

# Design of Orion Soil Impact Study Using the Modern Design of Experiments

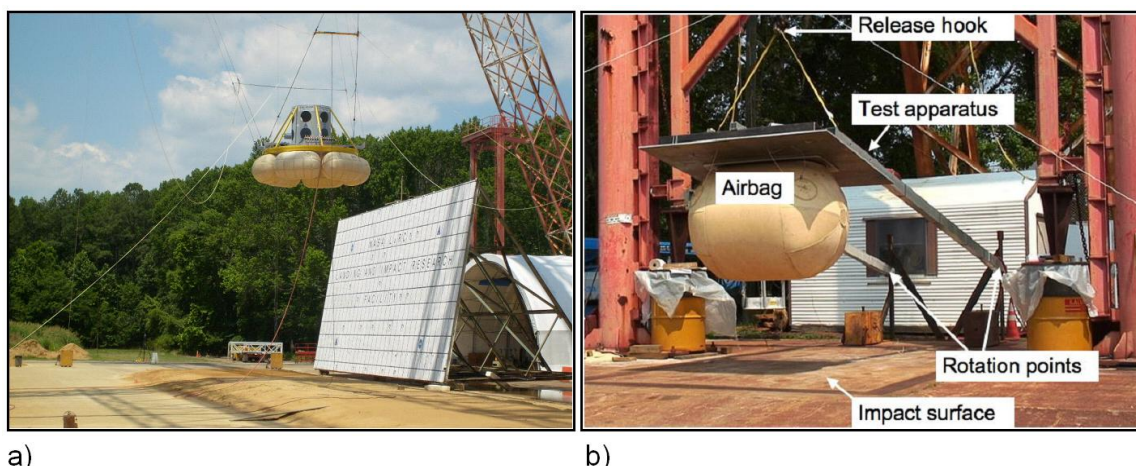
Richard DeLoach\*

NASA Langley Research Center, Hampton, Virginia, 23681

Two conventional One Factor At a Time (OFAT) test matrices under consideration for an Orion Landing System subscale soil impact study are reviewed. Certain weaknesses in the designs, systemic to OFAT experiment designs generally, are identified. An alternative test matrix is proposed that is based in the Modern Design of Experiments (MDOE), which achieves certain synergies by combining the original two test matrices into one. The attendant resource savings are quantified and the impact on uncertainty is discussed.

## I. Introduction

NASA has engaged in studies of a number of land-based return scenarios for the Orion Crew Exploration Vehicle. A land-based return would cost less, and require less complex operations than sea-based recoveries such as those used in the *Mercury*, *Gemini*, and *Apollo* programs. Among other concepts have been considered are airbags and crushable subsystems. Reference 1 describes early experiments in support of one of these concepts, the airbag landing scenario. Figure 1 shows various stages of airbag empirical studies conducted at Langley Research Center.



**Figure 1. Orion airbag landing dynamics studies at Langley Research Center; a) a six-airbag configuration attached to a full-scale Orion boilerplate capsule, b) a single airbag assembly in a test apparatus.**

The airbag approach proved to have certain weight and volume disadvantages, but efforts to evaluate other alternatives have continued. Part of this effort entails an investigation of landing dynamics and loads associated with impacts on different variations of soil. Two Orion Landing System subscale soil impact experiments at Langley Research Center were proposed in support of this effort, the first of which involved a number of vertical drops of a scaled Orion boilerplate test article onto two different soil surfaces, and the second of which added a horizontal component of velocity to the first test. Both tests were designed using a conventional experimental method popular in aerospace testing, known as the One Factor At a Time (OFAT) method. The OFAT method has certain

\* Senior Research Scientist, Aeronautical Systems Engineering Branch, NASA Langley Research Center, MS 238, 4 Langley Blvd, Hampton, VA 23681, Associate Fellow.

productivity and quality limitations that are described in more detail below. Resource constraints were in fact identified as a problem in both of these experiments, with dates in the originally proposed test matrices described as “a best case scenario with no lost days due to rain or other testing problems.” There were also cost constraints.

The author was asked to review the initial test matrices with a view to evaluating the potential of formally designed experiments in this application. An alternative test design was developed that exploits certain elements of testing technology known collectively at Langley Research Center as the Modern Design of Experiments (MDOE). It was constructed to provide relief to the test schedule and certain other benefits, including the addition of an adequate number of replicates to assess uncertainty as well as certain quality assurance measures designed to reduce experimental error. The basic principles of MDOE testing for aerospace application, as well as selected examples, are provided in the references<sup>2-18</sup>. This paper had its genesis in an informal internal report to the Langley Orion landing dynamics team, but is published here as a tutorial example of “how the sausage is made” in a practical application of formal experiment design principles.

Section II of this paper describes the original OFAT test matrices and discusses ways in which they can be improved. Section III presents the MDOE test matrix and describes the rationale for its construction. Section IV discusses certain quality issues and presents a smaller MDOE test matrix that circumvents some of these issues while also conserving run count and therefore cycle time and direct operating costs. Section V contains some summary remarks.

## II. The Original OFAT Test Plan and Possible Improvements

The first of two originally proposed tests consisted of a number of vertical drops of a scaled Orion boilerplate test article onto two sandy surfaces that differ by moisture content and therefore density. The lower-density, dryer surface has a moisture content of 2.87% and a density of 80.0 lbs/ft<sup>3</sup>. The higher-density, moister surface has a moisture content of 16.66% and a density of 100.3 lbs/ft<sup>3</sup>. Two vertical velocities were planned, 25 ft/sec and 35 ft/sec, as well as two test article pitch angles, 28° and either 23° or 33°, to be determined. Table 1 presents the original drop test schedule of runs.

**Table 1. Original Test Matrix for Vertical Drop Experiment.**

Date	Vertical Speed (fps)	Test Article Pitch (deg)	Sand Moisture (%)	Sand Density (lbs/ft <sup>3</sup> )
6/22/09	25	28	2.87	80.0*
6/23/09	25	28	2.87	80.0
6/24/09	35	28	2.87	80.0
6/29/09	25	23/33	2.87	80.0
7/7/09	35	23/33	2.87	80.0
7/10/09	25	28	16.66	100.3
7/13/09	25	28	16.66	100.3
7/14/09	35	28	16.66	100.3
7/17/09	25	23/33	16.66	100.3
7/21/09	35	23/33	16.66	100.3

Table 2 list the runs originally proposed for a second test, in which the test article would be suspended via a cable system and would swing in an arc that intersects the ground. This swing test would therefore impart a horizontal component of velocity to the test article as well as the vertical component of the drop test in Table 1.

As noted in the introduction, both the original drop test (Table 1) and swing test (Table 2) are examples of an experiment design methodology that is common in aerospace research. Known as One Factor At a Time (OFAT) testing, this method is characterized by the fact that in successive data points the levels of all independent variables (or factors) except one are held at a constant level. The OFAT practitioner changes only one factor at a time in progressing from point to point.

Considerable efficiency can be achieved by a kind of “multitasking” in which we change more than one factor level at a time as we progress through the test matrix from point to point. Each point effectively works harder when multiple factors are changed at a time, by inducing compound changes in the response variables attributable not simply to the change in a single factor, but to changes in more than one factor at a time. Understandably, sensible researchers who encounter this concept for the first time tend to be concerned that individual factor effects cannot be distinguished from each other if data are acquired in this way. If impact velocity and pitch angle are both changed

before the next run, for example, how is it possible to say how much of the resulting change in impact g-loads is due to the velocity change, and how much is due to the change in pitch?

**Table 2. Original Test Matrix for Swinging Impact Experiment.**

Date	Vertical Speed (fps)	Horizontal Speed (fps)	Test Article Pitch (deg)	Sand Moisture (%)	Sand Density (lbs/ft <sup>3</sup> )
7/31/09	25	30	28	2.87	80.0*
8/5/09	25	30	28	2.87	80.0
8/7/09	25	40	28	2.87	80.0
8/12/09	35	40	28	2.87	80.0
8/14/09	35	50	28	2.87	80.0
8/19/09	25	30	28	16.66	100.3**
8/21/09	25	30	28	16.66	100.3
8/26/09	25	40	28	16.66	100.3
8/28/09	35	40	28	16.66	100.3
8/30/09	35	50	28	16.66	100.3
8/31/09	35	60	28	16.66	100.3

Fortunately, it is neither impossible nor particularly difficult to segregate factor effects in response data acquired when multiple factors are changed on successive runs, if the experiment is designed according to a few fundamental principles. It is therefore possible to enjoy the efficiency that accrues from making each data point “work harder,” without confounding the effects of one factor change with those of another. This fact is exploited in the MDOE test matrices offered below.

The two OFAT test matrices in Tables 1 and 2 have two factors in common: vertical speed and soil type. Both factors are set at two levels. The drop test varies pitch angle but not horizontal speed, while the swing test varies horizontal speed but not pitch angle. There is an opportunity to achieve some savings in the number of runs, as well as to gain some otherwise unavailable insights, by combing the two three-factor tests into a single four-factor test.

Note that both the main vertical speed effect and the soil-type effect would both be known after the first OFAT test—the drop test of Table 1. Changing these variables again in the swing test is therefore not necessary, except to reveal interaction effects involving horizontal speed (to reveal how horizontal speed effects change from one soil or one vertical speed to another). However, this segregated design forecloses options to examine what is potentially an equally important interaction between pitch angle and horizontal speed, since one or the other is held constant in each of the two OFAT tests. That is, the original two-test approach cannot detect if the effect of changing pitch angle depends on horizontal speed, and conversely.

Another advantage of combining the two OFAT tests is that it effectively increases the number of pure-error degrees of freedom available to assess random error. For small sample sizes this can substantially improve precision.

In each of the OFAT tests, one point is replicated for each soil type. Since data acquired for each soil type is analyzed separately in an OFAT test, the empirical estimates of standard deviation will feature only one degree of freedom in each case. Variance estimates based on such a small sample are notoriously poor estimators of the true population variance.

If the OFAT tests were combined so that g-loads could be modeled as a function of all four independent variables at once, the same number of replicates would translate into a four degree-of-freedom estimate of pure error variance, reducing the uncertainty by almost 80% compared to the single degree-of-freedom case.

Combining the two tests into one means there is an opportunity to save some runs by examining soil type and vertical speed effects once instead of twice, and it also means that some additional insights might be had by examining the interaction between pitch and horizontal speed. Precision estimates can be improved, and there is also an opportunity for improving the accuracy of the test, as will now be outlined briefly.

The unexplained variance in a sample of experimental data is assumed to be randomly distributed about some mean that is stable with time. Unfortunately, under commonly occurring conditions this mean tends to vary systematically with time due to effects that induce non-random variations in the data, resulting in a net bias shift that changes slowly with time. This is often the dominant source of uncertainty in high-precision tests for which the random error is small, yet systematic changes are often ignored under the assumption that, except for random error, the only changes that occur in measured response data are those that are induced by the experimenter making changes in the independent variables.

Environmental effects are one potential source of this time-varying bias shift in outdoor testing. In this test, for example, the sand that may be dry on day one will likely absorb water over time from humidity and dew, to say nothing of rain. Instruments will drift, operators will tire or get rushed at the end (“fatigue effects”) or get more into a groove through repetition (“learning effects”), the sand containment system may deform over time, and there will be scores of other effects that induce systematic (not random) changes in the data that will not be detected. Such errors tend to go in one direction for prolonged intervals, with ordinary random error superimposed.

Consider the potential impact of such systematic variations on the original swing test. In this test, the horizontal speed is changed systematically over a two-week period from 30 fps to 50 fps for the dry sand runs and over a comparable period from 30 fps to 60 fps for the moist sand runs. Let us imagine that over the testing period, some combination of systematic changes in moisture content and other effects result in a gradual reduction in impact g-loads compared to what would have been measured absent the systematic effects. Because horizontal speed is changed systematically with time, there is no way to distinguish between the effects of the systematic speed changes prescribed in the test matrix, and the systematic effects of gradually changing moisture, etc., that were not prescribed, but that occurred anyway. Stated differently, the g-loads would have changed over time whether speed was changed or not, even if all other known factors in the test matrix were held constant.

By attributing the sum of speed effects plus the effects of changing moisture and other systematic errors to horizontal speed, one obtains an incorrect perception of how horizontal speed actually affects g-loads. This is not a simple reduction in precision caused by ordinary random error, but rather an error of a more serious kind. Systematic errors degrade the *accuracy* of the result, not just the precision; we do not get a somewhat less precise estimate of what is essentially the right answer (“within experimental error”), we get the wrong answer altogether, and have no way of knowing it at the time. The effects of systematic error generally surface only when there is an attempt to reproduce the experimental results, sometimes months or years later, in an independent test in which data are inevitably acquired under a different set of unexplained systematic variations.

Systematic errors may or may not be in play in the soil impact modeling tests, and they may or may not be significant if they are. Prudence dictates that we defend against such errors, however, since their effects can be so serious if they do occur. There is thus an opportunity to improve the reliability of the original OFAT test results by employing an inexpensive quality assurance tactic that defends against systematic unexplained variance.

There is a widely recognized defense against systematic error that consists of simply randomizing the run order in which points in the test matrix are acquired. This ensures that not all low-speed data are acquired early, for example, when responses might be biased systematically in one direction, and not all the high-speed data are acquired later, when responses might be biased the other way. This in turn ensures that we see only the systematic effects of changes in impact speed, because by randomizing the run order, we induce an equal probability of a positive or a negative systematic error contribution at any given horizontal speed. That is, randomizing the run order simply converts systematic error effects into another component of error that is randomly distributed about the true functional relationship between independent variables and response variables. This random error component is easy to detect and easy to quantify. The chief virtue of randomizing the run order, however, is that it preserves the true functional relationship between independent variables and the responses that depend upon them.

### **III. A Two-Soil MDOE Test Plan**

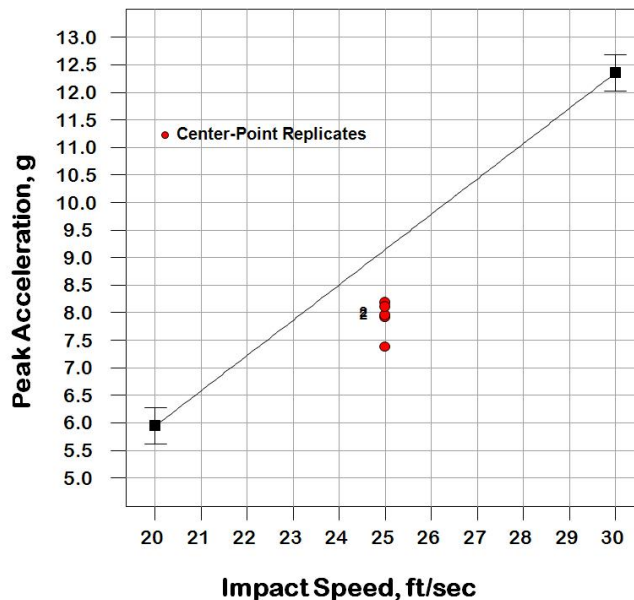
Among many ways that MDOE and OFAT test methods differ is the response surface modeling (RSM) perspective of an MDOE design. The goal is always to establish an empirically-derived mathematical relationship between each response of interest (horizontal and vertical impact g-loads in this test), and the independent variables upon which they depend (horizontal speed, vertical speed, pitch angle, and soil type in this test). Since there is seldom any prior knowledge of the true functional form of the response function, it is approximated by a polynomial Taylor series of sufficient order to adequately represent the response over the range of independent variables of interest. This RSM orientation informs the experiment design process, in that the test matrix is constructed with a view to maximizing the quality of response predictions made with such a model.

#### **A. Order of Model**

The order chosen for the Taylor-series approximation to the true (but unknown) response function is a key determinant of the adequacy of the empirical response model. The model will feature lack-of-fit errors if the order is too low. If the order is too high, there can be large prediction errors near the design-space boundaries (often where the greatest accuracy is needed), and resources can be wasted acquiring the extra degrees of freedom needed to fit a model with superfluous higher-order terms.

Because the true underlying relationship between each response variable and the factors upon which it depends is generally unknown, there is inevitably a certain amount of guesswork in selecting the order of the polynomial function that will be used to approximate it. The decision was guided in this test by the number of distinct levels of each factor selected by the subject-matter experts in the original OFAT design. For example, the fact that only two levels of vertical speed were included in either the OFAT drop test or the OFAT swing test suggests that the relationship between impact g-loads and vertical impact speed can be adequately represented as first-order over the limited speed range of interest (25 fps to 35 fps).

Figure 2 displays how peak impact g-loads varied with vertical impact speed in the Orion airbag test illustrated in Fig. 1. In this test, resource constraints dictated the decision to reject an earlier quadratic design and to model the responses as first-order functions of the independent variables with factor interactions. A number of replicates were acquired at the center of the design space to test for curvature, among other reasons. The center-point replicates are the red points in Fig. 2. The spread in center-point replicates suggest the degree of scatter in the data.



**Figure 2. Impact g-loading as a function of impact speed for an Orion airbag landing.**

Some curvature is indicated, with the entire sample of center points located below the straight-line response approximation. On the other hand, the mean of the center-point sample is not substantially below the straight line, given the spread in the center points and the fact that there is also some uncertainty in the response model itself (indicated by the Least Significant Difference (LSD) bars at the two ends of the model).

The soil impact study will not involve airbags and the g-loads are likely to be somewhat different, but Fig. 2 does suggest some small second-order effect in the relationship between g-load and vertical speed. The original OFAT drop test matrix features only two levels of pitch angle (28° and either 23° or 33°, TBD), suggesting that a response model that is first-order in pitch would be adequate, but based on the experience displayed in Fig. 2, a response model that is second-order in pitch angle may provide a better response representation.

The OFAT swing test features three levels of horizontal speed for the dry-soil runs and four levels for moist soil, suggesting that a response up to third order might be necessary for the latter case. However, subsequent discussions among the principals have focused on only three levels of horizontal speed for either soil type, suggesting that a second-order response model would be adequate.

Soil type is a *categorical* variable, meaning that it can only be set at discrete levels. In this test there are only two levels: “dry” and “moist”. A response surface model representing g-loads as a function of this factor cannot have quadratic or higher-order terms involving this factor.

## B. Scale of the Experiment

Scaling the experiment refers to the process of determining how many runs to acquire. Since each run costs money and consumes cycle time (for which some cost can be attributed), costs are minimized in an MDOE experiment design by prescribing the fewest runs adequate for the experiment. We note in passing that this philosophy conflicts with the common OFAT perception of productivity, which equates productivity with data volume. The OFAT practitioner is inclined to acquire *the most* runs that resources permit, while the MDOE practitioner seeks to manage with *the fewest* runs that are adequate to achieve the objectives at hand.

For reasons discussed in the previous subsection, a second-order polynomial response model was selected to be fitted to data acquired in the Orion soil impact experiment. This key decision determines the minimum number of runs to be acquired, since there must be at least one degree of freedom (one run) for each term in the response model.

In general, a full  $d^{\text{th}}$ -order polynomial in  $k$  factors has  $p$  terms, including the intercept term, where

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

A full 2<sup>nd</sup>-order polynomial in  $k=4$  factors would therefore have  $(2+4)!/2!4! = 15$  terms. Since soil type is a categorical factor with only two levels, there can be no quadratic term for this variable and thus the polynomial response model for this experiment will have only 14 terms.

Let us make the following assignment of independent variables:

- $x_1$  = horizontal speed
- $x_2$  = vertical speed
- $x_3$  = pitch angle
- $x_4$  = soil type

The response model can then be represented as follows, where the  $b$ 's represent model coefficients determined by regression, with obvious subscripts:

$$\begin{aligned} & b_0 \text{ (one intercept term)} \\ & + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 \text{ (four linear terms)} \\ & + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4 \text{ (six interaction terms)} \\ & + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 \text{ (three quadratic terms)} \end{aligned} \quad (2)$$

Note that there are  $1+4+6+3 = 14$  terms in this model, as Eq. (1) predicts after correcting for the lack of a quadratic  $x_4$  term, and thus the minimum number of runs to fit this model is 14. However, it would be unwise to specify only 14 runs for this experiment, because fitting the model given in Eq. (2) would exhaust all the available degrees of freedom, leaving no residual degrees of freedom to assess experimental error. We therefore specify some additional so-called *pure error* degrees of freedom, consisting of replicates of some subset of the 14 points specified to fit the model.

There is an element of judgment in specifying the number of replicates, but the quality of an experimental estimate of standard deviation degrades substantially as this number gets smaller. The 95% precision-interval half-width approaches the well-known “two-sigma” level (actually, 1.960 sigma) as the number of replicates approaches infinity, but it is a common convention to say that a standard deviation based on 10 or more replicates is sufficient to invoke the “large sample” approximation by which a 95% precision interval half-width may be said to be “two sigma.” For sample sizes smaller than 10, a small-sample adjustment is applied that results in a larger number of standard deviations corresponding to a 95% precision interval half-width.

Table 3 reveals how the 95% precision interval depends on the number of replicates used to estimate the standard deviation. Five replicates are specified for this test, bringing the total number of runs to 19. However, as Table 3 indicates, fewer replicates can be specified at the expense of additional uncertainty.

When sufficient information is available to do so, the scaling process takes into account precision goals and the quality of the measurement environment. For a given standard deviation in replicated runs, the required volume of data depends on the precision requirement and the number of terms in the response model, as Eq. (3) indicates. In this equation,  $t_\alpha$  is the number of standard deviations associated with a  $(1-\alpha)\%$  precision interval half-width,  $\sigma$  is the

standard deviation of replicated points,  $\gamma$  is the precision level, and  $p$  is the number of terms in the response model, as given in Eq. (1).

$$n = p \left( \frac{t_\alpha \sigma}{\gamma} \right)^2 \quad (3)$$

Equation (3) indicates that the minimum number of runs required to meet precision requirements increases with the complexity of the response model ( $p$ ) and the inherent variability of the test environment ( $\sigma$ ). Also, by Table 3 the minimum volume of data increases as the number of replicates decreases ( $t_\alpha$ ) and as the precision requirement gets more stringent.

**Table 3. 95% Precision Interval Half-Widths for Various Numbers of Replicates, in Standard Deviations.**

Replicates	95% PIHW
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
$\infty$	1.960

No precision requirements were specified by the principals for this test, and the standard deviation is unknown. For these reasons the scaling was based simply on the term count in the response model (14) and a number of replicates judged to be reasonable (5), for a total of  $n=19$  runs. The plan is to estimate the precision associated with this run count after a five degree-of-freedom estimate of the standard deviation has been obtained. We know that for this test,  $n=19$ ,  $p=14$ , and (from Table 3),  $t_\alpha=2.571$ . Inserting these numbers into Eq. (3) yields the following result:

$$\gamma = \left( \sqrt{\frac{p}{n}} \right) t_\alpha \sigma = \sqrt{\frac{14}{19}} (2.571) \sigma = 2.21 \sigma \quad (4)$$

That is, the 95% precision interval half-width associated with response predictions made with Eq. (2) will be 2.21 times the standard deviation estimated from five replicated runs. The Orion airbag impact study mentioned earlier was characterized by a standard deviation in peak vertical g-load of 0.31 g. While g-loads in the soil impact test are likely to be different, it is possible that the run-to-run repeatability might be comparable to the airbag test. If that is the case, we will be able to report model predictions  $\pm(2.21)(0.31) \text{ g} = \pm 0.68 \text{ g}$ , with 95% confidence.

Note from Eq. (3) that the minimum number of runs is a sensitive function of the precision requirement. Let us assume for a moment that the  $\pm 0.68 \text{ g}$  precision we are anticipating is deemed insufficient, and that an uncertainty of no more than  $\pm 0.5 \text{ g}$  is required for this test at the 95% confidence level. If the five degree-of-freedom estimate of standard deviation is still 0.31 g, then by Eq. (3) and Table 3 the minimum number of runs required to deliver the specified  $\pm 0.5 \text{ g}$  precision level would be

$$n = p \left( \frac{t_\alpha \sigma}{\gamma} \right)^2 = 14 \left[ \frac{(2.571)(0.31)}{0.5} \right]^2 = 35.6 = 36 \quad (5)$$

Thirty-six runs far exceed what is currently planned. From a “cup half empty” perspective, it might be accurately stated that a relatively small improvement in precision (from  $\pm 0.68 \text{ g}$  to  $\pm 0.5 \text{ g}$ ) would require a considerable increase in the number of runs. From the “cup half full” point of view, this illustrates that substantial resource savings can be achieved at the expense of a relatively minor compromise in the specified level of precision. In this example,  $36-19=17$  runs can be saved by relaxing the precision requirement a mere  $0.68-0.5=0.18 \text{ g}$ .

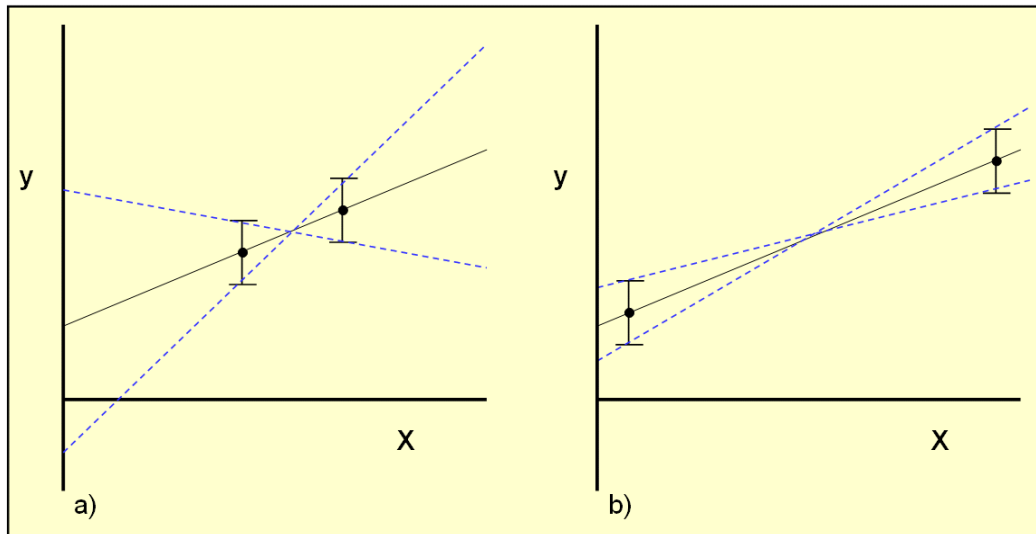


The discussion of how specifications of precision are related to minimum data volume requirements is provided simply as tutorial background. As previously noted, there are no specified precision requirements in this test, and no a-priori estimates of the standard deviation. The experiment design calls for 19 runs based solely on the quadratic response model that has been specified and a decision to include five replicates. Rather than scaling the experiment to meet specified precision requirements, the precision achievable with the experiment scaled as described will be quantified and documented, by solving for  $\gamma$  in Eq. 3 once  $\sigma$  has been empirically determined.

### C. Site Selection

Given that 19 runs will be acquired, it is necessary to decide *which* runs. That is, it is necessary to decide which combinations of factor levels will comprise the 19 runs. This process is called “site selection,” because each data point can be represented graphically as a location (or “site”) in a design space constructed by assigning each independent variable to one axis of a Cartesian coordinate system. Every point in such a design space represents a unique combination of independent variable values.

OFAT site selection decisions are often based on operator convenience, or they are made to maximize data acquisition rate or to provide the most uniform coverage of the design space that is possible. Site selection decisions in an MDOE experiment design are made to maximize quality. It turns out that the uncertainty associated with response surface model predictions can be influenced by the selection of sites within the design space where the data are acquired to which the model is fitted. Figure 3 illustrates the basic concept for the simple case of a first order response function of one independent variable.



**Figure 3. First order function of one variable fitted to data acquired at different design space sites. a) Nearer center of design space, b) Nearer design-space boundaries.**

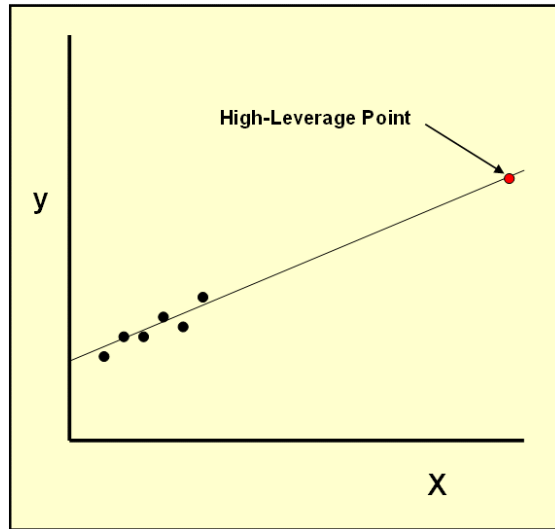
The solid line in each part of Fig. 3 represents the best fit to two data points, each featuring the same degree of experimental error as indicated by the error bars on each point. The dashed lines represent extreme values of straight lines that might have been fitted if other points had been acquired that were within experimental error of the points actually acquired.

Even though the experimental data feature the identical amount of uncertainty in Figs. 3a and 3b, there is a much greater range of possible slopes and y-intercepts in Fig. 3a than in Fig. 3b. That is, there is more uncertainty in the slope and y-intercept estimated in Fig 3a than in the slope and y-intercept estimated in Fig 3b, notwithstanding the same experimental error in both cases. The improvement in Fig 3b relative to Fig 3a can be explained entirely by the difference in site selection. When fitting the data to a response model of the form  $y = b_0 + b_1x$ , the uncertainty in  $b_0$  and  $b_1$  depends on the sites selected to acquire the data.

A simple first-order function of one variable was used in Fig. 3 to illustrate the relationship between site selection and uncertainty in the coefficients of the fitted response model, but this is a general phenomenon that extends to functions of any order, fitted to any number of independent variables. In the MDOE test matrix presented below, the 14 unique design space sites required for this design were selected to provide the smallest error in the coefficients of a second-order model fitted to four independent variables, absent the quadratic term of one of the



variables (soil type). These site selection decisions require a substantial number of calculations that are typically performed with experiment design software dedicated to this specific task.



**Figure 4. High Leverage Point**

Having chosen the 14 unique sites that minimize the coefficients in the response model of Eq. (2), it remains only to decide which five of these points to replicate in order to be able to make a reasonable estimate of random error. In the MDOE test matrix presented below, this decision is based on the *leverage* of each data point.

A data point is said to possess high leverage if a given experimental error in that point would have a significant degree of influence on the coefficients of the fitted model. Figure 4 illustrates this concept. An experimental error in the large-x data point would have a greater influence on the slope and y-intercept of the fitted line than the same error in any of the other points.

**Table 4. Two-Soil MDOE Test Matrix.**

Run Order	H-Vel	V-Vel	Pitch Angle	Soil Type
	fps	fps	deg	
1	20	35	33	Dry
2	40	35	23	Dry
3	20	35	23	Moist
4	60	35	28	Moist
5	40	35	33	Moist
6	60	35	33	Dry
7	40	30	28	Dry
8	60	25	33	Moist
9	20	25	23	Dry
10	60	35	28	Moist
11	60	30	23	Moist
12	40	30	28	Dry
13	60	25	23	Dry
14	40	25	33	Dry
15	20	25	33	Moist
16	40	25	23	Moist
17	60	30	23	Moist
18	20	25	23	Dry
19	60	25	33	Moist

A data point is not suspect simply because it has high leverage. The high leverage simply means that if there are any problems with that point, the consequences can be relatively severe. As long as the experimental error in a high-leverage point is not extreme, the fact that it has high leverage presents no special problem.

The leverage associated with a data point acquired at a given site can be reduced by replicating the point acquired at that site. A single replicate reduces the leverage at a given site by a factor of two, with each point now sharing leverage equal to half the leverage of the point before it was replicated. That is, replication provides a kind of diversification, in which the risk of a bad data point is distributed over two or more points. The more often the same point is replicated, the lower the leverage will be for each point acquired at that site.

Leverage was computed for each of the 14 unique sites necessary to fit Eq. (2) with the smallest uncertainty in the regression coefficients. These points were then rank-ordered by leverage, with the five highest-leverage points selected for replication. In this way, the final five points were selected for the MDOE test matrix of Table 4.

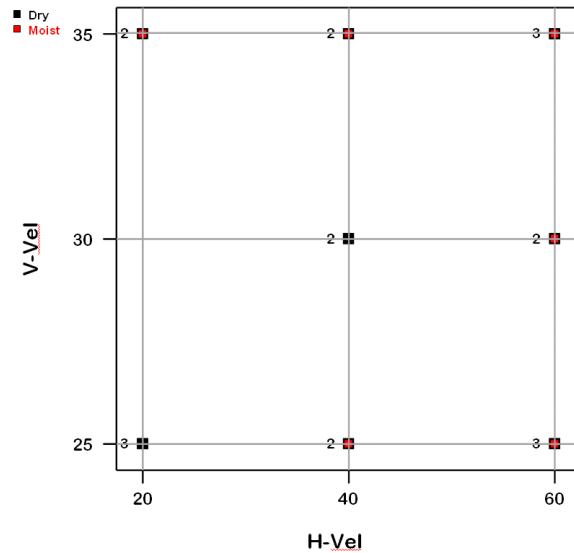


Figure 5a. Design space in horizontal and vertical velocity, fps.

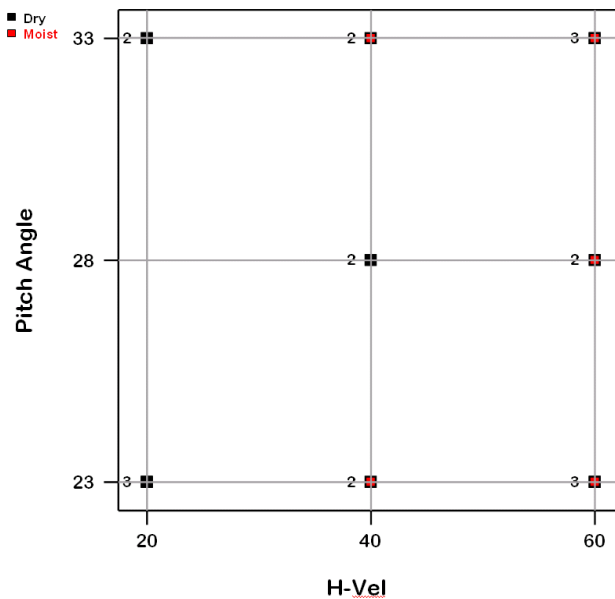
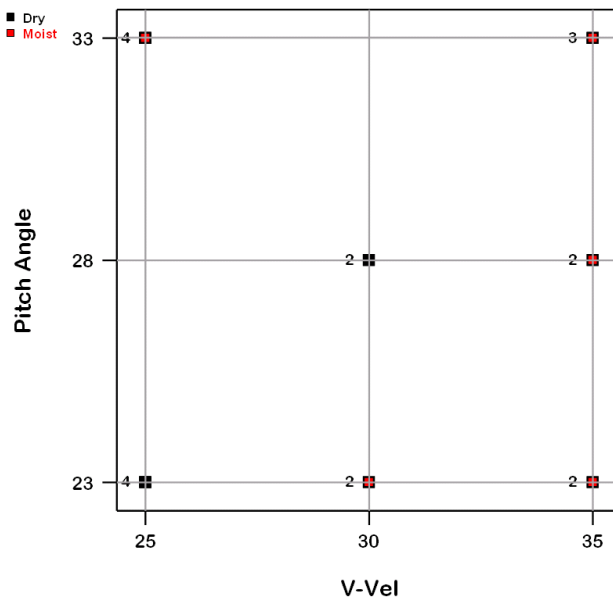


Figure 5b. Design space in horizontal velocity (fps) and pitch angle (deg).



**Figure 5c. Design space in vertical velocity (fps) and pitch angle (deg).**

In Table 4, runs 4, 7, 8, 9, and 11 are replicated by runs 10, 12, 19, 18, and 17, respectively. All other points are unique. Figures 5a, 5b, and 5c show the test matrix in graphical form, displaying the design space from three perspectives. Numbers next to various sites indicate how many points total are acquired at that site. Note that these are not necessarily replicates. Each view shows only two numerical variables. Some sites have points differing by the value of the third numerical variable.

#### IV. Discussion

The run order is randomized to defend against systematic variation that is likely to occur over such a long-duration outdoor experiment; however, the difficulty of randomizing on soil type is acknowledged, due to the practical problems of drying a large mass of moist sand. If all the dry runs are executed before any of the moist runs, then a key factor in the test; namely, the difference in impact loading due to soil type, will be confounded with the sum of all systematic variations that occur over the duration of this test. There is always the option to ignore the possibility of systematic error and execute all the dry runs first and then the moist runs, but the quality of the test result will be degraded if unexplained systematic variations are in play. There are alternative experiment designs available to cope with restrictions on randomization (a class of designs known as “split-plot design,” for example), but these require a larger number of runs than resource constraints in the current test can accommodate.

Note in Table 4 that it is not necessary to change from dry to wet or from wet to dry on every new run. There are in fact some streaks, so that there are only nine transitions. Note also that transitions from dry to moist present no particular problem. It is only transitions from moist to dry that are problematical, as these are the only transitions that would require that moisture be removed from the sand. There are only four such transitions in Table 4, between runs 5 & 6, 8 & 9, 11 & 12, and 17 & 18. One way to maintain the defense against systematic variation provided by randomization while avoiding the practical difficulties of drying out the sand between runs would be to replace the moist sand of runs 5, 8, 11, and 17 with fresh, dry sand before executing runs 6, 9, 12, and 18. There would be some added expense, which could be charged to the cost of maintaining quality in the test.

Another factor to consider that bears on the question of randomizing on soil type is that there are significant differences in horizontal speed for each of the moist-dry transitions. This means that the horizontal impact location will differ considerably from run to run. It may not be necessary to replace all of the moist sand with dry sand on the four moist-dry transitions, but only that portion of the landing area where the impact will occur on the next dry-soil run. One other point to be made is that by planning to replace moist sand with fresh dry sand at prescribed intervals, potential complications can be avoided that are associated with the dry sand gradually picking up moisture over the duration of the dry runs.

Note that the negative consequences of doing all dry runs and all moist runs in two groups have nothing to do with the MDOE design. The original OFAT drop and swing test matrices also featured runs grouped by dry and moist soil, and they, too, would be vulnerable to systematic variation. There is always the option of trading some quality for some convenience, by executing all of the dry runs in Table 4 before any of the moist runs.

One other option, already under discussion by the principals, is to conduct the experiment for one soil type only. The effect of soil type on landing loads could not be quantified under this option but some additional schedule relief could be achieved with fewer runs, and also some additional quality assurance could be derived from the fact that the other three independent variables are all easy to change and therefore present no impediment to run order randomization.

Table 5 below is a one-soil, fully randomized MDOE test matrix, designed to provide a fit to the response surface model of Eq. (6) that minimizes uncertainty for all of the regression coefficients.

$$\begin{aligned}
 & b_0 \text{ (one intercept term)} \\
 & + b_1x_1 + b_2x_2 + b_3x_3 \text{ (three linear terms)} \\
 & + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 \text{ (three interaction terms)} \\
 & + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 \text{ (three quadratic terms)}
 \end{aligned} \tag{6}$$

Equation (6) is a second-order polynomial in the three numerical variables: horizontal and vertical speed, and pitch angle. It differs from Eq. (2) in that the linear soil-type variable is eliminated, as are three of the interaction terms involving the interaction of soil type with each of the numerical variables. Since four terms are eliminated, only 10 data points are required to fit this model, down from 14 for the two-soil case, although some residual degrees of freedom must be added to assess uncertainty. (See Eq. (1) with  $d=2$  and  $p=3$ ).

**Table 5. Single-Soil MDOE Test Matrix**

Run Order	H-Vel fps	V-Vel fps	Pitch Angle deg
1	20	25	23
2	40	25	28
3	20	35	28
4	60	25	33
5	40	35	33
6	40	30	23
7	60	35	23
8	20	35	33
9	60	25	23
10	60	25	23
11	20	25	23
12	20	25	33
13	20	35	23
14	60	35	23
15	60	30	28
16	60	25	33
17	30	30	30.5

Five replicates were added to the 14 points required to fit Eq. (2) in the two-soil MDOE design, bringing the total point count to 19 (Table 4). Because the original OFAT run count of 21 runs was considered problematical, no additional residual degrees of freedom were specified for the 19-run MDOE two-soil design. Instead, the decision was made to provide a two-run cushion relative to the OFAT plan. Had the resources been available, lack-of-fit degrees of freedom would have been specified in addition to the pure error (replicate) degrees of freedom that were included in the design. Given the reduction of five runs afforded by dropping soil type as a variable, there is an opportunity to add lack of fit degrees of freedom as well as replicates.

“Model” degrees of freedom represent the minimum number of points required to fit a given model [Eq. (1)]. “Pure error” (PE) degrees of freedom are comprised of replicates of model points. “Lack of fit” (LOF) degrees of freedom consist of addition points that are acquired at unique sites in the design space; that is, they are not replicates

of any other points. LOF degrees of freedom serve three useful functions. First, if the response function is more complex than originally assumed, LOF degrees of freedom provide additional points to which higher-order terms in the model might be fit. But even if the original response model is adequate, the LOF points serve to reduce the average leverage of the fitted points, making the response model less vulnerable to experimental error at any given point. Finally, LOF degrees of freedom facilitate certain goodness-of-fit tests that indicate when a quality result has been achieved.

For the single-soil MDOE design of Table 5, four replicates were added to the 10-point minimum run count required to fit Eq. (6). In addition, three LOF degrees of freedom were added, bringing the total run count to 17. The addition of three LOF degrees of freedom reduced the average leverage from 0.737 to 0.588, more than a 20% reduction. A leverage of 1 is the maximum value a point can have, and corresponds to a case in which the response is forced through that point. If a straight line is fitted to two points, for example, the leverage of each point is 1. Nine of the points in the two-soil design had a leverage of 1. Because of the addition of LOF degrees of freedom in the one-soil design, none of the points in that design has a leverage of 1, and the overall design is much less vulnerable to experimental error in any one or two points.

The single-soil design achieves a four-run savings compared to the original 21-run OFAT plan. Seven of these runs—four PE and three LOF—are for quality assurance and quality assessment, but are not absolutely essential. It would be possible to execute the single-soil MDOE test in as few as 10 runs, if resource constraints required it.

## V. Summary Remarks

The OFAT experiment designs for two Orion landing system subscale soil impact experiments have been reviewed, with a view to determining if they might be improved by the application of MDOE testing methods. The OFAT tests included some provision for assessing pure error, and the drop test especially displayed a laudable symmetry that would have permitted the estimation not only of main effects for each of its independent variables, but interaction effects among all the variables as well. The swing test had to make some concessions to good experiment design structure to accommodate what was initially believed to be a requirement for a relatively large number of horizontal speed levels, but it too provided for some replication to assess pure error.

Both OFAT tests were susceptible to unexplained systematic variation. There was a significant degree of unnecessary duplication across the two tests, with no option to assess interactions between independent variables that were changed in one test but not the other. There was no provision to assess any nonlinear dependence on any of the independent variables except horizontal speed. More runs were specified than necessary to obtain the information available from either OFAT test. With some alterations, considerably more information could be acquired in fewer runs, and with less uncertainty.

The two OFAT tests were combined into a single MDOE test matrix with the following characteristics:

- Main effects for horizontal speed, vertical speed, pitch angle, and soil type, as well as interaction effects among all six pairwise combinations of these variables can be quantified.
- Second-order effects (curvature) in all three of the numerical variables can be quantified.
- The minimum number of runs necessary to quantify main effects, interaction effects, and curvature has been computed and documented.
- An improved estimate of random error has been incorporated.
  - More degrees of freedom to produce a more reliable estimate
  - More representative distribution of replicated points throughout the test period
- A proactive defense against potentially serious systematic trends in the unexplained variance has been invoked (run-order randomization) that ensures the assumptions of statistical independence are met, which minimizes uncertainty.
- Replicates were selected to reduce the greatest instances of leverage, thereby minimizing the potential adverse impact of significant experimental error.
- Restrictions on randomization by soil type were discussed, with certain workarounds proposed.
- An alternative single-soil test matrix was developed with the same quality assurance and assessment features of the two-soil design.
- The two-soil MDOE design achieves the above improvements with a reduction of two runs compared to the OFAT approach (19 runs vs. 21, or 10%).

The single-soil MDOE design achieved the above improvements with a reduction of four runs compared to the OFAT approach (17 runs vs. 21, or 19%).

## References

- <sup>1</sup>DeLoach, R., and Lyle, K. H., “An Airbag Landing Dynamics Experiment Using the Modern Design of Experiments,” AIAA-2009-0622, 47th AIAA Aerospace Sciences Meeting and Exhibit, Orlando, Florida, January 5–8, 2009.
- <sup>2</sup>DeLoach, R., “Tailoring Wind Tunnel Data Volume Requirements Through the Formal Design Of Experiments,” AIAA-98-2884, 20th Advanced Measurement and Ground Testing Conference, Albuquerque, New Mexico, June 1998.
- <sup>3</sup>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, Nevada, January 2000.
- <sup>4</sup>DeLoach, R., “A Factorial Data-Rate and Dwell-Time Experiment in the National Transonic Facility,” AIAA 2000-0828, 38th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, January 2000.
- <sup>5</sup>DeLoach, R., Hill, J. S., and Tomek, W. G., “Practical Applications of Blocking and Randomization in a Test in the National Transonic Facility (Invited),” AIAA 2001-0167, 39th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, January 2001.
- <sup>6</sup>Morelli, E.A., and DeLoach, R., “Response Surface Modeling Using Multivariate Orthogonal Functions (Invited),” AIAA 2001-0168, 39th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, January 2001.
- <sup>7</sup>Underwood, P., Everhart, J., DeLoach, R., “National Transonic Facility Wall Pressure Calibration Using Modern Design of Experiments (Invited),” AIAA 2001-0171, 39th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, January 2001.
- <sup>8</sup>Parker, P., and DeLoach, R., “Response Surface Methods for Force Balance Calibration Modeling,” 19th International Congress on Instrumentation in Aerospace Simulation Facilities, Cleveland, Ohio, August 2001.
- <sup>9</sup>DeLoach, R., “Applications of the Modern Design of Experiments at NASA Langley Research Center (Invited),” *Proceedings of the American Statistical Association, Section on Physical and Engineering Sciences* [CD-ROM], American Statistical Association, New York, 2002.
- <sup>10</sup>DeLoach, R., “Tactical Defenses Against Systematic Variation in Wind Tunnel Testing,” AIAA 2002-0885, 40<sup>th</sup> AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, January 2002.
- <sup>11</sup>Danehy, P. M., DeLoach, R., and Cutler, A.D., “Application of Modern Design of Experiments to CARS Thermometry in a Supersonic Combustor,” AIAA 2002-2914, 22nd AIAA Aerodynamic Measurement Technology and Ground Testing Conference, St. Louis, Missouri, June 24–26, 2002.
- <sup>12</sup>DeLoach, R., “Formal Experiment Design as a Tool to Automate Aerospace Ground Testing (Invited),” Tenth Annual Spring Research Conference on Statistics in Industry and Technology, University of Dayton, June 4–6, 2003.
- <sup>13</sup>Morelli, E.A., and DeLoach, R., “Ground Testing Results Using Modern Experiment Design and Multivariate Orthogonal Functions (Invited),” AIAA 2003-0653, 41st AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, January 6–9, 2003.
- <sup>14</sup>Dowgillo, R. M., and DeLoach, R., “Using Modern Design of Experiments to Create a Surface Pressure Database From a Low Speed Wind Tunnel Test,” AIAA 2004-2200, 24th AIAA Aerodynamic Measurement Technology and Ground Testing Conference, Portland, Oregon, June 28–30, 2004.
- <sup>15</sup>DeLoach, R., “The Modern Design of Experiments: An Emerging Knowledge Management Approach to Aerospace Testing,” AIAA-2005-7076, AIAA Engineering InfoTech@Aerospace Conference, Arlington, Virginia, September 26–29, 2005.
- <sup>16</sup>Albertani, R., Stanford, B., DeLoach, R., Hubner, J. P., and Ifju, P. S., “Wind Tunnel Testing and Nonlinear Modeling Applied to Powered Micro Air Vehicles with Flexible Wings,” AIAA Journal of Aircraft, Summer 2007.
- <sup>17</sup>Erickson, G. E., and DeLoach, R., “Estimation of Supersonic Stage Separation Aerodynamics of Winged-Body Launch Vehicles Using Response Surface Methods,” 26th International Council of Aeronautical Sciences, Anchorage, Alaska, September 14–19, 2008.
- <sup>18</sup>DeLoach, R., Marlowe, J. M., and Yager, T. J., “Uncertainty Analysis for the Evaluation of a Passive Runway Arresting System,” AIAA-2009-1156, 47th AIAA Aerospace Sciences Meeting and Exhibit, Orlando, Florida, January 5–8, 2009.