AIAA 98-0713

Applications of Modern Experiment Design to Wind Tunnel Testing at NASA Langley Research Center

Richard DeLoach
NASA Langley Research Center
Hampton, VA

36th Aerospace Sciences Meeting & Exhibit
January 12–15, 1998 / Reno, NV

For permission to copy or republish, contact the American Institute of Aeronautics and Astronautics
1801 Alexander Bell Drive, Suite 500, Reston, Virginia 20191-4344

APPLICATIONS OF MODERN EXPERIMENT DESIGN TO WIND TUNNEL TESTING AT NASA LANGLEY RESEARCH CENTER

Richard DeLoach*
NASA Langley Research Center
Hampton, VA 23681-0001

Abstract

A "modern design of experiments" (MDOE) approach to wind tunnel testing is under evaluation at NASA Langley Research Center which differs from conventional "one factor at a time" (OFAT) pitch-polar test designs in certain fundamental ways. This paper outlines the differences, both philosophical and procedural, in the two design methodologies, and compares results obtained using both methods in specific wind tunnel tests. Comparisons are made of the relative costs to achieve the same technical objective, where comparisons are made in terms of wind-on minutes, data volume, and electrical power consumption. Designed experiments appear to have the potential for saving as much as one-third to one-half of the wind on minutes of conventional pitch-polar tests, and one quarter to one third of the wind-on costs. At the same time they increase precision by removing "block effects" from the unexplained variance in test results and illuminate interaction effects that cannot be quantified when variables are changed one at a time.

Introduction

In wind tunnel testing as in all other experimental disciplines, aspects of a system are manipulated in order to quantify how that system responds. In the case of wind tunnel testing, changes are made in such independent variables as model attitude, control surface configuration, and flow field conditions. Resulting changes are noted in such response variables as the aerodynamic forces and moments acting on the model, the pressure and temperature distributions over its surfaces, etc. The objective of such testing is to model the response variables in terms of the independent variables so as to be able to predict with confidence the behavior of full-scale flight systems. In short, we model in order to predict, and we seek reliable cause-effect relationships through testing in order to model reliably.

The specific relationships we infer between stimulus and response in wind tunnel testing can be confounded by variables covarying with the ones we intend to change as we execute the test plan. For example, if the sideslip angle changes slightly (due to flow-field dynamics, say) while angle of attack (AoA) is systematically varied, then the effects of AoA on such responses as lift and drag can be contaminated by the unintended simultaneous changes in sideslip angle. The traditional defense against such complications has been to vary one factor at a time while endeavoring to "hold everything else constant". Experiment designs featuring this error control strategy are commonly described as "classical designs" in the literature of designed experiments and the general test procedure is called "One Factor at a Time (or "OFAT") testing.

Advocates of an alternative approach to experimentation recognize the futility of efforts to "hold constant" all potential covarying variables in a system as complex and energetic as a wind tunnel, especially in the face of precision requirements anticipated for 21st-century aeronautical research. The term "Modern Design of Experiments" (MDOE) is used to contrast this alternative approach with the "classical design" methods widely used in present-day wind tunnel testing. The method is so-named because of its relatively late introduction (compared to classical designs practiced since the Renaissance), dating from the original 1935 publication by R.A. Fisher of his seminal book, "The Design of Experiments". This book has since been released in several updated editions (ref 1).

The experimental testing technology community at NASA Langley Research Center (LaRC) is proposing that wind tunnel researchers consider modern design methods as an alternative to the classical experiment designs associated with OFAT testing. Four modern design wind tunnel tests were conducted at LaRC in 1997 to demonstrate for the LaRC wind tunnel research community the relative costs and benefits of this approach compared to classical wind tunnel test designs. This paper contrasts the basic approaches of OFAT and MDOE testing, outlines the general procedures used when modern design methods are applied in wind tunnel research, and summarizes the

* Senior Research Scientist

Copyright © by the American Institute of Aeronautics and Astronautics, Inc. No copyright is asserted by the United States under Title 17, U. S. Code. The U. S. Government has a royalty-free license to exercise all rights under the copyright claimed herein for Government Purposes. All other rights are reserved by the copyright holder.

1997 MDOE testing experience at LaRC. We begin with remarks on the fundamental philosophical difference between MDOE experimentation and OFAT data collection.

Modern Experiment Design – A Process-Oriented Approach to Aeronautical Research

The modern design of experiments (MDOE) approach to wind tunnel experimentation differs from classical one-factor-at-a-time (OFAT) data collection because of fundamental difference in the philosophies of the two methods. MDOE emphasizes a process-oriented approach while OFAT methods are generally centered on the individual tasks that comprise the overall experimental research process.

As is true of any process, experimental aeronautical research is comprised of an interrelated set of tasks that produce some end result of value. It can be viewed as a system by which inputs – in the form of questions about a flight system – are converted into outputs that have value – in the form of answers to scientific/engineering questions that lead to insights into the characteristics of some flight system. Scores of individual tasks comprise this process, from the definition of research requirements through all the steps in the execution of the test to the analysis and reporting of results. While each of these tasks is critical to a successful outcome of the process, no single one of them creates the desired end value by itself. We do not achieve insights into the performance characteristics of a flight system by "taking data", for example, or by any of the other individual tasks in the research process, but only as a result of performing the entire process. This process-centered focus, which characterizes the MDOE approach to testing, is especially relevant today because challenges facing the aeronautical research community at the dawn of the 21st century are process-oriented rather than task-oriented. Several examples come readily to mind:

The cost of experimental aeronautical research is higher than it might otherwise be, not because individual researchers are inefficient or that specific tasks in the research process are exorbitantly expensive, but because quantifiable criteria for a successful process outcome are seldom articulated before the experiment is conducted. With no ready metric by which to determine when there has been sufficient information gathered in an experiment to successfully complete the overall research process, the researcher opts to expend all available test resources rather than to risk stopping prematurely.

Wind tunnel results sometimes include errors that are unacceptably large for a particular purpose, not because the measurement systems are imprecise or the operators are incompetent, but because a task-oriented approach encourages us all to focus only on the errors our tasks contribute to the final result. Facility personnel dwell on experimental and measurement errors introduced in the data acquisition. Analysts dwell on lack of fit and other errors associated with data reduction and analysis. The total uncertainty of the end-to-end research process is often no one's responsibility.

One reason that design cycles are longer than we would like is not that researchers are inefficient or inherently slow in performing individual tasks. It is that there is often poor linkage between the original research requirements (the input of the research process) and the character and volume of information needed to provide missing knowledge about a system (the output of the process). Absent such linkage the researcher is likely to err on the side of excess, spending as much time meticulously exploring one region of the inference space as another, unable to clearly identify how to allocate wind-on minutes most efficiently in the tunnel.

OFAT experimental procedures are often rigid and predictable, but not because wind tunnel engineers lack imagination. It is because the relationship between the details of the data acquisition task and the desired end product of the research process are seldom well understood by all participants. If you believe your task is to execute some number of pitch polars in a fixed timeframe, you have an entirely different outlook than if you view yourself as responsible for defining the relationship between stability axis drag coefficient and model attitude over some range of independent variables. The former is a task-oriented perspective while the latter perspective is process oriented. In the latter instance you are much more likely to measure your performance by more relevant metrics than in the former, and if you produce a result that leads to reliable flight predictions, no one will inquire (or care very much) if you met some rigid pitch polar production quota. (The exception may be the project accountant, who is likely to have a higher regard for those who get results with minimum resources than those who use "exhaustion of all resources" as the sole stopping criterion for a test.)

These examples illustrate a basic concept that W. Edwards Deming expressed succinctly when he said: "It is a mistake to assume that if everyone does his job it will be all right. The whole system may be in trouble." In the case of experimental aeronautical research, the task-oriented focus of classical research and OFAT testing procedures can be a significant contributor to such systemic "trouble". MDOE methods offer a process-oriented alternative to the OFAT focus on individual task excellence, which has so often failed to achieve entirely satisfactory results. This alternative perspective recognizes that all that

really matters in research are the inputs – the research questions – and the outputs – scientifically defensible answers to those questions.

This seemingly obvious observation – that an experiment is successfully completed when questions posed at the start of the process are satisfactorily answered – has enormous consequences. It implies that effective wind tunnel research is more than a compartmentalized series of rote data collection and post-hoc aeronautical data analysis tasks. It relies critically on contributions not only from the discipline-specific researcher but also from those who can accurately characterize the facility measurement environment and the limitations it poses, and those who can provide scientifically defensible, real-time reliability assessments as data are acquired. For without such extensions to classical wind tunnel research it is impossible to recognize when questions posed at the beginning of the test have been satisfactorily answered. Yet this necessary blurring of the lines of demarcation that have historically separated research, facility, and testing technology responsibilities defies a culture and tradition of task-based aeronautical research that has been entrenched for most of the history of flight.

The OFAT method places special emphasis on tasks that impact high-volume data collection, which serves as a common productivity metric. The assumption is that the customer will be well served if each task is performed well, especially if data are produced in high volume. But no one involved in the process – including the customer, quite often – seems to pay much attention to whether all the individual tasks are providing specific answers to relevant research questions. The MDOE method imposes a discipline on aeronautical research activities which results in a rigorous process perspective and an unwavering focus on process outcomes. It organizes the aeronautical research process into design, execution, and analysis sub-processes, each of which will be described in some detail in the sections that follow.

The Design Phase of the MDOE Process

The formal design of a wind tunnel experiment, like any other process or sub-process, begins with inputs and generates outputs from them. The output of the design process is a run schedule. Inputs to the design process include a clear statement of the objective of the experiment, expressed in terms of specific response variables to be measured, specific independent variables to be manipulated, and the precise range over which each independent variable is to be varied. Additional inputs include information needed to estimate the volume of data necessary to achieve the required precision in the experimental results, including quantified estimates of the response variance, the required resolution of the experiment, and the researcher's inference risk tolerances (to be described below). Other inputs that are exceedingly useful in the design process but not absolutely required are any prior data, past experience, or other information that sheds light on the general shapes of functions relating the treatment and response variables.[1] The design process ultimately reduces to finding answers to these two questions: 1) "How many data points do I need?" and 2) "Which points?"

How Many Data Points?

For scientific and engineering experimentation two considerations govern the required volume of data. First, sufficient data must be acquired to create a mathematical model that can describe each response variable in terms of all the treatment variables of interest upon which it depends, over some prescribed range of each such variable. Secondly, enough additional data must be acquired to assess the adequacy of the model. Specifically, we must acquire enough additional data to ensure that the model satisfies design precision requirements and is capable of reliable predictions. We begin by considering the model and the requirements for constructing it.

The Math Model – Estimating the Regression Degrees of Freedom: We generally do not know the exact functional form of the relationship between responses such as lift or drag, and the variables that influence them such as AoA, sideslip angle, Mach number, and control surface deflection angle. That is, in most cases we cannot write down a mathematical description that will predict precisely how a change in the level of one or more treatment variables will affect the forces, moments, and other aeronautical responses that interest us. Such a mathematical relationship is, of course, the Holy Grail of experimental aeronautical research, in that with such a perfect model we would know everything about how the aircraft responds to manipulations of the treatment variables. While we cannot know the exact nature of this function, we can use curve-fitting techniques to approximate it arbitrarily closely, at least within limited ranges of the independent variables.

Figure 1 illustrates the general idea of a "graduating function" as an approximation to an

[1] Independent and dependent variables are commonly called "treatment" and "response" variables respectively, a practice with roots in early medical applications of designed experiments in which the "response" of patients to various medical "treatments" was the object of formal study.

unknown true function, obtained by fitting a curve to data acquired over some restricted range of the independent variable. If this so-called "inference subspace" (also called the "design subspace") is small enough, then an arbitrarily complex true function can be represented adequately with a relatively simple curve fit. This simple curve behaves like a mathematical French curve, fitting the unknown true function well over some limited range.

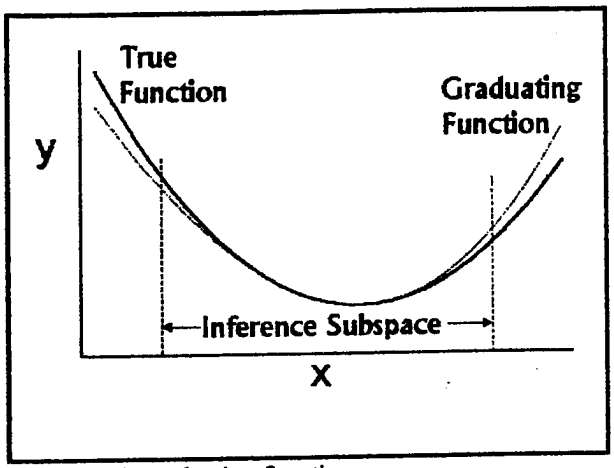


Figure 1: A graduating function

The true function can be well approximated throughout the full range of all independent variables by a family of such graduating functions, each applying to a different (adjacent or overlapping) inference subspace. As a minimum, a sufficient number of data points (regression degrees of freedom) must be acquired to fit each graduating function in this family of curves. A low-order Taylor series serves as a convenient general form for the graduating function.

The number of parameters in a Taylor series of order d in k independent variables is easy to compute using equation 1:

$$p = \frac{(d+k)!}{d!k!} \qquad (1)$$

Since each parameter in the Taylor series requires one degree of freedom, equation 1 also defines the number of regression degrees of freedom necessary to fit a full d^{th}-order model in k independent variables. This general polynomial representation of the graduating function therefore defines the minimum number of data points required in an experiment. If there are l inference subspaces then there will be lp regression degrees of freedom required.

While p points in each inference subspace are all that are required to estimate a d^{th} order model in k variables, an experiment design featuring this minimum number of points would provide no additional degrees of freedom to assess the quality of the model (goodness of fit, uncertainty, etc.) To be reassured that a proposed model represents an adequate basis for making predictions requires that sufficient additional data be acquired to meet at least four conditions.

First it is necessary to ensure that the structure of the model itself does not contribute significantly to the uncertainty in predictions made with the model. We require that "pure error" effects dominate whatever uncertainty is in the model – the effects of ordinary chance variations in the data only. We reject as inadequate any model for which so-called "lack of fit" errors are the dominant source of uncertainty. These are errors introduced when the model simply does not fit the data properly. Examples include efforts to fit a low-order polynomial model over an inference subspace in which the response function is rich in significant higher-order terms. To determine if there is significant lack of fit error, a number of "genuine replicates" must be acquired in addition to the minimum number of regression degrees of freedom. The use of genuine replicates to quantify contributions of lack of fit to the total uncertainty will be described in some detail below. For now, suffice it to say that "genuine replicates" are defined as repeated data points acquired in such a way as to permit any source of error to have its effect. Such points are therefore acquired at random intervals throughout the test, with intervening changes in the independent variables having taken place. Note in particular that we do not consider a simple "repeat point" as a genuine replicate when it merely represents an immediate sequential acquisition of data with no intervening changes in set point.

Even if it can be demonstrated that model prediction errors are due only to intrinsic scatter in the data used to create the model and are not due to lack of fit errors caused by the functional shape of the model itself, it is still necessary to determine that the uncertainty is small enough to satisfy design precision goals. For example, if it is required that the drag coefficient be known with 95% confidence to within a half drag count over some prescribed range of treatment variables, then it is necessary to acquire sufficient data to ensure that, given the intrinsic variability in the data, the uncertainty in model predictions can be said to be no greater than this with some level of confidence.

A model with no significant lack of fit errors and with a total uncertainty that satisfies design requirements may yet be inadequate if the range of predicted response variables is not large compared to the uncertainty in individual response variable measurements. It is possible to develop a model with arbitrarily small lack of fit errors and with arbitrarily low total uncertainty, for example. by fitting over suitably restricted ranges of the independent variables.

In such cases the model may be simply fitting "noise", in which case it will be of little utility for predicting responses as a function of treatment variable levels. Tests exist to determine when there is sufficient "signal to nose ratio" in the model, as will be described presently.

Even if a model is constructed with no significant lack of fit errors, with sufficiently low total uncertainty to satisfy design criteria, and with enough range that it is unlikely to be fitting noise only, there is a fourth criterion that an adequate model must meet. It is necessary to *demonstrate* that the model can predict response variables at combinations of the treatment variables that were not used to fit the model. This entails acquiring a set of additional data beyond those points used to construct the model. Some number of these so-called *confirmation points* are acquired and compared with model predictions to confirm the predictive power of the model. If no significant difference can be resolved between values of the confirmation points and predictions made with the model, the model may be declared to be adequate. The number of confirmation points to be acquired must strike a balance between the researcher's "comfort level" and the project's budget constraints but 25% more than the number needed to define the model and quantify pure error and lack of fit error is a common selection.

The total number of data points required for a given experiment is defined, then, by a degrees of freedom budget with a number of line items. The minimum number of points needed to estimate the regression coefficients in the model (the "regression degrees of freedom") depends on the number of treatment variables, the nature of the model, and the number of inference subspaces over which graduating functions are to be fit. This number can be computed precisely for a polynomial linear regression model as described in equation 1.

It can be shown that equation 2 gives the minimum number of residual degrees of freedom needed to satisfy the experimental quality constraints discussed earlier.

$$N = p\left(\frac{\sigma}{\delta}\right)^2\left(t_{\alpha/2} + t_{\beta}\right)^2 \qquad (2)$$

Equation 2 expresses the required number of residual degrees of freedom, N, in terms of resolution (δ), measurement environment (σ), and acceptable levels of risk for Type I and Type II inference errors (α and β). Note also that the more complex the model; i.e. the larger the number of parameters, p, to be fit, the more data that must be acquired to meet design goals. This is because each regression coefficient carries with it some added uncertainty, which must be offset by acquiring additional information about the system. For this reason among others, one should generally resist the temptation to gratuitously fit high-order models to the data. Better results are usually to be had if the response or treatment variables are transformed to permit an adequate fit to a lower-order model, or if the range of treatment variables is restricted so that first- or second-order models provide an adequate fit.

It is clear from equation 2 that for a given model, the inferential validity of the experimental result is determined entirely by the volume of data acquired. That is, for a given resolution requirement and a given standard deviation in the response variable, the data volume determines the probability of both Type I and Type II errors. This means that there is some minimum volume of residual degrees of freedom (specified by equation 2) for which Type I and Type II inference error probabilities are compatible with customer-specified risk tolerances as quantified by α and β.

In summary, if the experiment produces a model capable of predicting system responses with specified uncertainties quoted with a confidence of $1-\alpha$ (where both the uncertainty and α are specified by the customer), and the data were acquired in sufficient volume to resolve a change as small as δ with a probability of $1-\beta$ (where δ and β are both specified by the customer), then the experiment is a valid one. The total volume of data required in each inference subspace is the sum of the regression degrees of freedom computed with equation 1 and the residual degrees of freedom computed with equation 2, under the assumption that lack of fit errors are not significant and a sufficient signal to noise ratio exists for the model (to be discussed in further detail below.)

These discussions are intended to illustrate that some finite volume of data is all that is required to achieve a particular research outcome. The acquisition of less than this volume of data invalidates the result because either α or β then exceeds customer-specified validity thresholds. But the acquisition of substantially more that the minimum required volume of data is a waste of money and time. A particularly wasteful element of classical OFAT wind tunnel testing is the practice of acquiring data in volumes limited only by the exhaustion of (more or less arbitrarily allocated) resources. The MDOE methodology puts considerable emphasis on defining specific objectives in the early design phase of the experiment and then rationally quantifying the volume of data needed to achieve those objectives. The intent is not to limit the volume of data to the absolute minimum needed to get the job done; it is perfectly acceptable to acquire some reasonably

generous margin of additional data to compensate for uncertainties in original estimates of the measurement environment, for example. The guiding principal is simply that the researcher should be able to explain why a particular volume of data is necessary. Inherent in this principal is that the research must *know* why a particular volume of data is necessary, which requires an intimate knowledge of the exact purpose of the test and the criteria by which the successful completion of specific test objectives can be identified. This requires a level of preparation in the experiment design phase of the research process – well in advance of the execution phase – that is not commonly encountered in classical OFAT wind tunnel testing.

The Selection of Data Points

The prior section focused on how the number of data points relates to design goals. The specific choice of independent variables is also important. The uncertainty in regression predictions depends not only on chance variations in the data upon which the regression is based, but also on the relative location of points where the predictions are made and points where the regression data were acquired. We can exploit this fact by basing the regression on particular points in the inference space selected to maximize one or more measures of excellence.

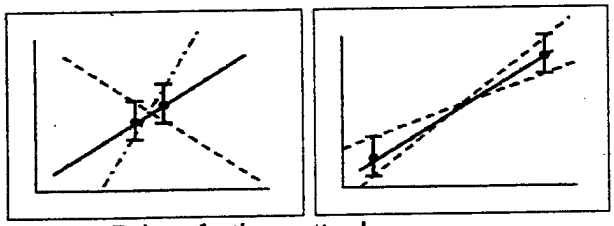


Figure 2: Point selection matters!

Figure 2 is a very simple example that illustrates this important principal. A minimum volume of data (two points) is used to estimate a linear function of one variable. On the right, the two points are acquired relatively close to each other. On the left, the same experiment is conducted with data points acquired further apart. Because of the finite error bars about each data point, there is less uncertainty in the estimate of the slope of this curve if the independent variable levels are further apart than if they are closer together. This elementary case illustrates a general principal, which is that the uncertainty in the regression coefficients (and therefore the uncertainty in model predictions) depends on the inference space geometry of the independent variable levels used in the regression. This is true in general, and applies to models of higher order and with more independent variables than the simple case illustrated in figure 2.

An enormous library of general-purpose and specialty designs can be found in the literature with point distributions that satisfy various requirements. While a survey of such designs will not be undertaken here, we will describe one particular design due to Box and Wilson (ref 6). The Box-Wilson, or "Central Composite Design" (CCD) is the most widely used general design for fitting second-order response surface models and is the design that has been most commonly used in the author's MDOE applications at NASA Langley Research Center.

The Box-Wilson design is comprised of a symmetric distribution of variables illustrated in figure 3 for the case of three independent variables.

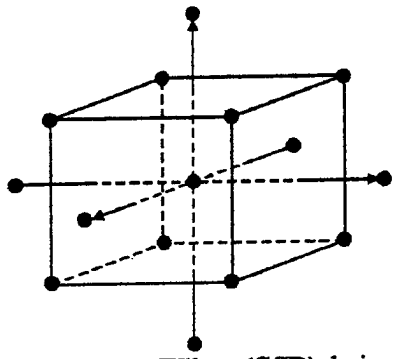


Figure 3: Box-Wilson (CCD) design

Each independent variable is associated with one axis in an inference space so that every point in that space represents a unique combination of independent variables. The adaptation of the design to any number of variables is straightforward.

The CCD is an orthogonal design so for a given spread in the treatment variables it features minimal errors in the regression coefficients and ensures that there will be no confounding by regressors orthogonal to the rest of the model. The degree of replication and the distances between axial points and the center of the can be adjusted to achieve certain benefits such as rotatability and orthogonal blocking. Rotatability is the property that model prediction errors depend only on the distance from the center of the design and not the direction, a desirable property if not an absolutely critical one. Orthogonal blocking insures that the values of the fitted regression coefficients are not affected by systematic changes in the tunnel that may occur between the time one block of data is acquired (the star points, say) and the time another block of data is acquired (the corner points, say). This provides protection against systematic additive tunnel effects, insuring that they affect at most the y-intercept (constant) term of the model and have no effect on the coefficients that reveal causal treatment-response relationships involving the independent variables under investigation. This structure provides similar protection

against between-tunnel effects in the event that the same experiment is conducted in another facility.

To summarize the experiment design process, we begin with a set of specific design objectives, including explicit listings of the treatment variables and response variables of interest. We specify the range of each treatment variable and the resolution/precision requirements for each response variable, including quantitative statements of acceptable inference error risks for both Type I and Type II errors. We also include as input to the design process any prior information or the benefit of any prior experience with similar experiments to help partition the inference space into regions throughout which it is likely that a low-order model can be fit in the (possibly transformed) treatment variables without generating significant lack of fit errors. We use this information to estimate the volume of data needed to achieve the stated objectives with acceptable levels of inference error risk, including some reasonable margin of additional data points to check model predictions and to augment the model if necessary. Having defined the necessary volume of data, the points are distributed in the inference space according to designs that optimize one or more of a number of excellence metrics. The output of the design process is then a run schedule consisting of data in sufficient volume to achieve design goals.

The Execution of a Modern Design Wind Tunnel Test

The output of the design process – a carefully designed run schedule – is the input to the execution process in which the experiment is actually conducted. The output of the execution process is a set of certified regression coefficients for graduating functions that adequately predict specified response variables over defined ranges of the treatment variables.

A critical difference between the execution of modern and classical designs is the order in which the data points are acquired. A classical design typically features the sequential setting of independent variable levels, often at constant intervals. The weakness of this sequential variation policy is that observed response functions are forced to change not only as a function of the level of the independent variables, but also as a monotonically increasing function of time. In a facility as complex as a modern wind tunnel it is not unlikely that other factors are also undergoing subtle but systematic changes with time. These changes are in addition to the random, chance variations in the data that are often assumed to be the only cause of uncertainty. Flow causes frictional heating, for example, which may cause mechanical expansion that might in turn influence wall effects and flow angularity. Solar heating throughout the day or radiative cooling at night may affect the rate of such effects. Transducers and signal conditioning are prone to drift, as is the data acquisition system. Operator fatigue might lead to subtle changes in technique, which result in systematic performance changes over time. There is an unknown (and unknowable) number of such sources of systematic change which undoubtedly come into play at some level over the days and weeks of a typical wind tunnel test. The apparent dependence of response functions on systematically varied independent variables will be influenced by these systematic nuisance-variable changes as well. OFAT sequential testing provides no ready means to separate these effects.

It is because of the vulnerability of OFAT sequential testing to systematic bias errors and unknown sources of precision error that modern design methods rely on randomization and replication to control error rather than the OFAT strategy of "holding everything constant". The schedule of independent variables is executed in a random sequence so that it is no more likely for low values to be run before high values than for high values to be run before low ones. With this error control strategy, the independent variables are not varied systematically with time. Therefore, any systematic change in response with the level of the independent variable can be attributed to the influence of that variable alone, and not to the effects of nuisance variables changing systematically with time. In effect, randomization converts undetectable bias errors to precision errors that can be quantified and controlled by replication.

Another characteristic that distinguishes the execution phase of an MDOE wind tunnel experiment from an OFAT data collection exercise is that the MDOE approach entails significantly more quantitative analysis of the data during the acquisition phase. The OFAT practitioner may examine line graphs and similar output during the test to get a general sense of trends and to identify potential trouble spots but in an MDOE experiment the researcher continually makes quantitative assessments of the quality of the results in hand. This is necessary to determine when sufficient information is in hand in a given inference subspace to justify moving on to other regions.

Lack of fit is quantified through a conventional analysis of variance in which the residual sum of squares associated with an ordinary linear regression is partitioned into pure error and lack of fit components. (This requires that genuine replicates be included in the design. See, for example, references 2-5.) The residual degrees of freedom are likewise partitioned and the pure error and lack of fit variances are thus estimated. A simple F-test then reveals if the lack of fit variance component is significant relative to the pure error component. If so, additional data can be acquired to permit a graduating function to be fit over a suitably

reduced range of the treatment variables to improve the fit, or a more complex model may be adopted. Such a model may feature additional independent variables, a different function of the existing variables, or higher order terms. An analysis of the residuals provides the only guidance available from the data as to how best to proceed. See references 4 and 5 for a comprehensive discussion of residual analysis. An F-statistic that exceeds a certain threshold also indicates that sufficient signal to noise ratio exists to ensure that the graduating function is not simply fitting chance variations in the data. Without such an indication the model would not be likely to make adequate predictions.

Analysis in the MDOE Process

Analysis is the third element in the MDOE research process. The design element produces a run schedule from inputs describing design goals and research questions, and the execution element produces a set of mathematical treatment-response models from the run schedules. The purpose of the analysis element is then to formulate answers to the original research questions using the response surface models developed during test execution.

A maxim among MDOE practitioners is that a well-designed experiment practically analyzes itself. One reason is that during the execution phase considerable analysis will have already been performed in the various quality assurance tests that are continually applied to assess the quality of intermediate results while the tunnel is running and there is still time to take any necessary corrective actions. By the time the MDOE researcher exits the tunnel, he will have already performed numerous computations to assess lack of fit, total residual error, signal-to-noise ratio, model prediction adequacy, and so on as the data are being acquired. Furthermore, as a byproduct of carefully formulating specific research questions and quality requirements in advance of the test, the researcher will have already identified much of the analyses to which the data are to be subjected. The software to perform those analyses will typically already have been written and tested a-priori with simulated data and much of it will have been exercised in the execution phase on intermediate results. After the test it is largely a matter of running the final test data through the same software.

In classical OFAT testing the analysis phase can be one of the most labor-intensive elements of the process. It is not uncommon for researchers to have rather incompletely formulated plans for precisely what analysis they intend to perform until they have had the opportunity to "look at the data". Much effort can be squandered in an ad-hoc examination of relationships in the data that the research finds interesting or entertaining but which have little to do with the research questions that the test was conducted to answer. This is especially true when the researcher does not have a clear idea of exactly what the specific research questions were in the first place. In this regard the author has little patience with the protestation often heard from OFAT researchers that a disciplined approach to research constrains "creativity" in the examination of experimental results and forecloses options to discover "interesting things" in the data. In the first place, the curtailment of a certain amount of the kind of "creativity" that sometimes occurs in the analysis of experimental data is not such a bad thing. An undisciplined approach to analysis can leave the researcher vulnerable to his own (generally inevitable) prejudices about what the answer "ought to be". It is characteristic of human nature to find in the numbers confirmation of one's own forecasting brilliance unless some discipline is imposed by defining carefully in the planning process (*before* the data are in hand) what the various decision criteria will be for interpreting the results and how they will be recognized. In the second place, it simply is not true that a carefully defined analysis plan forecloses options to find interesting and relevant but unanticipated results in the data. The author has conducted many designed research experiments without as yet encountering a single one that did not feature surprise results of one kind or another. Far from diverting one's attention from unanticipated nuggets in the data, a rigorous analysis plan tends to highlight patterns inconsistent with a-priori expectations, making them *more* evident than they would otherwise be.

The general procedures used in the design and execution phases of the MDOE research process are common to most experiments. Regardless of the specific objectives, it is usually necessary to define the volume of data and selection of points in the experiment design, and to invoke the same general procedures and precautions in the execution of the test. One will always want to randomize the run schedule to defend against systematic error insofar as it is practical to do so, for example, and to quantify lack of fit, chance variations, and other quality metrics as the data are acquired. Unlike the design and execution elements of the process, the specific details of the analysis sub-process tend to be a function of individual test objectives, but certain general types of analysis can be identified.

General Response Surface Analysis

A response surface is formed when response variables are modeled in terms of all independent variables simultaneously, and a general response surface analysis is the most common type of analysis in

scientific and engineering MDOE experiments. By the time the execution phase is completed, the researcher will have in hand a set of regression coefficients defining graduating functions for every response variable of interest, over one or more limited ranges of the independent variables. These graduating functions will have been checked with a sufficient number of confirmation points for the researcher to have confidence in response-variable predictions made for combinations of independent variables that are *between* the points acquired in the test. Extrapolation beyond the range of independent variables tested is generally a dangerous practice but with an adequate response surface model one should be able to interpolate with confidence. (Additional confirmation points should be acquired in the execution phase until this level of confidence is reached.) Thus, response surface models can be used to generate estimates of response variables (with acceptable uncertainty levels) at any combination of treatment variables within the inference space of the test, not just the ones for which set-points were physically established in the execution phase. In that sense the regression coefficients of the response surface graduating functions can be considered as a compact version of – and entirely equivalent to – an infinite-resolution database of response variable values throughout the entire range of independent variables tested.

Analysis of Derivative Functions

The explicit MDOE focus on modeling facilitates an analysis of derivative functions that can be especially useful in certain aeronautical research applications. For example, it is often desirable to quantify the effect of control surface deflections on various forces and moments on an aircraft model. In such cases the researcher is less interested in absolute roll moment per se, for example, than in the change in roll moment for a unit change in some control surface deflection angle. In other words, it is the derivative of roll moment with respect to control surface deflection that is of primary interest.

The number of discrete configurations that can be set is usually quite limited in OFAT full-factorial designs in which each new control surface setting doubles the volume of data if all combinations of the independent variables are to be examined. The typical OFAT solution is to limit the combination of conditions that is tested at each configuration and to let resource constraints dictate the number of configurations tested in a given experiment. An important contributing reason for this is that configuration changes tend to be labor-intensive and time-consuming, typically requiring the tunnel operating crew to dump flow, open the tunnel, and physically change model components.

In an MDOE experiment in which control surface deflection angles are among the treatment variables, the response surface models will be a function of those variables as well as all others. Common response surface designs such as the Box-Wilson or Central Composite Design described above do not entail a full factorial array of independent variables and are capable of adequately estimating the dependence of various response variables on multiple independent variables – including control surface setting – with a relatively parsimonious data set. The polynomial graduating functions obtained in a typical regression analysis can then be differentiated easily with respect to the deflection angle of various control surfaces. The result is a function that describes the rate of change in roll moment, say, with change in portside outboard elevator deflection angle as a function of angle of attack, sideslip angle, and Mach number. The possibility to perform such operations as differentiation on the original response surface models, generating additional response surfaces in the process, can greatly leverage the results of an MDOE experiment.

Interactions

Because OFAT testing methods require that only one variable be changed at a time, only main treatment variable effects can be directly illuminated. It is necessary to change more than one variable at a time to quantify the all-important *interactions* that influence response variable levels. For example, it is possible with OFAT methods to determine how drag forces vary with changes in angle of attack. This is an example of a typical main effect. But the change in drag caused by a unit change in angle of attack is different at low Mach number than at high Mach number. This is an example of an interaction effect, where the response of the system to changes in one treatment variable depends on the level of another. Important interaction effects are detected in a number of ways. If the response surface is simply plotted as a function of two potentially interacting variables, interactions are easily detected as a twist or warp in the response surface. The relative magnitude of the regression coefficients corresponding to interaction terms also give direct evidence of the relative contribution of interactions to the overall system response.

The wind tunnel is an extraordinarily interactive environment in which coupled relationships among independent variables are the norm, not the exception. And yet the true multivariate nature of flight systems can be only crudely estimated with experimental procedures in which only one variable is changed at a time. Only single-factor effects can be quantified directly. Subtle curvature and interaction effects, which

may have considerable impact on the optimization of flight system designs, can go undetected. Performance characteristics are maintained at sub-optimal levels, increasing cost and surrendering potential competitive advantages.

Patterns

The inference space is typically partitioned into subspaces over which response variables are fit. This is done primarily to reduce the lack of fit errors that would otherwise be unacceptably large if low-order graduating functions were fit over wide ranges of several independent variables at once. This provides an opportunity to gain additional insights into the behavior of the system. Simple bar charts displaying the magnitude of regression coefficients in different inference subspaces can provide surprising clear insights.

Examples will be taken from the most recent MDOE test at Langley as of this writing, which was a test of the X-33 technology demonstrator vehicle in the LaRC 14x22 Subsonic Tunnel on August 5-6, 1997, designated T461. The X-33 is a lifting body design that will be used to demonstrate certain technologies in propulsion, materials, thermal systems, etc, as an intermediate step in the development of a next-generation space transportation system vehicle. While the X-33 design is evolving rapidly, figure 4 illustrates the essential features of the model tested at Langley in T461.

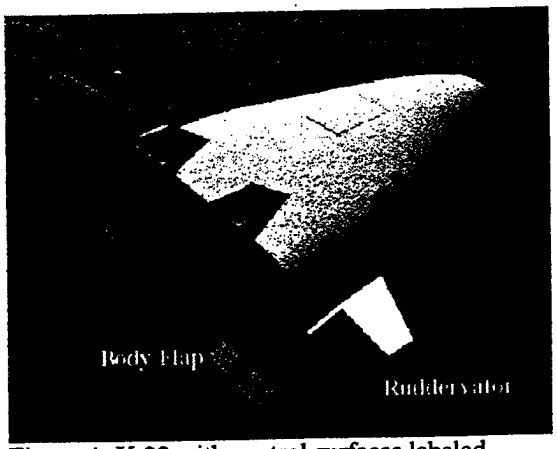


Figure 4: X-33 with control surfaces labeled

Standard aerodynamic forces and moments were measured in this test as a function of angle of attack, height (to assess ground effects), and deflection of the two right-side control surfaces labeled in figure 4 as "body flap" and "ruddervator".[2] The corresponding left-side control surfaces were fixed with zero deflection, as were two rudders mounted on the vertical tail structures shown in figure 3 on top of the vehicle.

Figure 5 displays first-order angle of attack regression coefficients for the X-33 stability axis coefficient of lift, from test T461. This figure illustrates how regression coefficient patterns can provide certain insights. The four bars on the left in this chart correspond to the four combinations of positive and negative control surface deflections for two surfaces, the starboard-side body flap and the starboard-side ruddervator. The four bars on the right are independent replicates on the four on the left.

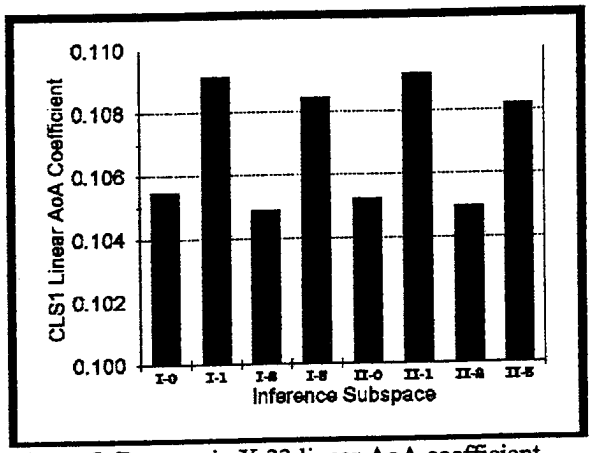


Figure 5: Patterns in X-33 linear AoA coefficient

A clear pattern is evident in the left four bars, which is repeated in the four bars on the right. Within each foursome, the odd-numbered inference subspaces have considerably larger AoA coefficients than the even-numbered coefficients. These happen to correspond to the inference subspaces in which the ruddervator deflections were positive (trailing edge down). Also, within each pair of positive or negative ruddervator subspaces, a subtle effect is clearly repeated; namely, that the lower-numbered subspace has the greater coefficient for each pair. These correspond to negative excursions of the body flap. So an overall picture emerges that a change in ruddervator has a greater effect on lift when the trailing edge of the ruddervator is down than when it is up, and that in either case the effect is greater if the body flap is negative than if it is positive. Because the regression

[2] The ruddervator is so named because the angle of the wing-like structure to which it is attached makes it behave partly as a rudder and partly as an elevator.

coefficients are quantitative, the relative magnitudes of these effects can be easily determined.

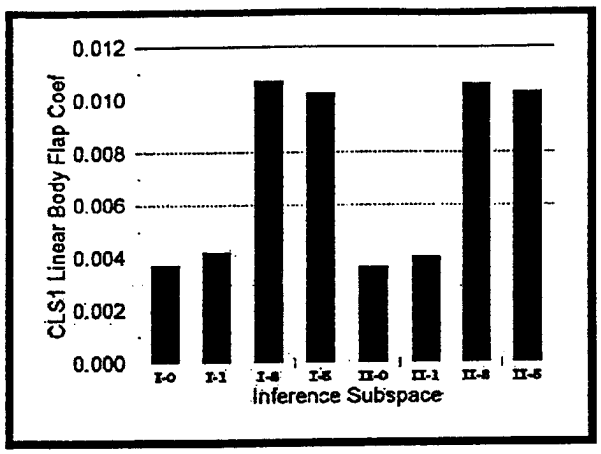


Figure 6: Patterns in X-33 Lift Body Flap Coefficients

Figure 6 is a display of X-33 regression coefficients similar in format to figure 5 but displaying a different pattern. Again the four bars on the left and right are independent replicates of regression coefficients corresponding to all four combinations of positive and negative body flap and ruddervator deflection. First-order body flap regression coefficients for lift are displayed in figure 6, giving the change in stability axis lift coefficient corresponding to a unit change in coded body flap deflection angle for the four control surface regimes. Note that within the left and right foursomes, the two bars on the right are significantly higher than the two on the left. These correspond to positive body flap deflection (trailing edge down). Within each pair of positive or negative body flap subspaces, another subtle effect is clearly repeated. For the case of the negative body flap deflection, the higher-numbered subspace (positive ruddervator deflection) has the greater coefficient. For the case of the positive body flap deflection, the lower-numbered subspace (negative ruddervator deflection) has the greater coefficient. So a pattern emerges in which it is clear that body flap changes have a greater influence on lift when the trailing edge is down than when it is up, and that the effect of body flap changes depends on the ruddervator setting. For positive body flap deflections, a unit change in body flap has a greater effect on lift when the ruddervator is negative (trailing edge up) than when it is positive, and conversely for negative body flap deflections. As is the case with AoA coefficients, the relative effects can be easily determined because of the quantitative nature of the regression coefficients. Also, the orthogonal (Box-Wilson) design employed in the X-33 test ensures that the estimates of AoA and body flap coefficients are not confounded by component contributions from other treatment variables in the design, but rather are "pure" effects.

Identification of Underlying Models

Underlying theoretical non-linear models relating response and treatment variables are generally unknown but their functional form can sometimes be inferred by examining the graduating functions. Consider the case in which a logarithmic transformation of independent variables improves the model. That is, assume that a better fit is obtained when the response is fitted in terms of the log of certain independent variables rather than the variables themselves. Suppose one observes that the magnitudes of the regression coefficients of two such transformed variables are nearly the same. If the regression coefficients are statistically significant at a customer-specified significance level, the numerical values do not have to be identical but simply close enough so that no significant difference exists between them given the standard errors in the coefficients. In such a case, the terms can be combined as the logarithm of a ratio or a product. For example, consider a well-fitted graduating function of the form $y = 1.98z_1 - 2.03z_2$ where $z_1 = \ln(x_1)$ and $z_2 = \log(x_2)$. It is not unreasonable to conclude that the true underlying relationship is probably of the form $e^y = (x_1/x_2)^2$.

In a recent MDOE experiment at Langley involving an advanced slender wing-body-tail configuration, a better model for normal force coefficient could be obtained by trigonometric transformations of certain model attitude variables, including angle of attack. The inference subspace was partitioned into smaller regions as is customary to improve fit and it was observed that at high angles of attack the normal force coefficient's graduating function contained a significant $\sin(\alpha)$ term. At low angles of attack the $\sin(\alpha)$ term was replaced by an interaction term between $\sin(\alpha)$ and $\cos(\alpha)$. That is, at high α the normal force featured a $\sin(\alpha)$ term but at low α that term became $\sin(\alpha)\cos(\alpha) \sim \sin(2\alpha)$. This suggests that the true underlying relationship between normal force coefficient and angle of attack may feature a term of the form $\sin(\alpha + \alpha_0)$, where α_0 is some small constant. At high α, the α_0 term is insignificant but at low α (where $\alpha \sim \alpha_0$), there is a better fit with a $\sin(2\alpha)$ term than a $\sin(\alpha)$ term. This illustrates how low-order polynomials, especially in suitably transformed treatment variables, can guide the researcher to

consideration of more complex underlying non-linear treatment-response relationships.

Multiple Objective Optimization

It is not uncommon for aircraft designers to want to satisfy several objectives at one time. For example, it may be desirable to determine combinations of control surface settings that simultaneously satisfy multiple constraints or achieve multiple objectives. One may wish to know what configuration simultaneously provides lift greater than some lower limit, drag less than some upper limit, and a rate of change of roll moment with control surface deflection angle that exceeds some threshold, for example. In analogous industrial applications this is known colloquially as "sweet spot analysis", in which the manufacturer seeks some combination of operating conditions that optimizes operations by some criterion. All combinations of independent variables that simultaneously satisfy all conditions and constraints are said to reside in the "sweet spot" of the inference space. Sweet-spot analysis has the potential for being a useful analytical tool in flight system design when coupled with MDOE testing methods.

Facility Performance Assessments

A rich array of analytical methods can be brought to bear to assess the state of the wind tunnel facility itself from information obtained in a formally designed experiment. These include augmentations of the original regression models to include qualitative "decision" or "dummy" variables to distinguish between data points acquired in one block of time from those acquired in an earlier or later block. By assessing the statistical significance of regression coefficients associated with these variables, it is possible to determine if significant "block effects" exist in the facility. Block effects are present when levels of the response variables are different in one block than another, even when the treatment variables are the same. This can be caused by countless effects, including drift in the instrument and data systems, thermal effects in the tunnel, operator fatigue, and so on. By arranging the blocking variables to be orthogonal to the rest of the design, it is possible in some circumstances to recover the coefficients of the graduating functions as if the blocking effects were not present. This has considerable potential for ameliorating between-facility differences and ultimately differences between ground testing and flight.

Summary of Experiences with MDOE in LaRC Wind Tunnel Tests

While the advantages of MDOE experimentation over classical OFAT high-volume data collection have been recognized by elements of the Langley testing technology community for some time, relatively recent external economic pressures have provided a catalyst for initiating this inevitably difficult transition to a modern design laboratory. An initial pilot MDOE wind tunnel experiment was conducted in January of 1997 in which model deformation was quantified as a function of angle of attack, Mach number, and Reynolds number in the Langley Unitary Plan Wind Tunnel using an advanced supersonic transport model. This test was conducted in both the classical OFAT tradition and using MDOE methods. The OFAT design featured 330 data points. The corresponding MDOE design required only 20 data points to obtain information of comparable or higher quality (as assessed in terms of 95% confidence interval half-widths). The MDOE process relies on randomization as an error control mechanism and therefore the average elapsed time per data point is greater than with the OFAT method in which data points are acquired in sequences that maximize data acquisition rate. But because of the substantial reduction in the number of data points required, the overall elapsed time is significantly reduced. There were 60% fewer wind-on minutes in the MDOE version of this initial model deformation test than in the OFAT version, with corresponding reductions in the levels of consumables such as electrical megawatt hours.

This initial success garnered the support of Center management for a series of additional tests of expanding complexity in a number of Langley facilities. A total of four MDOE experiments were conducted throughout the rest of the year. In addition to the initial model deformation experiment these included an experiment to quantify forces and moments as a function of angle of attack and Mach number across configuration changes in a high speed research model, a test to assess axial and normal forces as a function of model attitude and remotely commanded control surface changes on an advanced-concept commercial jet transport, and an experiment to quantify ground effects on stability axis force and moment coefficients as a function of angle of attack, sideslip angle, and the deflection angles of two control surfaces on an X-33 single-stage-to-orbit (SSTO) space transportation system technology demonstrator model. These experiments were conducted in a variety of Langley tunnels: the supersonic Unitary Plan Wind Tunnel (UPWT), the subsonic 14x22 tunnel, and the transonic 16-Ft tunnel. Two more MDOE experiments have been designed by the author for execution in January 1998. These are both supersonic stability experiments

designed to quantify roll and yaw moment as a function of the angles of attack and sideslip at fixed Reynolds numbers and Mach numbers of special interest. One involves a recent high speed research model configuration and one is for a research porous leading edge delta wing configuration.

The purpose of the tests conducted throughout 1997 has been to demonstrate MDOE test methods and results for the Langley wind tunnel community and to quantify resource savings afforded by this method in comparison with OFAT testing methods. Each test consisted of an extension of one to two days of a previously scheduled OFAT test. Evaluation ground rules requested by the author included no negative impact on the previously scheduled test, which in each instance was conducted in its entirety with no deviation from its original test plan. Some subset of each original OFAT test was reproduced in each MDOE extension, and resources required for both methods were carefully metered. To prevent inadvertent biases from affecting the resource assessments, personnel from the research facilities "kept the books" for the author, recording for both the OFAT and MDOE version of each test the wind-on minutes, megawatt hours consumed, and number of data points.

While the results varied somewhat from test to test, substantial resource reductions were observed in every instance. The MDOE method reduced wind-on minutes by an average of 64% across all tests. The minimum saving was 60% and the maximum was 67%, so the distribution was relatively narrow. Electric power consumption was reduced by 55% in the MDOE method relative to the OFAT method. The fact that this is somewhat less than the savings in wind-on minutes is attributed to the fact that the MDOE method, with its reliance on randomization, featured very non-systematic changes in model attitude, fan speed, etc, which may have contributed to additional power requirements. In this regard it is noted that additional long-term maintenance costs which may accrue from some of the MDOE execution tactics are not known and therefore are reflected in these resource comparisons. The greatest resource saving associated with the designed experiment approach was in data volume, where 82% fewer data points were required to obtain results of equal or better quality than with the OFAT high-volume data collection method. This last saving is especially significant since it reflects not only reduced execution costs in time and money but also substantially less cost of analysis. In the full-cost accounting environment of the 21st century, the reduction in salary costs for highly skilled (and thus highly paid) analysts can be significant if they are required to deal with substantially smaller data sets, not to mention contributions to the reduced production cycle time due to shortened analysis times. Data storage costs can also be expected to be less.

The author has been pleasantly surprised by the magnitude of the savings quantified in a year of MDOE testing at Langley Research Center. However, notwithstanding the effort to test the method in a wide range of facilities and for a wide range of test objectives, he is reticent to guarantee comparable savings on average for general wind tunnel testing until/unless considerably more testing has been done. Having said that, experience to date suggests that it is not unreasonable to anticipate that MDOE methods can reduce wind-on minutes by a third to a half (i.e. to a range of about 50% to 67% of the OFAT wind-on minutes), and that wind-on costs can be reduced by a quarter to a third. These forecasts are well below savings actually achieved to date in pilot tests at Langley, and represent what is believed to be a conservative estimate of what might be expected in the conversion from an OFAT high-volume data collection test philosophy to a test philosophy based on formal experimentation methods.

Concluding Remarks

MDOE methods have been fully developed since before World War II. Their adoption in aeronautical research applications represents a relatively low-risk proposition in that there is nothing inherent in the requirements of aeronautical research that taxes the method. On the contrary, wind tunnel applications with their relatively small number of clearly identified treatment and response variables, and generally well-understood cause-effect relationships, may be said to represent rather an easier application of the method than many medical and sociological problems in which it has been successfully applied, for example. (Compare the problem of quantifying drag as a function of Mach number and angle of attack, to the problem of quantifying heart disease risks as a function of all factors thought to be related.) Advocates of modern design are therefore obligated to explain why the method is no more widely used than it is in aeronautical research applications (which is to say, essentially not at all).

One of the contributing reasons is no doubt that aeronautical research laboratories have not been incentivised properly to seek efficient testing methods. In the private sector the major airframe manufacturers may spend a million dollars on a wind tunnel test in which they could save half with an efficient design, but they will use this information to build a product that will sell for $100 million dollars each. The potential savings have been in the noise floor, and would require an unacceptable culture shift. In the public sector, the taste for fiscal responsibility may be keenly felt but it is

rather recently acquired; the federal government has not traditionally been among the leaders in cost consciousness, not to put too fine an edge on the point. MDOE tends to be embraced first by those who need it most, and historically this has not included organizations with pockets as deep as those who do aeronautical research. Times are changing, however, and in this era of increasing competitive and economic pressures, a half a million dollars saved here and there may start to look more attractive. Beyond the cost savings, though, aeronautical research laboratories are likely to look at the technical benefits of designed experiments independent of their costs. It is the author's opinion that the technical advantages of the MDOE method would justify paying a substantial premium over the OFAT method and the fact that the MDOE method costs less is a significant but secondary consideration.

Another important reason that MDOE methods are not widely used is that aeronautical researchers have been able to produce such spectacular results without them. The number of advances made in the science of flight since Kitty Hawk boggles the mind, and essentially every one of them has been made without the benefit of modern design methods. Having said that, the pressure of ever increasing performance requirements is a strong driver to seek alternatives to procedures that may have carried us as far as they can. Past success provides no guarantee, and progressive research laboratories are not likely to ignore forever methods that can only be helpful.

It is the author's view that decisions to transition from one-factor-a-time high volume data collection to methods based on the modern design of experiments have already been made for all of us. They have been imposed by the same external, global forces that have been compelling all public- and private-sector interests to downsize, become more efficient, and compete more efficiently for the last decade. The only unresolved questions are who will lead, and who will follow in this transition. Those who recognize earliest the competitive advantages that accrue from well-designed experiments will benefit the most from them. The leaders will enjoy a competitive advantage that is likely to be brief, but may be significant.

Acknowledgements

The author is pleased to acknowledge his debt to many colleagues at NASA Langley Research Center. First among these is Mr. Alpheus Burner, who had the courage to surrender half of his allotted test time to an unscheduled MDOE experiment in the January 1997 model deformation test that lead ultimately to such widespread management support for further evaluation of MDOE methods at Langley. The essential contributions of the test engineers in Langley's Research Facilities Branch are gratefully acknowledged. Special thanks go to Messrs. Gary Erickson, Wes Goodman, and Dan Neuhart, who served as primary liaison between the author and the wind tunnel community in the design and execution of tests at UPWT, 16-FT TT, and 14x22 ST, respectively. The author is also grateful for many hours of stimulating exchange with Dr. Michael J. Hemsch of Lockheed Martin Engineering & Science Services of Hampton, VA.

References

1. Fisher, R. A. (1966). *The Design of Experiments*, 8th ed. Edinburgh: Oliver and Boyd.

2. Box, G. E. P., and N. R. Draper: *Empirical Model Building and Response Surfaces.* New York: John Wiley & Sons. 1987

3. Myers, R. H., and D. C. Montgomery: *Response Surface Methodology. Process and Product Optimization Using Designed Experiments.* New York: John Wiley & Sons. 1995

4. Draper, N. R., and H. Smith: *Applied Regression Analysis*, 2nd ed. New York: John Wiley & Sons. 1981

5. Montgomery, D.C., and E. A. Peck: *Introduction to Linear Regression Analysis*, 2nd ed. New York: John Wiley & Sons. 1992

6. Box, G. E. P., and K. B. Wilson (1951). On the experimental attainment of optimum conditions, *J. Roy. Stat. Soc., Ser. B*, 13, 1.