AIAA 2000-0828

A Factorial Data Rate and Dwell Time Experiment in the National Transonic Facility

R. DeLoach
NASA Langley Research Center
Hampton, VA

38th Aerospace Sciences Meeting & Exhibit
10–13 January 2000
Reno, Nevada

A FACTORIAL DATA RATE AND DWELL TIME EXPERIMENT IN THE NATIONAL TRANSONIC FACILITY

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

Abstract

This report is an introductory tutorial on the application of formal experiment design methods to wind tunnel testing, for the benefit of aeronautical engineers with little formal experiment design training. It also describes the results of a study to determine whether increases in the sample rate and dwell time of the National Transonic Facility data system would result in significant changes in force and moment data. Increases in sample rate from 10 samples per second to 50 samples per second were examined, as were changes in dwell time from one second per data point to two seconds. These changes were examined for a representative aircraft model in a range of tunnel operating conditions defined by angles of attack from 0° to 3.8°, total pressure from 15.0 psi to 24.1 psi, and Mach numbers from 0.52 to 0.82. No statistically significant effect was associated with the change in sample rate. The change in dwell time from one second to two seconds affected axial force measurements, and to a lesser degree normal force measurements. This dwell effect comprises a "rectification error" caused by incomplete cancellation of the positive and negative elements of certain low frequency dynamic components that are not rejected by the one-Hz low-pass filters of the data system. These low frequency effects may be due to tunnel circuit phenomena and other sources. The magnitude of the dwell effect depends on dynamic pressure, with angle of attack and Mach number influencing the strength of this dependence. An analysis is presented which suggests that the magnitude of the rectification error depends on the ratio of measurement dwell time to the period of the low-frequency dynamics, as well as the amplitude of the dynamics. The essential conclusion of this analysis is that extending the dwell time (or, equivalently, replicating short-dwell data points) reduces the rectification error.

Introduction

There is considerable emphasis on minimizing data acquisition time in the National Transonic Facility (NTF) at NASA Langley Research Center (LaRC) because of the relatively high cost of cryogenic operations required to achieve flight Reynolds numbers. The NTF data system currently uses a sample rate of 10 samples per second, for a one-second dwell time per data point. The resulting 10 frames of data are averaged to produce a single point for each channel of data. The selection of sample rate and dwell time was based in part on a study of electrical noise reduction techniques performed for NASA in 1988 by Wyle Laboratories.[1] The one-second dwell and 10-samples/second rate were identified as the most economical combination that would produce acceptable data quality in the presence of line noise and other sources of variation in the instrumentation and data system that existed at that time. Because numerous changes have been made in the data system and instrumentation in the years since this study, NTF personnel decided to reevaluate the sample rate and dwell time settings. Research customers of the NTF have also requested higher sample rates and longer dwell times. The potential benefits, if any, of such proposed changes in sample rate and dwell time have not been as easy to quantify as the costs, which for cryogenic operations are known to be high. So it has not been clear whether the contemplated changes in data system sample rate and dwell time would represent an attractive tradeoff between costs and benefits. The experiment described in this paper was conducted to address this question by examining the effects on forces and moments of increasing the sample rate by a factor of five and the dwell time by a factor of two.

There is little time available in the NTF schedule to devote to process experimentation because of the demands of research applications which comprise the facility's raison d'être. This circumstance is not unique to NTF, but is common in many industrial applications where facility operating time is precious and operating cost is high. Formally designed experiments are especially useful in such situations. They employ highly efficient experimental techniques capable of providing reliable answers from relatively small data sets. For this reason, and as part of a larger effort at Langley Research Center to evaluate the utility of formally designed experiments in wind tunnel research,[2] NTF personnel requested that the LaRC Experimental Testing Technology Division (ETTD) design a relatively compact experiment to examine the rate/dwell alternatives under consideration. The objective of the experiment was to evaluate these combinations of sample rate and dwell time over a representative range of tunnel operating conditions, while minimizing the NTF schedule interruption required for such an evaluation. This paper describes this effort, and provides some tutorial information on an evolving approach to wind

* Senior Research Scientist

tunnel research that is centered on the design, execution, and analysis of formally designed experiments.

The first section describes the design of the replicated, five-variable, two-level factorial experiment used in this study. This section includes a discussion of considerations that influenced the *scale* of the experiment (volume of data specified), and defines the quantitative criteria by which potential rate and dwell effects were evaluated. The second section outlines the execution of the experiment, touching on certain procedures used to simplify and expedite the data acquisition. The next section describes the analysis procedures and key results. An unanticipated discovery related to dwell effects is analyzed in the following section. Finally, concluding remarks address the consequences of the rate/dwell changes under consideration.

Design of Experiment

The purpose of this section is to describe the formal design of the NTF Rate/Dwell experiment in a way that provides some tutorial information on the general process of formal experiment design.

Conventional test design

Formal experiment design procedures might best be described by contrasting them with conventional test techniques widely practiced in late-20th-century wind tunnel testing. Conventional or "classical" experiment design procedures are known in the literature of experiment design as One Factor at a Time (OFAT) methods. This term derives from a shared characteristic of such experiment designs; namely, the practice of holding all independent variables constant save one, and changing that one variable systematically over some prescribed range. Any experimental aerodynamicist who has ever structured a wind tunnel test in terms of angle-of-attack "polars" is familiar with OFAT testing.

OFAT methods are popular in wind tunnel testing because they permit data to be acquired at the fastest possible rate. This generates the greatest volume of data for a fixed amount of resources (wind tunnel time, liquid nitrogen budget, etc.). High rates of data collection ("polars per hour", etc.) and high total volumes of data ("polars per facility per year") are popular productivity metrics in late-20th-century wind tunnel testing.

While OFAT testing does generate data in high volume, it has important shortcomings. A thorough description of the relative merits of formal experiment design and OFAT testing is beyond the scope of this paper, but a rich literature on the subject exists. References 3–5 are general texts on formal experiment design. Reference 6 compares OFAT methods and formal experiment design techniques in terms of wind tunnel research quality.

To structure the NTF Rate/Dwell experiment as a classical OFAT test, one might proceed as follows: A likely early step would be to identify a range of sample rates to study, including some convenient step size. For example, sample rates of 10, 20, 30, 40, and 50 samples per second might be selected. Likewise, a number of dwell times would be selected; say, 1.00, 1.50, and 2.00 seconds. For each combination of rate and dwell, the OFAT designer would probably prescribe a standard angle of attack polar (15 AoA levels, typically). Polars at these rate/dwell combinations would typically be repeated for some relevant set of Mach numbers; say, 0.7, 0.8, 0.82, 0.9. If total pressure (P_t) was of interest, then these points would be repeated for some reasonable number of P_t levels, say, four. This design would feature 15 combinations of rate and dwell investigated at 240 combinations of AoA, Mach, and P_t, for a total of 3,600 data points. While specific OFAT test designs would vary from this one in the specific ranges and increments of the independent variables, the scale of this design does not entail an atypical volume of data for 1990s wind tunnel testing.

If additional resources were available (more tunnel time, say), then additional data points might be considered, consistent with general OFAT productivity concepts equating high data volume with high productivity. An additional Mach number could be added to the OFAT design if resources were available to support 4500 points instead of 3600, say. Similarly, if fewer resources were available than a 3600-point design requires, some portion of the original test matrix might be eliminated. If one of the P_t levels were dropped, for example, the scale of the experiment could be reduced from 3600 points to 2700 points.

Comments on research productivity

The OFAT design process generally features the kind of trade-off decisions described above, in which the test matrix is scaled to produce the largest number of independent variable combinations that available resources will permit. This approach is intended to ensure that the researcher obtains the greatest amount of information possible, given existing resource constraints. On the face of it, this approach does not seem at all unreasonable. Why *not* acquire the greatest volume of data that available resources will permit? What is there to recommend the acquisition of *fewer* data points than resources permit?

Practitioners of formal experiment design would respond to these questions by noting that there are in fact *two* ways to optimize the unit cost of information from an experiment. The classical OFAT procedure of maximizing the data volume for a fixed test budget involves increasing the denominator of the "cost-benefit ratio" while holding the numerator fixed. But it is also possible to maximize this ratio the other way—by minimizing the numerator for a given denominator. That is, it is possible to define a specific benefit to be derived from an experiment—some specific information that is being sought, say—and then to design the experiment in such a way as to achieve that specifically-defined benefit with the smallest possible cost. In practical terms, minimizing the cost of data acquisition generally translates into minimizing the scale of the experiment—the volume of data acquired. Reference 7 describes the matching of data volume requirements to specific technical objectives, and this important idea will also be discussed below.

In some cases the operational costs might be negligible compared to the initial set-up costs, in which case there is little to recommend restraint in the scale of the experiment. If data points are essentially "free" after the set-up costs have been incurred, then the more data the better. This is seldom the case in a full-cost accounting environment, however. For example, the process of bringing a new airplane to market typically requires enormous sums of capital to be committed. The daily cost of that capital can dwarf more visible test operating costs such as electric power consumption for the fan motor or even the cost of liquid nitrogen for cryogenic high Reynolds number research. Other potential costs of extending the test cycle can be even greater. The cost of bringing a product to market after a competitor has already captured that market with an earlier entry can be substantial. There is only one 747.

Selection of variables

Formal experiment designs are generally structured in terms of "treatment variables" (more commonly called "independent variables" outside of formal experiment design circles) and "response variables" (read "dependent variables"). The 'treatment" and "response" notation is rooted in the early history of formal experiment design, in which agricultural and medical experiments were among the first successful applications of the method.

In the case of the NTF Rate/Dwell experiment, sample rate and dwell time are readily identifiable as candidate treatment variables. Angle of attack, Mach number, and total pressure were also selected as treatment variables. Adding these tunnel state variables to the experiment design achieved two benefits. First, it enabled the examination of possible *interactions* between rate/dwell effects and the tunnel state variables. That is, it enabled NTF to examine if rate or dwell effects were more pronounced at one combination of AoA, Mach, and P_t than at another. It also provided what is known in the language of formal experiment design as a "wider inductive basis" for any conclusions that might be drawn about changes in sample rate or dwell time. That is, rather than examining the rate and dwell effects at a single combination of tunnel conditions (cruise conditions for a typical commercial jet transport, say), we seek to examine these effects over some meaningful range of conditions. This increases confidence about the range of tunnel conditions over which general rate/dwell conclusions might be applied.

In an initial experiment such as this, it is usually neither necessary nor desirable to set multiple levels of the independent variables. At this stage we are simply asking whether a rate effect or dwell effect exists at all. Setting only two levels of the independent variables, a relatively low level and a relatively high one, will produce the most unambiguous answers to these questions. If we are unable to detect an effect over relatively wide changes in rate and dwell, there is no need to incur the cost of examining small incremental changes. Since the purpose of the experiment was to determine if increases in the sample rate and dwell time had any significant effect on the data, "low" levels for rate and dwell were selected to be the existing levels of 10 samples per second and one second, respectively. The following "high" levels were selected: 50 samples per second and two seconds. Two levels were selected for each of the three tunnel state variables as well: 0° and 3.8° for angle of attack, 0.52 and 0.82 for Mach number, and 15.0 pounds per square inch (psi) and 24.1 psi for total pressure. These levels represent a practical compromise between a relevant range of tunnel states to examine and our desire to change combinations of variables quickly.

After some discussion, axial force, normal force, and pitching moment were identified as specific response variables for which the experiment would be designed. Further discussions identified axial force as the response variable that NTF test engineers believed would be most likely to be susceptible to rate/dwell effects if they existed.

Statement of objectives

In formally designed experiments, it is usually convenient to express specific objectives in terms of "null hypotheses" and "alternative hypotheses." In this experiment, the null hypothesis for axial force sample rate, for example, is that there is no significant difference between axial force measurements made with a sample rate of 10 samples per second and axial force measurements made at 50 samples per second. The alternative hypothesis is that axial force measurements made at 10 samples per second and 50 samples per second are different. We can write these symbolically as follows:

$$H_0 : \mu_{AF,10} = \mu_{AF,50}$$
$$H_A : \mu_{AF,10} \neq \mu_{AF,50}$$

where $\mu_{AF,10}$ represents the mean of some number of axial force measurements made at a sample rate of 10 samples per second, and $\mu_{AF,50}$ represents the mean of some number of axial force measurements made at a sample rate of 50 samples per second. The corresponding hypotheses for axial force dwell time are:

$$H_0 : \mu_{AF,1} = \mu_{AF,2}$$
$$H_A : \mu_{AF,1} \neq \mu_{AF,2}$$

with obvious notational changes. Hypothesis pairs similar to these for axial force are constructed for the other response variables of interest as well.

Null and alternative hypotheses provide a compact means of describing the objectives of the experiment. Because it is relatively easy to disprove a null hypothesis when it is false, this is the normal line of attack. We reject the alternative hypothesis only if we are unable to disprove the null hypothesis, and we accept the alternative hypothesis only if we can disprove the null hypothesis.

Resolution and inference error risk tolerance

This focus on "proving" and "disproving" formal hypotheses leads to the concept of inference error risk, which must also be taken into consideration during the experiment design process. By definition, the existence of a rate or dwell effect means that if force and moment measurements are made at one sample rate or dwell time and then repeated later at a different sample rate or dwell time, they will be different the second time than the first. Unfortunately, forces and moments measured later will be different *even if the sample rate and dwell time are unchanged*, simply because of ordinary chance variations in the data (measurement error). Therefore, "zero change" cannot be the criterion for rejecting the existence of rate and dwell effects, nor can the existence of such effects be established simply on the basis of non-zero differential measurements. Since "there are no zeros" in real-world differential measurements (except by pure chance), it is necessary to define the size of rate or dwell effect that is important. This introduces an important formal design concept called the "resolution" of the experiment design, and the companion concept of "inference error risk tolerance." Both concepts are crucial to defining the *scale* of the experiment, or the volume of data required.

Because chance variations are usually just as likely to cause positive errors as negative ones and both are typically of similar magnitude, replication can increase the precision of experimental results because the positive and negative errors will tend to cancel as larger numbers of points are averaged. This suggests that we can achieve arbitrarily high precision levels if we are willing to incur the cost of a sufficiently high volume of data, assuming the means are stable. Equivalently, this implies that we can quantify the volume of data required for some specifically prescribed level of precision.

If rate and dwell changes caused effects in the forces and moments that were large compared to the uncertainty in making individual force and moment measurements, the effects would be obvious and there would be little need for a special experiment to investigate their existence. From the perspective of an experiment designer, however, the opposite extreme is much more interesting. What if the rate/dwell effects are real but subtle? That is, what if they are large enough to be a troublesome component of the total experimental error, but not so much larger than the uncertainty associated with individual response measurements that they are readily detected? It is in these kinds of circumstances, when it is necessary to make subtle inferences with high confidence, that formal experiment design methods are especially useful.

After discussions of the potential significance of rate/dwell effects, a consensus emerged that systematic changes of 0.20 pounds or more in axial force, 2.0 pounds or more in normal force, and 5.0 inch-pounds or more of pitching moment would constitute response changes large enough to be important. Changes less than these amounts would be treated as indistinguishable from zero for practical purposes. That is, if a sample rate change from 10 sec^{-1} to 50 sec^{-1} caused the axial force to change by less than 0.20 pounds, no rate effect would be inferred for axial force. Likewise, no dwell effect would be inferred if a dwell time change from one second to two seconds caused the axial force to change by less than 0.20 pounds. Similar statements apply to normal force and pitching moment.

Because of experimental error there is always some probability, however small, of making an inference error. For example, experimental error might lead us to conclude that a sample rate effect exists when in fact it does not (a "Type I" or "alpha" error in the language of formal experiment design). In terms of our formal statement of objectives, this type of inference error would cause us to accept the alternative hypothesis when the null hypothesis was in fact true. Conversely, we might fail to detect a true effect (a "Type II" or "beta" error), leading us to accept the null hypothesis when the alternative hypothesis was the correct choice.

If we conclude erroneously that there is a dwell effect, the facility operators may be influenced to increase the time over which all future data points are acquired. This would reduce data production rates and increase costs unnecessarily. On the other hand, if we fail to detect a true effect, the quality of all future research results might be adversely impacted. Each of the two types of inference errors therefore has consequences, but the impacts may be different. These considerations influence the (necessarily non-zero) inference error probabilities that must be established as acceptable in the experiment design in order to scale the experiment properly.

After considerable discussion, a consensus emerged that inference errors of the second type (failing to detect effects that could compromise the integrity of the NTF data) were more serious than Type-I errors, which could lead to higher operating costs but which would not degrade data quality. An inference error probability of 0.05 was finally selected as acceptable for Type-I errors. That is, we wish to acquire a sufficient volume of data to ensure 95% confidence in our result if we infer that there is no significant rate or dwell effect. (Note that with a finite volume of data, we cannot establish that no effect exists at all, however small. If we accept the null hypothesis, we mean simply that if an effect does exist, it is smaller than the resolution for which we have designed the experiment. For example, if we accept the null hypothesis for axial force sample rate, we mean simply that if a change from 10 sec^{-1} to 50 sec^{-1} changes the axial force at all, the change is less than 0.2 pounds, and similarly for dwell.) For Type-II errors, 0.01 was selected as the acceptable level of inference error risk. That is, we want to acquire sufficient data to ensure that there will be at least a 99% probability of detecting any rate or dwell effects that actually do exist.

Three points are especially relevant to this discussion of inference error risk. The first is that acquiring additional data can reduce the probability of making an inference error. The second is that inference error probabilities cannot be driven to zero with any finite volume of data, no matter how large. The third is that data points are not free. These three points lead to a useful interpretation of the cost of data acquisition; namely, that it is *the premium we pay for insurance against inference errors*. It is dangerous to be under-insured, but it is costly (not to say foolish) to be over-insured as well. We seek a volume of data that will give ample protection against inference errors—certainly no less, but also no more data than that.

Data volume estimates

The volume of data necessary to drive inference error probabilities to acceptably low levels can be

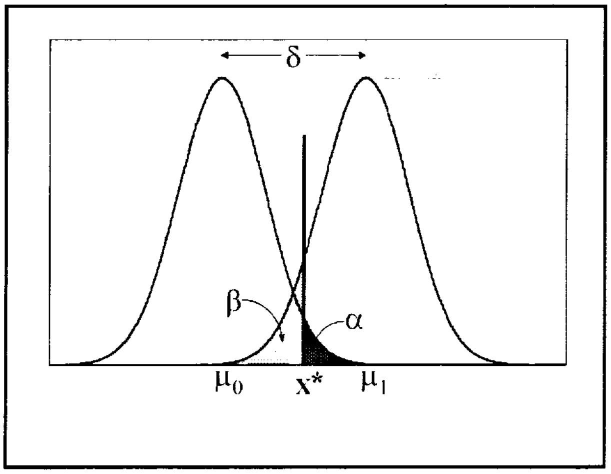

Figure 1. Distribution of mean effects.

estimated with the aid of figure 1. Let us assume that we have acquired some number of differential response measurements made at sample rates of 10 sec^{-1} and 50 sec^{-1}. That is, we have acquired some number of pairs of points, one at each of the two sample rates, and have subtracted the response acquired at 10 sec^{-1} from the response acquired at 50 sec^{-1} for each of the pairs of points. Each such differential measurement is an estimate of the sample rate effect. We average the number of rate effects that we have measured.

If the distribution on the left in figure 1 is centered on $\mu_0=0$, this would represent the distribution of rate effects if the null hypothesis were true—no change in mean response associated with a change in sample rate. The distribution on the right, centered δ units away at μ_1, would represent the distribution of rate effects if the average rate effect was δ. If δ is greater than or equal to the minimum response difference we have defined as comprising an important rate effect, the alternative hypothesis would be said to be true.

The quantity x* in figure 1 represents a threshold criterion by which we will infer whether an important effect has or has not occurred. If the mean of all the individual effects we measure exceeds the threshold criterion, x*, we will accept the alternative hypothesis (effect exists). Otherwise, we will accept the null hypothesis (no effect). The area of the left distribution to the right of x* in figure 1, labeled α, represents the probability of committing a Type-I inference error—claiming an effect when there is none. The area under the right distribution that is to the left of the x* criterion, β in figure 1, represents the probability of committing a Type-II inference error—failing to detect a true effect of δ.

Since figure 1 represents distributions of sample *means*, the variance (width) of both distributions in the figure will decrease with increasing data volume. This will result in corresponding reductions in the areas labeled α and β. That is, increasing the volume of data will decrease the probabilities of committing both types of inference errors, as expected. We wish to know the volume of data for which α=0.05 and β=0.01 in this figure when δ just does equal the size of effect that we consider important. This will represent the minimum volume of data necessary to resolve an effect large enough to be important, consistent with our inference error risk tolerances. The experiment will be a failure if we acquire less than this minimum volume of data because we would not be able to report our conclusions with an acceptably small probability of inference error (α or β).

Note, however, that acquiring any *more* than this amount of data represents wasted resources, because it incurs more costs than necessary to drive the inference error probability to acceptably low levels. Having defined inference error probabilities of 0.05 and 0.01 as acceptable, there is no need to "pay the premium" for additional insurance against inference error. We might decide to "err on the side of the angels" by acquiring somewhat more data than the absolute minimum needed to be consistent with our inference error risk tolerance, but there is no need to substantially exceed this amount simply because all available resources are not yet consumed. It is from this perspective that the OFAT inclination to arbitrarily maximize data volume (in the name of "productivity", no less), is seen as most suspect.

We compute the optimum volume of data by representing the criterion level, x*, in terms of its distance both from μ_0 and from μ_1 in figure 1:

$$x^* = \mu_0 + \frac{t_\alpha \sigma_e}{\sqrt{N}} \tag{1}$$

$$x^* = \mu_1 - \frac{t_\beta \sigma_e}{\sqrt{N}} \tag{2}$$

where σ_e represents the standard deviation in the sample distribution of individual effects, t_α and t_β are tabulated t-statistics associated with the α and β inference error probabilities respectively, and N is the number of rate effect estimates. Note that $\sigma_e/\sqrt{N}$ is simply the standard error in the distribution of mean rate effects, which the distributions in figure 1 represent.

Subtract equation (1) from equation (2), noting that $\delta = \mu_1 - \mu_0$:

$$\delta = \frac{\sigma_e}{\sqrt{N}}\left(t_\alpha + t_b\right) \tag{3}$$

Solve this for N:

$$N = \left(t_\alpha + t_b\right)^2\left(\frac{\sigma_e}{\delta}\right)^2 \tag{4}$$

Recall that figure 1 represents distributions of mean effects, where an effect is defined as the difference between responses measured at two different levels of some treatment variable such as sample rate or dwell time. Let y_{-1} and y_{-1} represent response measurements acquired at low and high levels of some treatment variable, and let y represent the effect estimate obtained by subtracting y_{-1} from y_{+1}. Then:

$$y = y_{+1} - y_{-1} \tag{5}$$

The variance in the rate effect estimate, y, can be described in terms of the variance in y_{-1} and y_{+1} by employing standard error propagation methods:[8]

$$\sigma_e^2 = \left(\frac{\partial y}{\partial y_{-1}}\right)^2 \sigma^2 + \left(\frac{\partial y}{\partial y_{-1}}\right)^2 \sigma^2 \qquad (6)$$

where the unsubscripted σ represents the standard deviation in individual response measurements. The two squared derivatives each have a value of 1, so that (6) reduces to

$$\sigma_r^2 = 2\sigma^2 \qquad (7)$$

which, when inserted in (4), yields:

$$N = 2\left(t_\alpha + t_\beta\right)^2 \left(\frac{\sigma}{\delta}\right)^2 \qquad (8)$$

In equation (8), N represents the number of *effects* that must be averaged to satisfy inference error requirements. That is, N represents the number of *pairs* of response measurements that are to be made.

The quantities α, β, and δ in equation (8) are specified by the researcher. The quantity σ is an inherent characteristic of the measurement environment, which is provided by facility operators or estimated from previous experience. Table I lists pre-test standard error forecasts (σ) and resolution requirements (δ) for the three response variables of interest; namely, axial force (AF), normal force (NF), and pitching moment (PM). Also listed in Table I are the standard errors that were measured in this test, confirming that forecasts made by NTF test personnel based on their experience and examination of data from similar previous tunnel entries were quite adequate for data scaling purposes. The quantities α and β, specified as 0.05 and 0.01 respectively, have corresponding t-statistics of t_α=1.960 and t_β=2.327. These numbers were used to compute the data volume requirements in Table I, which represent the number of high-low *pairs* of each independent variable level to be set, based on results from equation (8) rounded to the next largest integer.

The t-statistic for β is always drawn from a double-sided table of t-statistics. We obtain t_α from either a single-sided or double-sided table, depending on whether the alternative hypothesis is single-sided or double-sided. For this experiment, the alternative hypothesis states that responses measured at high/low levels of a given independent variable are *different* from each other. For such a case we draw t_α from a double-sided table. For example, we have as the alternative hypothesis for the axial force rate effect:

$$H_A : \mu_{AF,10} \neq \mu_{AF,50}$$

If we were only interested in discovering whether $\mu_{AF,50}$ was *larger* than $\mu_{AF,10}$ by δ, say, then the alternative hypothesis would be single-sided and we would have drawn t_α from a single-sided table. The value of t_α would have been 1.645 in that case instead of 1.960, and so the minimum number of measurements required [from equation (8)] would have been reduced by $[(1.645 + 2.327)/(1.960 + 2.327)]^2 = 0.858$ compared to the number required with a double-sided alternative hypothesis. A moment's reflection reveals why fewer data points are required for a single-sided alternative hypothesis than for a double-sided one. In the case of a single-sided alternative hypothesis, we only need to acquire enough information to correctly place the result in one of *two* categories—greater than some amount, or not greater. If the alternative hypothesis is double-sided, we must be able to correctly place the result in one of *three* categories—greater than some amount, less than some amount, or between those two amounts. This will require more data for a given inference error risk.

We conclude from Table I that the most severe data volume requirements are associated with axial force. We must acquire 30 pairs of axial force measurements at high/low levels of sample rate and 30 pairs at high/low levels of dwell time. Similarly, we must acquire at least the number of pairs of normal force and pitching moment data indicated in Table I in order to accept either the null hypothesis or the alternative hypothesis with the prescribed degree of inference error protection. We would like these pairs of force and moment measurements to be made over of a range of AoA, Mach, and P_t levels as quantified earlier.

It is instructive to compare the scale of this design with conventional experiments in which only one variable is changed at a time. Data acquisition cost comparisons will vary depending on the details of the particular experiment and many formally designed experiments will require significantly more data that this one. There are even some circumstances in which a formal experiment design might specify a greater volume of data than a conventional OFAT experiment.[9] However, experience obtained at Langley Research Center in a wide range of formally designed wind tunnel tests suggests that formally designed experiments require on average about a 80% less data volume than classical OFAT test designs.[2]

Formal design advocates will also emphasize the specificity of objectives expressed in terms of null and

Table I. Data Volume Estimates

Response	Estimated Sigma	Resolution	Measured Sigma	Point-Pairs
AF (lbs)	0.18	0.20	0.20	**30**
NF (lbs)	1.5	2.0	1.7	**21**
PM (in-lbs)	4.0	5.0	3.5	**24**

alternative hypotheses, with their well-defined, quantitative values of required resolution and inference error risk tolerance. OFAT objectives are often expressed in somewhat more nebulous terms, such as, "The objective of this test is to *get [lots and lots of] data* on rate and dwell effects," or, "The objective is to *study* the rate/dwell issue." The difficulty with objectives stated this way is that it is not very easy to know when the objective has been achieved, and therefore when it is appropriate to declare victory and terminate the experiment. At what point in the OFAT process do we claim to have "studied" the rate/dwell effect, and when do we say that the objective of gathering data on the phenomenon has been successfully completed? It is not uncommon for these questions to be answered in terms of the exhaustion of resources, especially tunnel time. All too often "we have acquired data" and "we have studied the effect" when we have no more tunnel time available for the experiment.

Test Matrix

There are 32 unique combinations of the five independent variables in this experiment at two levels each. A full factorial test matrix (all combinations of all levels of all variables) would feature 16 measurements of each response at high sample rate, say, and 16 otherwise similar response measurements at low sample rate, allowing 16 estimates of the rate effect. From Table I, this is not sufficient to meet design requirements for resolution and inference error protection. We therefore specify a replicated full factorial design in five variables at two levels each, as in Table II. This permits 32 estimates of each effect, which satisfies our data volume requirements with a total of 64 data points. The column headers P, M, A, R, and D in Table II represent total pressure, Mach number, angle of attack, sample rate, and dwell time, respectively. The table entries are in coded units, with "+1" representing the higher of the two levels of each variable and "-1" representing the lower.

Table II. Replicated full factorial test matrix.

Case	P	M	A	R	D
1	-1	-1	-1	-1	-1
2	-1	-1	-1	-1	-1
3	-1	-1	-1	-1	1
4	-1	-1	-1	-1	1
5	-1	-1	-1	1	-1
6	-1	-1	-1	1	-1
7	-1	-1	-1	1	1
8	-1	-1	-1	1	1
...	...	...	...	...	...
...	...	...	...	...	...
63	1	1	1	1	1
64	1	1	1	1	1

Execution of the Experiment

To minimize conflicts with NTF's production schedule, the rate/dwell experiment was designed as a target of opportunity process improvement experiment that could be executed any time that would be convenient. An opportunity to execute this design occurred at the end of a lateral controls test to study the effects of Reynolds number and Mach number on the control power (rolling moment per degree of deflection) of ailerons and spoilers on a generic subsonic wing. This test involved a vehicle designated as the "Pathfinder-I fuselage with Controls Wing B." Before removing the model from the facility, the 64-point rate-dwell test matrix was executed using this model. The rate/dwell experiment was completed in two hours.

The test matrix is illustrated in Table II in what is called "standard order." Standard order reveals the basic underlying structure and pattern of the test matrix, but this is not the order in which the points are acquired. The actual run order is randomized. Randomization is one of the most visible operational distinctions between formal experimentation and the high-volume data collection activities that characterize OFAT designs. It exploits the fact that systematic facility variations can most efficiently confound the true dependence of responses on the treatment variables whose levels are changed in the experiment only if those levels are also varied systematically. That is, in order for time-varying systematic errors to wreak the greatest possible havoc with experimental results, it is necessary that independent variable levels also be varied systematically with time.

Randomization allows the researcher to circumvent the practical difficulties and expense of eliminating systematic errors as a prerequisite for high quality research results. It converts known sources of systematic error to easily handled random error. Moreover, it does the same thing even for systematic errors that remain *undetected* and *unknown*. It is not necessary to find the systematic errors in order to defend against them. No effort or expense is required to establish whether the system is in statistical control or not (the savvy experimenter simply stipulates that it is not), and no effort or expense is required to then remove specific sources of systematic error or to correct for them. See reference 6 for further discussion of randomization and other techniques for defending against systematic error in the absence of statistical control.

A randomized order was prescribed for the tunnel state variables in the NTF rate/dwell experiment—angle of attack, total pressure, and Mach number. Rate and dwell combinations were randomized post-test, via software. At the time of this experiment, sample rate and dwell time variations required troublesome alterations to the control system source code. These difficulties were avoided by acquiring all data points at high dwell and high sample rate. That is, each point was acquired at a 50-sample/second rate for two seconds, resulting in 100 frames of data for each data point.

Other combinations of sample rate and dwell size were easily generated in the post-test data reduction. For example, the "high rate, low dwell" points were created by simply deleting the last 50 frames, retaining only the first 50 that would have been acquired at a 50 sample/second rate for one second. Similarly, the "low rate, high dwell" combination was created by retaining only every fifth data frame, producing the same result that would have been obtained if the data had been acquired at 10 samples per second for two seconds instead of 50 samples/second. Finally, the "low rate, low dwell" combination was produced by first creating "low rate, high dwell" points and then retaining only the first half of the frames.

Analysis of Experiment and Results

The factorial design in Table II is a very compact arrangement by which several effects can be estimated conveniently with the same set of data. Compare Case 1 with Case 3 in this table, for example. In both cases the levels for total pressure, Mach number, angle of attack, and sample rate are the same. The two cases differ only in dwell time. The normal force measured for Case 1 was 175.12 pounds. For Case 3 it was 175.19 pounds. This suggests that a change in dwell time from one second to two seconds results in an increase in normal force of 0.07 pounds. However, this is only a single estimate of the dwell effect for normal force. The factorial design generates a total of 32 such estimates. These estimates are made two times independently at all 16 different combinations of high and low sample rate, AoA, Mach, and P_t. This range of non-dwell variables over which dwell is estimated is called the "inductive basis" in the language of formal experiment design. Full factorial designs such as this offer a wide inductive basis; that is, the dwell effects are estimated over *all* combinations of the other variables in the experiment, and are not limited arbitrarily to a subset of available comparison points.

We define the dwell effect for normal force as the average of all 32 independent estimates of this effect. (By Table I we know that this exceeds the minimum volume of data required by our inference error risk tolerances.)

Table III. Rate/Dwell Effects and Inference Criteria

Response	Measured Rate Effect	Measured Dwell Effect	Inference Criterion
AF (lbs)	0.059	-0.097	0.072
NF (lbs)	-0.450	0.480	0.613
PM (in-lbs)	0.410	0.860	1.262

The normal force dwell effect was determined in this way to be 0.48 pounds.

At first glance, Table I seems to suggest that we are in a position now to accept the null hypothesis for normal force rate effect, because the 0.48-pound effect we detected is less than the two pounds prescribed in the design as large enough to be important. However, figure 1 reveals that the true situation is somewhat more complicated. The 0.48-pound measured effect might come from a distribution such as the one on the left in figure 1, for which the true population mean is $\mu_0 = 0$ pounds. The 0.48-pound estimated value could be due simply to a small positive experimental error. In this case we would be justified in accepting the null hypothesis—no significant dwell effect for normal force. However, the observed effect could just as easily have come from a distribution such as the one on the *right* in figure 1, centered at $\mu_1 = 2$ pounds, with the 0.48-pound value due to a *negative* departure from the two-pound mean caused by experimental error. In that case, a proper inference would result in accepting the alternative hypothesis for normal force dwell effect.

We use the criterion level x*, illustrated in figure 1 and quantified in equations (1) and (2), to decide whether to accept the null hypothesis or the alternative hypothesis. Inserting into equation (1) the measured normal force standard error value from Table I of $\sigma = 1.7$ pounds and using $N = 32$ and $t_\alpha = 2.040$ (the double-sided t-statistic corresponding to $\alpha = 0.05$ with 31 degrees of freedom), we obtain $x^* = 0.61$ pounds. Since our measured effect—0.48 pounds—is less than this, we accept the null hypothesis with at least 95% confidence and conclude that a change in dwell from one second to two seconds does not result in a normal force change of two pounds or larger. We use this same procedure to compare measured rate and dwell effects for all three response variables—normal force, axial force, and pitching moment—to inference criteria computed as in this example. Table III lists the values of the effects and inference criteria.

The information in Table III is displayed graphically in figure 2, where the absolute value of the ratio of each

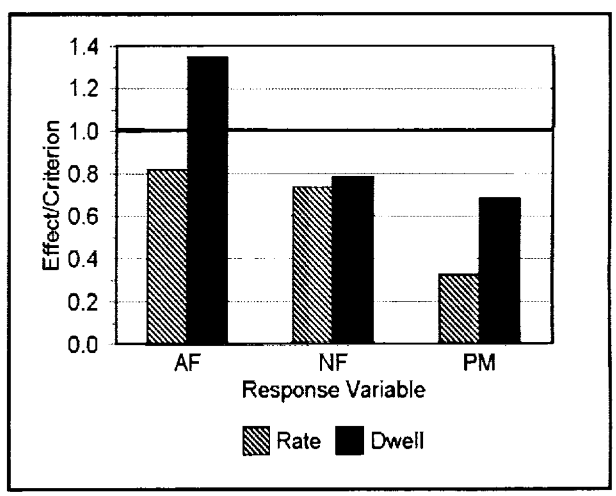

Figure 2. Rate and dwell effects relative to inference criteria

effect to its corresponding inference criterion is displayed. Ratios less than one imply that the proper inference is to accept the null hypothesis. Ratios greater than one imply that the alternative hypothesis should be accepted.

Table III and figure 2 suggest that the proper inference in the case of sample rate effect is to accept the null hypothesis for all of the response variables, with 95% confidence. We conclude that increasing the sample rate of the NTF data system from 10 samples/second to 50 samples/second has no important effect on axial force, normal force, or pitching moment.

We likewise accept the null hypothesis for the case of normal force and pitching moment dwell effects. However, the measured axial force dwell effect does exceed the corresponding inference criterion (a dwell effect with a magnitude of 0.097 pounds and an inference criterion of 0.072 pounds). We therefore accept the alternative hypothesis for axial force dwell effect with 99% confidence, and conclude that there are axial force dwell effects of 0.2 pounds or more.

The observed dwell effect for axial force stimulated further investigation, beginning with an examination of interaction effects. An axial force dwell interaction effect would be said to exist if the magnitude of the dwell effect depended on the level of some other variable. For example, if the axial force dwell effect were larger at high Mach number than at low Mach number, we would say that an interaction exists between dwell and Mach number. Interaction effects can provide insights into the underlying causes of effects that are observed in an experiment. Factorial designs permit interaction effects to be quantified as readily as main effects, another advantage they have over conventional OFAT test designs.

To quantify the interaction between dwell and Mach, say, in this two-level factorial design, we would first quantify the dwell effect for the higher Mach number (Mach 0.82 in this experiment). The factorial geometry of the design ensures that of the 32 Mach 0.82 points, 16 will have been acquired with a dwell of two seconds and 16 will have been acquired with a dwell of one second. Thus there will be 16 axial force dwell effect estimates available from the 32 data points acquired at Mach 0.82. The first step is therefore to compute the average of these 16 high-Mach axial force dwell effects.

The next step is to repeat the process for the 32 low-Mach points (Mach 0.52). By symmetry there will be 16 independent axial force dwell estimates available from these points, from which the average is to be computed as before. The interaction between dwell and Mach is then defined simply as the difference between the average dwell effect at high Mach and the average dwell effect at low Mach. The Dwell/Mach interaction for axial force was computed in this way to be –0.028.

We wish to choose between null and alternative hypotheses with respect to this possible effect. That is, we wish to determine with some confidence if the magnitude of the true effect is zero or non-zero. As always, experimental error introduces some uncertainty into the process, and is also the key to making the determination. We will declare the –0.028-pound axial force Mach-dwell interaction to be real (that is, we will accept the alternative hypothesis for this interaction) if its magnitude is large compared to the uncertainty in estimating it. Otherwise we will attribute the non-zero numerical value to experimental error and embrace the null hypothesis—"no interaction effect observed."

We will derive a simple formula for the uncertainty of this or any other effect in a two-level factorial design. Begin by noting that the Mach-dwell effect is computed by taking the difference between the dwell effect estimated at high Mach and at low Mach. Each of these in turn represents some linear combination of the measured axial force values that are added or subtracted in some pattern. The general form of the interaction is represented in equation (9). The average of half of the response measurements, combined algebraically in some pattern, is subtracted from a corresponding arrangement involving the other half of the data to form what is called a "contrast," C:

$$C = \left(\frac{y_1 \pm y_2 \pm \cdots \pm y_{N/2}}{N/2} \right) - \left(\frac{y_{N/2+1} \pm y_{N/2+2} \pm \cdots \pm y_N}{N/2} \right) \quad (9)$$

The y_i are individual response measurements (axial force, say), acquired at different combinations of the independent variables. We can utilize ordinary error propagation methods (see, for example, reference 8) to describe the variance in the contrast, C, in terms of the variance associated with the individual response measurements, y_i. Assuming the individual response measurements are independent, we have

$$\sigma_C^2 = \left(\frac{\partial C}{\partial y_1} \right)^2 \sigma_{y_1}^2 + \left(\frac{\partial C}{\partial y_2} \right)^2 \sigma_{y_2}^2 + \ldots + \left(\frac{\partial C}{\partial y_N} \right)^2 \sigma_{y_N}^2 \quad (10)$$

Note that:

$$\left(\frac{\partial C}{\partial y_i} \right)^2 = \left(\pm \frac{2}{N} \right)^2 = \frac{4}{N^2} \text{ for all } y_i.$$

Assume that the variance is the same for each response measurement so that the subscript notation can be dropped. Then

$$\sigma_C^2 = \sum_{i=1}^{N} \left(\frac{2}{N} \right)^2 \sigma^2 = N \left(\frac{2}{N} \right)^2 \sigma^2$$

and

$$\sigma_C^2 = \frac{4\sigma^2}{N} \quad (11)$$

Equation (11) is quite general; it applies to interactions of any order—not simply two-way interactions but also n-way interactions where n can be up to the total number of variables in the design. It applies equally well to main effects ("one-way interactions").

Recall that the Mach-dwell interaction was –0.028 pounds. Inserting into equation (11) the value of 0.20 pounds for σ from Table I, and noting that in this case N is not the number of *pairs* of points but the *total* number of points, 64, we compute 0.05 pounds for the standard error for axial force effects. The effect we estimated from the data is thus not much more than half a standard deviation away from zero—much too close to zero for us to accept an alternative hypothesis for the Mach-dwell interaction. We therefore accept the null hypothesis and conclude that we are unable to detect a significant interaction between Mach number and dwell time. That is, we can see no difference in the dwell effects of data acquired at our high-Mach level and those acquired at our low-Mach level.

We obtained a similar result when we examined the interaction between dwell and AoA. The dwell-AoA effect was –0.072 pounds, again not large compared to the 0.05-pound standard error in the estimate. So again we accept the alternative hypothesis and conclude that no dwell-AoA interaction can be inferred from the data.

The dwell-P_t interaction was –0.13 pounds, between two and three standard deviations away from zero. Under the generally reliable assumption (rooted in the Central Limit Theorem) that net experimental errors are normally distributed with mean of zero, the probability of such an outcome occurring purely by chance, given that the standard error is 0.05 pounds, is 0.0114. That is, there is a probability in excess of 98.8% that this interaction is not due to simple chance variations in the data, and is therefore a real effect. We therefore reject the null hypothesis for the pressure-dwell interaction and conclude that dwell effects at higher pressures are in fact different than dwell effects at lower pressures.

The probabilities that the interactions are real and not zero can be expressed equivalently as "odds" against a zero value, expressed in the form "M to one." For example, if the probability of a real effect is 95%, the probability that the effect is not real is 5% or one chance in 20. The "odds" are therefore "19 to one" of a non-zero effect. Odds favoring a real effect are presented in figure 3 for the dwell-Mach interaction ("DM"), the Dwell-AoA interaction ("DA"), and dwell-P_t interaction ("DP").

Note that the odds favoring a real effect must exceed 19 to one before such a claim can be made with at least 95% confidence. Figure 3 suggests that while the dwell effect has little to do with angle of attack or Mach number over the ranges that those variables were changed in this experiment, total pressure does clearly seem to influence the dwell effect. That is, the magnitude of the dwell effect is apparently different at 24.1 psi than at 15 psi.

To gain further insight into the axial force dwell effect, individual frame time histories were plotted for axial force. Each data point featured 100 frames of data acquired at 50 sec^{-1} for two seconds. These 100-point, two-second axial force time histories revealed an interesting pattern. Figure 4 is an example.

The time history in figure 4 provides some insight into the axial force dwell effect. The general assumption has been that the time-history is approximately flat, with only small, random variations about a constant level due to inevitable measurement error. The objective in averaging several frames of data over the dwell-time of the data point is to average out these relatively small chance variations in order to obtain a reliable estimate of the underlying constant value.

Figure 4 illustrates that this view is entirely wrong for axial force. There are significant departures from the value of axial force that would be obtained by averaging the individual frames. The peak-to-peak variation in figure 4 is approximately 1.4 pounds, which is substantially larger than the 0.2-pound level documented in Table I as large enough to be of concern for axial force. Furthermore, the variations from the mean are anything but random. There is a pronounced sinusoidal variation in the axial force. (The small step-changes in the axial force time history are attributed simply to least-significant-bit effects associated with the finite resolution of the digital data system.)

The period of this disturbance is a bit longer than two seconds. This corresponds roughly to the NTF circuit

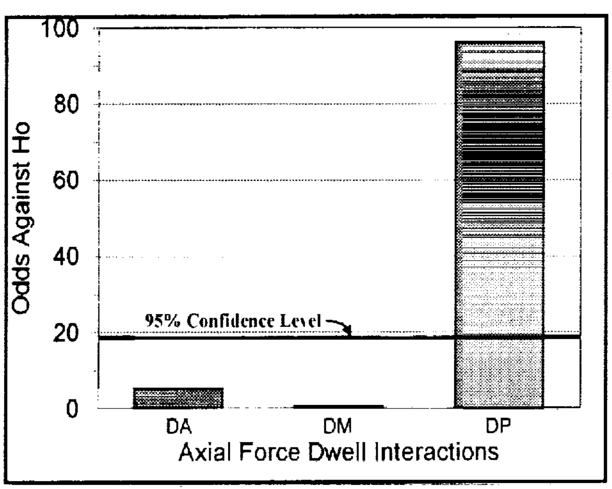

Figure 3. Odds against the null hypothesis for interactions of dwell effect with angle of attack, Mach number, and total pressure.

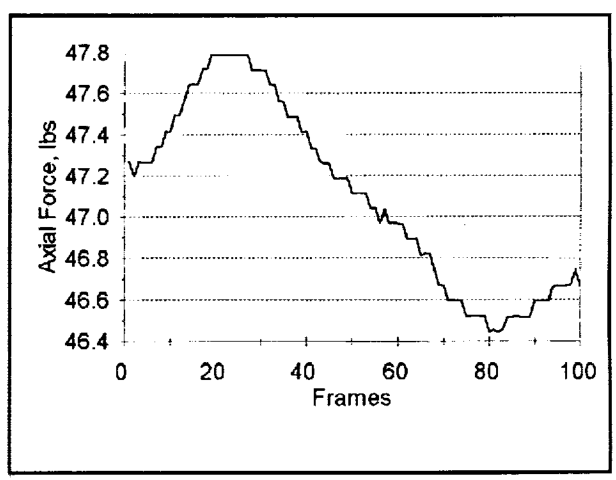

Figure 4. Representative axial force time history for a two-second dwell.

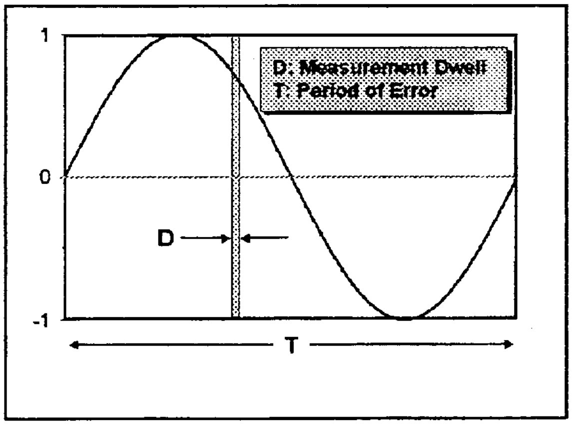

Figure 5. Errors from long-period variations and relatively short-period dwell times.

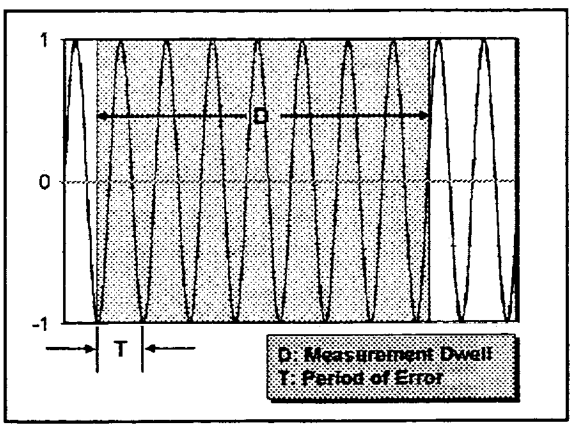

Figure 6. Errors from short-period variations and relatively long dwell times.

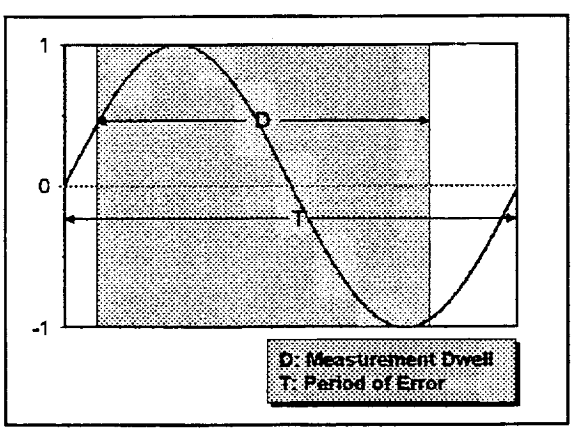

Figure 7. Schematic representation of dwell effect. Dwell neither long nor short compared to period of variation.

time—the time for a pressure disturbance to propagate all the way around the tunnel circuit. The clear evidence detected in the experiment of an interaction between dwell effect and total pressure suggests that the dwell effect ma be related to a tunnel circuit phenomenon that depends on pressure levels. Disturbances caused by flow blockage changes that accompany changes in angle of attack are one possible candidate, among many.

A picture begins to emerge of the physical phenomenon underlying the dwell effect. Pressure variations propagating around the tunnel circuit generate low-frequency disturbances that penetrate the one-Hz low-pass filters of the data system. Positive and negative departures from the true mean response are only completely canceled when the pressure disturbance is centered in the dwell period. At other times the positive and negative excursions only partially cancel, generating the equivalent of a "rectification error" that causes an effective bias shift in the average of all the frame data. Assuming these disturbances are approximately sinusoidal as figure 4 suggests, the bias error induced in each point can vary from zero for the case of the zero-crossing occurring halfway through the dwell time (complete cancellation) to something approaching the peak amplitude of the disturbance for low ratios of dwell time to period of the disturbance. The next section briefly describes an analysis of the dwell-time rectification error detected in the NTF rate/dwell experiment.

Dwell-Time Rectification Error Analysis

The impact that unexplained response variations have on measurements is a function of the period of the variation and the dwell time of the measurement, among other things. Figures 5 and 6 illustrate two extremes.

In figure 5, the period, T, of some error is long compared to a measurement dwell time, D. This might represent a seasonal variation, say, in which the data are acquired over a period that is much too short to permit any cancellation of the long-period effect. The result is a bias error that will have a different impact if the experiment is repeated at a different place in the cycle of the low-frequency variation.

Figure 6 represents the opposite extreme. In this figure the frequency of some unexplained variation is high enough that its period, T, is very small compared to the dwell time, D. Because the dwell time is long enough to encompass several cycles, this type of error can be reduced to arbitrarily small levels by replication.

While figures 5 and 6 represent common bias and precision error scenarios, respectively, the dwell effect observed in this experiment and represented schematically in figure 7 is "neither fish nor fowl." The period of the unexplained variance is neither very long compared to the dwell time nor very short, so that common descriptions of this effect as a "bias error" or as a "precision" error do not apply well.

To quantify the dwell effect error of figure 7, we assume that the dwell period starts at some time delay, τ, after the start of a new cycle. We integrate the sine wave, assumed to have amplitude A, from τ to $\tau + D$. We then integrate the resulting function over all values of τ and normalize appropriately to obtain an expression for the expectation value of the average bias error, equation (12).

$$\varepsilon_{bias} = \sqrt{\left(\frac{A^2}{2}\right)\left(\frac{\sin \xi}{\xi}\right)^2} \qquad (12)$$

where

$$\xi = \frac{\pi D}{T}.$$

Figure 8 shows how this error depends on the D/T ratio for a unit-amplitude error (A=1). It displays the standard "squared sin(x)/x" behavior of rectification errors as a function of cycles sampled. Note that the overall trend is for the dwell-effect error to decrease with longer dwell times, as would be expected—the longer the dwell, the more cycles of unexplained variation there are and therefore the more cancellation there can be. The error goes to zero for dwell times equal to integer multiples of the period, T, for which case there is perfect cancellation. Equation (12) can be used to compute the dwell time necessary to achieve an average rectification error due to dwell effect that is sufficiently small to meet the researcher's requirements, given a particular amplitude and frequency of disturbance. The dwell time required to drive the average axial force rectification error below the inference criterion threshold of Table III, given an amplitude and period as in figure 4, is about 2.2 seconds.

Note that equation (12) only describes the *average* rectification error over all phase relationships between the start of the dwell and the start of the disturbance. The actual error will vary. Also, it was well beyond the scope of this experiment to catalog rectification errors for anything even approaching the full range of operating conditions at NTF. It is likely that pressures higher than those covered in this experiment will result in greater rectification errors, and there may be significant interactions with Mach, AoA, and possibly other

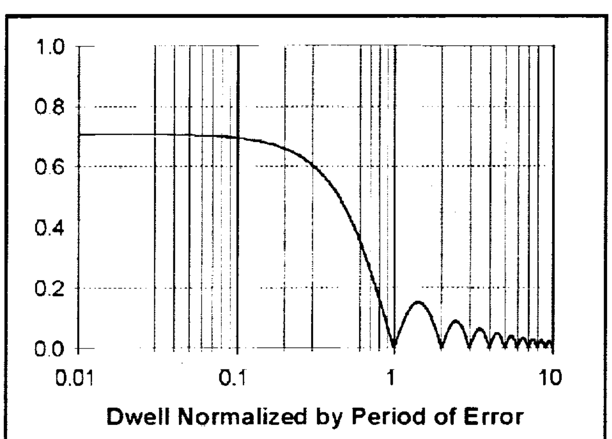

Figure 8. Average unit-amplitude bias error induced by dwell effect.

variables over different ranges of those variables. The essential finding of this study is that longer dwell periods will increase axial force data quality, at the expense of increased operating cost.

Concluding Remarks

A formally designed experiment (five-variable, two-level, replicated full factorial) has been executed in the National Transonic Facility to quantify the effects of increasing the data system's sample rate from 10 samples/second to 50 samples/second, and/or increasing dwell time from one second to two seconds. The experiment required 64 data points and was executed in two hours.

The magnitudes of rate and dwell effects large enough to cause concern ("significant" effects) were quantified and documented explicitly in the experiment design process before the experiment was executed. Significant effects were quantified for normal force, axial force, and pitching moment. The data volume was specified to ensure a probability in excess of 99% of detecting significant effects in the presence of worst-case anticipated experimental error. That is, the data volume was sufficient to ensure a probability of less than 1% that a true significant effect would go undetected, given worst-case experimental error conditions. The data volume was also great enough that the probability of incorrectly inferring a significant effect under worst-case experimental error conditions was less than 5% (at least 95% confidence in any inference of no effect).

No significant rate effects were observed for normal force, axial force, or pitching moment. No significant dwell effects were observed for normal force or pitching moment; however, a significant axial force dwell effect was observed. Furthermore, the axial force dwell effect was shown to interact significantly with total pressure. This dwell effect was subsequently attributed to a propagating pressure disturbance influenced by total pressure levels and resulting in a rectification error that depends on the ratio of dwell time to the period of the propagating disturbance.

The essential conclusions of this report are that over the ranges of variables studied, the proposed increase in sample rate would have no important influence on axial force, normal force, or pitching moment. The proposed increase in dwell time would have no important influence on normal force or pitching moment but would have a positive impact on the quality of axial force data by reducing the net effect of rectification errors associated with propagating pressure disturbances.

Acknowledgments

The author is indebted to Ms. Jean Foster and Messrs. Jerry Adcock and Charles Bobbitt of the National Transonic Facility for essential contributions to the design, execution, and analysis of the experiment described in this report. Mr. Tom Finley of the Experimental Testing Technology Division suggested the idea of post-test rate/dwell assignments and provided helpful critical reviews of the early data analysis.

References

1) McPhee, J. R.: "Electrical Noise Reduction Techniques Contributing to Improved Data Quality at the National Transonic Facility." NASA Contractor Report 4193, Contract NAS1-18304. Nov. 1988.

2) DeLoach, R. "Applications of Modern Experiment Design to Wind Tunnel Testing at NASA Langley Research Center." AIAA 98-0713. 36th Aerospace Sciences Meeting and Exhibit, Reno NV. Jan 1998.

3) Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). *Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building*. New York: Wiley.

4) Montgomery, D. C. (1997). *Design and Analysis of Experiments, 4th ed.* New York: Wiley.

5) Diamond, W. J. (1989). *Practical Experiment Designs for Engineers and Scientists, 2nd ed.* New York: Wiley.

6) DeLoach, R. "Improved Quality in Aerospace Testing Through the Modern Design of Experiments." AIAA 2000-0825. 38th Aerospace Sciences Meeting and Exhibit, Reno NV. Jan 2000.

7) DeLoach, R. "Tailoring Wind Tunnel Data Volume Requirements through the Formal Design of Experiments." AIAA 98-2884. 20th AIAA Advanced Measurement and Ground Testing Technology Conference, Albuquerque, NM. June 1998.

8) Coleman, H. W. and Steele, W. G. (1989). *Experimentation and Uncertainty Analysis for Engineers*. New York: Wiley.

9) DeLoach, R. "Split-Plot Experiment Designs in Wind Tunnel Tests with Hard to Change Variables." AIAA 2000-0827. 38th Aerospace Sciences Meeting and Exhibit, Reno NV. Jan 2000.