Within-Tunnel Variations in Pressure Data for Three Transonic Wind Tunnels

Richard DeLoach[1]
NASA Langley Research Center, Hampton, Virginia, 23681

This paper compares the results of pressure measurements made on the same test article with the same test matrix in three transonic wind tunnels. A comparison is presented of the unexplained variance associated with polar replicates acquired in each tunnel. The impact of a significance component of systematic (not random) unexplained variance is reviewed, and the results of analyses of variance are presented to assess the degree of significant systematic error in these representative wind tunnel tests. Total uncertainty estimates are reported for 140 samples of pressure data, quantifying the effects of within-polar random errors and between-polar systematic bias errors.

I. Introduction

This paper presents an analysis of pressure data acquired on the same test article in three US transonic wind tunnels in which nominally similar test matrices were executing in each facility. The test article was the AEDC 16T check standard model, a modified 5% scale model of an F-111.The participating tunnels were the National Transonic Facility at Langley Research Center (LaRC), the 11-Ft Unitary Plan wind tunnel at Ames Research Center (ARC), and the 16T wind tunnel at the Arnold Engineering and Development Center (AEDC).

An analysis of pressure data is presented here that uses methods similar to those used previously to analyze within-facility[1] and between-facility[2] force and moment data from the same test. The author was not involved in the original test and did not participate in its design or execution. Apparently, no resources were provided in the original funding of the study to support a comprehensive analysis of the data by its participants. General observations about facility-to-facility testing methodology differences have been discussed in workshops convened since the tests for that purpose, but the hundreds of thousands of measured aerodynamic forces, moments, and pressures recorded in this test were never methodically examined by the original test personnel to quantify within- and between-facility characteristics of the measurement environment. The author offered to provide such an independent analysis of the data after the fact.

The analysis of pressure reported in this paper, like the referenced analyses of force and moment data, differs from a typical analysis of wind tunnel data in one important respect: Wind tunnel tests commonly focus on the relationships between selected response variables and the independent variables that influence them. The intent is usually to acquire enough new knowledge about the test article that similar responses can be adequately predicted in the future for independent variable combinations of interest, assuming only that those variables were examined in the test and that they were varied over an adequate range of levels. Attention is therefore focused, quite properly, on the test article. In this analysis we are less interested in the test article than the facilities in which it was tested. We wish to examine the measurement environments of the participating wind tunnels. We will do so by considering *variance* in the data acquired in each facility.

Variance often has an unnecessarily restrictive association with experimental error because of an industrial model of wind tunnel testing that has evolved over time. By this model, *data* are regarded as the product of what is essentially an industrial process in which the tunnel is seen as a "data factory," the purpose of which is to produce this product in high volume. Concepts of quality have been borrowed by the experimental aeronautics community directly from industrial engineering to support this model; including the notion that minimal variance among similar units of product (replicates, in wind tunnel testing) is a prerequisite for quality.

We will proceed from a somewhat expanded view of variance by which the term is not automatically associated with a defect in the data sample. Rather, a distinction is made between what we will describe as *explained* variance and what we will describe as *unexplained* variance. Consider that every sample of wind tunnel data does in fact

[1] Senior Research Scientist, NASA Langley Research Center, MS 238, 4 Langley Blvd, Hampton, VA 23681, Associate Fellow, AIAA.

feature *variance*, by which we mean simply that the individual measurements in that sample are bound to differ from each other. Ordinary experimental error ensures that even replicated measurements will differ to some degree from each other, but by far the greatest component of total variance in a wind tunnel data set is not due to error at all, but to changes in the independent variables that are made intentionally. Indeed, it is only by analyzing such intentionally induced variance that we are able to make progress in experimental aeronautics.

We say that this large component of the total variance in a wind tunnel data sample is *explained* by the known independent variable changes that cause it. In addition to this explained variance, however, there is always an inevitable residual variance after we have accounted for all known changes, which we describe as *unexplained* variance. This unexplained variance is responsible for all of the experimental uncertainty in a wind tunnel test that is not attributable to constant bias errors, so it should be of great interest especially to facility personnel. It is this unexplained variance that is the topic of the current paper.

In a perfect world 100% of the variance would be explained, and even in the imperfect "real world" almost all of the variance in a wind tunnel data set is due to known independent variable changes and is therefore "explained" according to the nomenclature we are introducing. Since the unexplained variance is relatively small, often representing only a few parts per million of the total variance, it can be hard to detect, much less to quantify and otherwise study. To more clearly reveal the nature of the unexplained variance, it is therefore necessary to isolate it from the explained variance. Considerable preliminary data reduction is undertaken in this paper to isolate the unexplained variance for study, as will be discussed in the sections that follow.

Section II describes the test article and identifies the location of pressure taps. It also outlines the general plan of test for this experiment, including a constructive critique of the experiment design. Section III provides some tutorial background on variance-based methods of analyzing wind tunnel data and demonstrates how the total variance can be partitioned into explained and unexplained components. A further partitioning of the unexplained variance into constituent components is also described in Section III, as is a method for estimating total uncertainty in the presence of both random and systematic errors. Methods described in Section III were used to produce results that are presented in Section IV for within-facility variance estimated that are generated and compared for the three facilities. A discussion of independent measurement errors is presented in Section V. Section VI presents a summary and concluding remarks.

II. Test Article and Plan of Test

The test article and plan of test have been described in the papers cited earlier[1,2] in which within- and between-facility variance levels were presented for force and moment data. This information is summarized here for the convenience of the reader.

A. Test Article

Figure 1 shows the planform of the test article, the AEDC check standard model consisting of a modified 5% model of the F-111. The wings were modified to provide a 48-inch span at a fixed wing sweep angle of 35 degrees.

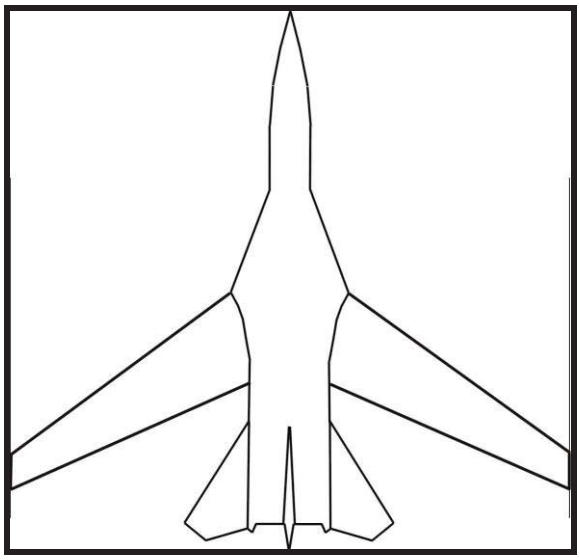

Figure 1. Planform of test article

Trip dots of the same size were applied at the same location in all facilities. They were located on the nose and upper and lower surfaces on the wing strake, wing, and horizontal and vertical tails.

Two control surface configurations were tested, designated Configuration 0 and Configuration 1. The horizontal tail was not deflected in Configuration 0, but it was deflected 10^0 in Configuration 1.

There was a requirement that all tests use the same instrumentation. Once the sting, balance, model, and pressure tubing were assembled the test article remained as one unit for the completion of the tests executed in each facility. This was to ensure that the bridging of the balance, the mounting of the balance to the model and sting, the routing of pressure tubing, and the build-up of the model did not changed from facility to facility.

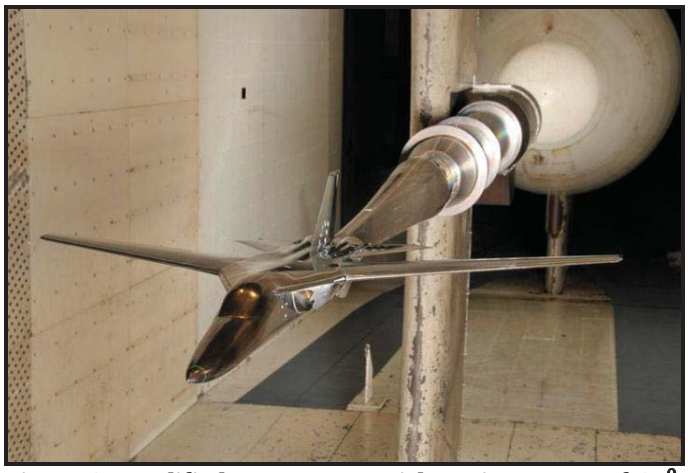

Figure 2. Modified F-111 test article. Wing sweep of 35^0, 48-inch wing span

The concepts of explained and unexplained variance were mentioned in the introduction. Generally, the details of the test article impact only the explained variance. The unexplained variance can be regarded more as a function of the measurement environment, and should therefore be facility-specific. Strictly speaking, it therefore would not have been entirely necessary to use the same test article in all three facilities for this study. Doing so, however, did help minimize the probability that different measurement systems might contribute differently to the unexplained variance in each facility, complicating between-facility comparisons. On the other hand, it might have been useful to be able to correlate levels of unexplained variance in each facility with the measurement systems used in them. In any case, the test was designed to eliminate the effects effects of measurement system differences insofar as it was possible to do so, with the result that differences in unexplained variance characteristics are more likely to be attributable to other causes.

Figure 2 shows the test article as mounted for testing in the AEDC 16T tunnel. There are fifty pressure taps in this model, with forty-four distributed on the wings and six on the underside of the fuselage. The forty-four wing taps are arranged in four linear chord-wise arrays of eleven taps each, ranging from near the leading edge to near the trailing edge, as indicated in Fig 3. There was an upper-surface chord-wise array and a corresponding lower-surface chord-wise array at each of two different span-wise distances from the fuselage. The upper and lower chord-wise tap arrays at the lesser span-wise distance, described as the "inboard taps," were on the starboard wing. The upper surface and lower surface chord-wise tap arrays at the greater span-wise distance are described in this paper as the "outboard taps" and were on the port wing. The pressure coefficient responses estimated for the upper-surface inboard array were labeled Cp(1) through Cp(11), with Cp(1) nearest the leading edge and Cp(11) nearest the trailing edge. They will be so referenced subsequently in this paper. Similarly, the pressure coefficients for the lower-surface inboard array were labeled Cp(12) through Cp(22), with Cp(12) nearest the leading edge and Cp(22) nearest the trailing edge. For the outboard wing tap arrays, responses measured on the upper surface are designated Cp(23) to Cp(33) from leading edge to trailing edge, and Cp(34) to Cp(44) from leading edge to trailing edge of the lower surface.

Four of the six fuselage taps are on the underside centerline. Two, designated FL4 and FL5, are on the port and starboard edges of the fuselage underside, respectively. See Fig. 3. One of these lateral fuselage taps, FL5, was labeled "Plugged" in the data sample from AEDC. The same tap was labeled "Leak" in the data from LaRC and

ARC, so no results are reported in this paper for that tap. Tap FL6, the aft-most centerline fuselage tap, was not connected and so yielded no data.

Two of the lower-surface wing taps generated obviously bad data. One of these was the second tap from the leading edge on the inboard array, designated Cp(13), with data labeled "BAD" from all three facilities. The other was the second tap from the *trailing* edge for the same lower-wing inboard array, designated Cp(21). Data from LaRC was labeled as "bad" for this tap.

In addition to the two wing taps and two fuselage taps that clearly did not produce usable data, four other taps were labeled as having "slow leaks," but no significant effects could be detected in the data acquired from those taps. They were taps Cp(16) and Cp(34), each with "slow leaks" at Ames and Langley, tap Cp(39) with a "slow leak" at Langley, and tap Cp(42) with a "leak" at Ames. Data from these taps were included in the analysis for this paper, since no significant leak effects were evident in the data.

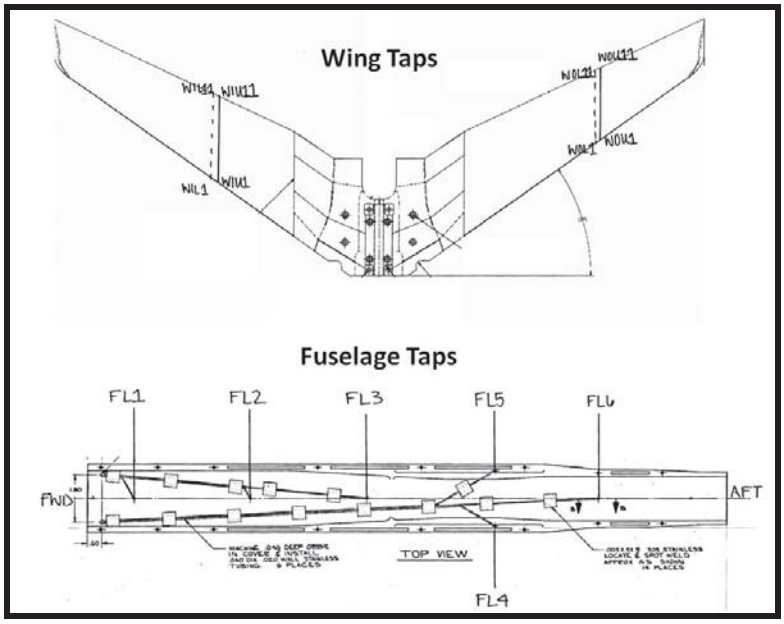

Figure 3: Pressure tap locations

B. Plan of Test

It should be noted at the outset that the current test was conducted with a conventional emphasis on the test article. It was not designed entirely as if the objective was to examine the facilities, notwithstanding the fact that this was the actual objective of the test. Instead, the tests conducted in each of the three participating facilities were executed as typical high-volume data acquisition exercises with the usual objective of acquiring as many response measurements on the test article as resource constraints would permit. The chief concessions to the true test objective were that nominally similar test matrices were executed in all facilities, and that the same test article—including the sting and balance assembly and all pressure tubing—was used in each test.

Site Number	Configuration	Reynolds Number per foot x 10^6	Mach Number
1	0	3.85	0.60
2	0	4.50	0.85
3	0	5.50	0.60
4	1	4.50	0.85

Table 1. Design Space Sites Where the Same Within-Facility Replicates Were Reproduced in All Three Tunnels

Unfortunately, while nominally similar test matrices were executed in each facility, the selection of specific sites within the operating envelope of each tunnel were in many cases unique to the interests of that specific facility. There are therefore numerous cases in which conditions were not reproduced across every participating facility, and

for which even *within*-facility replicates were not consistently acquired. The ideal scenario would have been for within-facility replicates to have been acquired for the same test conditions in each tunnel. Figure 4 identifies sites within the Mach/Reynolds number design space for which data were acquired in at least one tunnel. There are over sixty such sites. Eleven of these featured combinations of Mach number, Reynolds number, control surface configuration, and sideslip that were common to all three tunnels.. Four of these eleven featured within-facility polar replicates.

Sites within the relatively small subset for which within-facility replicates were acquired *in all three tunnels* are circled in red in Fig. 4. In some cases, the red circles seem to include more than one site each, but this is due only to the proximity of sites. It is the site nearest the center of each circle where within-facility replicates were acquired, and reproduced in all three tunnels. Table 1 lists the sites for which two or more polar replicates were reproduced in all three facilities. Figure 5 highlights these graphically.

Note in Table 1 and Fig. 5 that there are actually two sites that have the same Mach and Reynolds numbers per foot (0.85 and 4.50 x 10^6, respectively). They differ only in configuration number. One is Configuration 0, signifying no control surface deflections. One is Configuration 1, indicating a deflection of +10° in the horizontal tail control surfaces.

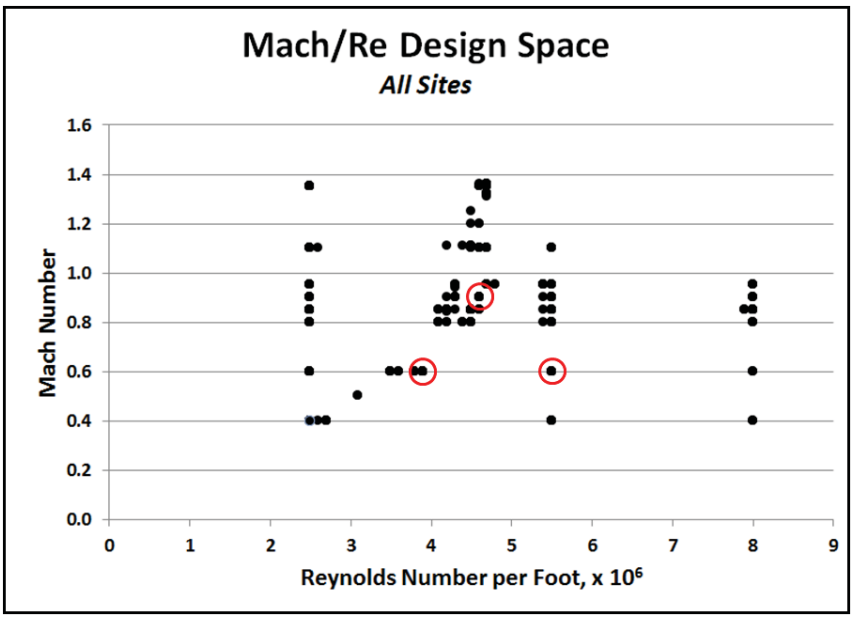

Figure 4. Test sites within the Mach/Reynolds number design space. Circles identify sites for which within-facility replicates were acquired in all tunnels.

While the choice of design space sites seems to have been made with the intent to acquire as much data of a general kind as possible, sites were not chosen that would be especially useful in finding answers to specific questions. For example, some questions that one might ask in an examination of measurement environments are these: Does the magnitude of unexplained variance change with Mach number? Does it change with Reynolds number? If it does change with Reynolds number, is that change different at higher Mach numbers than at lower Mach numbers? How do such Mach and Reynolds numbers effects change when the control surfaces are deflected, or are control surface deflections irrelevant? To what extent do these Mach number, Reynolds number, and Configuration number effects change from tunnel to tunnel?

Many other such questions could be addressed by arranging the sites in the design space judiciously, rather than simply distributing them to capture individually interesting conditions. For example, by comparing Sites 1 and 3 it would be possible to glean some information about Reynolds number effects at Mach 0.60 for control surfaces that are not deflected, but these conditions are so restrictive as to render unattractive the substantial data reduction and analysis effort necessary to achieve so little insight. Likewise, Sites 2 and 4 could be compared to quantify configuration effects for a single Mach/Reynolds number combination (0.85/4.5E06, respectively), but it would be hard to justify the analyses of data from nominally fifty pressure taps for such a meager result.

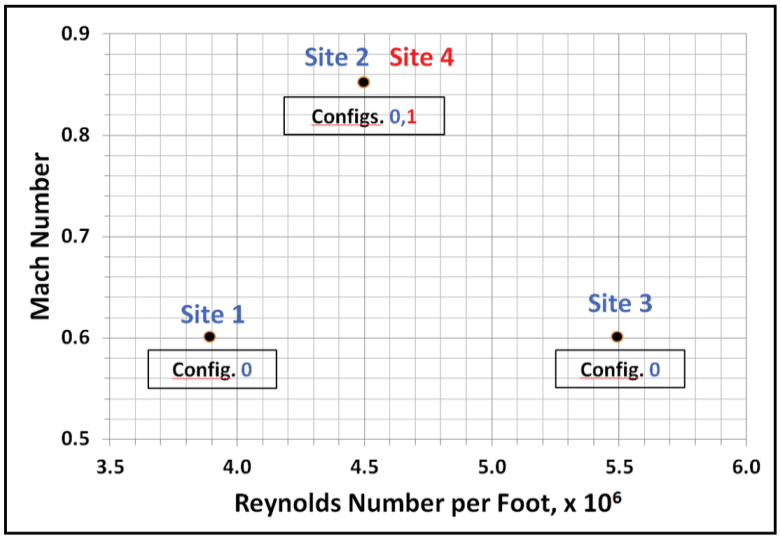

Figure 5. Four sites within the design space for which data were replicated both between and within all three facilities.

Since the experiment was not designed to quantify the effects of such factors as Mach Number, Reynolds number, and Configuration on the unexplained variance in pressure coefficient measurements, this paper focuses on Site 1 of Table 1. That is, all results apply for the case of Mach 0.60, Reynolds number of 3.85 x 10^6 per foot, and a clean configuration—no control surface deflections. While similar anecdotal evidence could be presented for three other arbitrarily selected sites within the design space, the design of the experiment mandates that the reader use his own judgment to decide how various features of the unexplained variance might systematically vary with site location. The results obtained at Site 1 are not uninteresting in and of themselves, and provide a good illustration of certain aspects of the unexplained variance that are likely to be in play at all sites in the design space.

III. Variance

We begin this section with a tutorial review of variance and how various categories of variance are computed. Recall that in the Introduction, variance in wind tunnel data was described as having explained and unexplained components. We have noted that most of the variance is explained by intentional independent variable changes, and is therefore unrelated to experimental error. This implies a distinction between what might be regarded as "good variance" and "bad variance," as we will show.

A. Overview of Variance in Data Samples

Variance, be it "good" or "bad," is expressed as the ratio of two quantities, a "sum of squares," SS, and the minimum volume of data required to compute the sum of squares. This later quantity, known as "degrees of freedom," is traditionally represented by the symbol "ν," and often simply as "df."

"Variance" implies "difference," and the SS calculation therefore requires us to specify some reference from which differences in each of the data points might be measured. For example, if we are computing the total variance in a sample of data points, that reference is often the sample mean, although in other circumstances it might be convenient to use another reference.

For an *n*-point sample, only $n - 1$ data points are needed to compute the SS if we know the sample mean. We actually use the values of all *n* data points in the SS calculation, but if we know the mean and $n - 1$ of the points we can always extract the value of the n^{th} point by subtracting the sum of the $n - 1$ points we know from the total of all *n* points. The total is simply the product of *n* and the *n*-point mean, both of which we know, so the smallest number of points needed to compute the SS, given the mean, is $n - 1$ in this example, and we have

$$df = n - 1 \tag{1}$$

Each data point is said to carry one degree of freedom. We say in this case that one of the *n* degrees of freedom in the sample is consumed by estimating the mean, leaving $n - 1$ to estimate the variance about that mean.

To compute the sum of squares, we begin by computing the *residual* for each data point. The residual of the i^{th} point, y_i, is just the difference between that point and the reference, y_0. These residuals are each squared and the squared residuals are added to produce the "sum of squares."

$$SS_{Total} = \sum_{i=1}^{n} \left(y_i - y_0 \right)^2 \tag{2}$$

The subscript, "*Total*," distinguishes the SS corresponding to the variance of the entire data sample from the SS for explained and unexplained components that will be introduced shortly. In general, the total variance or any component element of it is defined as the ratio of its sum of squares to the corresponding degrees of freedom. The total variance, therefore, is computed as follows:

$$\sigma_{Total}^2 = \frac{\sum_{i=1}^{n} \left(y_i - y_0 \right)^2}{n-1} \tag{3}$$

The variance is expressed in *squared* physical units because it is computed from the SS. It is often more convenient to express the spread of data in a sample by the *square root* of the variance, to return this measure to more meaningful physical units:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - y_0 \right)^2}{n-1}} \tag{4}$$

The quantity "sigma" from Eq. 4 is called the "standard deviation" for the example described here, in which the reference is the sample mean. For the general case, which includes situations in which the reference is something other than the sample mean, sigma is often called the "standard error."

We should note in passing that one rigorous convention is to use Latin letters to represent sample statistics such as those described here, reserving Greek letters to represent the corresponding population parameters where n tends to infinity. The sample standard error is represented by "s" by this convention and only the population standard error (large n) is represented by σ. The author follows the practical convention of using the terminology in greatest common use, which is to represent the standard error by "σ" rather than "s," accompanied by an unambiguous description of the context to indicate when the symbol is associated with a finite sample and not an infinite population.

Table 2 displays a small sample of data, illustrating how the sample variance is computed. The mean of this 7-point sample, -0.3694, is subtracted from each number in the "Value" column to produce the residuals, which are then squared and summed. The resulting sum of squares has a value of 0.0771.

Point	Value	Residuals	Squared Residuals
6	-0.4726	-0.1045	0.0109
2	-0.2625	0.1055	0.0111
4	-0.3699	-0.0018	0.0000
7	-0.5256	-0.1575	0.0248
1	-0.2106	0.1575	0.0248
3	-0.3143	0.0538	0.0029
5	-0.4210	-0.0529	0.0028
Mean =	**-0.3681**	**SS =**	**0.0774**

Table 2. Calculating the total sum of squares for a 7-point data sample

There are $n = 7$ points in this sample and thus $n - 1 = 6$ degrees of freedom, so the total variance is computed as follows:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(y_i - y_0)^2}{n-1} = \frac{0.0774}{6} = 0.0129 \tag{5}$$

from which we extract the standard deviation by taking the square root: $\sigma = 0.1136$.

These tedious details of computing the standard deviation are probably understood by most readers but they are presented here as a review for any who may not be current on this topic, and also as a starting point for describing how to partition this total variance into explained ("good") variance and unexplained ("bad") variance components. This latter topic is more likely to be new to experimental aerodynamicists who are quite accustomed to acquiring such data in high volume, but less accustomed to performing the kind of analysis that will now be described.

The data in Table 2 are displayed graphically in Fig. 6. The mean of -0.3681 from Table 2 is marked with a dashed line.

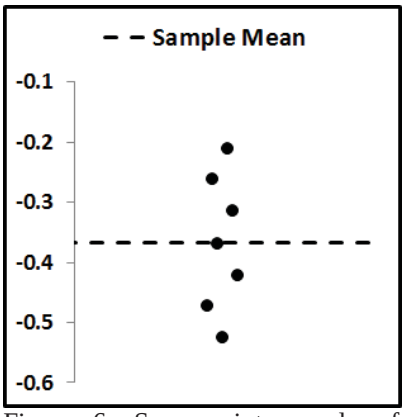

Figure 6: Seven-point sample of data, displaying variance

We now reveal that the data displayed in Fig. 6 are pressure coefficients comprising a pitch polar acquired from the outboard upper wing surface of the wind tunnel model used in the present test, as measured in one of the three tunnels cited above. The seven points correspond to angles of attack in the range of -3° to +3° in 1° increments.

If we translate each data point left or right as appropriate to ensure that its lateral position is proportional to angle of attack, we obtain Fig. 7.

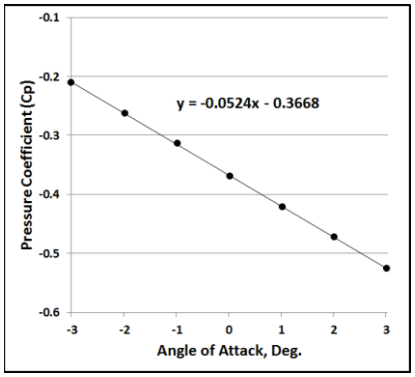

Figure 7: Seven-point sample of data, variance explained by changes in angle of attack.

Figures 6 and 7 display identically the same data. The points were shifted left and right according to the angles of attack where they were acquired, but there were no changes in the Cp values of the data. The sample mean is therefore identical, and the points from Fig. 7 are precisely the same points used in Table 2 to compute the sum of squares. The number of points, $n = 7$, is the same in both cases and therefore both cases have the same six degrees of freedom. The variances are thus identical, as are the standard deviations that customarily quantify the "scatter" in the data. Any yet the data displayed in Fig. 7 appears in some sense to have "less scatter" than the data as displayed in Fig. 6.

The appearance of reduced variance in Fig. 7 is an optical illusion created when our eyes are drawn to the straight line fitting the data in that figure. We see that there is less scatter *about the fitted line*, even though the scatter about the *sample mean* is the same. We say that the first-order polynomial response model displayed in Fig. 7 "explains" much of the total variance in the data sample.

We can quantify the SS associated with this explained component of the total variance by following a procedure analogous to the one used to quantify the SS for the total variance about the mean. We proceed exactly as before except that we use *computational* estimates of Cp instead of *empirical* estimates. The response model is used to generate these computational estimates.

Let $\hat{y}_i$ ("y-hat") represent the response predicted by the response model at the i^{th} angle of attack. Eq. 2 becomes

$$SS_{Explained} = \sum_{i=1}^{n} \left(\hat{y}_i - y_0 \right)^2 \tag{6}$$

The explained SS quantifies how much variation there is in *predicted* responses about the sample mean, rather than how much variation there is in *measured* responses. A detailed example is presented in Table 3.

For a response model with p parameters (polynomial model terms), there are $p - 1$ degrees of freedom associated with the explained variance, so we can compute the explained variance by dividing $SS_{Explained}$ by $p - 1$ if we wish. However, in a study of measurement quality, there is less interest in quantifying the explained variance than in quantifying the unexplained variance. The total and explained SS values are useful to know, because together they can be used to compute the unexplained SS, as will be described presently.

AoA	Computed Value	Residuals	Squared Residuals
-3	-0.2095	0.1585	0.0251
-2	-0.2635	0.1046	0.0109
-1	-0.3159	0.0521	0.0027
0	-0.3683	-0.0002	0.0000
1	-0.4208	-0.0527	0.0028
2	-0.4730	-0.1049	0.0110
3	-0.5256	-0.1575	0.0248
		Explained SS =	**0.0774**

Table 3. Calculating the explained sum of squares for a 7-point data sample

Tables 2 and 3 show that the explained SS is almost as large as the total SS in the current example, suggesting that the response model does in fact explain most of the variance in the data sample. The ratio of the explained SS to the total SS is a common metric for assessing the degree to which a proposed response model fits the data. It is called the *coefficient of determination*, R^2. The coefficient of determination in this example is the square of the correlation coefficient between Cp and AoA, which ranges from -1 to +1 for perfect negative and positive correlation, respectively. The R^2 statistic therefore ranges from 0 to a value of +1 for a perfect fit. It is less than 1 in practical circumstances because absent perfect data and a perfect response model, some portion of the total variance always remains unexplained by even the best of response models. However, extremely good response models can

often be generated from data acquired in modern wind tunnels. In the current example, the total sum of squares, SS_{Total}, is 0.0773735, and the explained sum of squares, $SS_{Explained}$, is 0.0773661. Therefore

$$R^2 = \frac{SS_{Explained}}{SS_{Total}} = \frac{0.0773735}{0.0773661} = 0.999904 \tag{7}$$

The explained SS differs from the total SS by only 96 parts per million, indicative of the high precision commonly achieved in modern wind tunnel testing. Notwithstanding how little of the total SS remains unexplained, this miniscule component is responsible for much of the uncertainty in a wind tunnel test. It is therefore of paramount interest in a study such as the present one, which focusses on the measurement quality that can be achieved in different facilities.

We can compute the unexplained SS using a formula similar to Eq 2:

$$SS_{Unexplained} = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \tag{8}$$

Here, the squared residuals each represent the difference between measured and predicted responses at the i^{th} angle of attack. Instead of the sample mean, we use the computational response estimate at each AoA as the reference by which to estimate the residual , so these residuals simply represent the distance each measured point is away from the line that fits the data. Table 4 illustrates how the unexplained SS is computed for the current example.

For a polynomial response model with p coefficients in the model (including the intercept), There are $n - p$ degrees of freedom associated with the unexplained variance. The present example features a first-order polynomial with $p = 2$ coefficients (slope and intercept) fitted to $n = 7$ data points, so there are five degrees of freedom associated with the unexplained variance. We therefore have

$$\sigma^2_{Unexplained} = \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2}{n - p} = \frac{7.297 \times 10^{-6}}{5} = 1.479 \times 10^{-6} \tag{11}$$

AoA	Measured Value	Computed Value	Residuals	Squared Residuals
-3	-0.2106	-0.2095	-0.0010	1.098E-06
-2	-0.2625	-0.2635	0.0009	8.598E-07
-1	-0.3143	-0.3159	0.0016	2.722E-06
0	-0.3699	-0.3683	-0.0016	2.530E-06
1	-0.4210	-0.4208	-0.0002	6.249E-08
2	-0.4726	-0.4730	0.0004	1.239E-07
3	-0.5256	-0.5256	0.0000	1.685E-09
			Unexplained SS =	**7.397E-06**

Table 4. Calculating the unexplained sum of squares for a 7-point data sample

It is not actually necessary to independently compute the total, explained, and unexplained SS as in Tables 2-4 because the sums of squares and their corresponding degrees of freedom are each additive. That is,

$$SS_{Total} = SS_{Explained} + SS_{Unexplained} \tag{9}$$

and likewise

$$df_{Total} = df_{Explained} + df_{Unexplained} \tag{10}$$

Therefore, given the total SS and the explained SS, the unexplained SS can be determined by simple subtraction. The unexplained df can be also be determined by subtracting the explained df ($p - 1$) from the total df given the mean ($n - 1$), to yield $n - p$ df.

The standard error is the square root of the unexplained variance, which for pressure coefficient data in this example is 0.0012. This compares favorably with a common industry tolerance of 0.0050 for "two sigma," although it overstates the precision with which individual measurements are likely to be made at this facility. The reason is that 0.0012 is the standard error associated with an estimate of pressure coefficient that is based on a response model, not a single measurement. The uncertainty in a response model prediction reflects the effect of "hidden replication" by which random variations about the fitted model partially cancel. It can be shown[3] that this hidden replication causes the average standard error across all points used to fit a polynomial regression model to be smaller than the standard error associated with an individual measurement by a factor of the square root of p/n. Since there must be at least as many fitted points as parameters in the model ($n \geq p$), the standard error of prediction cannot be greater than the standard error of measurement and can be significantly less, depending on the volume of fitted data and the complexity of the response model.

$$\sigma_{Model} = \left(\sqrt{\frac{p}{n}} \right) \sigma_{Measured} \tag{12}$$

In the present example for which $p = 2$ and $n = 7$, the standard response model prediction error of 0.0012 is therefore smaller than the standard error associated with individual measurements at this facility by a factor of 0.535. If we multiply 0.0012 by the reciprocal of this factor (1.871) to estimate the standard error for individual measurements, we find the value is 0.0023. This is consistent with common industry tolerance levels for individual Cp measurements, but almost twice as great as could be achieved in this example by response surface modeling.

B. Partitioning the Unexplained Variance

The standard error for the Cp data displayed in Fig. 7 has been estimated to be 0.0012 using response surface modeling methods, and to be 0.0023 if conventional OFAT (One Factor At a Time) data collection methods were used. This standard error, whether estimated by the direct measurement of replicated data points or from the residual variance of a response surface model, is proportional to the width of an uncertainty band that can be constructed about the plotted data as in Fig 8.

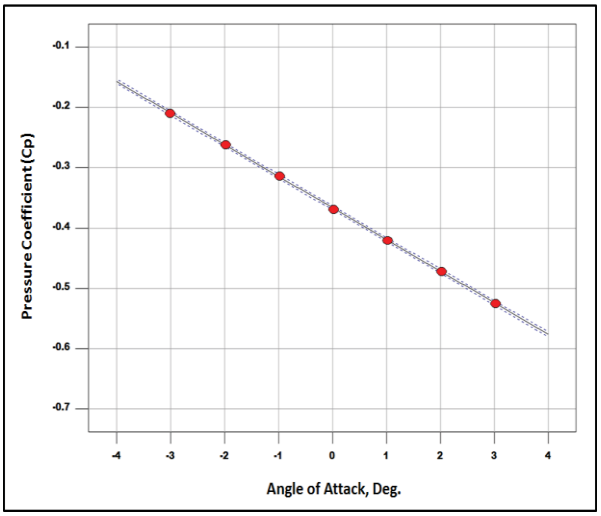

Figure 8: Representative Cp polar with 95% Prediction Interval limits.

The dashed lines in Fig. 8 represent the upper and lower limits of a 95% prediction interval. Given the precision of the data in Table 2 and the first-order polynomial response model displayed in Fig. 7,there is a 95% probability that an individual Cp measurement within the [-3°,+3°] AoA range that is drawn from the same population as the points in Table 2 will fall within these limits.

Consider now what happens if this same Cp polar is replicated a few minutes after it is acquired. Because all conditions are assumed to be unchanged, replicated polars are expected to be indistinguishable. Figure 9 displays the polar of Fig. 8 with two additional replicate polars, one acquired two minutes later and one acquired 1:17 later.

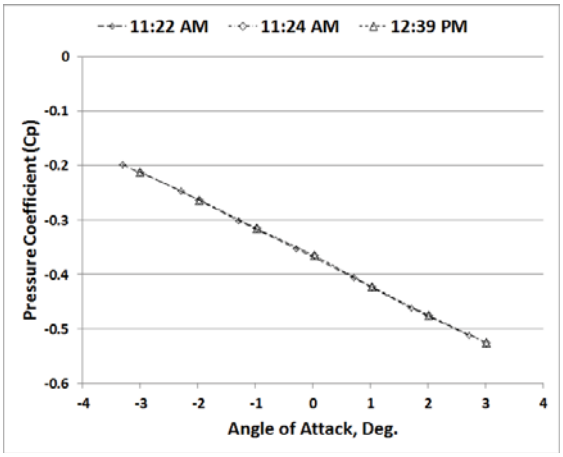

Figure 9: Three replicated Cp polars.

The unexplained variance associated with these three polars cannot be readily resolved in a plot such as Fig. 9 because the dynamic range of the figure is so much greater than meaningful experimental error kevels. It is not uncommon for a pressure coefficient error of 0.0025 to completely exhaust the error budget of a precision wind tunnel test, for example, but in Fig. 9 such an error would represent less than half of one percent of the full y-axis, or less than 1/25th of one of the horizontal divisions. We cannot see the small, unexplained variance that interests us because of the large component of total variance that can be explained by changes in angle of attack.

To isolate the unexplained variance, we can express Fig. 9 not as three convention polars superimposed, but as three *differential* polars, where instead of plotting Cp at each angle of attack, we plot the *difference* between CP and the average of the three Cp measurements at each angle of attack.

Figure 10 illustrates schematically the kind of result we would expect to see if all three polars coincided within random experimental error. At any given angle of attack, there are three Cp measurements, but there is no systematic relationship among them. If the error at any one angle of attack is known, it provides no information about the error from either of the other two polars at that angle of attack, or the error from the same polar at any other angle of attack. It would be a coin toss (three-sided coin!) as to which polar would produce the largest Cp measurement at a given angle of attack and which would produce the smallest. We say in such a case that all of the Cp measurement errors are random and independent of each other.

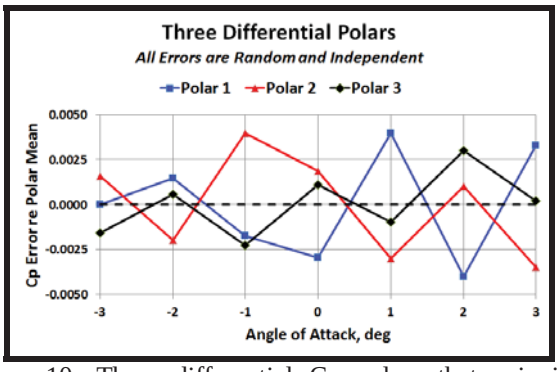

Figure 10: Three differential Cp polars that coincide perfectly except for ordinary chance variations in the data.

Contrast the situation represented in Fig. 10 with the one represented in Fig. 11. Unlike Fig 10, in which random errors are in play but there is no systematic bias from one polar to the next, Fig 11 represents a similarly idealized case in which there is no random error, but the three polars differ from each other by a systematic bias error.

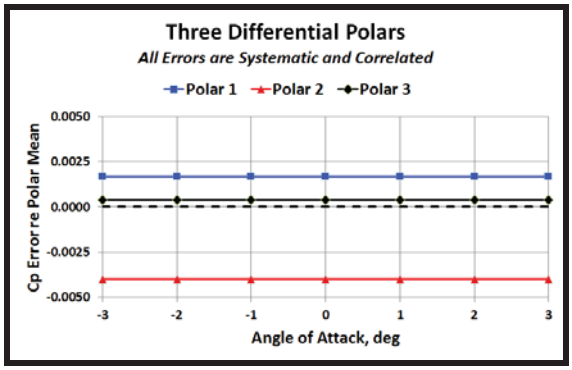

Figure 11: Three differential Cp polars that differ due to a systematic bias error.

The differential polars displayed in Figs. 10 and 11 were artificially generated to have identical sums of squares. They span nominally the same range of Cp error values, and the data samples each contain an identical number of points (21), but is the uncertainty in Cp the same in both cases? The uncertainties are in fact different for reasons associated with how uncertainty is reckoned, as will now be reviewed briefly.

Uncertainty in empirical estimates is commonly represented as the product of a standard error and a "coverage factor," *k*. The standard error is defined as the ratio of some sum of squares and the associated degrees of freedom, as described above. The coverage factor has associated with it a probability representing the likelihood that the error in an empirical response estimate will lie within a range specified by the uncertainty. It also depends on the number of degrees of freedom used to estimate the standard error.

The ISO Guide[4] recommends using values from the Student t-distribution for the coverage factor. For any specified probability, this t–statistic rapidly approaches an asymptotic limit as the volume of data increases[5]. For example, for 95% probability it approaches 1.960 in the limit, and is close enough to 2 for $\nu \geq 9$ that this value is commonly associated with data samples of such size. However, the coverage factor can be considerably larger than 2 for small data samples, reflecting the increased uncertainty in estimating the standard error itself. For only one degree of freedom, for example, the t-statistic has a value of 12.706. The 95% confidence interval for a two-point sample with $n - 1 = 1$ degree of freedom available to estimate the variance (the other degree of freedom having been consumed in estimating the mean) is thus ±12.706 standard deviations, over six times wider than it would have been if the standard error had been estimated from a large data sample. This is the reason that the uncertainty associated with Fig. 10 is different than the uncertainty associated with Fig. 11. Notwithstanding the fact that the SS is identical in both cases, there is a different number of error degrees of freedom in each case. This causes the coverage factor to be different as well as the variance.

Consider Fig. 10. At each angle of attack there are three empirical Cp estimates, so $n - 1 = 2$ degrees of freedom to estimate the standard error. Since there are seven angles of attack, there are thus 14 total degrees of freedom available to estimate the standard error. We say the 14 df are *pooled* over the angles of attack.

We compute the sum of squares at each individual angle of attack as described previously (summing the three squared differences between each data point and the three-point mean for that angle of attack), and add the individual sums of squares for all seven angles of attack to obtain a sum of squares that is likewise pooled over angles of attack.

It was noted above that the data displayed in Figs. 10 and 11 have the same SS. The numerical value of the SS is 1.241×10^{-4} for both figures. We divide this SS by the pooled df to obtain a 14-df estimate of the random error variance displayed in Fig. 10, and take the square root to obtain the corresponding pooled standard error.

$$\sigma_{Error} = \sqrt{\frac{SS_{Pooled}}{df_{Pooled}}} = \sqrt{\frac{1.241 \times 10^{-4}}{14}} = 0.0030 \tag{13}$$

Because we have more than 14 df to estimate the error, we can use a coverage factor of 2 to obtain the 95% prediction interval half-width. We can therefore quantify the uncertainty for Fig. 10 in terms of the 95% prediction interval half width as follows:

$$95\%\ \mathrm{PIHW}_{14df} = 2 \times \sigma_{Error} = 0.0060 \tag{14}$$

Now consider Fig. 11. We proceed initially just as with Fig. 10. Note, however, that not only is the SS for Fig 11 the same as the SS for Fig. 10, the per-AoA SS will be identical for every AoA. It is simply $SS_{Pooled}/7 = 1.774 \times 10^{-5}$. Call this SS_0. Likewise, let $df_0 = 2$ represent the number of degrees of freedom at each angle of attack. For Fig. 11 we therefore have that $SS_{Pooled} = 7 \times SS_0$, and $df_{Pooled} = 7 \times df_0$, so Eq. 13 becomes

$$\sigma_{Error} = \sqrt{\frac{7 \times SS_0}{7 \times df_0}} = \sqrt{\frac{1.774 \times 10^{-5}}{2}} = 0.0030 \tag{15}$$

The numerical value of the standard error is identical for the data displayed in Figs. 10 and 11, but there is this crucial difference: The data in Fig. 10 permits a 14 df estimate of the standard error but the data in Fig. 11 only permits a 2 df estimate. Even though the two standard errors are the same in this example, we cannot have as much confidence in the one based on 2 df as in the one based on 14 df. This is reflected in the difference in coverage factors. We can use a value of 2 as the coverage factor for the 14-df case, but we must use the t-statistic as the coverage factor for the 2-df case. The 2-df t-statistic for 95% confidence has a value of 4.303, so

$$95\%\ \mathrm{PIHW}_{2df} = 4.203 \times \sigma_{Error} = 0.0128. \tag{16}$$

Even though the volume of data displayed in Figs 10 is the same as displayed in Fig.11 and even though both data samples have the same variance, the uncertainty for Fig. 11 is twice that of Fig. 10. The reason is that there are fewer degrees of freedom available to assess uncertainty for Fig. 11 than for Fig. 10.

Because the data in Fig. 11 are perfectly correlated, once we have analyzed the variance about the mean at the first angle of attack, there is no further information to be had about variance from any of the other angles of attack. By contrast, because the measurements in Fig. 10 are all independent of each other, the variance at each new angle of attack is different, and therefore contributes something new to the overall understanding of the uncertainty. Stating this slightly differently, we have the equivalent of seven independent 2-df estimates of uncertainty in Fig. 10, but only one in Fig. 11. Therefore we simply have more information in Fig. 10 than in Fig. 11.

By acquiring data with correlated errors, we reduce the information available to assess the uncertainty because, by virtue of the correlation, new data points provide us with less new information about the scatter. Our estimate of the uncertainty is greater not because the physical scatter in the data is greater, but because we have more uncertainty in what the scatter actually is.

Summarizing, the uncertainty in the Cp data acquired at each pressure tap depends on whether the errors are random, as in Fig. 10, or systematic, as in Fig. 11. These two cases represent extremes that were artificially constructed to illustrate the effect of correlated errors, Fig. 12 represents actual Cp data from the present study as differential polars. These are data displayed earlier as conventional polars in Fig. 9. They were acquired from "Tap 29," which was located at the outboard spanwise row of taps on the upper surface of the wing, as the seventh tap from the leading edge in an 11-tap row.

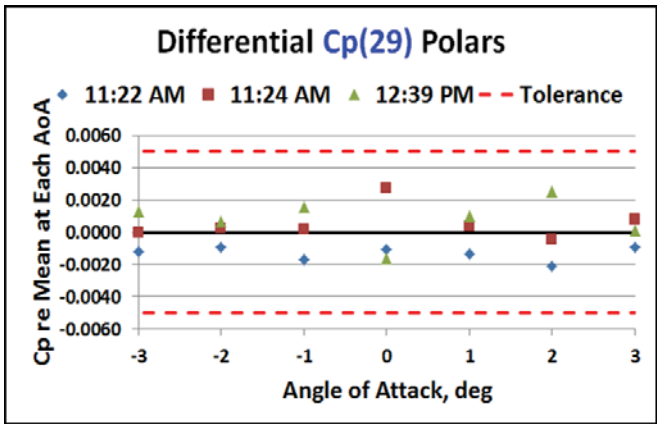

Figure 12: Three differential Cp polars. Red dashed lines assume a 95% precision interval tolerance for Cp of ±0.0050.

We need to know if the errors in Fig. 12 are independent or correlated in order to properly assess the uncertainty, but the situation displayed in Fig. 12 is not as clear as the idealized extremes presented for illustration in Figs. 10 and 11. The distributions of points do seem to differ somewhat with angle of attack, both in the amount of the scatter and the order of the points. These are characteristics of independent measurements. On the other hand, there is also fairly strong evidence of systematic between-polar differences.

For example, the error in every single data point acquired in the 11:22 AM polar is more negative than its corresponding point from the 11:24 AM polar acquired minutes later. If the errors have some given relationship for the first of seven angles of attack—the error from the first polar being more negative than the error from the second polar as in this case—the odds of the errors having the same relationship at six out of six subsequent angles of attack when there is an equal probability that the order would be reversed, is just one in 64. It is not impossible, but it is rather unlikely.

Similarly, errors in the points comprising the 12:39 PM polar seem to be generally biased higher than corresponding errors from the 11:22 polar. However, errors from the 11:24 AM polar and the 12:39 PM polar seem to more randomly distribute with respect to each other, so this visual examination is rather inconclusive. There seems to be both random and systematic components in the unexplained variance of this three-polar data sample. This suggests that we have more than two degrees of freedom available to assess uncertainty, but possibly less than 14, since some of the unexplained variance seems to be random and some seems to be systematic. We can clarify this using analysis of variance (ANOVA) methods to objectively partition the unexplained variance into random and systematic components. The ANOVA methods used for this purpose will now be briefly outlined.

Consider Table 5, which lists the data displayed in Figs 9 and 12. The 21 points in this data sample are presented as seven rows by three columns, with each row corresponding to a different angle of attack and each column corresponding to a different polar. We can calculate an $n - 1 = 20$ df estimate of the total variance of this data sample using Eq. 3, but we are more interested in partitioning the total variance into row-wise and column-wise components.

AoA	11:22 AM	11:24 AM	12:29 PM
-3	-0.2120	-0.2106	-0.2121
-2	-0.2628	-0.2625	-0.2630
-1	-0.3155	-0.3143	-0.3159
0	-0.3678	-0.3699	-0.3656
1	-0.4218	-0.4210	-0.4221
2	-0.4748	-0.4726	-0.4757
3	-0.5254	-0.5256	-0.5251

Table 5. Three Cp polar replicates

We expect most of the variance in this sample to be row-wise; that is, associated with the angle of attack changes we made for the express purpose of inducing this variance in order to reveal how Cp changes with AoA. In a perfect experiment, 100% of the variance would be row-wise, and there would be no systematic column-wise variation, as in Fig. 10. In reality, we anticipate some level of systematic column-wise variation, however small. We wish to determine if this column-wise variation is sufficiently large that it cannot be attributed to ordinary chance variations in the data. If so, it suggests that there are systematic differences between at least two of the polars, and possibly among all three.

There are several variations of ANOVA (analysis of variance) that can partition the variance of a data sample into constituent components. We use what is known as "two-way ANOVA without replication" in this analysis. Standard references such as Ref. 6 describe the computational details, and Ref 7 discusses some practical aerospace applications, but any ANOVA generates estimates of the sums of squares, degrees of freedom, and variances for constituent components of the total variance as well as for the total variance itself. In the present analysis, the ANOVA produces these results for the row-wise and column-wise variance components, as well as for a residual component of variance that always remains after other components are taken into account. This residual variance is associated with ordinary random error.

ANOVA computational routines are built into many common software packages; the calculations were performed for the present analysis using the Excel spreadsheet program. This makes the analysis readily accessible, and quite easy to use. It is simply a matter of organizing the data into an array of rows and columns as in Table 5, highlighting this array in the spreadsheet, and indicating where the results are to be displayed.

The results of a standard ANOVA typically feature certain statistics for each row and column (e.g. sums, point counts, means, and variances), but the central product is called an ANOVA table, which is essentially a convenient bookkeeping device for displaying relevant statistics for each variance component. Table 6 is such an ANOVA table, generated from the data in Table 5.

The first column in Table 6 simply identifies the variance components that are quantified. The next three columns, labeled SS, df, and MS, display the sum of squares, degrees of freedom, and "mean square" (another name for variance) that have been described in detail above. Note that the 21 data points in our sample have $n - 1 = 20$ degrees of freedom available to assess variance after one df is consumed in computing the mean. Of these 20, for "a" rows and "b" columns (a = 7, b = 3), there are $a - 1 = 6$ row-wise df and $b - 1 = 2$ column-wise df. Subtracting these 8 df from the total of 20 that we have to assess variance given the mean, leaves 12 df that are associated with random error. These results are tabulated in the ANOVA table. The last three columns in the ANOVA table, labeled "F," "P-value," and "F crit," reveal the significance of the row and column variance components.

The "F" column displays the ratio of each component's MS (variance) to the random error MS. This statistic, labeled "F" to honor Ronald Fisher who developed the ANOVA method a century ago, is the key to determining how significant each variance component is. In this example, we can see that the row-wise variance is over twenty-six thousand times greater than the variance attributable to random error, supporting a not-altogether surprising inference that changes in angle of attack do cause changes in pressure coefficient. For our purposes, this result is too well known to be of much interest.

Source of Variance	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows (AoA)	0.2310	6	3.85E-02	26297.4	0.0000	3.0
Columns (Polars)	1.90E-05	2	9.51E-06	6.5	0.0123	3.9
Random Error	1.76E-05	12	1.46E-06			
Total	0.2311	20				

Table 6. ANOVA table for three column, seven row array

Of much greater interest is the fact that the column-wise variance exceeds the random error variance by a factor of 6.5. This suggests that between-polar differences may be too great to attribute to ordinary chance variations in the data. However, the F-statistic is a random variable constructed as the ratio of two other random variables, each of which can wax and wane with each sample of data. It is at least theoretically possible that this specific data sample has an F-statistic as large as 6.5 simply because it featured an unlikely combination of unusually small random error and uncharacteristically large between-polar shifts, and that this ratio would not be likely to occur in more representative data samples.

Fortunately, the F-statistic has associated with it a probability distribution that enables us to objectively assess whether or not a given F value is likely to occur. This distribution depends on the number of degrees of freedom associated with the numerator and the denominator of the F-statistic so there is a family of such distributions, but the ANOVA table reveals that the governing distribution in our case has 2 and 12 associated with the numerator and the denominator, respectively. This distribution is displayed in Fig. 13.

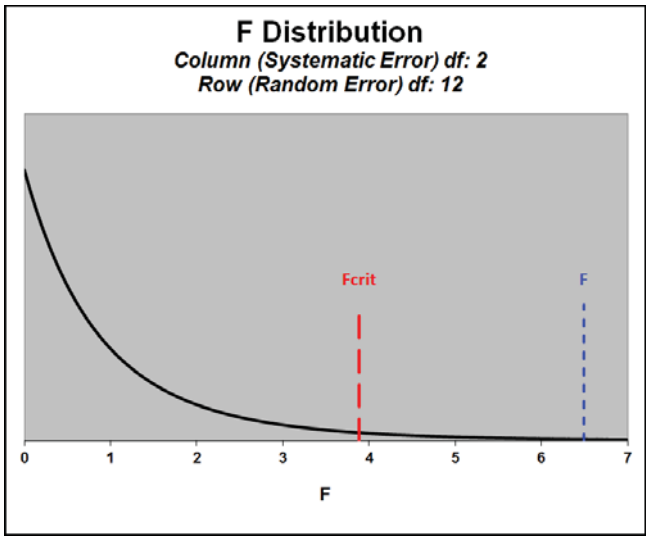

Figure 13: F distribution for 2 numerator df and 12 denominator df. Fcrit for $\alpha = 0.05$. Measured F of 6.5 from ANOVA Table 6.

The F statistic of 6.5 from ANOVA Table 6, quantifying the between-polars component of variance as a multiple of the random error variance, is also displayed in Fig. 13, as is the "Critical F value, Fcrit." Fcrit has the value of 3.9 in this case and is also taken from the ANOVA table. It has the property that the area under the F distribution to the right (greater than) Fcrit is equal to a prescribed probability value, α. In this case the value of α is 0.05, and we interpret this to mean that given the number of numerator and denominator degrees of freedom available to assess F, the probability that a particular instance of the F statistic will be less than Fcrit under the null hypothesis is $1 - \alpha$. The null hypothesis in this case is that there is no significant difference between the numerator and denominator variances used to compute the F statistic. If that is true, then the probability that ordinary chance variations in the data could conspire to generate an F statistic greater than 3.9 is no more than 5%. If the measured F is greater than Fcrit as in this example, we can reject the null hypothesis and infer with no more than a 0.05 probability of an inference error, that between-polar difference are in fact to large to be attributed only to chance variations in the data.

Each number in the ANOVA table column labeled "p-value" represents the probability that an F-statistic as large as the corresponding one in the "F" column could occur strictly due to an unlucky combination of chance variations in the data, if in fact there were no real row-wise or column-wise differences in the data. We see that this probability is represented as zero to four decimal places for the row-wise component of variance, suggesting that there is a very low probability indeed that changes in angle of attack impose Cp changes that are no greater than the random fluctuations that occur in such a data sample.

There is a p-value of 0.0123 associated with the column-wise variance component, suggesting that there is only a 1.23% probability that between-polar differences as large as were observed could have been due to nothing more than chance variations in the data. That is, we can say with 98.77% confidence that such between-polar differences were indeed systematic, and not due to random error.

In practice, we do not quote such probabilities so precisely, but simply test to see of the p-value fails to exceed some prescribed threshold such as "0.05." If it does fail to do so, we can infer with at least 95% confidence that the effect under consideration is due to some cause other than chance variations in the data; that is, that it is a real effect and not simply an artifact of noise.

Having used ANOVA to objectively infer that not all of the unexplained variance is random but that some of it is systematic, we are on the horns of something of a dilemma. Our objective is to quantify the uncertainty in our

measurements of pressure coefficient, but we have what almost seems like too much information to do this. We have variance estimates from the ANOVA table for ordinary random error and also for a component of uncertainty that can be attributed to the fact that the polar means do not seem to be time-invariant, even over relatively short time intervals. From these variance estimates we can compute standard errors for each error component, which in general we expect to be different. We also know the degrees of freedom for each type of error and can therefore generate a coverage factor for each that corresponds to a 95% precision interval, but these, too, will be different for each error type. How do we combine all of this information into a unified estimate of experimental uncertainty?

The solution to this problem lies in recognizing that ordinary chance variations in the data occur about mean values that are not stable, as is widely assumed in the experimental aeronautics community, but that actually change with time. Data points acquired over a very short interval, such as the time it takes to acquire a seven-point polar, share tunnel state conditions that will eventually change to some degree. Unless polars are acquired over an interval of time that is long enough to experience the full range of tunnel state conditions that will occur over the duration of a wind tunnel rest, we must expect some of these polars to be biased with respect to others. The experimental errors will be correlated in that case and as we have seen, the effect of correlated experimental errors is to inflate the estimates of uncertainty.

Fortunately, once we have been able to separately quantify by ANOVA the random and systematic bias contributors to uncertainty from a sample of polar replicates acquired at different times, we can use widely accepted methods for combining random error and systematic bias error. We seek pooled estimates of the standard random and bias errors, and a suitable corresponding coverage factor.

For combining random and systematic errors into a single estimate of uncertainty, the ISO Guide[4] recommends summing the mean squares (variances). Following this recommendation, we have

$$\sigma_U^2 = \sigma_S^2 + \sigma_R^2 \tag{17}$$

where σ_S^2 and σ_R^2 are the squares of the systematic and random standard errors; i.e. the systematic and random components of variance from ANOVA Table 6, and σ_U^2 is the square of the composite standard error for uncertainty. Inserting values from Table 6 into Eq. 17, we arrive at the following composite standard error for the current example:

$$\sigma_U = \sqrt{\sigma_S^2 + \sigma_R^2} = \sqrt{\left(9.51\times10^{-6}\right) + \left(1.46\times10^{-6}\right)} = 0.0033 \tag{18}$$

We follow Coleman and Steele[8] in developing a 95% coverage factor to use with the standard composite uncertainty, including their reliance on the ISO Guide[4] recommendation to use values from the t-distribution for this purpose. The t-statistic for a specified percent coverage (say, 95%) depends on the number of degrees of freedom used to estimate the standard error, as noted earlier. Our estimate of the standard systematic bias error is based on two df and our estimate of the standard random error is based on 14 df, but an effective number of degrees of freedom can be estimated with the Welch-Satterthwaite formula[4,8] as follows:

$$\nu_{WS} = \frac{\left(\sigma_S^2 + \sigma_R^2\right)^2}{\left(\sigma_S^4 \big/ \nu_S\right) + \left(\sigma_R^4 \big/ \nu_R\right)} \tag{19}$$

Here, ν_R and ν_S are the degrees of freedom used to estimate the random and systematic variance components. Inserting values from Table 6 into Eq. 19, we obtain the following effective number of degrees of freedom for the current example:

$$\nu_{WS} = \frac{\left[\left(9.51\times10^{-6}\right)+\left(1.46\times10^{-6}\right)\right]^2}{\left[\frac{\left(9.51\times10^{-6}\right)^2}{2}\right]+\left[\frac{\left(1.46\times10^{-6}\right)^2}{14}\right]} = 2.648 \tag{20}$$

One can consult standard statistical tables or use the functions built into common statistical software packages to determine that the t-statistic for 95% coverage and 2 df is 4.303, and for 3 df it is 3.182. Interpolating, we arrive at a coverage factor for 2.648 df of 3.716. Call this k_{WS}. Then the 95% prediction interval half-width for total uncertainty in this example is

$$95\%\ PIHW = k_{WS}\sigma_U = (3.716)(0.0033) = 0.0123 \tag{21}$$

If all of the errors had been random, the total unexplained variance could be quantified by adding the random and systematic sums of squares from the ANOVA table and dividing by the sum of the random and systematic degrees of freedom, to yield a 14 df estimate of the total unexplained variance, as follows:

$$\sigma^2 = \frac{\left(1.90\times10^{-5}\right)+\left(1.76\times10^{-5}\right)}{12+2} = 2.61\times10^{-6} \rightarrow \sigma = 0.0016 \tag{22}$$

The 95% coverage factor for 14 degrees of freedom can be taken to be 2, so the 95% prediction interval half-width is just 0.0032. By eliminating the systematic between polar shifts in this example, the total standard error could have been reduced from 0.0033 to 0.0016, a factor of 2.1. But because of the reduction in the 95% coverage factor from 3.716 to 2, the total uncertainty could have been reduced by an even greater margin, from 0.0123 (Eq.21) to 0.0032, a factor of 3.8.

IV. Results

Methods described in the previous section were used to analyze pressure coefficient measurements from each of nominally 50 taps in the test article described in Section II (excluding a few taps that failed). For each tap, there were three replicated Cp polars acquired within a time interval of less than an hour and a half on the same test article, at conditions that were reproduced as exactly as possible in three different facilities: AEDC's 16T, ARC's 11-FT, and LaRC's NTF. Those conditions were as follows: angles of attack from -3° to +3° in 1° increments, zero sideslip, zero roll, no control surface deflections, Mach 0.60 and a Reynolds number of 3.85×10^6 per foot.

Corrections for angle of attack set point errors were made to all data prior to comparing them. Figure 14 displays the magnitude of the AoA set point error, averaged over angle of attack, for each of the three 7-point polars analyzed in each of the three facilities.

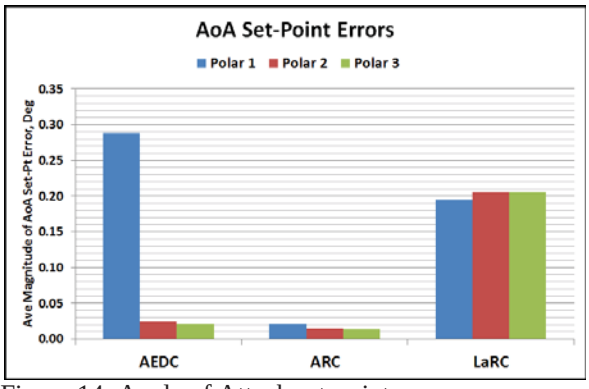

Figure 14: Angle of Attack set-point errors

AoA set-point errors are generally benign since for every data point, the actual measured angle of attack was recorded as well as the pressure coefficient, Cp. These set-point errors would not normally be problematical unless the data were analyzed under an assumption that all measurements were acquired at the AoA design points. For an analysis of within- and between-polar variation in response, however, it is necessary to normalize all of the data acquired at a given AoA design point to the same angle of attack.

The situation at AEDC would be the most problematical *for the current study* absent AoA set-point corrections. AoA set-point errors do not induce variance in the data simply because they are present, and not even because they might be relatively large. Artificial variance is induced in uncorrected data only if the set-point errors *differ*, as they did at AEDC. At LaRC, by contrast, even though the set-point errors were relatively large, they were generally consistent. Such errors would have relatively little effect on estimates of the variance in LaRC response data. Nonetheless, AoA set-point corrections were applied to all of the data before any further analysis, including the ARC data for which set-point errors were both consistent and relatively small.

As Fig. 8 and 9 suggest, the Cp dependence on angle of attack was very nearly first order over the range of AoA levels that was considered. A number of taps displayed very small second-order (curvature) effects, and a few even featured extremely small third-order (change in curvature) effects, but departures from a purely first-order relationship, especially over the small domain between adjacent AoA values, was considered small enough to justify linear interpolation as a means of normalizing the data to the same AoA design-points for all polars.

Analyzing the data began with the construction of differential polars as displayed in Fig. 12 for all of the pressure taps that yielded data, in all three of the facilities. There would have been a total of 150 such samples of data—one for each of 50 taps in each of the three facilities—if data had been available for all 50 taps. For various reasons ranging from “leaks” to “plugged taps” to “tap not connected,” 3 to 4 taps were unavailable in each tunnel, but across all three tunnels a total of 140 samples of data were analyzed, consisting of three 7-point polar replicates each.

Figure 12 is an example of one such differential polar plot, acquired in the AEDC 16T tunnel. For comparison, in Fig. 15 we also display the differential polars acquired under the same conditions from the same tap at Ames 11-FT and LaRC NTF.

Figure 15 is generally representative of the results obtained at other taps. Just as beauty is in the eye of the beholder, those who might hope that the unexplained variance would be dominated by relatively easy to analyze random fluctuations in the data will be able to find evidence supporting their case in these plots. Likewise, those who are especially wary of systematic errors will have no trouble seeing such effects in the same plots.

Because of the inevitable human inclination to see in the data support for a priori expectations and to discount contrary evidence as perhaps unrepresentative, it is particularly important in cases where the effects are as subtle as in the present analysis to rely on *objective* measures of the degree to which the measurement errors are independent. This is especially true given the high degree to which valid estimates of measurement uncertainty are contingent on such independence of individual measurement errors.

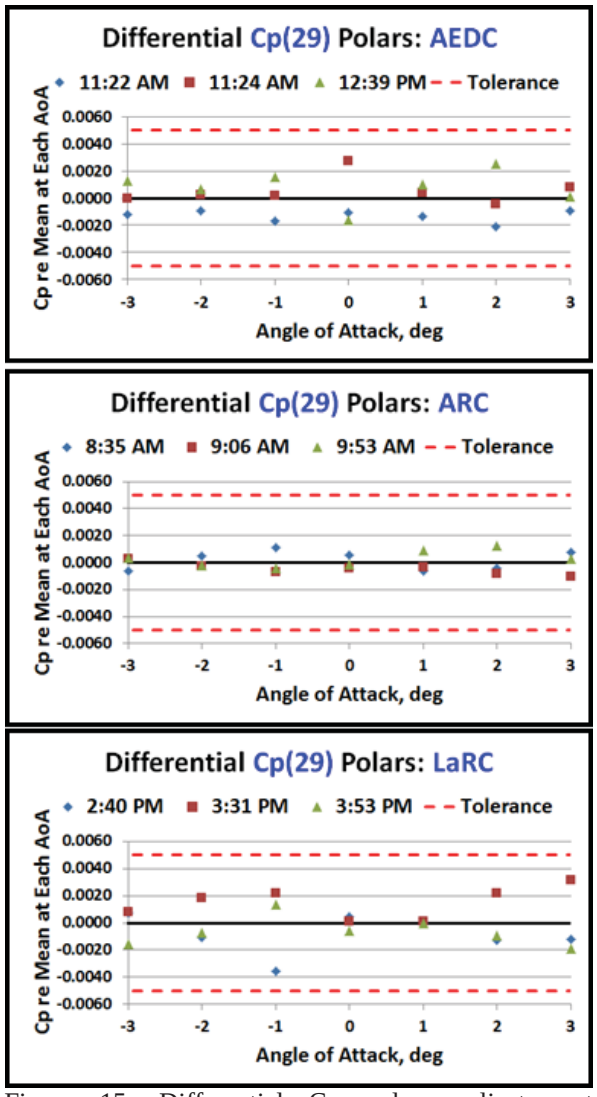

Figure 15: Differential Cp polar replicates at nominally identical conditions in three transonic tunnels. Red dashed lines assume a 95% precision interval tolerance for Cp of ±0.0050

Qualitative displays of measurement error provided by differential polar plots such as Fig. 15 were augmented in this study by objective ANOVA computations for each three-polar data sample. Figure 16 displays the ANOVA tables corresponding to the three cases in Fig. 15. Of particular interest is the p-value for columns (polars) in each ANOVA. Recall that for ANOVA tables such as these that are based on data arrays of the form of Table 5, p-values less than 0.05 for columns imply systematic column-wise variation that can be detected with at least 95% confidence. A glance at Fig. 16 shows that this is the case for the Tap 29 Cp data acquired at AEDC and at Langley, while we are unable to say with 95% confidence that for this tap the ARC polar means are displaced from each other by more than can be explained by ordinary chance variations in the data. We conclude that at Ames, the random component of the unexplained variance was dominant for this particular data sample, and that systematic between-polar bias shifts were insignificant.

	Source of Variance	SS	df	MS	F	P-value	F crit
AEDC 16T	Rows (AoA)	0.2310	6	3.85E-02	26297.4	0.0000	3.0
	Columns (Polars)	1.90E-05	2	9.51E-06	6.5	0.0123	3.9
	Random Error	1.76E-05	12	1.46E-06			
	Total	0.2311	20				

	Source	SS	df	MS	F	P-value	F crit
ARC 11 FT	Rows (AoA)	0.2365	6	3.94E-02	74526.2	0.0000	3.0
	Columns (Runs)	2.30E-06	2	1.15E-06	2.2	0.1565	3.9
	Error	6.35E-06	12	5.29E-07			
	Total	0.2365	20				

	Source	SS	df	MS	F	P-value	F crit
LaRC NTF	Rows (AoA)	0.2320	6	3.87E-02	16939.8	0.0000	3.0
	Columns (Runs)	2.34E-05	2	1.17E-05	5.1	0.0246	3.9
	Error	2.74E-05	12	2.28E-06			
	Total	0.2320	20				

Figure 16: ANOVA tables for nominally identical polar replicates acquired in three tunnels from Tap Cp(29). Compare with differential polar replicates displayed in Fig. 15.

The p-values for all 140 analyzed data samples were examined to see if any trends were evident in the role that systematic error played. It was of particular interest to see how the number of taps with significant systematic error varied by tunnel, and also by tap location on the test article.

Figure 17 is a schematic representation of tap location, indicating the degree of systematic unexplained variance at each tap location. It consists of a number of small rectangles that each represents a pressure tap. Forty-four of the rectangles are clustered into arrays representing the chordwise rows of taps located on the wing surfaces. See Fig. 3. For each wind tunnel, the 22 rectangles in the upper part of the array represent upper wing surface taps. Of those, the 11 on the left correspond to the more inboard spanwise location of a chordwise row of upper-surface taps and the 11 on the right correspond to the more outboard spanwise location of a similar chordwise row of upper-surface taps.

The 22 rectangles in the lower part of the array represent lower wing surface taps, with the 11 on the left and the 11 on the right again corresponding to inboard and outboard spanwise locations, as labeled in the figure. The 11 rectangles in each of the four upper quadrants are arrayed such that the upper-most rectangle corresponds to the tap closest to the leading edge of the wing, while the lower-most rectangle corresponds to the tap nearest the trailing edge. That is, the direction of flow is from top to bottom in Fig. 17.

There are similarly six small rectangles centered below the four wing-surface tap clusters, each corresponding to a tap on the lower fuselage surface as in Fig. 3. These are also arrayed such that the upper-most tap is forward-most and the lower-most tap is aft-most.

The rectangles are color-coded. Black taps generated no data. The green, yellow, and red taps are coded according to the p-value for the ANOVA executed on three Cp polar replicates acquired with that tap. The colors therefore indicate the degree to which systematic between-polar bias shifts dominated the unexplained variance.

Green taps signify p-values in excess of 0.05, as in the case of Ames 11-Ft in Fig. 16. For these taps, the observed between-polar variations are so small that the probability that they can be attributed to ordinary chance variations in the data exceeds 0.05. This means that for the data acquired from the green taps, there is less than a 95% probability that systematic between-polar differences are in play. In such cases, we cannot claim to have detected systematic polar shifts with at least 95% confidence, and we conclude therefore that the unexplained variance is dominated by random error, with insignificant systematic effects in play. Experimental errors from individual measurements acquired using the "green taps" are thus relatively independent, which is highly desirable.

For the yellow taps in Fig. 17, the ANOVA p-values are smaller than 0.05. For these taps, systematic between-polar bias shifts were large enough to detect with at least 95% confidence, but not with 99% confidence. We

describe any systematic component of unexplained variance that can be detected with at least 95% confidence as "significant," by which we mean *statistically* significant. That is, the bias shifts are large enough to detect unambiguously, but whether they are large enough to have *physical* significance is another matter. To address that issue, the uncertainty must be properly evaluated and compared to independently declared tolerance levels, about which more presently.

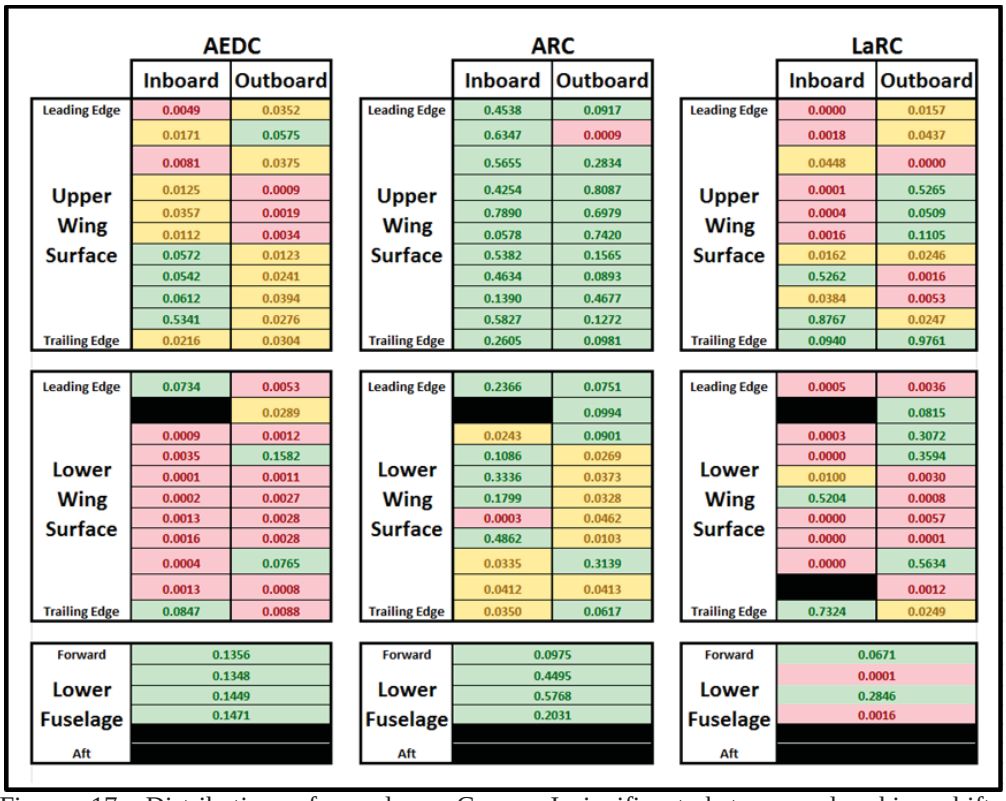

Figure 17: Distribution of p-values. Green: Insignificant between-polar bias shifts ($0.05 < p < 1$), Yellow: Significant between-polar bias shifts ($0.01 \leq p \leq 0.05$), Red: Very significant between-polar bias shifts ($0 < p < 0.01$).

The red taps in Fig. 17 have p-values smaller than 0.01. Data acquired from these taps feature between-polar systematic bias shifts that are large enough be detected with at least 99% confidence, a level of certainty that we describe as "very significant." In these cases there is less than one chance in 100—and depending on the magnitude of a specific p-value it could be *much* less than one chance in 100—that systematic between-polar bias shifts are the result of some unlucky combination of chance variations in the data, rather than real systematic state changes of some kind from polar to polar. It is therefore quite likely that not all experimental errors are independent in data acquired from the red taps of Fig. 17. The effective number of degrees of freedom available to assess uncertainty is reduced in those cases, which inflates estimates of the total uncertainty by increasing the coverage factor, as described above.

Perhaps the most obvious pattern that emerges from Fig. 17 is that the data from AEDC and LaRC seem to have more in common with each other than they do with the data acquired at Ames. Specifically, the Ames data features more green taps than either of the other two tunnels. This is especially true for the upper-surface taps, where 21 out of 22 taps are green for Ames, while only five upper-surface taps out of 22 are green for AEDC and only seven out of 22 for LaRC.

On the lower wing surfaces there are fewer green taps at Ames than on the upper wing surfaces, suggested that the upper-surface data are more readily reproducible, but even on the lower surface, there were roughly twice as many green taps at Ames than at either of the other two tunnels—ten lower wing surface taps at Ames were green vs. four at AEDC and six at LaRC.

Not only were there more taps at AEDC and at LaRC with systematic between-polar bias shifts than there were at ARC, the shifts were generally larger at AEDC and at LaRC than at Ames. There is greater certainty in detecting

between-polar bias shifts at taps color-coded red in Fig. 17 than at taps color-coded yellow. This increase in certainty is attributable to larger bias shifts that are easier to see. Of the 12 taps at Ames for which significant between-polar bias shifts were detected (yellow or red taps), only two were red—less than 20%. At AEDC there were 34 taps with significant between-polar bias shifts (yellow or red), and 20 of these were red—almost 60%. At LaRC there were 30 taps with significant between-polar bias shifts and 22 of these were red—over 70% of the taps with systematic between-polar bias shifts.

In short, the Ames Cp data seem to be more repeatable than the data acquired at AEDC and at LaRC. There are fewer cases of systematic between-polar bias shifts, and the shifts that are observed at ARC have a somewhat lower probability of detection, suggesting that they are generally smaller.

Figure 18 summarizes the data displayed in Fig. 17. Each pie chart shows for a given tunnel the fraction of pressure taps in this analysis that had a given likelihood of experiencing systematic, between-polar bias shifts. Round-off errors explain why the percentages do not sum to 100% in every case. As in Fig. 17, the green sectors correspond to cases in which no significant between-polar systematic bias shifts were observed, while the yellow and red sectors correspond to cases in which significant and very significant bias shifts were detected, respectively.

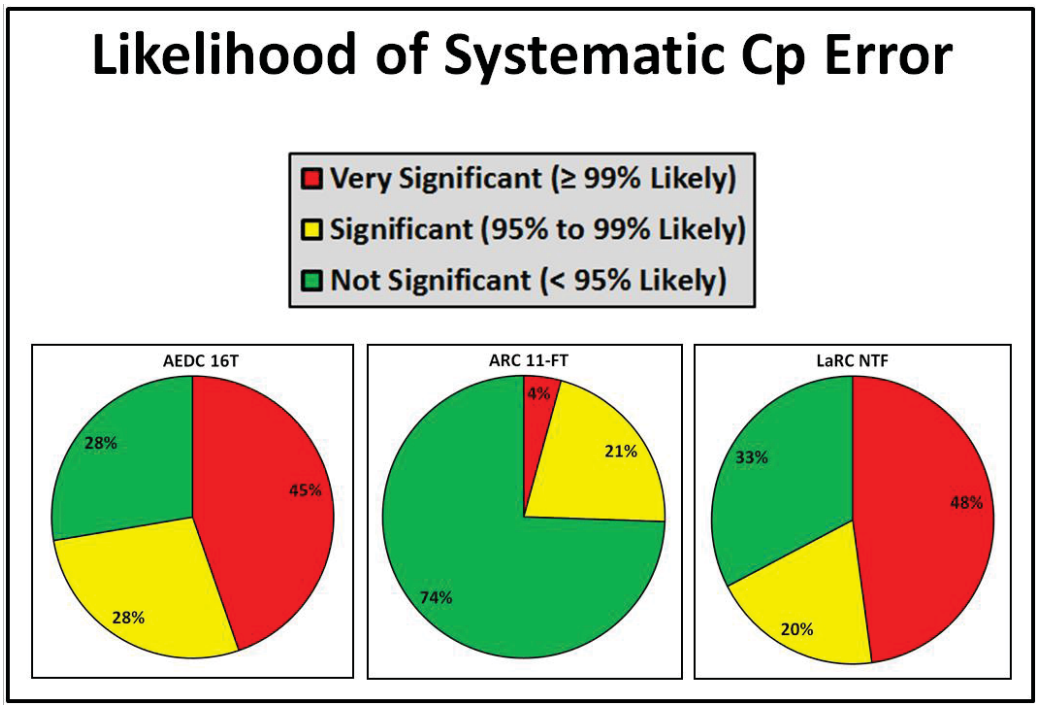

Figure 18: Magnitude and frequency of significant systematic between-polar bias shifts. Green: Insignificant between-polar shifts ($0.05 < p < 1$), Yellow: Significant between-polar shifts ($0.01 \leq p \leq 0.05$), Red: Very significant between-polar shifts ($0 < p < 0.01$).

Figure 18 further clarifies the difference between results obtained at ARC 11-FT on the one hand, and at AEDC 16T and LaRC NTF on the other. At AEDC and at LaRC, it not only appears that there are conditions for which experimental errors are not independent, it seems that this is the most likely scenario, with the occurrence of independent measurement errors as a relative exception. In almost half the cases (45% for AEDC 16T, 48% for LaRC NTF), systematic between-polar bias shifts were large enough to be detected with at least 99% confidence. They were large enough to detect with at least 95% confidence 73% of the time at AEDC and 68% of the time at LaRC. These systematic between-polar bias shifts result in reduced measurement independence, which effectively decreases the amount of information available to assess uncertainty. This can increase coverage factors, resulting in inflated estimates of total uncertainty. Even at ARC, where correlated measurement errors are less frequent, they still seem to occur too often to ignore—about one time in four.

Figures 17 and 18 provide objective evidence of the relatively frequent occurrence of systematic errors that are large enough to be detected unambiguously. This will be of special theoretical interest to quality assurance engineers who understand the implications, and who can therefore certify that interest in the independence of experimental errors is strictly pragmatic. It affects estimates of the total uncertainty, which we will now consider.

To review briefly, we express uncertainty as the product of some standard error (a "one sigma" value) and a coverage factor that depends on 1) a prescribed level of confidence, and 2) the amount of information upon which the standard error estimate was based. The ISO Guide[4] recommends using the Student t-distribution to quantify the coverage factor, so we compute the uncertainty as follows:

$$U_{1-\alpha} = t_{1-\alpha,df}\,\sigma_U \tag{23}$$

where α is some acceptable inference error probability and $1 - \alpha$ is the corresponding confidence level. A common value of α is 0.05, for which the confidence level is 95%. The quantity "*df*" is the number of degrees of freedom upon which the estimate of σ_U is based.

The standard error for total uncertainty, σ_U, was introduced in Eq. 17. It reflects the ubiquitous random error, as well as any significant systematic error that may be in play. While the standard error quantifies the actual physical scatter in the data, the coverage factor reflects how well the standard error is actually known. For a specified α, the coverage factor depends only on the number of degrees of freedom used to estimate σ_U, which we calculate with the Welch-Satterthwaite formula of Eq. 19 when systematic errors are in play.

We are concerned about the likelihood of significant systematic unexplained variance in the measurement environment not only because of the impact it can have on the scatter of the data, but because of the effect it has on reducing the number of degree of freedom available to estimate the total standard error, σ_U. This reduction is reflected in the Welch-Satterthwaite approximation of available degrees of freedom, which is less than the number that would otherwise be available in the absence of systematic bias error. The coverage factor, now comprised of a t-statistic with reduced degrees of freedom, is therefore inflated. This simply reflects the fact that we have greater uncertainty about the total standard error, σ_U. As a result, the total uncertainty is greater.

For these reasons, while it is interesting to assess facilities on the basis of how likely it is that the data points they generate feature independent experimental errors, it is more relevant to assess the *effect* of reduced independence, by quantifying the total uncertainty. We proceed exactly as we did with the detailed example presented earlier. We quantify the variance associated with within-polar variation (ordinary random error responsible for uncertainty limits around a single polar). We also quantify the component of unexplained variance that is associated with systematic between-polar bias shifts when there are multiple polars in the data sample. Both of these numbers come from the MS (Mean Square) column of an ANOVA table. Following the ISO Guide[4], we compute the square root of the sum of these two mean squares to generate a standard error that reflects both the random and systematic bias components of the unexplained variance.

We then multiply this composite standard error by a coverage factor drawn from the t-distribution for 95% coverage and a number of degrees of freedom that is computed with the Welch-Satterthwaite formula, Eq. 19. The Welch-Satterthwaite effective degrees of freedom has an upper bound equal to the number of degrees of freedom that would be available to assess uncertainty in the absence of correlated errors, and a lower bound equal to the smaller number of degrees of freedom that would be available if all of the unexplained variance was systematic. The actual value will depend on the ratio of systematic (between-polar) variance to the ordinary random error variance. That is, it will depend on the F-statistic from the ANOVA table. We can demonstrate this by simply substituting the definition of the F statistic into a slightly revised version of the Welch-Satterthwaite formula displayed in Eq. 19:

$$F = \frac{\sigma_S^2}{\sigma_R^2} \rightarrow \nu_{WS} = \frac{\left(\sigma_S^2 + \sigma_R^2\right)^2}{\left(\sigma_S^4 \big/ \nu_S\right) + \left(\sigma_R^4 \big/ \nu_R\right)} = \frac{\left[\sigma_R^2\left(1+F\right)\right]^2}{\sigma_R^4\left[\left(1 \big/ \nu_R\right) + \left(F^2 \big/ \nu_S\right)\right]} = \frac{\left(1+F\right)^2}{\left(1 \big/ \nu_R\right) + \left(F^2 \big/ \nu_S\right)} \tag{24}$$

By inspection one can see from Eq. 24 that, as expected, ν_{WS} approaches ν_R as F approaches zero, and ν_S as F approaches infinity.

Calculations as described here and demonstrated with the detailed example presented earlier were performed for all 140 three-polar data samples considered in this study, resulting in estimates of the 95% prediction interval half-width for each such data sample. Results for all taps are given in the Appendix but Fig. 19 displays representative results obtained for the upper wing surface at the inboard spanwise location at AEDC.

For each tap, two estimates of uncertainty are displayed. The first, in red, is typical of what is commonly reported for unreplicated polars. This would correspond to uncertainty limits around such a single polar, as indicated by the dashed lines in Fig. 8. This would also correspond roughly to the uncertainty band around multiple polars that

all coincided perfectly, except for ordinary random variations in the data. In either of these cases, bias errors that would tend to displace the polars with respect to each other are not normally taken into account.

In reality, replicated polars that might each display uncertainty bands as tight those in Fig. 8 can be displaced somewhat from each another. That is, the second polar can be biased slightly with respect to the first, and likewise for the third if there are three polars. This enables us to detect a bias error in addition to the single-polar random error seen in Fig. 8.

If we regard the polar mean for a sample of polar replicates as an unbiased estimator of the truth, then the displacement of each polar from that mean can be regarded as a sample drawn from some distribution of bias errors with a mean of zero and some particular standard error. This bias standard error will be different from the standard error associated with ordinary chance variations in the data. If the bias error is large enough compared to the random error, we say it is "significant" in the statistical sense, meaning that it can be detected with high confidence. Polar replicate data samples such as those analyzed in this study are said to display significant bias error when the ANOVA p-value is less than 0.05, as for the pressure taps represented with yellow and red color coding in Fig. 17, summarized in Fig. 18.

The addition of this bias error is reflected in the greater size of the green bars in Fig. 19, and in the appendix. The green bars are considerably larger in Fig. 19 than the red bars, due to two effects. There is an increase in the physical spread of the data due to the between-polar bias shifts. The coverage factors are also inflated because the loss of independence induced by a significant systematic component of the unexplained variance has the effect of reducing the number of degrees of freedom that are available to assess uncertainty, per Welch-Satterthwaite.

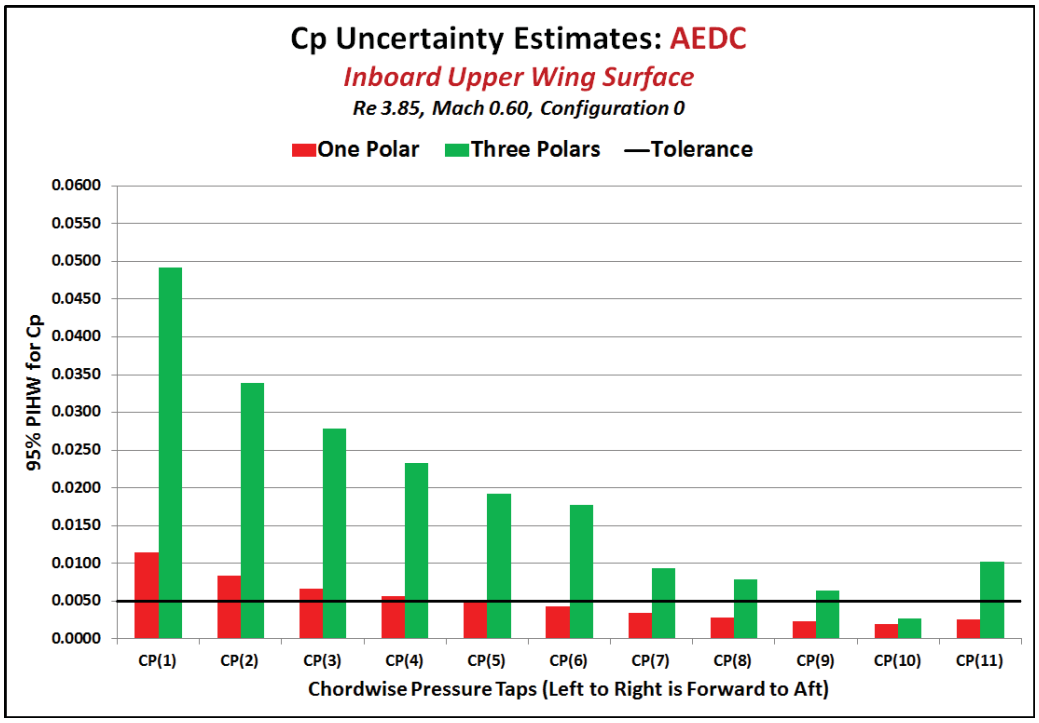

Figure 19: Inboard upper wing surface Cp uncertainty estimates at AEDC 16T, Mach 0.60, Reynolds Number 3.85×10^6, no control surface deflections. Red bars are estimates that would have resulted from a single polar. Green bars display increased uncertainty reflecting between-polar bias shifts among three polars. Tolerance of 0.0050 assumed.

Figure 19 reveals that the ordinary random errors we can see by analyzing only one polar (the red bars) are generally within tolerance if we adopt a common convention and declare 0.0050 as the tolerance limit for Cp. There also seems to be a modest trend in these results, by which random errors near the leading edge are somewhat greater than the random errors nearer the trailing edge. This trend is accentuated for the green bars, suggesting that not only is the random error slightly greater nearer the leading edge than the trailing edge for the data acquired at AEDC, but that systematic between-polar bias shifts are also greater near the leading edge. That is, it seems as if it is easier to duplicate a Cp polar near the trailing edge than near the leading edge. These same general trends are observed in the data acquired at LaRC in the NTF, Fig. 20, although not at the AEC 11-FT tunnel, Fig 21, where random error and

total random-plus-systematic error seems to be generally constant from leading edge to trailing edge. Figures 19-21 display uncertainties in Cp data acquired with the inboard upper wing surface pressure taps, but a comprehensive display of results also obtained elsewhere on the test article can be found in the Appendix.

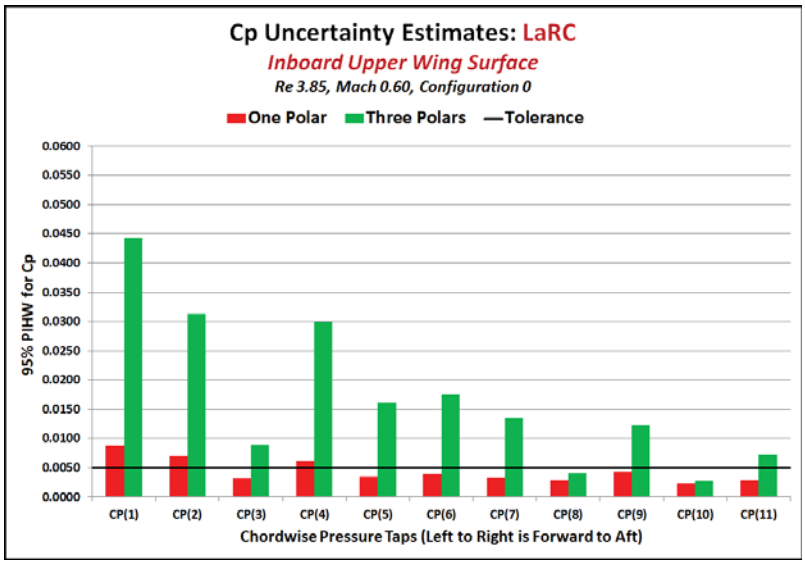

Figure 20: Inboard upper wing surface Cp uncertainty estimates at LaRC NTF, Mach 0.60, Reynolds Number 3.85 × 10^6, no control surface deflections. Red bars are estimates that would have resulted from a single polar. Green bars display increased uncertainty reflecting between-polar bias shifts among three polars. Tolerance of 0.0050 assumed.

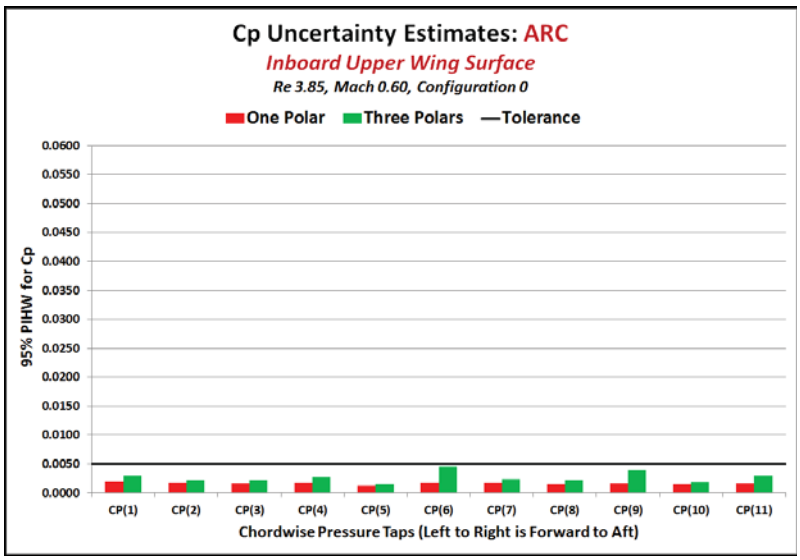

Figure 21: Inboard upper wing surface Cp uncertainty estimates at ARC 11-FT, Mach 0.60, Reynolds Number 3.85 × 10^6, no control surface deflections. Red bars are estimates that would have resulted from a single polar. Green bars display increased uncertainty reflecting between-polar bias shifts among three polars. Tolerance of 0.0050 assumed.

Figures 19-21 compare Cp uncertainties to a tolerance level specified for this analysis to be 0.0050. This is consistent with a common informal industry standard, but such tolerance levels are necessarily arbitrary. Obviously more of the measurements would have been within a more forgiving tolerance. Likewise, fewer would have been

within tolerance had it been more strict than 0.0050. It is the end user's prerogative (and his responsibility) to declare what level of tolerance is acceptable.

With these caveats, Fig. 22 reveals how often measurements were within a tolerance of 0.0050 in each of the tunnels considered in this analysis. The fraction of within-tolerance measurements is displayed for two error scenarios: 1) if only ordinary random error were in play (red bars), and 2) if replicated polars revealed bias shifts that were also taken into account (green bars).

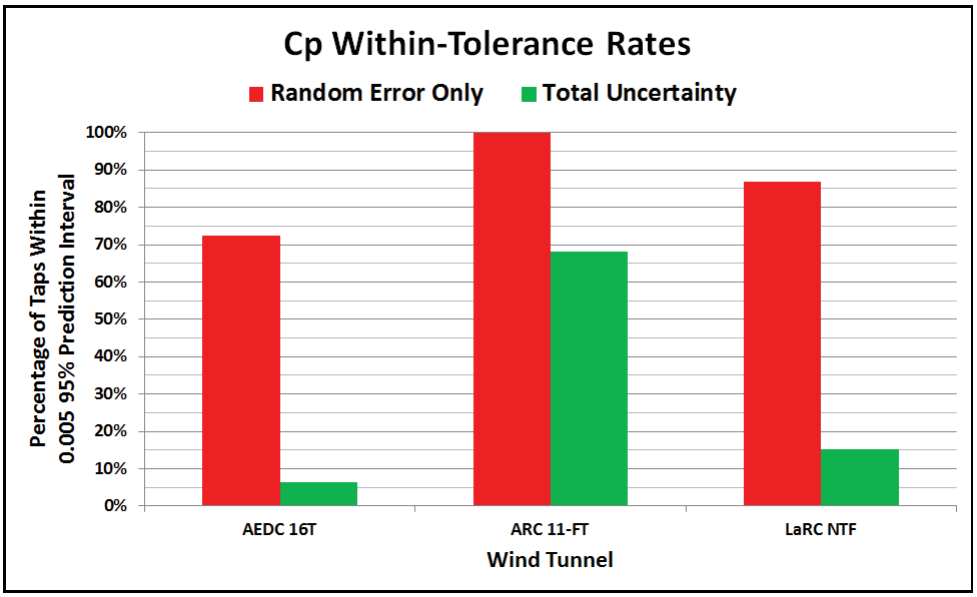

Figure 22: Frequency with which random errors (red bars) and total random plus systematic errors (green bars) were within a common 95% Prediction Interval Half Width tolerance limit of 0.0050.

It comes as no surprise that the estimated uncertainty is more likely to be within a specified tolerance if an important component of the total uncertainty is neglected. Figure 22 shows how much more often random-only errors were within tolerance than random-plus-systematic errors. At ARC 11-FT, all random errors were within tolerance. The random errors were within tolerance at AEDC in over 70% of the cases and at LaRC in over 85% of the cases.

The random error results of Fig. 22 describe individual Cp measurements. Multi-point sample means would have been within tolerance at AEDC and LaRC more frequently due to the cancelation of random errors. Nonetheless, when single polars are acquired as is the normal practice in current wind tunnel testing, no information is available to assess systematic between-polar bias shifts that cause the increases in true uncertainty that are revealed by the green bars. On those rare occasions when polar replicates are acquired, there is seldom any effort to distinguish between random and systematic components of the total unexplained variance; all variance is commonly regarded as random, which artificially inflates the number of degrees of freedom available to assess uncertainty.

The total random plus systematic errors at AEDC and LaRC were within tolerance much less frequently than the random-only errors; just over 5% of the cases at AEDC and about 15% of the cases at LaRC. These frequencies could have been improved by increasing the number of replicated polars, which would have added to the paltry number of degrees of freedom (two) available to assess the systematic component of the unexplained variance. Unfortunately, this is not a practical solution, since in a typical wind tunnel test resource constraints argue against acquiring polars even three times, much less substantially more often than that.

A detailed discussion of available quality assurance tactics is beyond the scope of the current paper, but the primary reason for the inflated total uncertainty observed at AEDC and LaRC is that the total unexplained variance was dominated in both cases by systematic, not random, errors. The key to reducing the total uncertainty is therefore to reduce the systematic error.

V. Discussion

A few further remarks are offered here on the role of independent measurements in experimental aeronautics. Some additional consequences of failing to ensure independence are addressed, and general approaches to ensuring independence are noted.

If the experimental errors associated with all measurements are independent, the experimental error of the mean of any sample is reduced with respect to the error of a single measurement by a factor of the square root of the sample size. In the limit as the sample size increases without bound, the experimental error of the sample mean approaches zero. The mean of a sample of any finite size will have associated with it a non-zero error, but that error can be driven below arbitrarily small tolerance levels acquiring *independent* measurements in sufficient volume.

This inverse relationship between experimental error and data volume is the basis for a common assumption that "more data is better than less data," and explains at least in part the focus on maximizing data volume in current wind tunnel testing. While the details are beyond the scope of this paper, we note in passing that the value added by each new point is actually a monotonically decreasing function of the volume of data currently in hand, so that the next point must eventually cost more to acquire than the value it adds; "more is better" is therefore only valid when the incremental cost of data acquisition is zero. Nonetheless, the error in the mean of a sample of independent measurements can in fact be driven as low as desired, if resources are available to acquire a sufficient volume of data.

It is common in experimental aeronautics for claims of measurement precision to rest on the fact that errors in the means of independent measurement samples decrease with the size of the sample. It is less common to make any effort to ensure that such independence actually exists in a wind tunnel test. To the extent that independence is considered at all, it seems to be frequently regarded as a default state of nature. That is, it appears to be widely believed that the independence of experimental errors can simply be *assumed*, absent any blunders that might otherwise destroy that independence.

Unfortunately, it is not always (or even usually) the case that experimental errors are truly independent of each other. It is much more likely that measurements acquired over a relatively short interval of time will have more in common with each other than with measurements made later or earlier, as in the current study. This is amply illustrated with the Cp ANOVA results presented here, in which polars frequently exhibited systematic biases with respect to each other that were large enough to be unambiguously detected, even when those polars were acquired in fairly rapid succession.

The presence of such systematic measurement differences destroys the independence of experimental errors in a given data sample. If all points in one polar are biased in the same direction relative to the points in another polar, the experimental error in one point is no longer independent of the error in the next. If the first error is positive, the second will be much more likely to be positive than negative. Absent independent errors, there is thus a reduced capacity for experimental errors to cancel, since increasing data volume in such circumstances does not have the anticipated effect of reducing uncertainty as the square root of the sample size.

In general, when the independence of experimental errors has not been ensured, sample statistics such as means and standard deviations that are used to characterize random variables of interest do not represent unbiased estimates of their corresponding population parameters. In short, correlated errors virtually guarantee the wrong answer in an experiment. It is simply a matter of degree. Note also that just as the uncertainty induced by random error exists whether replicated data points revealing it are acquired or not, uncertainty associated with the systematic component of unexplained variance is also present, whether replicated polars revealing it are acquired or not.

There are two ways to increase the independence of measurements acquired in a wind tunnel. The first, apparently implemented at ARC, is to physically eliminate much of the systematic error by various means, resulting in conditions in which replicated polars consistently fall nearly on top of each other. With much of the systematic error thus removed, most of what remains is random. We would characterize this approach as "perfecting the measurement environment."

An alternative approach that might be described as "coping with the measurement environment" is to accept it as is, but to design the wind tunnel test in such a way as to ensure that measurement errors are actually independent regardless of the environment. This is in fact rather easy to do, by experimental methods that have the effect of converting the natural systematic component of the unexplained variance into random error. The total uncertainty can then be driven below any specified tolerance limit by simple replication, which will cancel independent, random errors.

Obviously, any improvements in the measurement environment that can reduce systematic bias errors with reasonable cost and effort should be implemented. However, the author strongly recommends "coping" over "perfecting" when it comes to the measurement environment. The reason is that even in an environment such as

ARC 11-FT, in which systematic errors appear to have been substantially reduced, it is not possible to eliminate them entirely. This is evident from Fig. 21 and related plots presented in the Appendix, which shows that because of a non-zero bias error, the total uncertainty exceeds the random-only uncertainty at ARC in every case, if only by a small amount. In any case, it is never possible to be certain that the systematic component of the unexplained variance is negligible, even when it might actually be. The prudent approach is to assume that systematic errors are, or might be, in play, and to design the test accordingly, ensuring independent measurement errors by a variety of available quality assurance tactics.

The true quality of experimental aeronautics can be enhanced substantially by investing some time and effort in such tactics. The details are beyond the scope of the present paper, but the virtues and necessity of independent measurements have long been recognized outside of the aerospace industry, and techniques to ensure them are well documented in the literature of formal experiment design[9-12].

VI. Summary and Concluding Remarks

Pressure data acquired for the same conditions and on the same test article in three transonic wind tunnels have been analyzed to quantify experimental errors in each facility. The tunnels were AEDC 16T, ARC 11-FT, and LaRC NTF. Data samples were acquired in each of the three tunnels from nominally 50 pressure taps, although because a few of the taps did not yield data in each tunnel a total of 140 data samples were available for analysis instead of 150. Each data sample consisted of three 7-point pressure coefficient polars. Angles of attack ranged from -3° to +3° in 1° increments for all 420 polars included in this analysis. All data were all corrected for angle of attack set-point error using interpolation.

Plots of selected differential polars revealed systematic between-polar differences that exceeded the within-polar random error. This motivated an analysis of variance (ANOVA) to objectively partition the total variance in each data sample into components due to changes in angle of attack, changes from one polar to another, and ordinary chance variations in the data (random error).

The ANOVA results revealed significant systematic between-polar differences in over half of the 140 data samples. 68% of the samples acquired in the NTF had such differences, 72% of the samples acquired in AEDC 16T, and 25% of the samples acquired at ARC 11-FT.

The 21-point data samples had 14 degrees of freedom available to assess unexplained variance, after losing one to the sample mean and $n - 1 = 6$ to the seven levels of angle of attack. Of these 14 df, two were associated with systematic between-polar differences and the remaining 12 were associated with ordinary random error.

Because only two degrees of freedom were available to assess errors that systematically biased polars with respect to the polar mean of each sample, the associated 95% t-statistic had a value of 4.303. This relatively large coverage factor, coupled with the fact that the standard error ("one sigma") for systematic between-polar variations was often greater than the standard random error, resulted in systematic error estimates that were large compared to the random error.

Methods were used to determine 95% prediction interval half-widths for total uncertainty that were recommended in Coleman and Steele, by which random and systematic error components were combined using the Welch-Satterthwaite approximation for effective degrees of freedom to determine the t-statistic to use as a coverage factor. Results of this analysis revealed that 95% total uncertainty levels were greater in AEDC 16T and LaRC NTF than in ARC 11-FT. About 6% of the data samples acquired in AEDC 16T had total uncertainty levels that were within a tolerance level of 0.0050, and about 15% of the samples at LaRC NTF were within the same tolerance level, while 68% of the samples acquired at ARC 11-FT were within this tolerance.

If only random errors were considered and not between-polar systematic bias errors, 100% of the samples acquired at ARC 11-FT were within the 0.0050 tolerance. At LaRC NTF, the random error was within tolerance for 87% of the data samples, and for 72% of the samples at AEDC 16T.

Significant systematic error components were observed in the majority of data samples examined in this study. It was noted that the effect of such errors is to reduce measurement independence, which has the effect of reducing the information available to assess uncertainty. This reduction in available information translates into inflated coverage factors and greater overall uncertainty.

Two general strategies were described for ensuring measurement independence and thus reducing uncertainty for any given data sample. These were described as "perfecting the measurement environment," by physically eliminating sources of systematic error; and "coping with the measurement environment," by implementing quality assurance tactics that have the effect of converting existing systematic error into random error. Given how common it was to encounter data samples with reduced measurement independence in this study, and also given the relatively dramatic adverse impact that this loss of independence was observed to have on the estimates of total uncertainty, it is highly recommended that some test resources be devoted to actually assuring independence rather than simply

declaring it by assumption. Because “perfecting” the measurement environment can be difficult, and because it is also difficult to know when such efforts have been successful, the author recommends “coping” with real-world measurement environments by using quality assurance tactics that are widely available through the literature of formal experiment design.

Appendix

This appendix displays 95% prediction interval half-width estimates for Cp data acquired from multiple pressure taps on the same test article as tested under nominally identical conditions in three transonic wind tunnels: AEDC 16T, ARC 11-FT, and LaRC NTF. Test conditions were as follows: Mach 0.60, Reynolds Number 3.85×10^6, zero roll and sideslip, no control surface deflections. Red bars are estimates that would have resulted from a single polar. Green bars display increased uncertainty reflecting between-polar bias shifts among three polars. Tolerance of 0.0050 assumed.

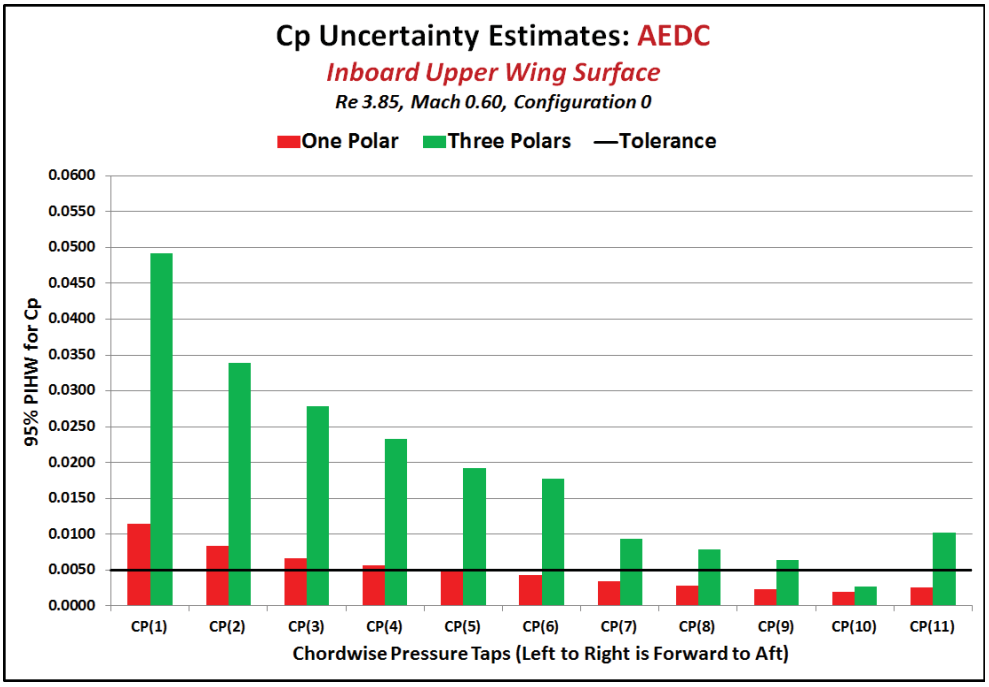

Figure A1: Inboard upper wing surface Cp uncertainty estimates at AEDC 16T.

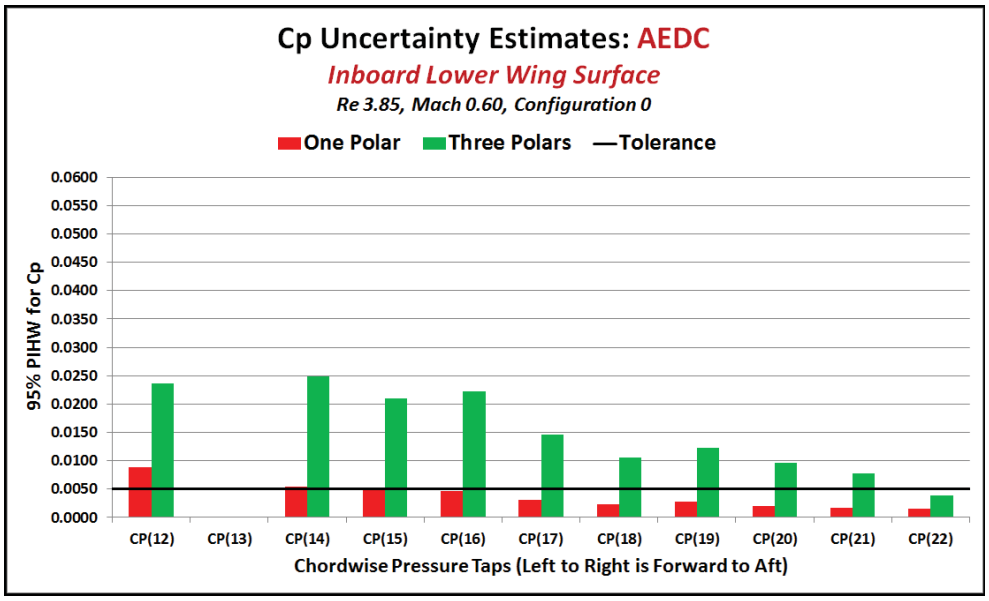

Figure A2: Inboard lower wing surface Cp uncertainty estimates at AEDC 16T.

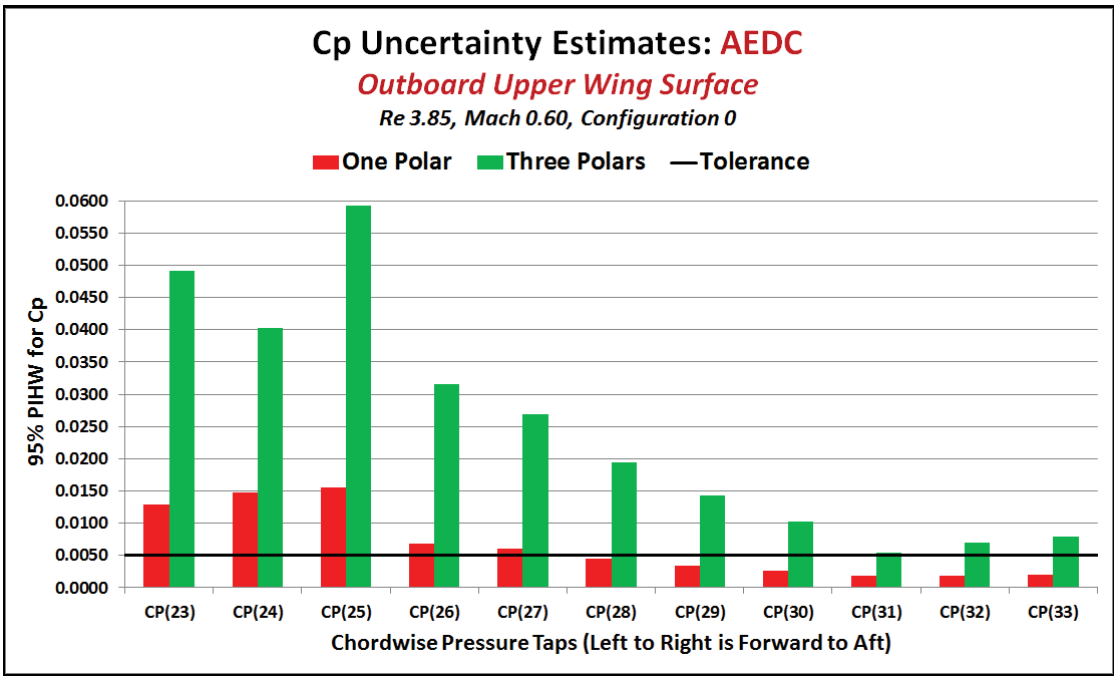

Figure A3: Outboard upper wing surface Cp uncertainty estimates at AEDC 16T.

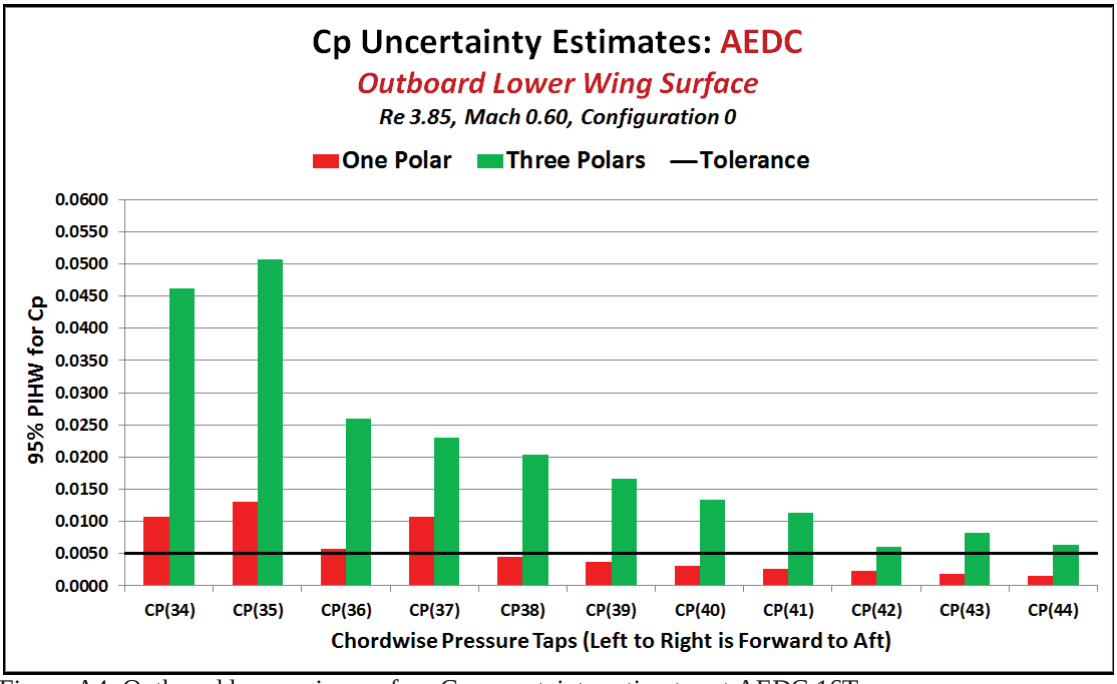

Figure A4: Outboard lower wing surface Cp uncertainty estimates at AEDC 16T.

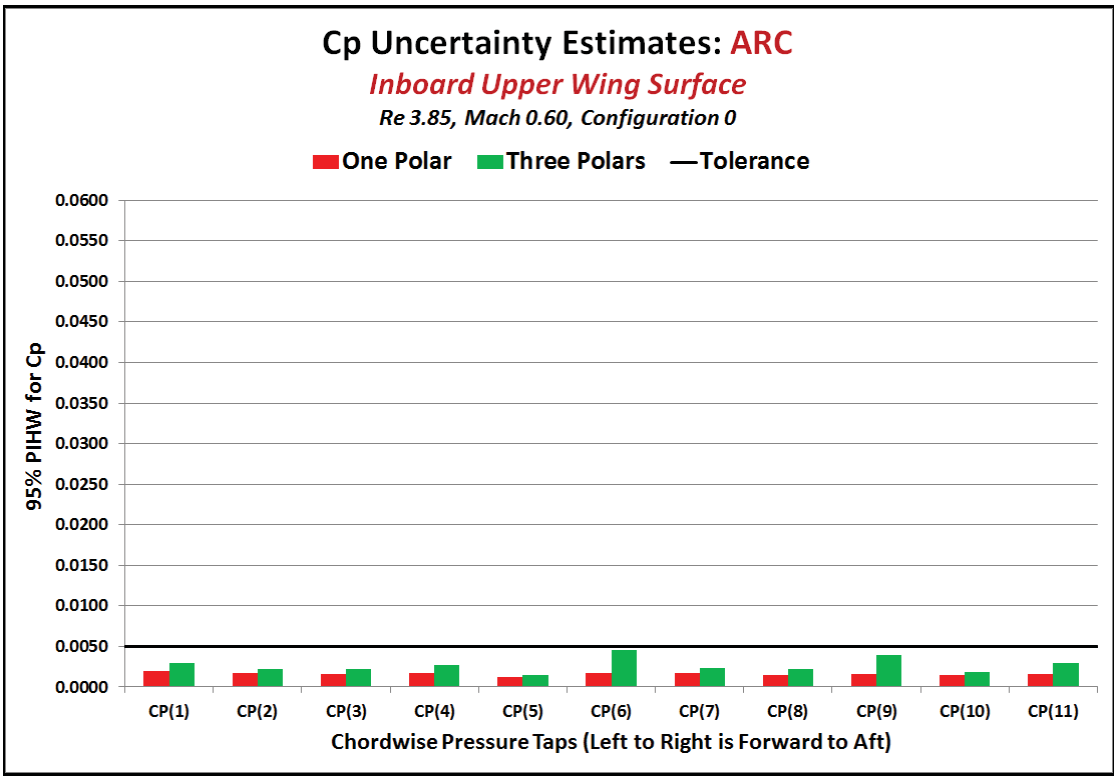

Figure A5: Inboard upper wing surface Cp uncertainty estimates at ARC 11-FT.

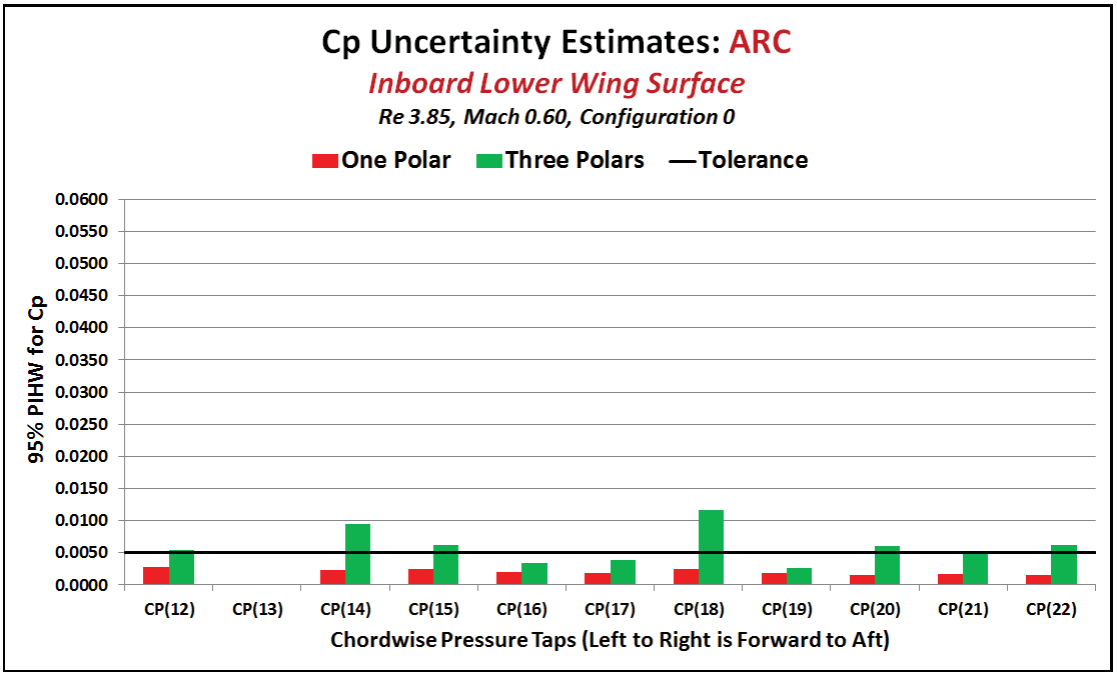

Figure A6: Inboard lower wing surface Cp uncertainty estimates at ARC 11-FT.

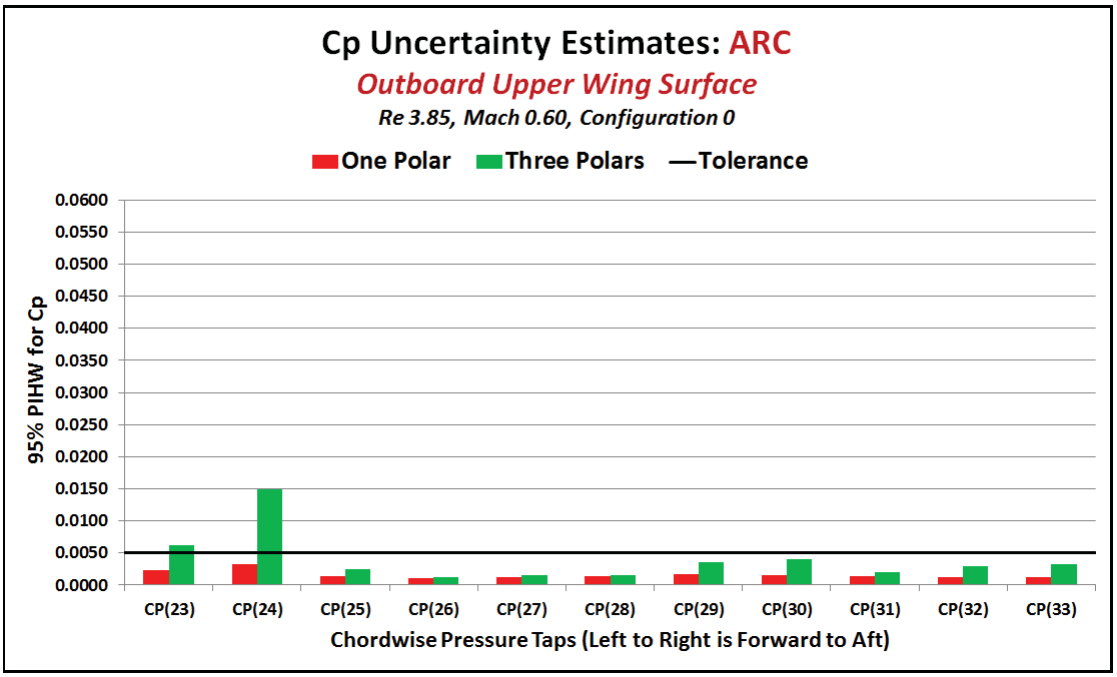

Figure A7: Outboard upper wing surface Cp uncertainty estimates at ARC 11-FT.

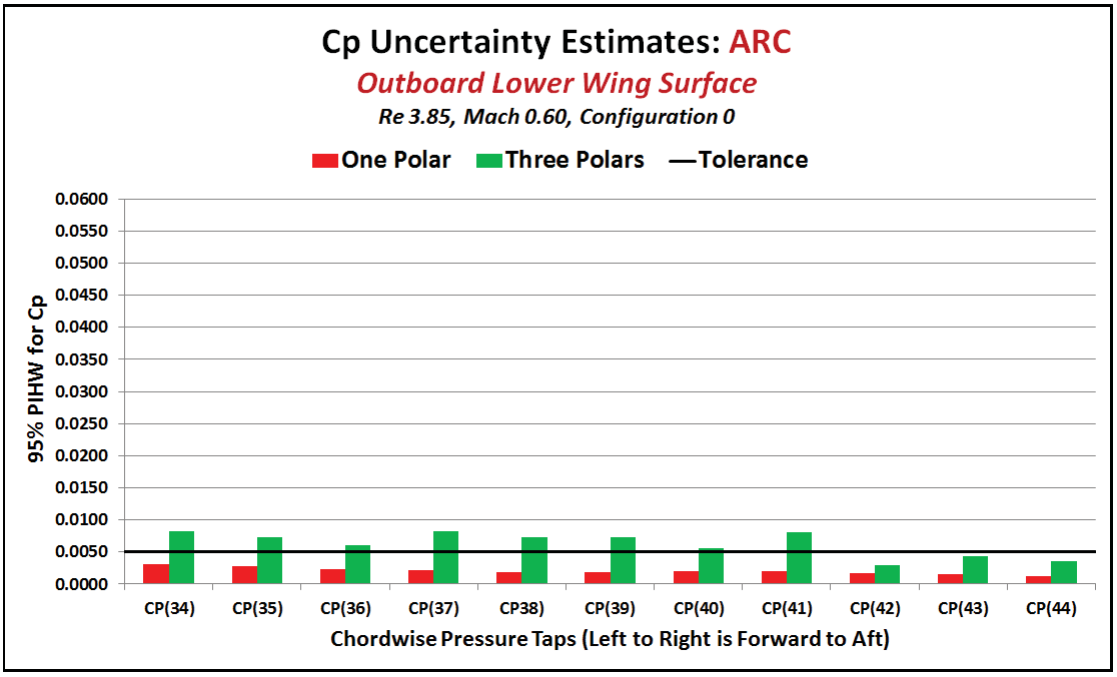

Figure A8: Outboard lower wing surface Cp uncertainty estimates at ARC 11-FT.

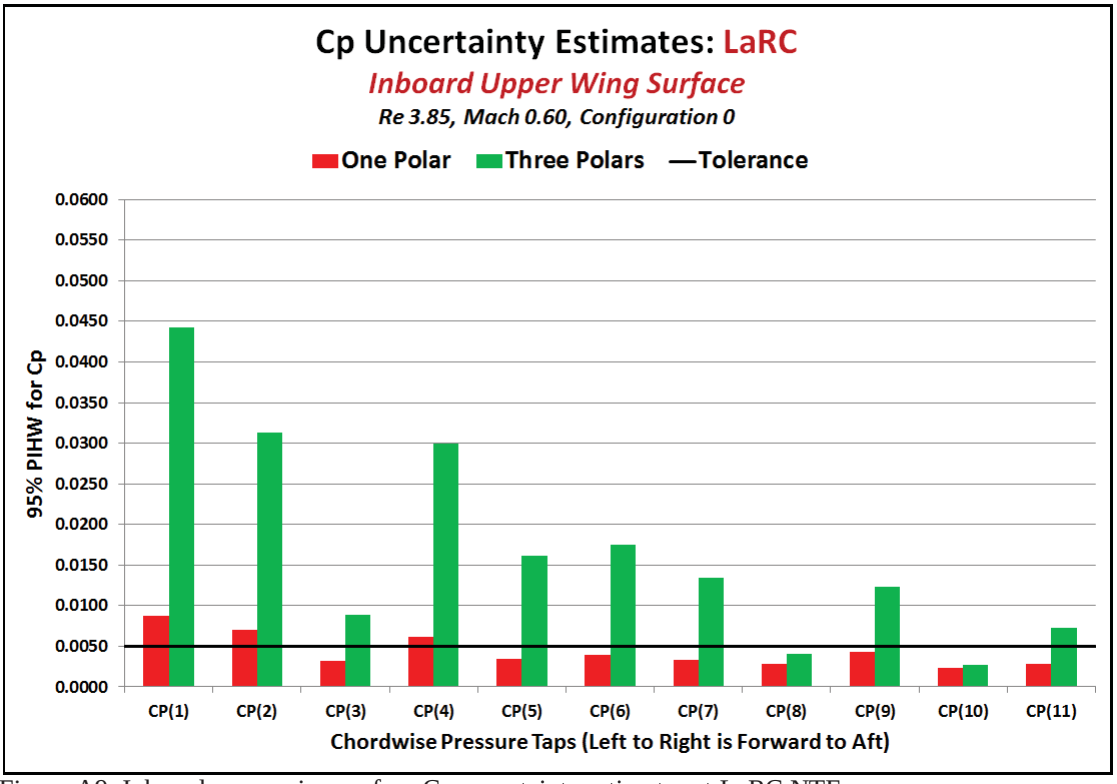

Figure A9: Inboard upper wing surface Cp uncertainty estimates at LaRC NTF.

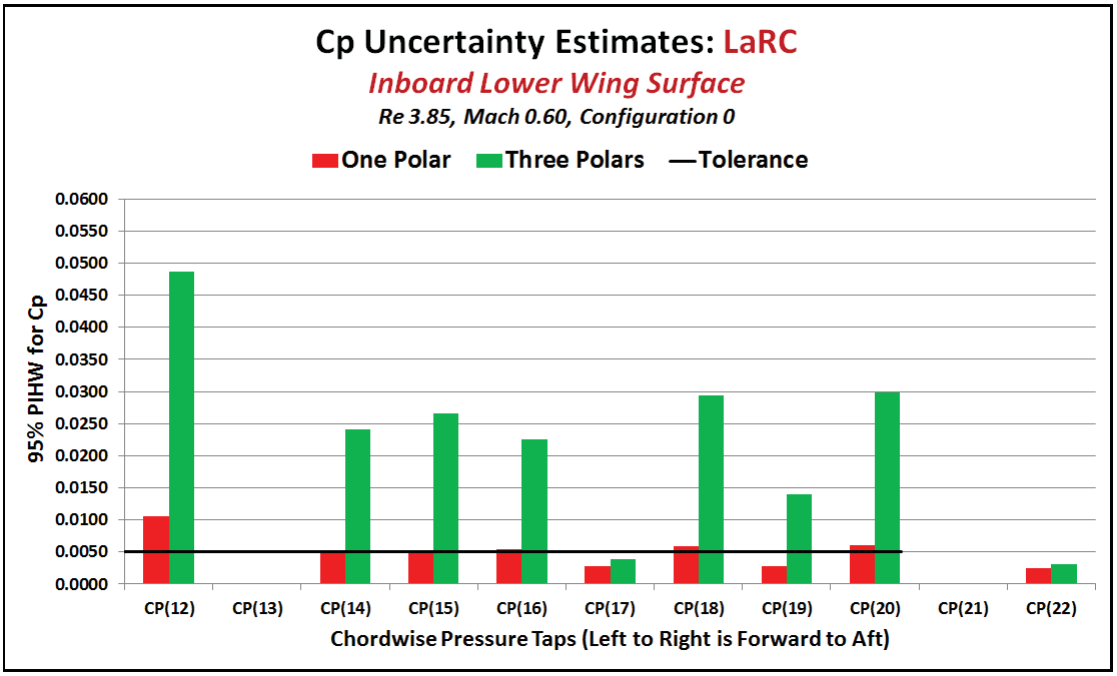

Figure A10: Inboard lower wing surface Cp uncertainty estimates at LaRC NTF.

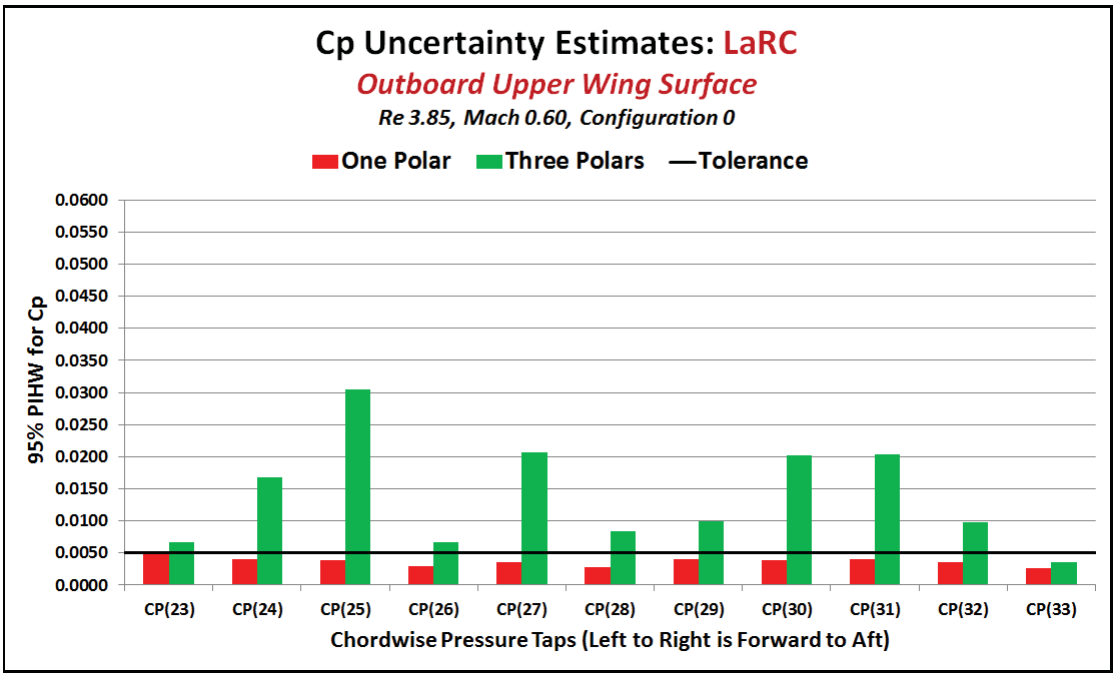

Figure A11: Outboard upper wing surface Cp uncertainty estimates at LaRC NTF.

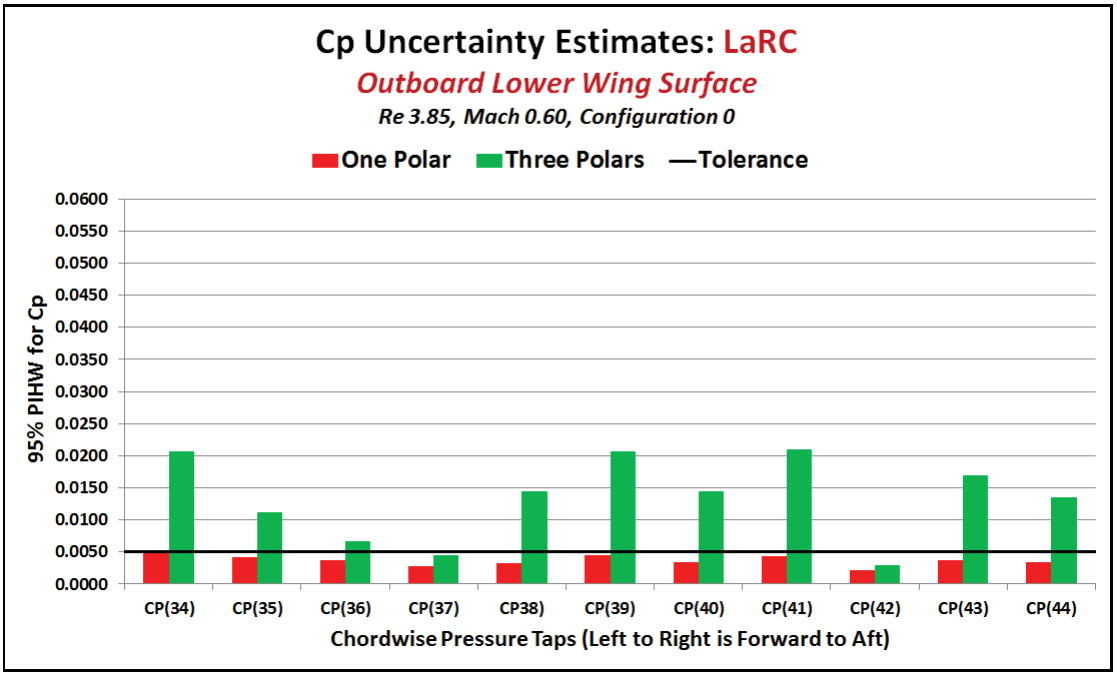

Figure A12: Outboard lower wing surface Cp uncertainty estimates at LaRC NTF.

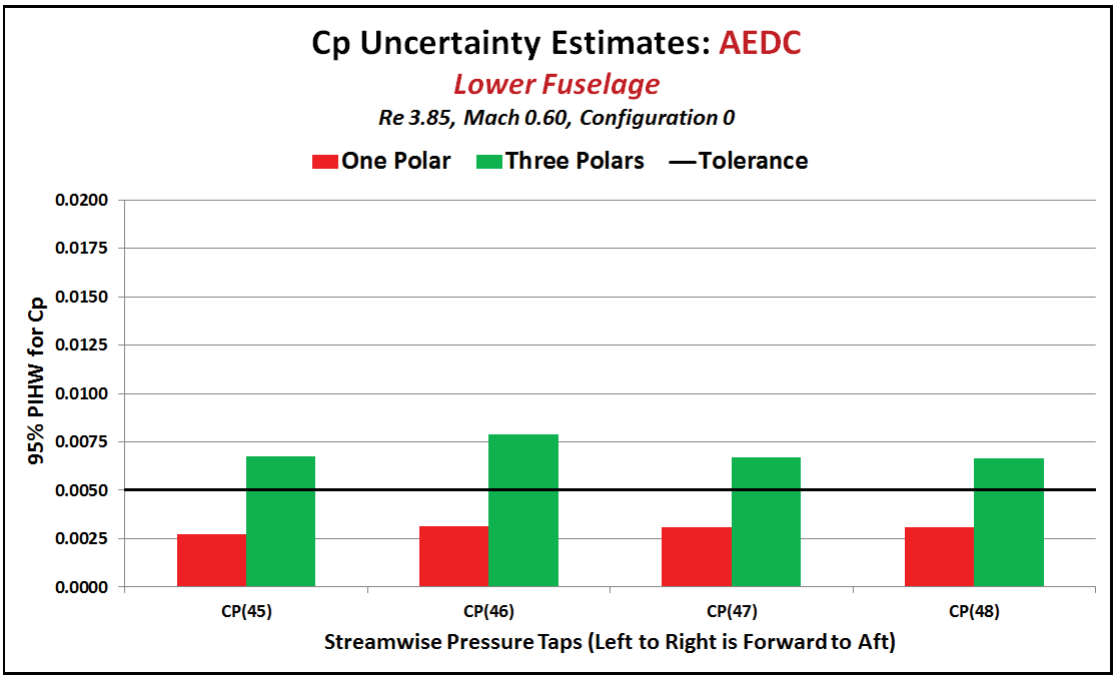

Figure A13: Lower Fuselage Cp uncertainty estimates at AEDC 16T.

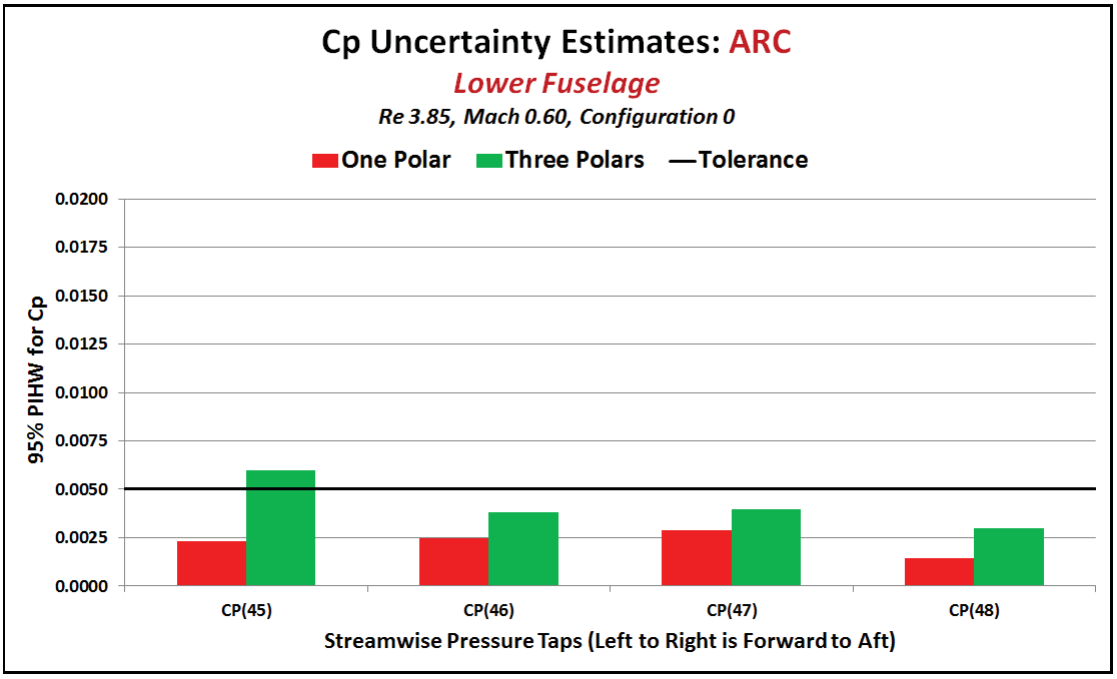

Figure A14: Lower Fuselage Cp uncertainty estimates at ARC 11-FT.

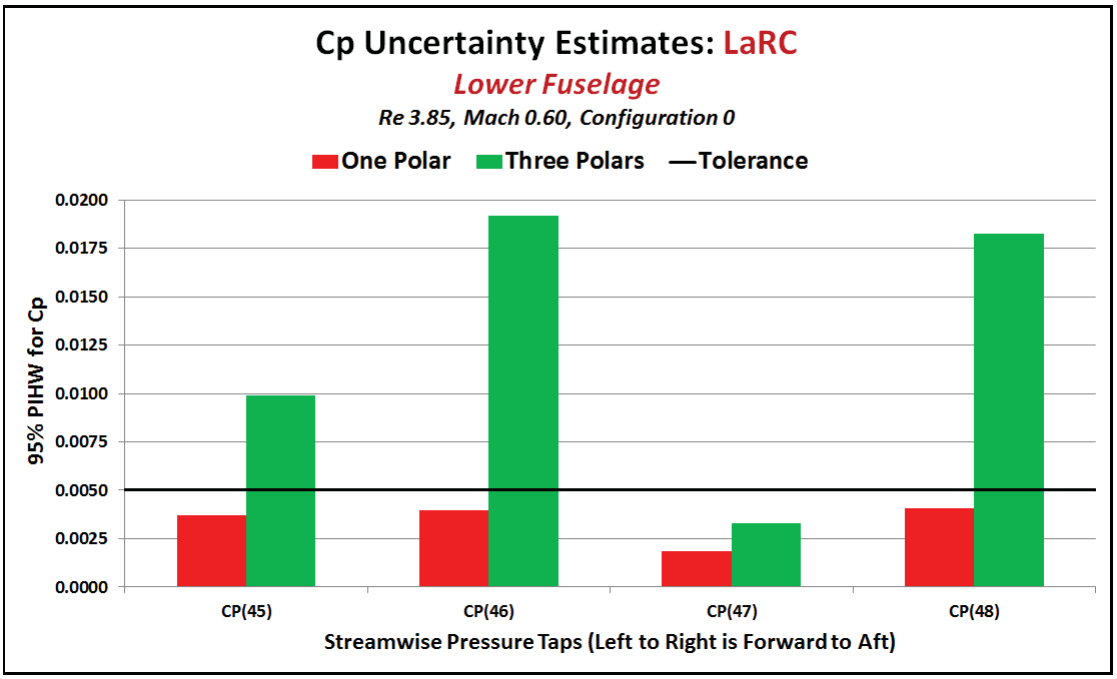

Figure A15: Lower Fuselage Cp uncertainty estimates at LaRC NTF.

Acknowledgements

This work was supported by the NASA Langley Center Innovation Fund. The author gratefully acknowledges Robert Dowgwillo and Dr. Mark Kammeyer of the Boeing Company for helpful discussions about aerodynamic pressure measurements.

References

[1]DeLoach, R., "Check-Standard Testing Across Multiple Transonic Wind Tunnels with the Modern Design of Experiments," AIAA 2012-3174, 28th Ground Testing Conference, New Orleans, LA, June 2012.

[2]DeLoach, R., "Comparison of Force and Moment Coefficients for the Same Test Article in Multiple Wind Tunnels," AIAA 2013-2490, AIAA Ground Testing Conference, San Diego, CA, June 2013.

[3]Box, G. E. P. and Draper, N. R., *Empirical Model Building and Response Surfaces*, John Wiley and Sons, New York, 1987.

[4]International Organization for Standardization, *Guide to the Expression of Uncertainty in Measurement*, ISO, Geneva, 1993.

[5]Steele, W. G., Furguson, R. A., Taylor, R. P., and Coleman, H. W., "Comparison of ANSI/ASME and ISO models for Calculation of Uncertainty," *ISA Transactions*, Vol. 33, 1994, pp. 339 – 352.

[6]Scheffe, H., *The Analysis of Variance*, John Wiley and Sons, New York, 1959.

[7]DeLoach, R., "Analysis of Variance in the Modern Design of Experiments," AIAA 2010-1111, 48th AIAA Aerospace Sciences Meeting and Exhibit, Orlando, Florida, January 4-7, 2010.

[8]Coleman, H.W., and Steele, W. G., *Experimentation and Uncertainty Analysis for Engineers 2nd ed*. John Wiley and Sons, New York, 1999.

[9]Fisher, R. A., *The Design of Experiments, 8th ed.* Oliver and Boyd, Edinburgh, 1966.

[10]Box, G. E. P., Hunter, W. G., and Hunter, J. S., *Statistics for Experimenters, An Introduction to Design, Data Analysis, and Model Building*. Wiley, New York, 1978.

[11]Montgomery, D. C., *Design and Analysis of Experiments, 5th ed*. Wiley, New York, 2001.

[12]DeLoach, R., "Improved Quality in Aerospace Testing through the Modern Design of Experiments (invited)," AIAA 2000-0825, 38th AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV, Jan 2000.