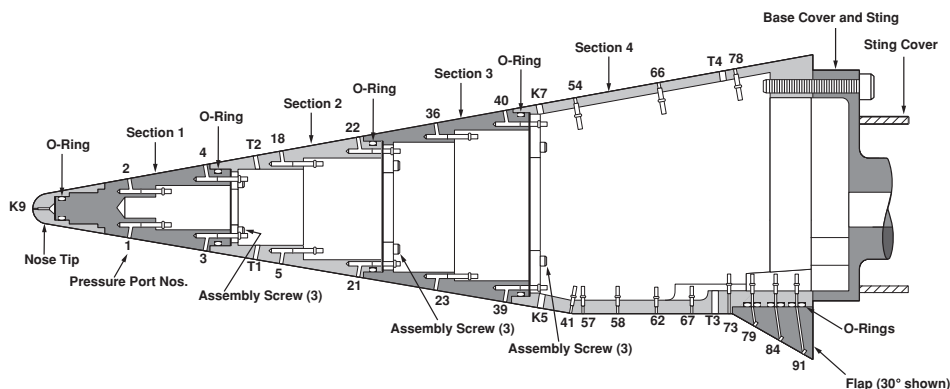


Figure 11.4 Longitudinal cross section of the pressure model showing some of the surface pressure ports denoted by a number, Kulite gauges denoted by K_n , and thermocouples denoted by T_n (Oberkampf *et al.*, 1995).

when the model is moved to a non-zero angle of attack, it remains in the center portion of the test section flowfield.

In preliminary tests of the pressure model, an amalgamation of pressure reference tubing, pressure module electrical cables, wires from the Kulite gauges, and thermocouple wires were taped to the outside of the sting. All of these tubes and cables needed to go from the base of the model to the aft portion of the tunnel where they could exit the downstream flowfield of the tunnel. In a project review of the preliminary JCEAP tests, it was pointed out by CFD analysts that this taped bundle of tubing and cables would cause a complex surface geometry for modeling in a CFD analysis. If a CFD analyst chose to simulate the flow in the base region of the model, near the model sting, then the CFD analysis would need to know the surface geometry of the bundle. After this was pointed out, we recognized that modeling the bundle was of *no interest* to the goals of the validation experiment. As a result, a cylindrical section that covered the bundle and sting was fabricated and attached to the sting, as shown in Figure 11.4. The cover provided an easily definable and repeatable geometry that could be much more easily meshed in a CFD solution. This simplification in base flowfield geometry exemplifies the principle: exclude any complex geometric features that *do not* provide value to the goals of the validation experiment.

The overriding consideration in the development of the surface pressure measurement system was the pressure lag time (commonly referred to as settling time) in the small diameter tubing connecting the model surface pressure ports to the two pressure transducer modules. If a pressure lag error existed, this could introduce a significant systematic uncertainty in the surface pressure measurements. To determine the actual lag characteristic of each port for the tunnel flow conditions of interest, the model was pitched from 0° angle of attack to 10° angle of attack in one continuous sweep, while pressure data were recorded for each of the 96 ports. The same type of sweep was made from 0° to -10° angle of attack. Typical pressure lags were 0.05–0.10 s to achieve a stable pressure within 0.1% of

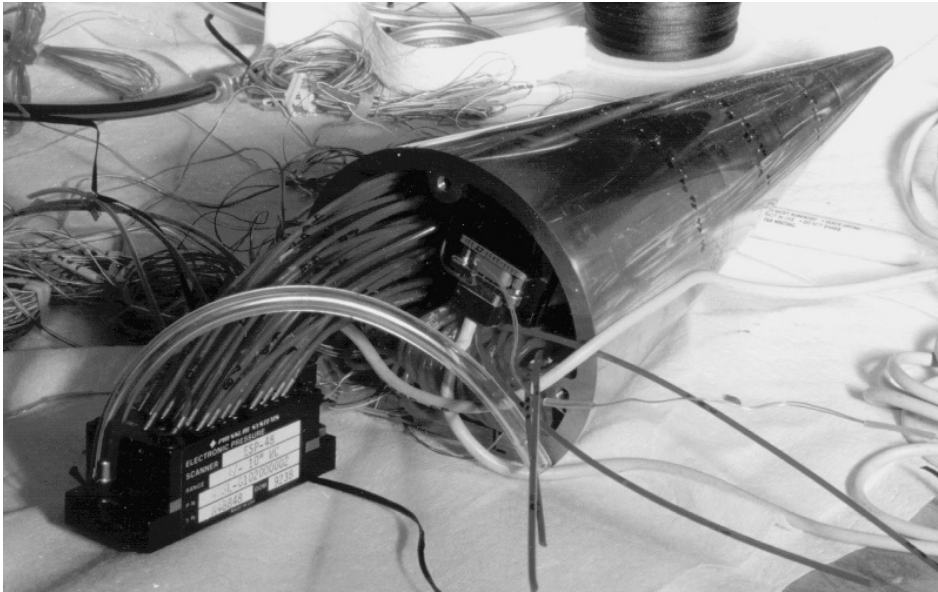


Figure 11.5 Assembly of the pressure model (Oberkampf *et al.*, 1995).

the final value. The largest lag time observed was 0.3 s, which occurred at one of the ports on the slice. To be conservative, a delay time of 0.5 s was used prior to recording pressure data following each change to a new angle of attack during data acquisition runs. The key to achieving this response frequency was mounting both pressure modules and the tubing, along with everything else inside the model. Figure 11.5 suggests the challenge of forcing everything inside the model during assembly, without kinking any of the pressure tubing or breaking any electrical connections.

11.2.3 Characterize boundary conditions and system data

After each model was fabricated, a detailed inspection of the surface geometry of the model was conducted. Although the model was precision made, the actual inspected dimensions of the model were used in future CFD simulations. For the pressure model, the location of the pressure orifices, the Kulite gauges, and the thermocouples were all inspected with regard to the quality of the sensor installation and the precise location of each. As stressed in Guideline 2, the geometry data that should be used in the simulation are the *as-tested* geometry. For the F&M model and the pressure model, they were built with very thick walls so that no deflection due to aerodynamic loads was possible. In addition, for the total temperature conditions available in the HWT, there was no change in geometry due to aerothermal heating of the model.

Preliminary wind tunnel tests were conducted to determine precisely what Reynolds number conditions would be used to attain laminar flow over the model. To aid in this

determination, shear-stress-sensitive liquid crystals were employed. With this technique, the liquid crystals are painted over the surface of the model before a wind tunnel run. The crystals change color, depending on the local surface shear stress applied to them. For the preliminary tests at moderate Reynolds numbers, the technique worked moderately well, but it was not definitive. For the lowest Reynolds number condition tested, however, the shear stress level was lower than the sensitivity range of the crystals that were available. The liquid crystals also provided streak visualization of flow on the surface of the model. By examining photographs of these surface flows, boundary layer transition waves due to cross-flow at angle of attack could be seen at the moderate Reynolds numbers tested. The existence of laminar boundary layer instabilities due to cross-flow had been computationally investigated as precursors to turbulent flow (Spall and Malik, 1991, 1992). At $R_L = 1.8 \times 10^6$ ($R = 6.82 \times 10^6/\text{m}$), however, no transition waves could be seen. As a result, all further experiments and computational simulations were conducted at this Reynolds number.

The wind tunnel flowfield had been calibrated on multiple occasions not only for the Reynolds number and total temperature of interest in JCEAP, but also for many other conditions in the operating envelope. Based on these calibrations, the spatially averaged free-stream Mach number across the central region of the test section was found to be 7.84. As in all wind tunnels, there is a slight variation in free-stream flow characteristics across the central portion of the test section. The spatially averaged conditions from the wind tunnel calibration were used as the inflow boundary conditions in the CFD simulations. It should be noted that the only time this Reynolds number and Mach number condition had been run before was during calibration of the wind tunnel. This exemplifies the common situation that experimental facilities are commonly used at unusual conditions for validation experiments, relative to the needs of project groups.

The run schedule for the combinations of angle of attack, model roll angle, and flap deflection angle was devised very differently from typical wind tunnel tests for project groups. For project groups, the goal is to obtain information at a wide range of operating conditions that are of interest for their system's performance. For validation experiments, the goal is to obtain information at a restricted set of conditions so that experimental uncertainty can be rigorously estimated for all input quantities needed for the simulation, as well as all SRQs that are measured. For JCEAP, the run schedule was constructed so that a number of replications were conducted, runs were randomized, and blocking techniques were incorporated. During each run the angle of attack was varied from -9 to 18° . The sequence of nominal angle of attack, α , during a run was 0, -9 , -6 , -3 , 0, 3, 6, 9, 12, 15, 18, and 0° . For each of these α s, the roll angle was set at 0 (slice on the windward side), 90, 180, and 270° . In addition, measurements were made with the model at two different axial locations in the test section by changing the length of the model sting. The forward axial location places the model essentially at the center of rotation of the arc sector drive. As a result, the model stays near the center region of the test section as it rotates. At the aft axial station, the model rotates and translates downward in the test section for positive α s; conversely, it rotates and translates upward for negative α s. Table 11.1 and Table 11.2

Table 11.1 *Run number schedule for the F&M experiment (Oberkampf and Aeschliman, 1992).*

Roll angle (deg.)	$\delta = 0^\circ$	$\delta = 10^\circ$	$\delta = 20^\circ$	$\delta = 30^\circ$
Model at forward tunnel station				
0	34, 36, 37, 73	63, 72	64, 71	65, 70
90	39	66, 67	69	68
180	40	55	56, 57	58
270	41	62	61	59, 60
Model at aft tunnel station				
0	74, 75	83	82	81
180	76, 77	78	79	80

Table 11.2 *Run number schedule for the pressure experiment (Oberkampf et al., 1995).*

Roll angle (deg.)	$\delta = 0^\circ$	$\delta = 10^\circ$	$\delta = 20^\circ$	$\delta = 30^\circ$
Model at forward tunnel station				
0	20, 22, 62	42, 43	48, 49	56, 57
90	24, 26, 59, 61	37, 39	46	54
180	30, 32, 58	35, 36	44, 45	50, 53
270	28, 29	40, 41	47	55
Model at aft tunnel station				
0	101, 102	118, 119	124, 126	131, 133
180	103, 112	115, 116	122, 123	127, 129

give the complete run schedule for the F&M experiment and the pressure experiment, respectively.

During the F&M experiment, the aerodynamic load on the model was measured. This measured load was then used to calibrate the change in angle of attack due to deflection of the mounting sting. While the tunnel was open, the model was loaded over the same range of aerodynamic loads and the change in angle of attack was measured. This angular deflection was then used to correct the indicated angle of attack from the arc sector mechanism. For the pressure experiment, the same procedure was used, except, because of the stiff sting, the deflections were much smaller than the sting used in the F&M experiment.

Finally, thermocouples were used to measure the surface temperature of the model during a run. The largest differences in the surface temperatures occur between the windward and leeward sides of the model at angle of attack. Consequently, the thermocouple measurements

were made 180° apart around the model. The measured temperature was then used as the surface temperature boundary condition in the CFD simulations.

11.2.4 Synergism between computation and experiment

Examples have been given of the synergistic interaction of computation and experiment in the design of the JCEAP experiment. Here a more detailed discussion will be given of an interaction that could not have happened without the real-time synergism of the computational and experimental team members. Stated differently, if the computational and experimental team members are diverted from a validation experiment for significant portions of their time, or worse, they do not even work on it at the same time, the extraordinary synergism that is possible will not be realized. In addition, it may not even be realized that it was lost.

After the F&M wind tunnel tests were completed, a detailed experimental uncertainty analysis (to be discussed) was conducted. This analysis revealed an anomaly that was most notable in the experimental data for the pitch moment at zero angle of attack for a flap deflection of 0° . The anomaly could also be seen for other flap angles and it could be seen in the normal force for zero angle of attack, but it was not as obvious. Based on the assumption of a uniform free-stream flowfield, it must be true that at zero angle of attack the pitch moment for $\phi = 0^\circ$ and $\phi = 180^\circ$ should be the same mean value (based on many statistical samples), after the appropriate sign reversal was made for symmetry. It was observed that the difference in pitch moment coefficient between $\phi = 0^\circ$ and 180° was 0.008. Although this may seem small, if this error is converted into an error in measuring the angle of attack, it would convert to an error of 0.6° . This was much larger than what the measurement error in angle of attack was demonstrated to be.

To resolve this issue, several possible sources of error were investigated. An analysis was conducted to determine if the nonuniform spatial distribution of base pressure could contribute to pitch moment. Based on five pressure measurements distributed over the base of the model, this effect was found to be negligible. Another possibility investigated was the misalignment of the strain gauge balance mounting taper with respect to the model centerline, which would result in an angle of attack error. Precision model inspection showed misalignment to be less than one minute; therefore, this was not a significant contribution to the discrepancy.

In parallel with the experimental investigation, a CFD effort was initiated to try and assist in understanding the situation. A number of additional CFD solutions were generated with very refined meshes for the zero angle of attack case. These solutions had the highest confidence in simulation accuracy because they were for $\alpha = 0^\circ$ and $\delta = 0^\circ$, where the flowfield is the simplest case of the entire experiment. The computed results for the pitch moment produced a value that did not match experimental results that were in question. The computational results were essentially halfway between each of the experimental results for $\phi = 0^\circ$ and 180° , respectively. Different members of the team, however, had different levels of confidence in the CFD results.

It was also suggested that this anomaly could be caused by operating the wind tunnel at a Reynolds number much lower than the designed Reynolds number of the nozzle. The nozzle contour was designed taking into account a boundary layer displacement thickness corresponding to operating the tunnel at a Reynolds number of $R = 19.7 \times 10^6/\text{m}$. At the low Reynolds number used in JCEAP, the displacement thickness would be larger and, consequently, this could produce slight flow non-uniformity in the test section. Therefore, it was decided to conduct another tunnel entry to reproduce the anomaly at the JCEAP Reynolds number and also to conduct F&M measurements at the designed Reynolds number of the nozzle. During this second tunnel entry at the JCEAP Reynolds number the pitch moment anomaly could *not* be reproduced. The mean value of the zero angle of attack pitch moment coefficient for $\phi = 0^\circ$ and $\phi = 180^\circ$ matched within 0.001, which was the expected level of uncertainty in the data. Then began an even more intensive investigation to discover what had changed since the first tunnel entry.

It was suggested that the test section of the tunnel might have moved slightly from the horizontal since the last tunnel entry. This was possible because all of the different Mach number test sections are on a large revolver-type mechanism. The test section was moved, reassembled, and checked several times. It was always found to be horizontal within one minute. During this checking of angles, however, it was found that the model pitched nose down when the tunnel is closed and locked in place using hydraulic cylinders. Repeated closures and measurements showed this movement to be a constant -0.08° . This consistent systematic error due to tunnel closure had, to our knowledge, existed in the facility since it was built. This constant -0.08° error correction in angle of attack was made to all of the data. The correction, however, was not sufficient in magnitude to explain the anomaly previously identified.

During the second tunnel entry, it was observed that our attempt at reproducing the anomaly did not include the base pressure instrumentation. When the base pressure transducers were installed and the pitch moment measured, the anomaly *was* precisely reproduced. Each of the five base pressures was connected to a transducer that had a full scale of 0.007 atmospheres; therefore, these transducers were grossly off-scale when the tunnel was opened to atmospheric pressure before and after a run. When all the strain gauge instrumentation was connected into the data system and the base pressure transducers were off-scale, in response to atmospheric pressure, those analog data channels were saturated. This introduced a systematic error to all data channels through the common signal return line. This systematic error was easily circumvented and eliminated from the data by using the second air-off zero taken after the tunnel was evacuated just before a run.

Two final observations were made. First, this systematic error due to an electrical ground loop had existed in the past any time base pressures were measured on a test item. It could not be classified as a major experimental error, but the error was definitely detectable by careful statistical checks of the data. These kinds of error are rarely reported anywhere, but good organizations learn from them. Second, the CFD results were exactly correct in predicting the final measured result after the ground loop bias error was removed from the

data. Members of the team who had not believed CFD could calculate some flowfield cases more accurately than they could be measured were stunned.

11.2.5 Independence and dependence between computation and experiment

The nature of the physics modeling for JCEAP lent itself to natural independence between computation and experiment. That is, for laminar flow of an ideal gas, there are essentially no other major modeling issues. Some of the assumptions in the submodels could be challenged, as will be discussed in Section 11.4.1, but these were not expected to produce significant changes in the simulation results.

As discussed in Guideline 4, the simulation must be provided with the input data measured during the experiments. After each of the models was geometrically inspected, these geometry measurements were used in all of the simulations. After the experiments were completed, the *as-tested* model input data from the experiments were used to compute new CFD simulations. For example, the measured average values for the entire run schedule for free-stream conditions and wall temperature boundary conditions were used in the new simulations. Since these data were very similar to what had been assumed in the initial simulations, there was very little change in the computational results.

Before these final simulations were computed, however, a formal mesh convergence analysis was completed. Richardson's extrapolation was used to estimate the mesh convergence error for the 3-D flowfield, as discussed in Chapter 8, Discretization error. This analysis was only done for the primary CFD code used, a parabolized Navier–Stokes code (Walker and Oberkampf, 1992). This code could accurately simulate all of the $\delta = 0^\circ$ cases at angles of attack up to about 16° . For the flap deflection cases, separated flow ahead of the flap did not permit this modeling approach to be used. For the separate flow cases, a full Navier–Stokes code was used. For this code, however, no mesh convergence analysis was conducted because of the computational resources required.

11.2.6 Hierarchy of experimental measurements

JCEAP was designed from the beginning to span two levels of the spectrum of difficulty (see Figure 11.1). Measurements and predictions were made of a dependent variable in the PDEs (pressure), as well as measurements and predictions of integrated quantities (F&Ms). In addition to these measurements, liquid crystal and Schlieren photographs were taken. The liquid crystals not only provided excellent information in the determination of laminar flow conditions, but they were also very valuable in surface flow visualization of separated flows over the geometry. Figure 11.6 is an example of one of these visualizations of the streak pattern on the surface of the geometry. Figure 11.6 is for $\alpha = 0^\circ$ and a flap deflection of 10° . The deflected flap causes a shock wave to form just ahead of the flap, generating an adverse pressure gradient along the slice portion of the geometry. The laminar boundary layer cannot tolerate the adverse pressure gradient, so the boundary layer separates on the slice region. Near the plane of symmetry through the slice the flow reattaches midway up the

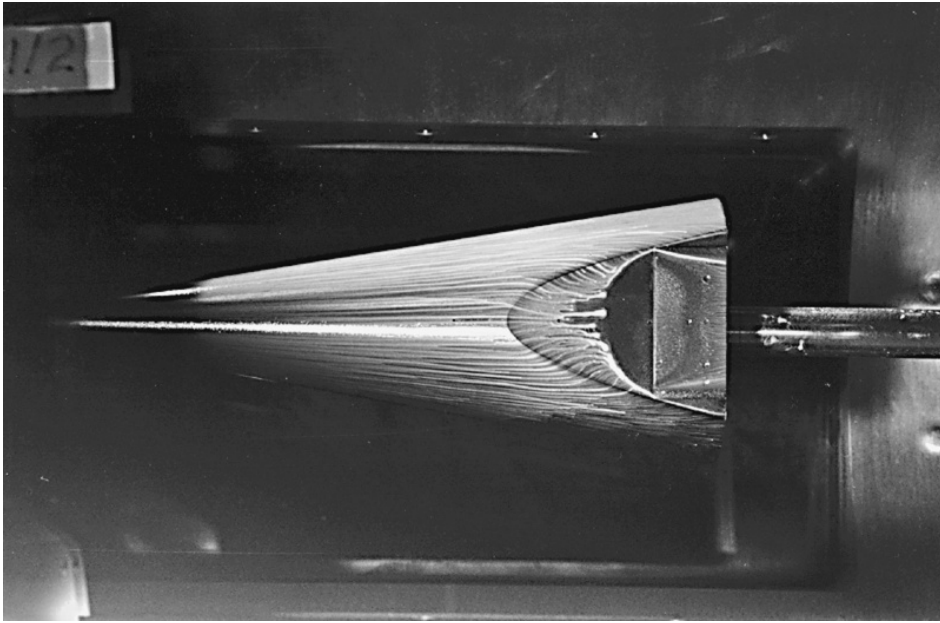


Figure 11.6 Surface flow visualization on JCEAP for $\alpha = 0^\circ$ and $\delta = 10^\circ$ (Oberkamp and Aeschliman, 1992).

length of the flap, but the reattachment point moves toward the front along the flap edges. Visualization photographs, such as these, can provide excellent qualitative information for determining if the physics modeling assumptions are appropriate for the particular conditions of the experiment.

Schlieren photographs show the shadow pattern due to the deflection of light by a refractive index gradient through the flowfield. Schlieren photographs were taken at every angle of attack on every wind tunnel run. These photographs were helpful and instructive in understanding the flowfield, particularly the shock waves in the field.

At one point during the F&M testing it was noticed that for certain model geometries at large angles of attack (15° to 18°) the normal force and pitch moment measurements deviated from the expected trend with α . The situations where this trend was particularly evident were large flap deflections for $\phi = 0^\circ$, large angles of attack, and when the model was at the aft tunnel station. It was suspected that wall interference could be causing the anomaly, so the Schlieren photographs were examined. In addition, a high-speed Schlieren video was taken of these suspect situations. Figure 11.7 is a Schlieren photograph of the model at $\alpha = 17.9^\circ$, $\phi = 0^\circ$, $\delta = 20^\circ$, at the aft tunnel station (Run 82). This figure graphically shows upstream interference in the form of a shock wave from the lower wall of the test section impinging on the underside of the model.

After various analyses and tests, it was concluded that the flow mechanism for wall interference upstream of the model was by way of the window cavity. The window cavity

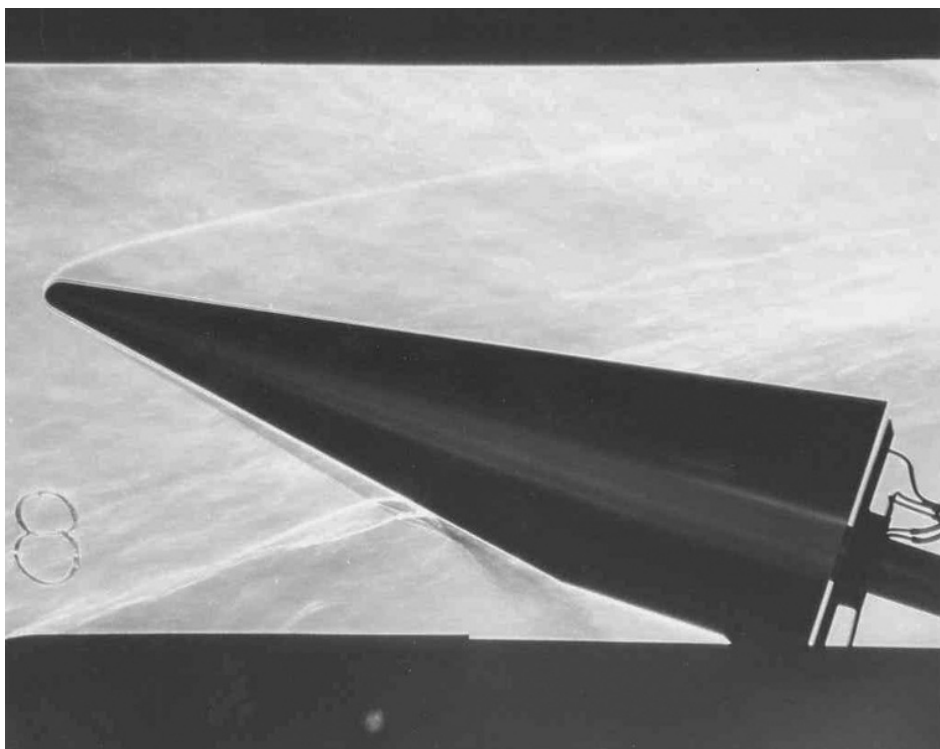


Figure 11.7 Schlieren photograph showing the shock wave from the window cavity impinging on the lower side of the model (Oberkampf *et al.*, 1993).

is the volume formed between the inside surface of the flat windows and the circular cross section of the wind tunnel. At large angle of attack and large flap deflection, the strong shock wave from the flap merged with the bow shock wave and impinged on the window cavity. The high pressure from the shock wave was fed forward in the subsonic separated flow of the window cavity. This, in turn, increased the angle of the free shear layer over the open cavity, which caused a shock wave to be generated at the *front* of the cavity. This shock wave then impinged on the bottom of the model, as seen in Figure 11.7. This phenomenon was exacerbated at aft tunnel stations because the model was significantly aft of the center of rotation of the arc sector mechanism. This caused the model to translate farther downward as the angle of attack increased, bringing it closer to the window cavity.

Methods were investigated to determine at what angle of attack wall interference began to affect the F&M data for any configuration and tunnel station. Using the Schlieren photographs for each angle of attack, we were able to identify most of the cases of when interference occurred. The most sensitive measure was found to be the trend of base pressure with angle of attack. When the wall interference shock wave began to impinge on the separated flow in the base of the model, the base pressure would begin to rise uncharacteristically. The base pressure for each orifice was plotted versus angle of attack for every wind tunnel run. When the base pressure began to abruptly increase with angle

of attack, it was concluded that the shock wave was near the base of the model. Although this in itself does not contaminate the data because the base pressure was eliminated from the data regardless of its value, it was decided not to use any F&M data for higher angle of attack for that particular run. Once this wall interference was recognized, the windows were removed and metal blanks were inserted that conformed to the circular inner surface of the test section. This removed any feature of the window with regard to the flow inside the test section. For the pressure experiment, the window blanks were installed for the entire experiment.

11.3 Example of estimation of experimental measurement uncertainties in JCEAP

Techniques for estimating experimental measurement uncertainty can be grouped into (a) the ISO/ANSI approach described in the *ISO Guide to the Expression of Uncertainty in Measurement* (GUM) (ISO, 1995, 2008) and the *US Guide to the Expression of Uncertainty in Measurement* published by the American National Standards Institute (ANSI, 1997), and (b) the statistical design of experiment (DOE) method (Montgomery, 2000; Box *et al.*, 2005; Hinkelmann and Kempthorne, 2008). Each method has strengths and weaknesses and situations where one has advantages over the other. Given the centuries that measurement uncertainty has been studied, one would think that the foundational issues have been established. Not so! There has been a major international debate within the metrology community for at least the last three decades concerning fundamental terminology and concepts. Although the ISO/ANSI technique is internationally sanctioned, there are a number of well-founded criticisms of the terminology and procedures. Four recent monographs have captured most of the criticisms of the ISO/ANSI approach (Grabe, 2005; Rabinovich, 2005; Drosig, 2007; Salicone, 2007). It is beyond our scope to discuss these criticisms, other than to say we agree with most of them. Our goal here is to discuss in some detail the DOE method and how it was effectively used in the JCEAP validation experiment. Our experience clearly demonstrated that the DOE method gave a much more representative estimate of experimental measurement uncertainty than the ISO/ANSI method. Given the focus on accurate estimation of measurement uncertainty in validation experiments, this is a major finding.

11.3.1 Random and systematic uncertainties

As discussed in Chapter 2, Fundamental concepts and terminology, we will use the terminology of Grabe (2005); Rabinovich (2005); and Drosig (2007):

Error in a quantity: a deviation from the true value of the quantity.

Referring to Eq. (2.2), we have

$$\varepsilon_m = y_m - y_T, \quad (11.1)$$

where (for our interest here) ε_m is the error in the measured quantity y , y_m is the measured value of the quantity, and y_T is the true value of the measured quantity. Here we assume y_T is a scalar quantity as opposed to a random variable. For a measured quantity, y_T is *never* known; it is only estimated (unless a reference standard is available). Similarly, ε_m is *never* known for a measured quantity, only estimated. Any estimate of ε_m is referred to as an *uncertainty estimate*, as opposed to an error estimate. Let the estimation of ε_m be given as ε . The estimation ε is always considered a positive quantity, since the sign of ε_m cannot be known. As a result, ε_m is estimated by the interval $[-\varepsilon, +\varepsilon]$. Whether the estimate is accurate or not, does not change the fundamental concepts or notation. For example, ε_m may or may not be in the interval $[-\varepsilon, +\varepsilon]$, depending on the accuracy of the estimate.

The traditional concept in metrology is to segregate contributors to uncertainty into two types according to their property: random and systematic (Grabe, 2005; Rabinovich, 2005; Drosig, 2007). The following definitions are central to metrology.

Random uncertainties: inaccuracies of a measurement that, in the course of multiple measurements of the same quantity under the same conditions, vary in an unpredictable way.

Systematic uncertainties: inaccuracies of a measurement that, in the course of multiple measurements of the same quantity under the same conditions, remain constant or vary in a predictable way.

Random and systematic uncertainties are *always* present in any measurement. Whether they are large relative to the magnitude of the true value, y_T , or which one dominates in magnitude, depend on many details of the measurement process.

The characterization of the random uncertainty requires that multiple measurements be made. This is usually referred to as *replication* of the experimental measurement. For example, suppose a fixed mass is being weighed on a mechanical scale with a dial indicator. Each time the mass is weighed, there is a random component to the measurement uncertainty due to the mechanical mechanism, laboratory environment conditions, as well as to the person reading the indicator. If one only weighed the mass once, but had several people read the indicator, this is not actually a replication of the experiment. If one person made multiple measurements using the same instrument, using the same procedure, in the same environment, this is sometimes referred to as a 0th order replication (Moffat, 1988; Coleman and Steele, 1999). These references suggest higher orders of replication are possible.

To quantitatively deal with these issues, certain definitions and concepts from probability and statistics need to be introduced. Let $y_i, i = 1, 2, \dots, n$, be the individual values obtained in a measurement process. There is no assumption made concerning how these individual measurements were made, e.g., they could be made using the same instrument or different instruments, by the same laboratory or different laboratories, or over a short period of time or a long period of time. Let \bar{y} be defined as the sample mean of these measurements,

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (11.2)$$

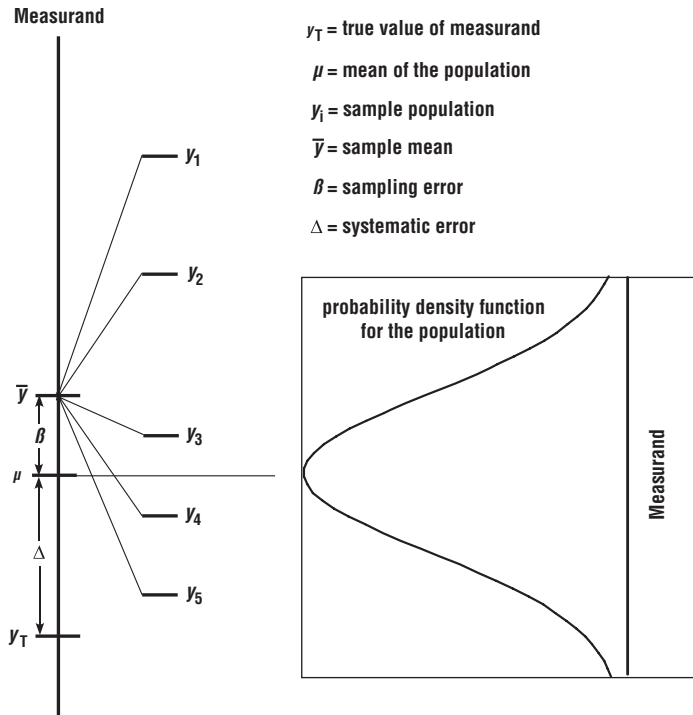


Figure 11.8 Random and systematic contributors to experimental measurement error.

The individual y_i s are assumed to be independent draws from the parent population. The mean of the parent population, μ , is defined as

$$\mu = \frac{1}{N} \sum_{i=1}^N y_i, \quad \text{where } N \text{ is countably infinite.} \quad (11.3)$$

Figure 11.8 depicts the various contributors to measurement error. Let β be defined as the error in \bar{y} , so that we have

$$\beta = \bar{y} - \mu. \quad (11.4)$$

Note that here we can refer to the contributors as *errors* because we are defining the quantities with respect to their true, but unknown, values. β is the error in \bar{y} due to taking a finite number of samples from the parent population. β is a random variable and we will refer to it as the *sampling error* for any set of n measurements.

Let Δ be defined as the *systematic error* in the experimental measurement, so that we have

$$\Delta = \mu - y_T. \quad (11.5)$$

The fundamental error equation can be written as

$$\beta + \Delta = \bar{y} - y_T. \quad (11.6)$$

The only quantity in Eq. (11.6) that is known is \bar{y} . Using DOE techniques we attempt to estimate β and Δ so that we can estimate y_T . Even though β and Δ are convolved in every measurement made, we know that $\beta \rightarrow 0$ as $n \rightarrow \infty$. Estimating Δ is much more difficult because we generally have poor knowledge of its possible causes. The exception to this, of course, is when we have a reference standard so that we know y_T . It should be noted that the ISO/ANSI technique does *not* make a distinction between the true value, y_T , and μ (Grabe, 2005). Quoting the ISO/ANSI Guide “the terms ‘value of a measurand’ (or a quantity) and ‘true value of a measurand’ (for a quantity) are viewed as equivalent.”

If the probability distribution for the random uncertainty is given by, or assumed to be, a normal distribution, then it is well known that a statistical confidence interval can be derived for β (Devore, 2007).

$$|\beta| \leq t_{\alpha/2, v} \frac{s}{\sqrt{n}}, \quad (11.7)$$

where $t_{\alpha/2, n-1}$ is the $1 - \alpha/2$ quantile of the t distribution for $n - 1$ degrees of freedom, and s is the sample standard deviation. s is given by

$$s = \left[\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2}. \quad (11.8)$$

Any level of confidence can be chosen, which is given by $100(1 - \alpha)\%$. The higher the confidence level chosen, the wider the bounds on $|\beta|$ because $t_{\alpha/2, n-1}$ increases for a fixed n . For n greater than 16, the cumulative t distribution and the cumulative standard normal distribution differ by less than 0.01 for all quantiles. In the limit as $n \rightarrow \infty$, the t distribution approaches the standard normal distribution.

The real difficulty in experimental measurement uncertainty is estimating the systematic (bias) uncertainty, Δ . The two strategies for dealing with systematic uncertainties are (a) to identify and attempt to reduce them, and (b) to estimate their magnitude. Systematic uncertainty is an epistemic uncertainty because it is due to lack of knowledge. Depending on the situation, it could be viewed as either a recognized epistemic uncertainty or a blind epistemic uncertainty. Recalling the definitions given in Section 2.3.2, we have

Recognized epistemic uncertainty: an epistemic uncertainty for which a conscious decision has been made to either characterize or deal with it in some way, or to ignore it for practical reasons.

Blind epistemic uncertainty: an epistemic uncertainty for which it is not recognized that the knowledge is incomplete and that the knowledge is relevant to modeling the system of interest.

In either case, adding knowledge to the measurement process (e.g., by using higher accuracy equipment or using different high quality measurement laboratories) can decrease or better quantify their magnitude.

The traditional emphasis and belief are that with very accurate reference standards and with careful calibration and laboratory procedures, systematic uncertainties would become insignificant. For simple measurement processes of the basic physical characteristics of a manufactured product, this is generally the case. However, for most measurements in engineering and the physical sciences, systematic uncertainties are *not* insignificant. In fact, they are commonly the *dominant* contributor to uncertainty because no reference standards are directly applicable. This situation is disconcerting to many experimentalists and not believed by some. However, it has been found in many situations to be the case and only rarely is it openly discussed (Youden, 1972; Morgan and Henrion, 1990). One of the reasons is that more creative designs for sampling must be devised in order to use the power of classical statistics in dealing with systematic uncertainties.

The sampling concepts in DOE methods can be employed to attack systematic uncertainties (Montgomery, 2000; Box *et al.*, 2005; Hinkelmann and Kempthorne, 2008). This is done by using randomization and blocking techniques in the design of the samples taken in an experiment. *Randomization* is designing the samples so that all factors that could possibly affect the measurement result are randomly included in the samples. *Blocking* is a design technique used to either: (a) improve the precision with which comparisons among the factors of interest can be made, or (b) identify the individual factors that contribute to uncertainty. In essence, these techniques use specially designed sampling to convert systematic uncertainties into random uncertainties. Then they can be estimated with traditional statistical techniques. As a result, when systematic uncertainties are quantified in this way they are referred to as *correlated systematic, or bias, uncertainties*, meaning that they are uncertainties that are correlated with specific sources identified through random sampling. Note that in attempting to identify systematic uncertainty sources, the random, or background, uncertainty is also present. The random uncertainty can be segregated from the systematic sources using blocking.

Using randomization to help quantify systematic uncertainties means devising different experiments so that one can expose the factors that could contribute to systematic uncertainty. There are many different kinds of factors that come into play in a complex experiment. Here we list a number of different types of factors and several examples.

- *Experimental instrumentation*: unknown bias in a sensor, measuring device, or transducer; hysteresis in a strain-gauge; shift in zero-reading due to thermal, pressure, acceleration, or radiation sensitivity; frequency response of a complete measuring system; unknown electrical ground loop; misalignment or miscalibration of optical equipment; drifting time reference; drifting frequency reference.
- *Experimental procedures*: calibration procedure; preparation procedure; measuring procedure.
- *Experimental hardware*: imperfections or asymmetry in test hardware; assembly of test hardware; installation of test hardware in the facility.
- *Facility characteristics*: preparation of the facility for a test; calibration of the facility; nonuniformity or asymmetry in facility hardware or characteristics.
- *Data recording and reduction*: analog measurement; analog-to-digital conversion; data reduction procedure such as a programming error in the software.

- *Experimental personnel*: individual technique in human controlled instruments or piece of equipment, such as individual operators of the equipment; individual technique in assembly of the test hardware or facility preparation.
- *Time of experiment*: dependence on time of day, day of week, week of month, or month of year.
- *Weather conditions*: dependence on atmospheric conditions including temperature, pressure, winds, humidity, dust, sunshine, and clouds.

As can be seen from the types of factors and examples shown above, the list of possible sources of systematic uncertainty is surprisingly long. After an experimentalist compiles a list of all of the possible sources for their experiment, he/she must attempt to determine which ones may be important and which ones can be varied so that their effect can be randomly sampled. When considering which ones may be important, one must resolutely keep an open mind. The common attitude that must be overcome is “We’ve checked all of these before and everything is fine.” The proper attitude is “What can possibly be changed so that we can compare our existing results with a new technique or procedure?” These ideas must, of course, be balanced with (a) the options that are possible within a facility; (b) whether qualified personnel needed to implement a new technique are available; (c) costs for new equipment, personnel, and their training; and (d) time needed for the additional experiments to be conducted.

Sometimes it may be impossible to make any significant changes in techniques or procedures within a facility. For this situation, one should consider using a different experimental facility. There is no better sampling technique than to conduct blind test comparisons using different facilities. Using different facilities, particularly if they have entirely different instrumentation techniques, addresses most of the types of factor discussed above, thereby assessing many of the most troublesome systematic uncertainties. If one is conducting relatively standard tests in commercial facilities, then there is typically a wide range of options for conducting the experiment. If one is using specialized facilities or one-of-a-kind facilities, then one has few options available.

11.3.2 Example of DOE procedure for JCEAP force and moment experiment

The following discussion is taken from Oberkampf and Aeschliman (1992); Oberkampf *et al.* (1993); and Aeschliman and Oberkampf (1998) dealing with the force and moment (F&M) measurement portion of the JCEAP experiment. The F&M quantities that were measured on the configuration were the normal force, the pitch moment, the axial center of pressure, and the forebody axial force. The forebody axial force is the total axial force measured on the model using the six-component strain-gauge balance, minus the axial force due to pressure on the base of the model. See these references for more details.

11.3.2.1 DOE principles

In the F&M experiment it was decided to identify only two blocks of measurements over which randomization of experimental runs would take place. The first block contains a

mixture of random uncertainties and systematic uncertainties. The random uncertainties are those that would normally be considered to contribute to variability in the measurement of F&M. The systematic uncertainties are those that would only be found through deliberate intent to sample their effect. Here we simply list them together because, depending on your perspective, they could be segregated differently:

- run-to-run variation in the F&M strain gauge output;
- hysteresis, nonlinearity, thermal sensitivity shift, and thermal zero shift in the F&M strain gauge output;
- run-to-run variation in base pressure transducers and instrumentation for eliminating the base drag;
- analog data-reduction system and analog-to-digital conversion;
- digital data recording system;
- model pitch, roll, and yaw alignment in the test section using different qualified technicians to make the measurements;
- imperfections and nonsymmetric features in the model geometry;
- assembly and reassembly of the model using different qualified technicians;
- run-to-run variation due to facility operating conditions over days, weeks, and months;
- run-to-run variation due to atmospheric condition effects on the facility;
- run-to-run variation in the wind tunnel operator setting free-stream conditions in the test section using different qualified operators.

The combination of all of these uncertainties is commonly referred to as end-to-end random uncertainty in the experimental result. Here we will refer to them simply as *random*.

The second block of samples was directed at identifying the systematic uncertainty due to nonuniformity of the flowfield in the test section of the wind tunnel. It has been long recognized that flowfield nonuniformity, or flowfield quality, is a significant contributor to uncertainty in wind tunnel measurements, but there has been only limited success to quantify this contributor (AGARD, 1994; AIAA, 1999). Flowfield nonuniformity in the test section of any type of wind tunnel can be caused by many sources. Here we list a few sources that are primarily associated with supersonic wind tunnels:

- poorly designed wall contours ahead of the test section, including the contraction section ahead of the nozzle throat, the nozzle expansion region, and the test section region;
- turbulence generated from the heater section of the wind tunnel ahead of the contraction section and the nozzle throat;
- poorly manufactured wind tunnel sections and their assembly for an experiment;
- operation of a fixed nozzle wall wind tunnel at a Reynolds number different from the design condition;
- slight changes in the location of boundary layer transition on the walls upstream or downstream of the nozzle throat due to slight variability in the wall temperature, acoustic environment, upstream flow control valves, and vibration of the walls.

All of the systematic uncertainties associated with the flowfield nonuniformity component will be referred to simply as *flowfield*.

The approach used in JCEAP was to devise a sampling procedure to test the presumption of a perfectly uniform free-stream flowfield in the test section. The fundamental idea is the following: what is the increase in uncertainty, beyond the random uncertainty in F&M measurements, due to testing the model at different locations within the test section? Stated differently, by placing the model at different locations in the test section and comparing this uncertainty to the uncertainty obtained from just testing at one location, what is the increase in uncertainty on the final measured quantities of interest? The number of possible locations for comparisons is greatly expanded by recognizing that the F&M measured at a positive angle of attack and $\phi = 0^\circ$ should be the same for a negative angle of attack of the same magnitude and $\phi = 180^\circ$, if the flowfield is perfectly uniform. Because of physical and mechanical limitations of the test section hardware, there were a limited number of locations that could be attained. All of these locations are within regions of the test section where the flowfield is considered *acceptable* for testing.

11.3.2.2 DOE analysis and results

Using the technique of blocking, comparisons are made between different wind tunnel runs and portions of different runs to segregate random and flowfield uncertainty contributors. To calculate the random uncertainty, one compares all possible combinations of F&M measurements that were made for the same physical location in the test section. In this way one can estimate the complete end-to-end random measurement uncertainty. By examining the run number schedule (Table 11.1) one can choose run pairs that have the same roll angle ϕ , the same flap deflection angle δ , and the same location in the test section. Then comparisons can be made between the same F&M quantities for the same angles of attack. The total number of run pairs that can be formed for quantifying the random uncertainty is 14. All of these run pairs are shown in the second column of Table 11.3. The number of angle of attack pairs for each run pair is shown in the third column of the table. The total number of angle of attack pairs constructed in this way is 160 for each F&M quantity.

To calculate the additional uncertainty contribution due to the flowfield, one examines Table 11.1 to determine which run pairs place the model at different locations in the test section. One finds two ways of attaining different locations. The first method forms run pairs that have the same ϕ and δ , but are at different axial stations in the test section. All of the run pairs that can be constructed in this way are shown in the fourth column in Table 11.3. The total number of run pairs that can be formed in this way is 20. Also shown in the fifth column of the table is the number of angles of attack that are in common for each run pair. The total number of angle of attack pairs constructed in this way is 220 for each F&M quantity.

The second method forms run pairs based on mirror symmetry between the model at a roll angle of 0° and pitched to a positive α , and the model at $\phi = 180^\circ$ and pitched to negative α , both conditions for the same δ . The mirror symmetry pairs are formed for both runs at the same axial station. As the maximum negative angle of attack was -10° , the residuals for individual angles of attack can only be calculated over the range -10° to

Table 11.3 Run pairs for F&M uncertainty analysis (Oberkampfer et al., 1993).

Pair number	Random		Axial location		Mirror symmetry	
	Run pair	No. of α s	Run pair	No. of α s	Run pair	No. of α s
1	(34, 36)	12	(34, 74)	12	(34, 40)	8
2	(34, 37)	12	(34, 75)	12	(36, 40)	8
3	(34, 73)	12	(36, 74)	12	(37, 40)	8
4	(36, 37)	12	(36, 75)	12	(73, 40)	8
5	(36, 73)	12	(37, 74)	12	(63, 55)	8
6	(37, 73)	12	(37, 75)	12	(72, 55)	8
7	(63, 72)	12	(73, 45)	12	(64, 56)	8
8	(64, 71)	11	(73, 75)	12	(64, 57)	8
9	(65, 70)	10	(63, 83)	11	(71, 56)	8
10	(66, 67)	11	(72, 83)	11	(71, 57)	8
11	(56, 57)	11	(64, 82)	10	(65, 58)	8
12	(59, 60)	11	(71, 82)	12	(70, 58)	8
13	(74, 75)	12	(65, 81)	10	(74, 76)	8
14	(76, 77)	10	(70, 81)	10	(74, 77)	8
15	—	—	(40, 76)	10	(75, 76)	8
16	—	—	(40, 77)	10	(75, 77)	8
17	—	—	(55, 78)	10	(83, 78)	8
18	—	—	(56, 79)	10	(82, 79)	8
19	—	—	(57, 79)	10	(81, 80)	8
20	—	—	(58, 80)	10	—	—

+10°. All run pairs that can be constructed for estimation of this uncertainty component are shown in sixth column of Table 11.3. The total number of run pairs that can be formed in this way is 19. Using interpolation for the angle of attack range to obtain four positive and four negative angles of attack, the seventh column in the table shows the total number of angles of attack that are in common for each run pair. The total number of angle of attack pairs constructed in this way is 152 for each F&M quantity.

Now that it is seen what types of different run pairs are needed for the comparisons shown in Table 11.3, some comments are needed concerning how the run schedule (Table 11.1) was constructed before the experiment. First, replicate runs are *not* afterthoughts in the design of randomization and blocking techniques. Careful thought must be devoted to the construction of the run schedule before the experiment begins to achieve the needed samples. Second, replicate runs should be done as far apart in the run schedule as possible so as to incorporate the most changes of all of the factors discussed above. For example, replicate runs for conditions associated with runs 63 and 72 are better than replicate runs for conditions associated with runs 66 and 67. Runs 66 and 67 were conducted on the same day, whereas runs 63 and 72 were conducted weeks apart.

For each run pair an average value, $(\bar{})$, of the measured F&M quantity can be defined as

$$\begin{aligned}(\bar{C}_n)_{\alpha_i} &= \left\{ [(C_n)_p + (C_n)_q]_{\alpha_i} \right\} / 2, \quad i = 1, 2, \dots, I, \\(\bar{C}_m)_{\alpha_i} &= \left\{ [(C_m)_p + (C_m)_q]_{\alpha_i} \right\} / 2, \quad i = 1, 2, \dots, I, \\(\bar{x}_{cp})_{\alpha_i} &= \left\{ [(x_{cp})_p + (x_{cp})_q]_{\alpha_i} \right\} / 2, \quad i = 1, 2, \dots, I, \\(\bar{C}_a)_{\alpha_i} &= \left\{ [(C_a)_p + (C_a)_q]_{\alpha_i} \right\} / 2, \quad i = 1, 2, \dots, I.\end{aligned}\quad (11.9)$$

C_n , C_m , x_{cp} and C_a are the normal force coefficient, pitch moment coefficient, axial center of pressure, and forebody axial force coefficient, respectively. p and q represent the run numbers from which measurements are taken, α_i refers to the angle of attack at which each F&M measurement was made, and I is the maximum number of angles of attack which are in common for both run pairs.

The pitch sector mechanism in the wind tunnel is programmed to yield a specified pitch angle according to the digital counter, given an initial starting angle. It lacks, however, a feedback control system for the pitch angle desired. As a result, the pitch increments programmed are not accurately repeatable from run to run because of slight variability in friction in the mechanism. As a result, the nonrepeatability in pitch angles ranges up to 0.5° . The actual pitch angle attained, however, is known to within $\pm 0.02^\circ$ based on the digital counter output. Therefore, in the average quantities calculated for each run pair in Eq. (11.9), it would be inappropriate to assign to measurement uncertainty the fact that identical pitch angles were not obtained from run to run. That is, the inaccuracy is not in the data itself, but caused by a requirement of the uncertainty analysis procedure. This inaccuracy can be eliminated by using a least squares fit of each F&M quantity for each pitch sequence for each run. Using these data fits, it is then possible to calculate the F&M measurements at precisely the same angle of attack for each run pair.

Let the difference between an individual F&M quantity and the average measurement at each angle of attack be defined as a local residual, so that

$$\begin{aligned}(\Delta C_n)_{\alpha_i} &= (C_n)_{\alpha_i} - (\bar{C}_n)_{\alpha_i}, \quad i = 1, 2, \dots, I, \\(\Delta C_m)_{\alpha_i} &= (C_m)_{\alpha_i} - (\bar{C}_m)_{\alpha_i}, \quad i = 1, 2, \dots, I, \\(\Delta x_{cp})_{\alpha_i} &= (x_{cp})_{\alpha_i} - (\bar{x}_{cp})_{\alpha_i}, \quad i = 1, 2, \dots, I, \\(\Delta C_a)_{\alpha_i} &= (C_a)_{\alpha_i} - (\bar{C}_a)_{\alpha_i}, \quad i = 1, 2, \dots, I.\end{aligned}\quad (11.10)$$

The total number of random residuals, i.e., measurements at each pitch angle (based on the second and third columns in Table 11.3) is 320 for each F&M quantity. Because of the definition for the run pair average given in Eq. (11.9), 160 of the residuals are positive and 160 are negative, i.e., mirror images. Based on the fourth and fifth columns in Table 11.3, the total number of axial location residuals was 440. Based on the sixth and seventh columns in Table 11.3, the total number of mirror symmetry residuals was 304. The total number of residuals for the F&M experiment was 1064.

Table 11.4 Summary of results for F&M uncertainty analysis (Oberkampf and Aeschliman, 1992).

Uncertainty component	C_n		C_m		x_{cp}/L		C_a	
	$\hat{\sigma}$	%	$\hat{\sigma}$	%	$\hat{\sigma}$	%	$\hat{\sigma}$	%
Random	0.474×10^{-3}	20	0.406×10^{-3}	19	0.413×10^{-3}	9	0.426×10^{-3}	63
Flowfield	0.941×10^{-3}	80	0.851×10^{-3}	81	1.322×10^{-3}	91	0.324×10^{-3}	37
Total	1.054×10^{-3}	100	0.943×10^{-3}	100	1.385×10^{-3}	100	0.535×10^{-3}	100

The sample variance for the random component and also for the total experiment (random and flowfield components combined) is calculated using the equation

$$\hat{\sigma}^2 = \frac{1}{2m} \sum_{j=1}^m (\Delta_1^2 + \Delta_2^2)_j, \quad (11.11)$$

where m is the number of local residuals and Δ_1 and Δ_2 are the residuals for each run pair from Eq. (11.10). (Note that the sample variance is an estimate of a population variance and is indicated by $\hat{\cdot}$.) Because of the independence between random uncertainty and flowfield uncertainty, the sum of the variances can be written

$$\hat{\sigma}_{\text{total}}^2 = \hat{\sigma}_{\text{random}}^2 + \hat{\sigma}_{\text{flowfield}}^2. \quad (11.12)$$

The sample variance due to the random component and that for the total experiment can each be computed separately. As a result, the component due to the flowfield can be found from

$$\hat{\sigma}_{\text{flowfield}} = \sqrt{\hat{\sigma}_{\text{total}}^2 - \hat{\sigma}_{\text{random}}^2}. \quad (11.13)$$

Table 11.4 gives the estimated standard deviation $\hat{\sigma}$ (square root of the variance estimate) due to random uncertainty, flowfield uncertainty, and the total for each F&M quantity in the experiment. Also shown in the table is the percent contribution due to each of the components identified. In the calculation for the standard deviation for x_{cp}/L , all residuals for angle of attack 0° were excluded. This was done because it is well known that the uncertainty in x_{cp} becomes indefinite as the normal force approaches zero. If the model geometry were axisymmetric, then the standard wind tunnel procedure of computing the ratio of $C_{m_\alpha}/C_{n_\alpha}$ for x_{cp} would be appropriate. For the present nonsymmetric geometry, however, this method is not applicable because it is not known at what angle of attack, if any, both C_m and C_n attain zero. Except for the case of calculating x_{cp} for $\alpha = 0$, all experimental F&M measurements were included in the results shown in Table 11.4, i.e., no measurement outliers were discarded in the entire uncertainty analysis.

From Table 11.4 it can be seen that the random uncertainty in C_n , C_m , and x_{cp} due to the entire wind tunnel system ranges from 9 to 20%, whereas that due to the flowfield

nonuniformity is 80 to 91%. The random component of uncertainty in C_a is 63% and that due to the flowfield is 37%. This reversal of uncertainty contributions compared to the other quantities is believed to be due to (a) the axial force being very *insensitive* to flowfield nonuniformity because it is the sum of the normal stress and shear stress acting along the axis of the body and (b) the difficulty in removing the base pressure component from the total axial force.

Although this type of uncertainty analysis is new to wind tunnel uncertainty estimation, the JCEAP F&M experiment clearly demonstrates that, for most measured quantities, the dominant contributor is the systematic uncertainty due to the flowfield nonuniformity; *not* the random uncertainty. The long established procedure for wind tunnel uncertainty estimation is based on the ISO/ANSI practice, which, as is well known, is focused on estimating the random component of uncertainty, *not* the flowfield uncertainty. Some will quickly jump to the conclusion that the obvious problem is the poor quality flow in the Sandia hypersonic wind tunnel. As will be seen in the discussion of the JCEAP pressure experiment in the next section, this is *not* the case.

11.3.3 Example of DOE procedure for JCEAP surface pressure experiment

The following discussion is taken from Oberkampf *et al.* (1995) and Aeschliman and Oberkampf (1998), dealing with surface pressure measurements in the JCEAP experiment. See these references for more details. The DOE technique was applied to the surface pressures measured by the two pressure modules, but not the high frequency measurements from the Kulite gauges.

11.3.3.1 DOE principles

In the JCEAP pressure experiment we quantified three contributors to measurement uncertainty. Two of the three contributors were discussed earlier in the F&M experiment: end-to-end random uncertainty and flowfield nonuniformity. The random component for the pressure experiment is nearly the same, except now it includes uncertainties associated with the surface pressure measurement system instead of those associated with the F&M measurement system. The flowfield component for the pressure experiment is the same as discussed earlier. Some examples of the pressure measurement contributors that will be grouped into random uncertainty are:

- run-to-run variation in the individual pressure transducers within each pressure module;
- hysteresis, nonlinearity, thermal sensitivity shift, and thermal zero shift in the pressure modules;
- run-to-run variation in the reference pressure that is fed to the pressure modules;
- analog data reduction system and analog-to-digital conversion;
- digital data recording system;
- model pitch, roll, and yaw alignment in the test section using different qualified technicians to make the measurements;
- run-to-run variation due to facility operating conditions over days, weeks, and months;

- run-to-run variation due to atmospheric condition effects on the facility;
- run-to-run variation in the wind tunnel operator setting free-stream conditions in the test section using different qualified operators.

In addition to the random and flowfield components, local measurements (such as surface pressure and shear stress) allow an additional component of systematic uncertainty to be identified. This component is the uncertainty due to (a) imperfections or asymmetry in the wind tunnel model, and (b) imperfections due to the individual pressure sensor in each module, their connecting tubing, and the associated orifice on the surface of the model. For model and pressure sensor imperfections to be identified, however, the model must have at least two planes of mirror symmetry. All of the systematic uncertainties associated with the model and pressure sensor imperfections will be referred to simply as *model*.

Examples of some of the sources that fall into the model component are:

- model fabrication deficiencies that cause the loss of mirror symmetry of the surface contour of the wind tunnel model;
- accidental damage to the model, for example during wind tunnel testing or handling, that causes the loss of mirror symmetry of the surface contour of the model;
- any deflection or distortion of the axis of the model, such as due to fabrication or aerodynamic or aerothermal heating of the model;
- any deflection or distortion of the model surface or lifting surfaces that causes the loss of mirror symmetry of the surface contour of the model, such as due to fabrication or aerodynamic or aerothermal heating of the model;
- a poorly fabricated pressure orifice, such as one containing a burr from machining;
- accidental damage to a pressure orifice, such as a damaged edge near the model surface that occurred any time after the model was fabricated;
- a pressure leak occurring anywhere between the pressure orifice and the connection of the tubing to the pressure module;
- any blockage or a kink in the connecting tubing anywhere between the pressure orifice and the pressure module.

Using blocking techniques, we are able to identify three components contributing to measurement uncertainty: random, flowfield, and model. We will briefly summarize the blocking strategy and discuss each one in more detail.

- Random: compare pairs of measurements that have the same physical location in the test section, the same pressure port on the model, and exposure to the same induced flowfield of the model.
- Flowfield: compare pairs of measurements that have the same pressure port on the model, exposure to the same induced flowfield of the model, but are at different physical locations in the test section.
- Model: compare pairs of measurements that have same physical location in the test section, exposure to the same induced flowfield of the model, but are different pressure ports on the model.

To calculate the random uncertainty, one compares pressure measurements from different runs for the same: location in the test section, pressure port, angle of attack, roll angle, and flap deflection angle. With these restrictions, each pair of ports compared will have the same location in the vehicle-induced flowfield. When differences in pressure port measurements

are computed, the uncertainty due to flowfield and model cancels out in computing the residual.

To calculate the flowfield uncertainty, a two-step process is needed. First, one compares pressure measurements from the same pressure port and exposure to the same induced flowfield, but at different locations in the test section. The same induced flowfield can be generated by having the same angle of attack, roll angle, and flap deflection angle, but it can also be generated by using mirror symmetry features as discussed with regard to the F&M uncertainty analysis. When the differences in pressure port measurements are computed, they will be a combination of the random and flowfield components, but the component due to the model cancels out in computing the residual. The second step is to subtract the variance due to the random component from the combined variance due to the random and flowfield components, leaving only the contribution due to the flowfield.

To calculate the model uncertainty, a similar two-step process is needed. First, one compares pressure measurements from the same physical location in the test section and exposure to the same induced flowfield, but using different pressure ports. Given these constraints, the only possibility for comparisons is to have two or more planes of mirror symmetry of the model geometry. For the JCEAP geometry, it only has one plane of symmetry aft of the beginning of the slice. As a result, comparisons of ports on JCEAP can only be made for ports on the conical portion of the geometry. This symmetry feature is also common with most aircraft geometries. For four-finned missile geometries, however, there are four planes of symmetry over the fin region so more comparisons could be made.

For JCEAP, we can only compare pressure ports that are sufficiently upstream of the slice on the geometry. In addition, we should only use data from deflected flap cases where we are confident there is no influence of the deflected flap upstream of the beginning of the slice. The case of $\delta = 0^\circ$ produces no axially separated flow on the model, so there is no concern that there is any influence upstream on the cone surface. For $\delta = 10^\circ$ and 20° , liquid crystal surface flow visualization showed that no axially separated flow progressed ahead of the beginning of the slice. A flap deflection of 30° , however, showed a separated flow influence over the entire slice region so this flap deflection was not used in these comparisons. Therefore, only run pairs with 0 , 10 , and 20° flap deflections were used. To further reduce the possibility of any upstream influence of the slice or flap on the ports used only axial stations up to 15.75 cm were considered. This is 2.73 cm ahead of the slice, or about 10 to 20 boundary layer thicknesses. The second step in the calculation process is to subtract the variance due to the random component from the combined variance due to the random and model components, leaving only the model component.

11.3.3.2 DOE analysis and results

To calculate the random uncertainty, one examines the run schedule (Table 11.2) and chooses run pairs that have the same roll and flap angles and have the same tunnel location. In addition, the run pairs chosen must allow comparison only between the same pressure ports on the model. Examples of run pairs that meet these conditions are (20,22), (24,61), (103,112), (42,43), (124, 126), and (131,133). It can be shown that a total of 29 run pairs

meet the required conditions. Pressure measurements were obtained for a total of 12 angles of attack for each run; nine distinct angles of attack and three measurements at $\alpha = 0$. As a result, there are a total of 18 combinations of α where pressure comparisons can be made (nine distinct α comparisons plus nine permutations of $\alpha = 0$ measurements). Therefore, the total number of pressure port comparisons is

$$(96 \text{ ports}) \times (29 \text{ run pairs}) \times (18 \alpha \text{ pairs}) = 50\,112 \text{ comparisons.}$$

The actual number of comparisons is slightly less than this estimate because some pressure ports were outside the calibrated range for certain conditions. As a result, the total number of pressure port comparisons for random uncertainty was found to be 48 164.

To make these pressure port comparisons, it is required that the α for each of the two runs is identical. If they are not the same, then part of the difference in the two measurements will be due to non-repeatability of α caused by the model pitch mechanism. As mentioned earlier, the pitch control mechanism does not have a feedback control system, so the repeatability of α from one pitch sweep of the mechanism to another can be as large as $\pm 0.5^\circ$. Since it was known that very similar angles of attack were needed for the pressure comparisons, roughly 10 to 20 runs had to be repeated because the resulting α deviated by more than $\pm 0.3^\circ$ from the nominal values. For the final run numbers shown in Table 11.2, α did not deviate from the nominal value by more than $\pm 0.28^\circ$. The average deviation from the nominal value of α for all runs listed in Table 11.2 was 0.057° .

Deviations in α of up to 0.28° from run to run would introduce an unacceptably large error in the estimated surface pressure uncertainty. To minimize this uncertainty in the analysis, all of the pressure measurements were interpolated to the nominal angles of attack. To accomplish this, a cubic spline interpolation with adjustable knot locations was computed for each pressure port as a function of α for each run.

Differences in pressure port measurements were computed with the following technique. Let the pressure measurement for port i and angle of attack j be denoted as $\left(\frac{p_i}{p_\infty}\right)_j^r$, where the superscript denotes the run number r . Then the average pressure of the port for the two runs being compared is given by

$$\left(\frac{\bar{p}_i}{p_\infty}\right)_j^{r,s} = \frac{1}{2} \left[\left(\frac{p_i}{p_\infty}\right)_j^r + \left(\frac{p_i}{p_\infty}\right)_j^s \right], \quad (11.14)$$

where $i = 1, 2, \dots, 96$ and $j = 1, 2, \dots, 18$, where 18 is the total number of α s. Let the absolute value of the difference between a pressure measurement and the average pressure be defined as the residual. Then the residual is given by

$$\left(\frac{\Delta p_i}{p_\infty}\right)_j^{r,s} = \left| \left(\frac{p_i}{p_\infty}\right)_j^r - \left(\frac{\bar{p}_i}{p_\infty}\right)_j^{r,s} \right|. \quad (11.15)$$

Note that the residual can be computed using either the pressure measurement from run r or s .

To calculate the flowfield uncertainty, one examines Table 11.2 and chooses run pair that have the same induced flowfield and the same flap deflection angle, but the model

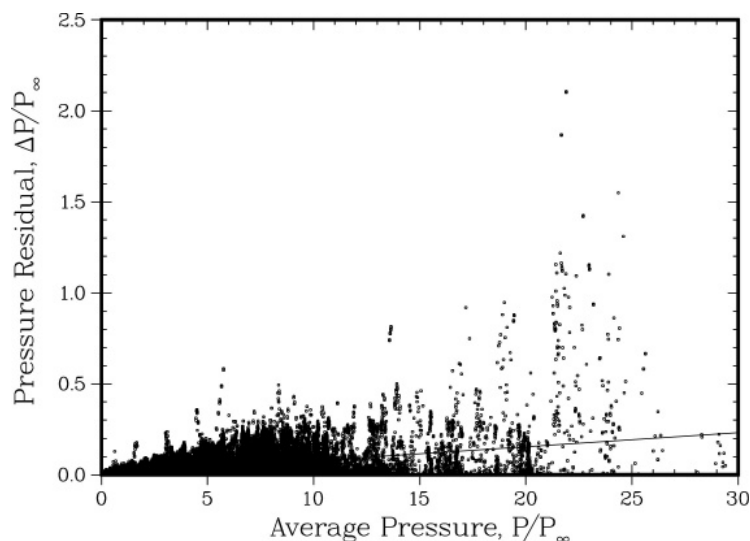


Figure 11.9 All pressure residuals for random, flowfield nonuniformity, and model geometry uncertainty (Oberkampf *et al.*, 1995).

has different locations in the test section. In addition, the run pairs chosen must allow a comparison only between the same pressure ports on the model. One finds four types of run pairs that will produce the types of comparisons desired. These are (a) comparisons between measurements made at different axial locations in the test section, (b) comparisons between different roll angles at $\alpha = 0$ at the same tunnel station, (c) comparisons between positive α with a roll angle of 0° and negative α with a roll angle of 180° , and (d) comparisons between positive α with a roll angle of 90° and negative α with a roll angle of 270° . Examples of run pairs for each of these types of comparison are, respectively, (20,101), (24,32), (35,43), and (46,47).

The total number of pressure port comparisons for these four types, minus the number of comparisons lost due to over-ranged ports, is 101 838 residuals. The residuals for flowfield and random uncertainty are computed by the same equations given above, but the number of angles of attack, j , for each of the types is different.

To calculate the model uncertainty, one examines Table 11.2 for combinations of roll angle, flap deflection angle, both runs at the same tunnel station, the same induced flow field, but different pressure ports are compared. One finds six types of run pairs that meet the required conditions. All of these pairs are formed by comparing different roll angles, the same flap angle, and the same tunnel location. The total number of pressure port comparisons for these six types, minus the number of comparisons lost due to over-ranged ports, is 24 196. The residuals for model and random uncertainty are computed using the same equations given above.

Plotted in Figure 11.9 are all of the residuals computed for random, flowfield, and model uncertainty. In the entire pressure uncertainty analysis no pressure measurement

outliers were discarded. It can be seen from this figure that the magnitude of the uncertainty steadily increases with the magnitude of the pressure measured. This characteristic is typical of nearly all experimental measurements. This trend is accounted for in the residuals by scaling the residuals with the magnitude of measured pressure. A least squares fit of the residuals was computed using a linear function constrained to pass through zero. The resulting fit was computed to be

$$\frac{\Delta p_{ls}}{p_{\infty}} = 0.00875 \frac{p_s}{p_{\infty}}, \quad (11.16)$$

where p_s is the surface pressure measured. This linear fit is also shown in Figure 11.9.

The sample variance is now calculated with each local sample scaled according to the least squares fit given above. The equation for estimating each type of variance, normalized by the least squares fit of the residuals, is given by

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{k=1}^N \left[\frac{(\Delta p / p_{\infty})}{(\Delta p_{ls} / p_{\infty})} \right]_k^2, \quad (11.17)$$

where N is the total number of residuals (i.e., pressure comparisons) and the subscript k indicates the k th residual. The sample standard deviation due to the flowfield and the model can then be calculated from

$$\hat{\sigma}_{\text{flow}} = \sqrt{\hat{\sigma}_{\text{flow+instrumentation}}^2 - \hat{\sigma}_{\text{instrumentation}}^2}, \quad (11.18)$$

$$\hat{\sigma}_{\text{model}} = \sqrt{\hat{\sigma}_{\text{model+instrumentation}}^2 - \hat{\sigma}_{\text{instrumentation}}^2}, \quad (11.19)$$

respectively. The sample standard deviation due to all of the uncertainty sources is given by

$$\hat{\sigma}_{\text{total}} = \sqrt{\hat{\sigma}_{\text{instrumentation}}^2 + \hat{\sigma}_{\text{flow}}^2 + \hat{\sigma}_{\text{model}}^2}. \quad (11.20)$$

Table 11.5 gives the summary statistics for the uncertainty estimates for the pressure measurements. It is seen from the table that the dominant contributor to uncertainty of pressure measurements, as was seen in the F&M measurements, is due to the nonuniformity of the flowfield in the test section. The second most important contributor was the model; the least important was the end-to-end random contribution due to the entire instrumentation system and wind tunnel facility.

The dominant contribution of non-uniform flow to uncertainty suggests the question: is this just characteristic of the present wind tunnel or is it characteristic of other hypersonic wind tunnels? An earlier application of the DOE uncertainty estimation procedure was applied to a similar surface pressure experiment in Tunnels A and B of the von Karman Gas Dynamics Facility at the US Air Force Arnold Engineering Development Center, Tullahoma, Tennessee (Oberkampf *et al.*, 1985). Tunnel A is a variable Mach number facility in the supersonic range and Tunnel B can be operated at Mach 6 and 8. The absolute magnitude (no normalization using the free-stream static pressure) of the present results

Table 11.5 *Summary of results for the surface pressure uncertainty analysis (Oberkampf et al., 1995).*

Component of uncertainty	Number of residuals	Normalized $\hat{\sigma}$	% of total RMS uncertainty
Random	48 164	0.56	12
Flowfield	101 838	1.28	64
Model	24 196	0.79	24
Total	174 198	1.60	100

for flowfield uncertainty was compared to those for Mach 8 in Tunnel B of the von Karman Gas Dynamics Facility (Oberkampf *et al.*, 1995). Hypersonic Tunnel B is recognized world wide for the quality of its flowfield. This comparison showed that the Sandia HWT and Tunnel B are *comparable* in terms of the effect of the flowfield non-uniformity on surface pressure measurements. As an independent check on the results of DOE procedure, it was also found that the random uncertainty computed using the DOE technique compared very closely to that computed by the traditional ISO/ANSI technique by the von Karman Gas Dynamics Facility staff (Oberkampf *et al.*, 1985).

In summary, the DOE approach using randomization and blocking techniques has been used on four different wind tunnel experiments: three of the data sets were for surface pressure measurements and one was for body forces and moments. The method has been applied to three wind tunnels: (a) two hypersonic wind tunnels, Tunnel B of the von Karman Gas Dynamics Facility and the HWT Facility at Sandia National Laboratories; and (b) one supersonic wind tunnel, Tunnel A operating at Mach 3. The method showed that, even in these high-quality flowfield facilities, the largest contributor was systematic uncertainty due to nonuniformity of the flowfield. It was shown that the flowfield nonuniformity was typically four to five times larger than the end-to-end random uncertainty of the facility. We strongly suspect that the largest contribution to experimental uncertainty in most, if not all, hypersonic wind tunnels is due to flowfield nonuniformity. Although this technique has not been applied, to our knowledge, to any other wind tunnel, we suggest the flowfield nonuniformity may also dominate in other facilities. A similar DOE technique has been applied recently to wind tunnel testing, but no quantitative assessments have been made with regard to the magnitude of flowfield nonuniformity to other contributors (DeLoach, 2002, 2003).

The natural question is: why hasn't this procedure been investigated by wind tunnel facilities over the past 14 years when it was published in the open literature? There are number of possible answers, some of which are: (a) natural resistance to new procedures and technology in organizations; (b) users of wind tunnels, i.e., paying customers, are more interested in quantity of data obtained during the experiment than careful uncertainty estimation of data; and (c) facility owners would rather not deal with this risky topic.

We argue that the experimental uncertainty issue, just as issues with code verification, solution verification, and uncertainty quantification, must be faced in the future if increased credibility is to be earned in both experimental measurements and computational simulations. We believe the customers and decision makers using these products must provide the long-term motivation for moving ahead because they are the ones most at risk. They are the ones with the most skin in the game.

11.4 Example of further computational–experimental synergism in JCEAP

Roughly five years after the completion of the JCEAP project, there was renewed interest at Sandia National Laboratories in the experiment (Roy *et al.*, 2000). The interest was primarily directed toward using the pressure data for validation activities related to the Sandia Advanced Code for Compressible Aerothermodynamics Research and Analysis (SACCARA) code (Payne and Walker, 1995; Wong *et al.*, 1995a,b). The SACCARA code evolved from a parallel-distributed memory version of the INCA code (Amtec, 1995), originally developed by Amtec Engineering. The SACCARA code employs a massively parallel-distributed memory architecture based on multi-block structured meshes. The Navier–Stokes equations are solved for conservation of mass, momentum, global energy, and vibrational energy (where applicable) in finite volume form. The viscous terms are discretized using central differences. The SACCARA code has two options for determining the inviscid interface fluxes, the Steger–Warming flux vector splitting scheme (Steger and Warming, 1981) and Yee’s symmetric TVD scheme (Yee, 1987). Second-order spatial accuracy is obtained with the former via MUSCL extrapolation of the primitive variables, while the latter is nominally second order in smooth regions of the flow. Both schemes employ a flux limiter that reduces to first-order spatial accuracy in regions with shock waves.

Initial comparisons were made between the SACCARA code and the JCEAP pressure results for low angles of attack and zero flap deflection. For these attached flow cases, there was high confidence that the predictions should be very accurate. It was found, however, there was larger than expected disagreement. So, improved simulations were computed for a specialized case: flow over the conical portion of the vehicle at zero angle of attack case so that a 2-D axisymmetric flow could be accurately computed (Roy *et al.*, 2000; Roy *et al.*, 2003a,b). Figure 11.10 shows the surface pressure vs. the nondimensional axial length, x/R_N , where R_N is the nose radius of the vehicle. The experimental measurement shown at each axial station is the mean value of all measurements taken during the experiment, along with $\pm 2\sigma$ of the estimated total uncertainty (random, flowfield, and model). The number of individual measurements at each axial station ranged from 48 (on the side of the body opposite to the slice) to 768 (on the slice side of the body at $x/R_N = 16$ and 26). The large number of samples is the reason the $\pm 2\sigma$ values are extremely small, since the uncertainty band decreases as $1/\sqrt{n}$. As can be seen in Figure 11.10, the computed pressures fall consistently below the experimental data, with the poorest agreement occurring just

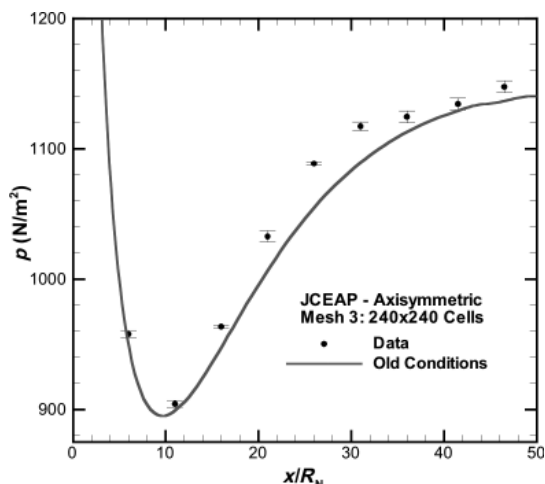


Figure 11.10 SACCARA simulation (curve) compared with the JCEAP pressure data for the surface pressure data on the cone, $\alpha = 0$ (Roy *et al.*, 2003).

upstream of the slice location, which begins at $x/R_N = 36.37$. The maximum difference, occurring at the $x/R_N = 26$ location, is 3.3%; well outside the estimated experimental 2σ uncertainty bounds.

As a result, an investigation was launched into trying to discover the cause of the unexpected disagreement. The first phase of the investigation was an extensive code verification and solution verification effort. These efforts were reported in Roy *et al.* (2000, 2003a), and discussed in Chapters 6–8. The results of the first phase showed no noticeable change in the results shown in Figure 11.10. The second phase of the investigation pursued a number of other possible sources for the disagreement (Roy *et al.*, 2000, 2003b). This second phase of the investigation is briefly reported here; focusing on the strategy that could be used when one is faced with the situation of unexplained disagreement between high confidence simulations and experimental results. This type of investigation almost always yields improved understanding of the simulation and experimental results, as well as unexpected synergism between each approach. See the cited references for the full details.

11.4.1 Assessment of computational submodels

In any modern scientific computing analysis there is a wide range of submodels that are combined to yield the complete physics model for the simulation. In an investigation such as this, it is important to critically examine (a) each submodel to try and determine if the assumptions in the submodel are appropriate for the experiment of interest and (b) evaluate important modeling assumptions of the flowfield to determine their effect. Stated differently, one conducts a quantitative sensitivity analysis with respect to the submodels and numerical approximations to determine if they might explain the disagreement between computation

and experiment. The following submodels and numerical approximations were evaluated and the result is summarized.

11.4.1.1 Transport property submodels

A study of the transport properties of nitrogen was undertaken in order to insure accuracy over the temperature range of interest (50 K to 650 K). Three different models for absolute viscosity were evaluated: Keyes' model, Sutherland's law, and a power law model. Each was evaluated using available experimental data. In critical evaluations, it is highly recommended to plot the estimated *error* in the model as a function of the independent variable, and *not* just the quantity of interest vs. the independent variable. A much more critical perspective can be gained of the model accuracy. Keyes' model was found to be the most accurate over the temperature range with a maximum error of 5% found in the low temperature range.

Keyes' model for the thermal conductivity takes the same form as that for the absolute viscosity; however, the constants used are different. Using the usual constants in Keyes' model for thermal conductivity and comparing with experimental data showed that the model performed poorly at both low and high temperatures. Better agreement with the data is found by simply assuming a constant molecular Prandtl number along with Keyes' model for viscosity. Using this model showed significantly smaller errors of 10% and 40% in the higher and lower temperatures, respectively. While the choice of transport models can have a large impact on the skin friction and heat transfer, the choice of the model was found to have negligible effects on the surface pressure, as would be expected.

11.4.1.2 Equation of state submodel

To test the validity of the ideal gas equation of state, the densities and temperatures from an ideal gas solution were used in an *a posteriori* calculation of the pressure using the more accurate Beattie–Bridgeman equation of state. These pressures were then compared to the ideal gas solution results, with maximum differences found to be less than 0.05% for the entire flowfield. Thus, the ideal gas equation of state was used.

11.4.1.3 Thermodynamic submodel

In order to determine the thermal state of the flow, i.e., vibrational equilibrium versus nonequilibrium in each section of the wind tunnel, calculations were performed for the HWT Mach 8 nozzle. These calculations employed the SACCARA code and assumed fully turbulent boundary layers on the tunnel wall from the heater section, through the contraction section, through the expansion section, and into the test section. This is a reasonable assumption based on the community of operators of hypersonic wind tunnels. The design specifications (pre-fabrication) were used for the geometry definition, as opposed to post-fabrication inspection of the wind tunnel. One difference between the two was that an inspection of the nozzle throat diameter indicated a diameter of 23.01 mm as compared to 22.70 mm in the design specifications. This difference is probably due to slight erosion of

the metal throat through the years of operation. Although this difference is small (1.37%), it could lead to Mach number overprediction by as much as 0.4% based on a simple isentropic analysis. Since a full inspection of the contour of the various sections was not feasible at the time, it was decided to use the design specifications for each section. Three axisymmetric mesh levels were employed in order to ensure mesh convergence of the solution, with the fine mesh having 280×120 cells in the axial and radial directions, respectively.

In order to determine the thermal state of the flow at the test section, the nozzle flow was simulated assuming thermal nonequilibrium using the standard Landau–Teller formulation for vibrational relaxation. Simulation results convincingly showed that the vibrational temperature freezes out very near the plenum stagnation temperature of 633 K, resulting in nonequilibrium flow through the contraction region, throat, and expansion region of the nozzle. This is significant because the nozzle flowfield was *calibrated* using the assumption of thermal equilibrium.

As a result, it was incumbent to investigate the effect of this inappropriate assumption in the calibration of the facility on the JCEAP results. To address this question, a one-dimensional analysis code was written for calculating the isentropic flow in the nozzle. This code integrates the adiabatic and isentropic relationships from the nozzle plenum conditions out to a specified static pressure, and assumes either (a) the gas is vibrationally frozen at a specified temperature, or (b) thermal equilibrium is modeled via a harmonic oscillator. Using this 1-D analysis, the effects of vibrational nonequilibrium on the free-stream conditions can be estimated. Relative to the thermal equilibrium case, the effects of vibrational nonequilibrium on Mach number, static pressure, static temperature, and velocity at the test section were found to be in error by +0.11%, −0.21%, −0.93%, and −0.35%, respectively. Since we were concerned with validation of the surface pressures, the effect on the free-stream pressure of only 0.21% was considered negligible.

11.4.1.4 Continuum flow assumption

In order to ensure that the assumption of continuum flow is valid for the wind tunnel nozzle in the low-pressure rapid expansion region, Bird's continuum breakdown parameter P was calculated. Continuum flow theory begins to break down for $P > 0.02$. The maximum values calculated in the Mach 8 nozzle simulations were approximately 2×10^{-5} . This result supports the use of the continuum flow assumption.

11.4.1.5 Outflow boundary condition assumption

The SACCARA simulations for the surface pressure on the conical portion of the vehicle used a zero gradient condition in the axial direction applied at the outflow boundary. The outflow boundary was specified as the baseplane of the vehicle, $x/R_N = 51.96$. This boundary condition is not appropriate in the subsonic portion of the boundary layer where pressure disturbances can travel upstream from the separated base flow of the vehicle.

In order to assess the effects of this boundary condition assumption on the pressure distributions, an axisymmetric case was computed, which included the base region. The

computational domain (without a support sting) was extended roughly 2 m past the baseplane of the model. The computational region was extended for such a long distance beyond the baseplane to ensure that supersonic flow existed everywhere across the outflow boundary of the new computational region. The wake was assumed to remain laminar over the entire region. A highly clustered mesh near the baseplane was used to capture the shear layer separating near the circular edge of the base. In the simulation, it was observed that the pressure drops dramatically as the boundary layer nears the baseplane. The upstream influence (from the baseplane forward) was $2.5R_N$, which was approximately four boundary layer thicknesses. Thus, the presence of the base flow affects the surface pressure no further forward than an axial distance of $x/R_N = 49.5$ (see Figure 11.10). Given that the experimental data comparisons only extend to $x/R_N = 46.5$, it is concluded that the original outflow boundary condition in SACCARA was adequate.

11.4.1.6 Axisymmetric flow assumption

In critically questioning the strength of each physical modeling assumption, the question was raised: could the presence of the slice on the opposite side of the body from the pressure orifices affect the pressure readings by way of the subsonic region of the boundary layer? The validity of the axisymmetric flow assumption was addressed by conducting full three-dimensional calculations of the JCEAP geometry including the planar slice on the aft end of the model. A symmetry plane was assumed, thus only half of the model was simulated. The three-dimensional mesh was based on the 240×240 axisymmetric mesh and employed 105 azimuthal mesh cells from the cone-symmetry plane to the slice-symmetry plane (180 deg apart). The wall normal spacing from the 240×240 axisymmetric mesh was retained over the entire surface of the three-dimensional mesh.

With the exception of the planar slice region, essentially no effect was found on the cone-side of the vehicle compared to the axisymmetric solutions.

11.4.1.7 Re-evaluation of the experimental data

After all of this computational effort, it was time to start critically examining the experimental data again. The first, and easiest, thing to check in the experimental data was the calculation of the average free-stream conditions from all of the 48 wind tunnel runs shown in Table 11.2. Surprisingly, an arithmetic error was found in computing the free-stream static pressure originally reported in the experiment. When the free-stream static pressure values were re-averaged, the resulting static pressure was found to be 1.4% larger than the static pressure initially reported in the experiment.

The corrected free-stream conditions were then run with the SACCARA code. The results are shown in Figure 11.11, along with the old conditions. As expected, the effect of increasing the free-stream static pressure by 1.4% is to increase the surface pressure by approximately 1.4%. The computational results with the new free-stream static pressure are now in better agreement with the experimental data. Accounting for the estimated

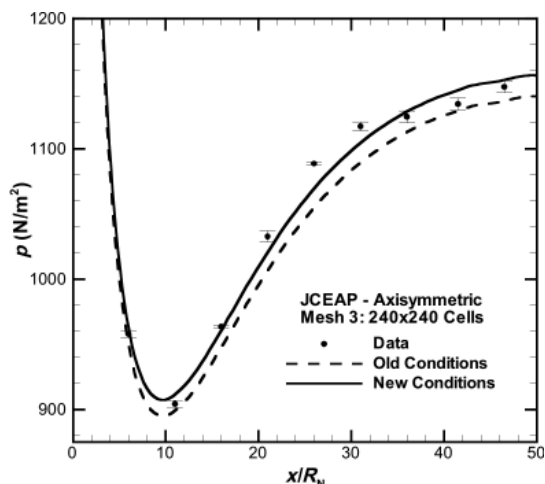


Figure 11.11 SACCARA simulation compared with the surface pressure data using both the old and corrected free-stream conditions, $\alpha = 0$ (Roy *et al.*, 2003).

experimental uncertainty bounds and an estimate of the numerical errors, the maximum error in surface pressure relative to the experimental data is now 1.5% at $x/R_N = 26$.

11.4.2 Simulation of the flowfield nonuniformities

Examining Figure 11.11, there appear to be two trends in the experimental data corrected for the error in free-stream static pressure that are not captured in the simulations. The first anomalous trend is that the experimental pressure is increasingly higher than the simulation as one moves from $x/R_N = 16$ to about $x/R_N = 30$. The second trend is that the experimental data has a marked decrease in slope beginning at $x/R_N = 30$. This latter feature is more disturbing because no such feature can occur on a blunted cone at zero angle of attack in a uniform, ideal gas, laminar hypersonic flow. In obtaining the experimental data for $\alpha = 0$, data were taken at various azimuthal angles around the model, various roll angles, and at two different axial locations in the tunnel. This type of data averaging procedure, however, will not account for the effects of axisymmetric nonuniformities, i.e., nonuniformities that are functions of the radial coordinate in the wind tunnel, because all the data were taken along the centerline of the test section. An attempt to quantify the axisymmetric nonuniformities in the test section was initiated.

11.4.2.1 Use of the flowfield calibration data

As part of the calibration procedure of the wind tunnel flowfield, the stagnation pressure is measured in the plenum, where the heaters are located. The stagnation temperature of the flowfield is determined from the measured plenum pressure and the mass flow rate that is measured upstream of the heaters. The local Mach number in the test section is calibrated

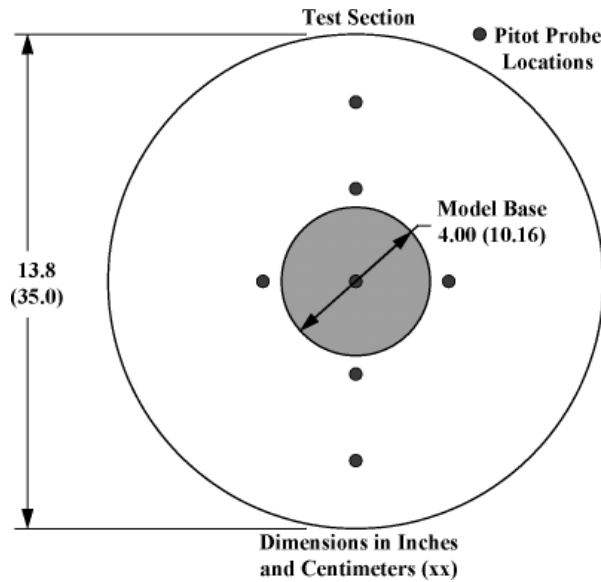


Figure 11.12 Pitot probe locations for calibration of the test section flowfield (JCEAP model base shown for reference) (Roy *et al.*, 2003).

using a seven-probe pitot pressure rake. The pitot probe locations are shown in Figure 11.12 along with the wind tunnel cross section (outer circle) in the test section. The JCEAP model base radius is shown (shaded circle) just for reference. The measured pressure points in the test section are at three radial coordinates: $y = 0$, 5.72, and 11.4 cm. The measurements from the four probes located 5.72 cm away from the centerline were averaged. Similarly, the measurements from the top and bottom probes, located 11.4 cm from the centerline, were averaged.

The pitot probe data of interest exist at seven axial stations in the test section. The pitot probe locations are presented in Figure 11.13 as the circular points. The flow direction is left to right, the zero axial station, $x = 0$, is located at the farthest upstream pitot probe location, and the radial coordinate is denoted by y . Only pitot data taken at Reynolds numbers within 15% of the JCEAP test Reynolds number were used, and these data were then corrected to the nominal Reynolds number. The corrections were considered reliable because flowfield calibrations were conducted at a range of Reynolds numbers.

The pitot probe data were then used in a newly written method of characteristics code to generate a complete flowfield in the vicinity of the JCEAP model locations. This complete flowfield, including axisymmetric nonuniformities, could then be used in a new simulation of the flow over the JCEAP geometry. The axisymmetric method of characteristics (MOC) scheme of Hartree was employed (Owczarek, 1964). This MOC implementation used an implicit integration scheme, that is, it determined the slope of the characteristic line as the average between the origination point and the destination point. For example, the C^- characteristic shown in Figure 11.13 uses the average slope found between points B and

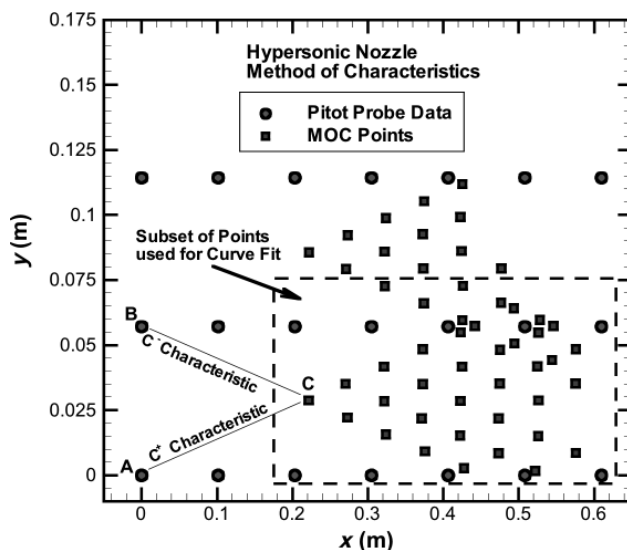


Figure 11.13 Axial and radial locations of the pitot probe calibration data and the data generated using the method of characteristics (Roy *et al.*, 2003).

C, and the C^+ characteristic uses the average of points A and C. In this manner, the characteristic network shown in Figure 11.13 is generated. The major assumption is that the radial velocity of the flow at the pitot probe locations is zero. This assumption must be made because the pitot probe data contain no information on the flow angularity.

Using the measured calibration points and the MOC generated data, an axisymmetric, nonuniform flowfield can be constructed. Figure 11.14 shows the three-dimensional surface of the computed free-stream pressure data as a function of axial and radial coordinates in the test section. A number of expansion and compression waves are evident, with a dominant axisymmetric wave-focusing effect occurring at $x = 0.52$ m at the centerline.

11.4.2.2 Simulation using the nonuniform flowfield

There is now sufficient data resolution to include the axisymmetric nonuniformities as a nonuniform inflow boundary condition for new simulations of the flow over the model. Since there were two different axial testing locations used for the vehicle, two different sets of inflow boundary conditions were used in the simulations. The simulation results accounting for axisymmetric flowfield nonuniformities at both the fore and aft model locations are presented in Figure 11.15. Up until the $x/R_N = 30$ axial location, the two simulations accounting for nonuniformities obtain even better agreement with the experimental data than the uniform flow free-stream conditions. However, for $x/R_N > 30$, the nonuniform boundary simulations over predict the pressure relative to the experimental data.

Although the computational construction of the nonuniform flowfield showed significant improvement with the measurements over most of the length of the model, the construction does have two weaknesses. First, there is very sparse radial resolution of the calibration data

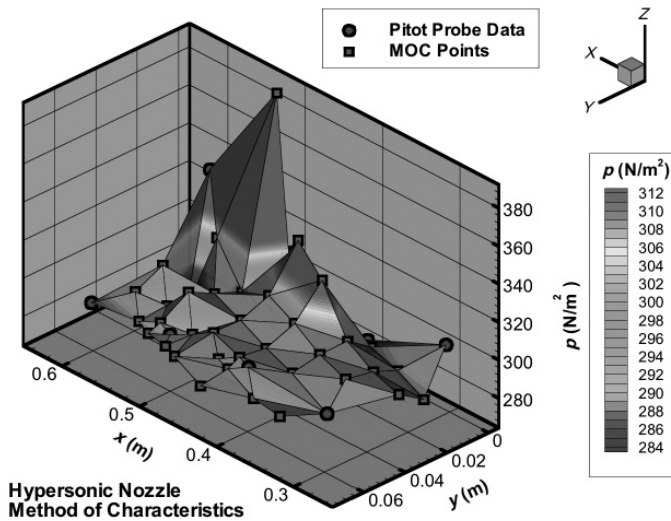


Figure 11.14 Free-stream static pressures from a combination of flowfield calibration data and MOC generated data (Roy *et al.*, 2003). See color plate section.

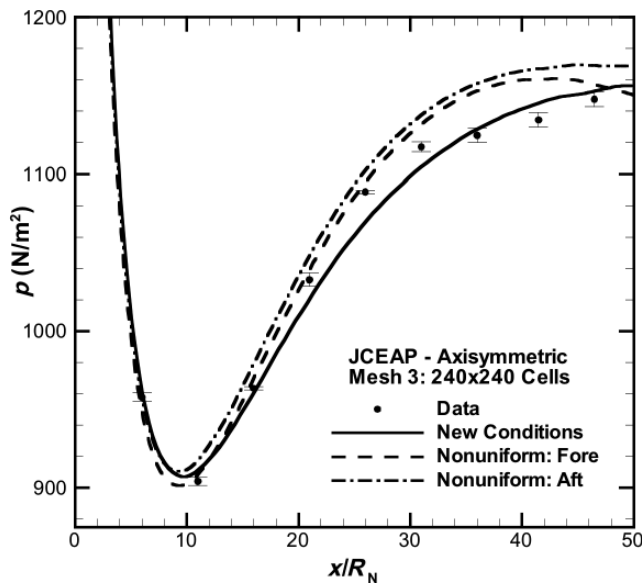


Figure 11.15 Comparison of the experimental data and SACCARA simulations accounting for axisymmetric nonuniformities, $\alpha = 0$ (Roy *et al.*, 2003).

over the test section of the wind tunnel. Second, the assumption of zero radial velocity at the calibration data points is only an approximation to the actual flowfield. A more spatially precise and definitive calibration of the test section, especially the beginning of the test section, would eliminate these weaknesses so that an accurate nonuniform flowfield could be used for future simulations.

11.4.3 Lessons learned for validation experiments

The JCEAP experiment and the subsequent computational analyses provided an exceptional opportunity to develop the fundamental ideas of what validation experiments should be. Not that the JCEAP experiment did everything correctly, but it was a valuable learning experience. Specific lessons were learned concerning validation experiments in wind tunnels, but many of these can be extended to validation experiments over a wide variety of physics. Here, we give a few summary comments for validation experiments applicable to essentially any technical field.

The value of joint planning and design in the JCEAP experiment was seen in almost every aspect of the project. There were two key features that allowed this joint activity to occur. First, there was the joint funding for both the experimental and the initial computational effort. The funding was from an internal research and development source, allowing much more of a research and discovery environment. Second, the experimental and computational team members were from the same organization and had worked together in the past. This was a luxury that is nearly impossible to achieve in large-scale validation activities, especially for projects of a national or international scale.

The level of characterization of an experimental facility that is needed for a validation experiment is remarkably higher than what is normally available from the facility. This high level of characterization is usually difficult and expensive to obtain. It requires research oriented personnel and those who question accepted assumptions concerning the facility, instead of more production-oriented personnel. In addition, this raises the question of the feasibility of conducting validation experiments in high-cost, high-volume data output facilities. These types of facility usually have testing capabilities well beyond small research facilities, but in these expensive facilities there is unrelenting pressure to keep data production high for the facility customer, instead of taking time to refine the facility characterization or investigate possible sources of uncertainty. As a result, to do validation experiments in high-cost facilities requires a much higher level of planning and preparation before the testing begins, as well as additional financial resources compared to traditional production testing in the facility. Finally, facility customers who want to conduct a high-quality validation experiment should clearly notify the facility operator during the planning stage that it desires personnel who have the appropriate experience and mindset for a validation experiment. Experimental facilities that embrace a more critical and comprehensive uncertainty estimation procedure could use this to their advantage with validation-type customers.

As commonly occurs in validation experiments, there are differences observed between computational results and experimental results. If there is an opportunity to pursue the possible causes of these differences, a great deal can be learned concerning both the simulation and the experimental measurements. For example in the follow-on computational work to JCEAP, even small differences between simulation and experiment were pursued. This resulted in a higher level of understanding of both the simulations and the experimental facility. Weaknesses in the facility were uncovered and specific recommendations were

made concerning how the characterization of the facility could be improved for future testing. Indeed, the next higher level of validation activities will involve accounting for the known imperfections in a facility in the computational simulations.

The use of DOE with randomization and blocking conclusively showed that in three different wind tunnel facilities the systematic uncertainties dominate the experimental measurement uncertainty. This is not welcome news to most wind tunnel operators because the experimental uncertainty that is traditionally quoted is only the random component of uncertainty. Various esteemed metrologists, such as Youden (1972), have called attention to the importance of systematic uncertainties in experimental measurements. However, the change in perspective in the experimental measurement community occurs at a glacial pace. More than two decades ago, McCroskey (1987) tried to call attention to the problem of systematic errors in wind tunnel testing by stating:

Reliable determination and assessment of the accuracy of aerodynamic data generated in wind tunnels remains one of the most vexing problems in aeronautics. Aerodynamic results are seldom duplicated in different facilities to the level of accuracy that are required either for risk-free engineering development or for the true verification of theoretical and numerical methods. This shortcoming is particularly acute with regard to today's rapid proliferation of new Computational Fluid Dynamic (CFD) codes that lack adequate validation.

As we have tried to point out, this lack of interest in seeking out and quantifying systematic errors is not just in the wind tunnel community, but it is a systemic problem in experimental facilities. It is our view that only when experimental facilities see that it is to their advantage will they change their procedures.

11.5 References

- Aeschliman, D. P. and W. L. Oberkampf (1998). Experimental methodology for computational fluid dynamics code validation. *AIAA Journal*. **36**(5), 733–741.
- Aeschliman, D. P., W. L. Oberkampf, and H. F. Henfling (1994). Fast-response, electronically-scanned multi-port pressure system for low-pressure hypersonic wind tunnel applications. *AIAA Aerospace Ground Testing Conference*, AIAA Paper 94-2580, Colorado Springs, CO, American Institute of Aeronautics and Astronautics.
- Aeschliman, D. P., W. L. Oberkampf, and F. G. Blottner (1995). A proposed methodology for CFD code verification, calibration, and validation. *16th International Congress on Instrumentation for Aerospace Simulation Facilities*, Paper 95-CH3482-7, Dayton, OH, ICIASF.
- AGARD (1994). *Quality Assessment for Wind Tunnel Testing*. NATO Advisory Group for Aerospace Research & Development (AGARD), AGARD-AR-304.
- AIAA (1999). *Assessment of Experimental Uncertainty with Application to Wind Tunnel Testing*. S-071A-1999, Reston, VA, American Institute of Aeronautics and Astronautics.
- AIAA (2003). *Assessing Experimental Uncertainty – Supplement to AIAA S-071A-1999*, Reston, VA, American Institute of Aeronautics and Astronautics.
- Amtec (1995). *INCA User's Manual*. Bellevue, WA, Amtec Engineering, Inc.

- ANSI (1997). *U.S. Guide to the Expression of Uncertainty in Measurement*. Boulder, CO, American National Standards Institute.
- Barber, T. J. (1998). Role of code validation and certification in the design environment. *AIAA Journal*. **36**(5), 752–758.
- Benek, J. A., E. M. Kraft, and R. F. Lauer (1998). Validation issues for engine – airframe integration. *AIAA Journal*. **36**(5), 759–764.
- Box, G. E. P., J. S. Hunter, and W. G. Hunter, (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*. 2nd edn., New York, John Wiley.
- Chapman, D. R., H. Mark, and M. W. Pirtle (1975). Computer vs. wind tunnels. *Astronautics & Aeronautics*. **13**(4), 22–30.
- Coleman, H. W. and W. G. Steele, Jr. (1999). *Experimentation and Uncertainty Analysis for Engineers*. 2nd edn., New York, John Wiley.
- Cosner, R. R. (1995). CFD validation requirements for technology transition. *26th AIAA Fluid Dynamics Conference*, AIAA Paper 95-2227, San Diego, CA, American Institute of Aeronautics and Astronautics.
- DeLoach, R. (2002). Tactical defenses against systematic variation in wind tunnel testing. *40th AIAA Aerospace Sciences Meeting & Exhibit*, AIAA-2002-0885, Reno, NV, American Institute of Aeronautics and Astronautics.
- DeLoach, R. (2003). Blocking: a defense against long-period unexplained variance in aerospace ground testing. *41st Aerospace Sciences Meeting and Exhibit*, AIAA-2003-0650, Reno, NV, American Institute of Aeronautics and Astronautics.
- Devore, J. L. (2007). *Probability and Statistics for Engineers and the Sciences*. 7th edn., Pacific Grove, CA, Duxbury.
- Drosg, M. (2007). *Dealing with Uncertainties: a Guide to Error Analysis*, Berlin, Springer-Verlag.
- Grabe, M. (2005). *Measurement Uncertainties in Science and Technology*, Berlin, Springer-Verlag.
- Hinkelmann, K. and O. Kempthorne (2008). *Design and Analysis of Experiments: Volume 1 – Introduction to Experimental Design*. 2nd edn., Hoboken, NJ, John Wiley.
- ISO (1995). *Guide to the Expression of Uncertainty in Measurement*. Geneva, Switzerland, International Organization for Standardization.
- ISO (2008). *Uncertainty of Measurement – Part 3: Guide to the Expression of Uncertainty in Measurement*. ISO/IEC Guide 98-3:2008, Geneva, Switzerland, International Organization for Standardization.
- Marvin, J. G. (1995). Perspective on computational fluid dynamics validation. *AIAA Journal*. **33**(10), 1778–1787.
- McCroskey, W. J. (1987). *A Critical Assessment of Wind Tunnel Results for the NACA 0012 Airfoil*. Washington, DC, National Aeronautics and Space Administration.
- Moffat, R. J. (1988). Describing the uncertainties in experimental results. *Experimental Thermal and Fluid Science*. **1**(1), 3–17.
- Montgomery, D. C. (2000). *Design and Analysis of Experiments*. 5th edn., Hoboken, NJ, John Wiley.
- Morgan, M. G. and M. Henrion (1990). *Uncertainty: a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. 1st edn., Cambridge, UK, Cambridge University Press.
- Oberkampf, W. L. and D. P. Aeschliman (1992). Joint Computational/Experimental Aerodynamics Research on a Hypersonic Vehicle: Part 1, Experimental Results. *AIAA Journal*. **30**(8), 2000–2009.

- Oberkampf, W. L. and F. G. Blottner (1998). Issues in computational fluid dynamics code verification and validation. *AIAA Journal*. **36**(5), 687–695.
- Oberkampf, W. L. and T. G. Trucano (2002). Verification and validation in computational fluid dynamics. *Progress in Aerospace Sciences*. **38**(3), 209–272.
- Oberkampf, W. L., A. Martellucci, and P. C. Kaestner (1985). *SWERVE Surface Pressure Measurements at Mach Numbers 3 and 8*. SAND84-2149, SECRET Formerly Restricted Data, Albuquerque, NM, Sandia National Laboratories.
- Oberkampf, W. L., D. P. Aeschliman R. E. Tate, and J. F. Henfling (1993). *Experimental Aerodynamics Research on a Hypersonic Vehicle*. SAND92-1411, Albuquerque, NM, Sandia National Laboratories.
- Oberkampf, W. L., D. P. Aeschliman J. F. Henfling, and D. E. Larson (1995). Surface pressure measurements for CFD code validation in hypersonic flow. *26th AIAA Fluid Dynamics Conference*, AIAA Paper 95-2273, San Diego, CA, American Institute of Aeronautics and Astronautics.
- Oberkampf, W. L., D. P. Aeschliman, J. F. Henfling, D. E. Larson, and J. L. Payne (1996). Surface pressure measurements on a hypersonic vehicle. *34th Aerospace Sciences Meeting*, AIAA Paper 96-0669, Reno, NV, American Institute of Aeronautics and Astronautics.
- Owczarek, J. A. (1964). *Fundamentals of Gas Dynamics*, Scranton, PA, International Textbook.
- Payne, J. L. and M. A. Walker (1995). Verification of computational aerodynamics predictions for complex hypersonic vehicles using the INCA code. *33rd Aerospace Sciences Meeting and Exhibit*, Reno, NV, American Institute of Aeronautics and Astronautics.
- Porter, J. L. (1996). A summary/overview of selected computational fluid dynamics (CFD) code validation/calibration activities. *27th AIAA Fluid Dynamics Conference*, AIAA Paper 96-2053, New Orleans, LA, American Institute of Aeronautics and Astronautics.
- Rabinovich, S. G. (2005). *Measurement Errors and Uncertainties: Theory and Practice*. 3rd edn., New York, Springer-Verlag.
- Roache, P. J. (1998). *Verification and Validation in Computational Science and Engineering*, Albuquerque, NM, Hermosa Publishers.
- Roy, C. J., M. A. McWherter-Payne, and W. L. Oberkampf (2000). Verification and validation for laminar hypersonic flowfields. *Fluids 2000 Conference*, AIAA Paper 2000-2550, Denver, CO, American Institute of Aeronautics and Astronautics.
- Roy, C. J., M. A. McWherter-Payne, and W. L. Oberkampf (2003a). Verification and validation for laminar hypersonic flowfields, Part 1: Verification. *AIAA Journal*. **41**(10), 1934–1943.
- Roy, C. J., W. L. Oberkampf, and M. A. McWherter-Payne (2003b). Verification and validation for laminar hypersonic flowfields, Part 2: Validation. *AIAA Journal*. **41**(10), 1944–1954.
- Salicone, S. (2007). *Measurement Uncertainty: an Approach via the Mathematical Theory of Evidence*, Berlin, Springer-Verlag.
- Spall, R. E. and M. R. Malik (1991). Effect of transverse curvature on the stability of compressible boundary layers. *AIAA Journal*. **29**(10), 1596–1602.
- Spall, R. E. and M. R. Malik (1992). Linear stability of three-dimensional boundary layers over axisymmetric bodies at incidence. *AIAA Journal*. **30**(4), 905–913.

- Steger, J. L. and R. F. Warming (1981). Flux vector splitting of the inviscid gasdynamic equations with applications to finite-difference methods. *Journal of Computational Physics*. **40**, 263–293.
- Walker, M. A. and W. L. Oberkampf (1992). Joint computational/experimental aerodynamics research on a hypersonic vehicle: Part 2, Computational results. *AIAA Journal*. **30**(8), 2010–2016.
- Wilcox, D. C. (2006). *Turbulence Modeling for CFD*. 3rd edn., La Canada, CA, DCW Industries.
- Wong, C. C., F. G. Blottner, J. L. Payne, and M. Soetrisno (1995a). Implementation of a parallel algorithm for thermo-chemical nonequilibrium flow solutions. *AIAA 33rd Aerospace Sciences Meeting*, AIAA Paper 95-0152, Reno, NV, American Institute of Aeronautics and Astronautics.
- Wong, C. C., M. Soetrisno, F. G. Blottner, S. T. Imlay, and J. L. Payne (1995b). *PINCA: A scalable Parallel Program for Compressible Gas Dynamics with Nonequilibrium Chemistry*. SAND94-2436, Albuquerque, NM, Sandia National Laboratories.
- Yee, H. C. (1987). *Implicit and Symmetric Shock Capturing Schemes*. NASA, NASA-TM-89464.
- Youden, W. J. (1972). Enduring values. *Technometrics*. **14**(1), 1–11.