JOURNAL OF AIRCRAFT
Vol. 47, No. 2, March–April 2010

Application of Design of Experiments to Flight Test: A Case Study

Aaron A. Tucker*
U.S. Air Force, Edwards Air Force Base, California 93524
Gregory T. Hutto†
U.S. Air Force, Eglin Air Force Base, Florida 32547
and
Cihan H. Dagli‡
Missouri University of Science and Technology, Rolla, Missouri 65409

DOI: 10.2514/1.40301

Modern flight testing tends to be a complex, expensive undertaking, so any increases in efficiency result in considerable savings. Design of experiments is a statistical methodology that enables a highly efficient investigation where only the samples needed are collected and analyzed. Increased information is garnered from the data collected, whereas the number of data points required to understand the system is reduced. In this limited effort, an actual flight-test program serves as a case study to compare and contrast five different designs to explore a flight-test envelope. The case-study data are analyzed using each designed experiment, and the results are compared and contrasted as a cost-benefit relationship between flight-test resources expended and system understanding gained (i.e., statistical confidence and power). The design of experiments methodologies, as applied to this case study, show a 50 to 80% reduction in flight-test resources yet produce similar levels of statistical confidence and power. In an era of restricted budgets and time lines, careful design and thoughtful analysis of flight-test experiments can make the difference between a failed or canceled flight-test program and the successful fielding of a needed capability.

Nomenclature

k	=	number of independent variables
M	=	Mach number
n	=	number of data conditions
p	=	statistical significance
α	=	angle of attack
α_{level}	=	angle of attack, level flight
β	=	angle of sideslip
σ	=	standard deviation

Introduction

MODERN flight testing tends to be complex and expensive, so any increases in efficiency would result in considerable savings. Design of experiments is a statistical methodology, which enables a highly efficient investigation where only the samples needed are collected and analyzed. The application of design of experiments to the design of flight tests can result in a substantial increase in test efficiency. Increased information is garnered from the data collected while the number of data points required to understand the system is reduced. Design of experiments has been used in wind-tunnel tests to determine basic stability and control derivatives [1] and optimize a leading-edge extension on an F/A-18 [2]. Simpson and Wisnowski presented a step-by-step approach to applying a designed experiment to a CV-22 terrain-following/terrain-avoidance radar test [3]. This paper will apply several experimental designs to a case study in the form of a flight research program. The results of each analysis are compared in terms of flight-test resources expended. A disciplined application of design of experiments techniques shows great promise to efficiently extract information from flight-test data and help guide the decisions of the test team and operational community.

A typical problem facing the designers of flight-test experiments is an investigation of an altitude/airspeed aircraft envelope (e.g., Fig. 1) [4]. Numerous safety and operational limitations (e.g., stall margin, minimum altitude, structural loads, and performance) must be honored throughout a typical test program and typically do not lend themselves to a simple, designed experiment. As a result, most test teams resort to a survey method of regularly spaced data points on a flight envelope constructed by their subject matter experts. The survey method, however, allows for a simple, descriptive study where only limited information can be extracted from a given data set. Generally, the tester knows only how the system responded under the conditions tested with no way to expand those data into a system-response model or measure the experimental error. In other words, the tester knows only what happened during the test and nothing about how likely those data reflect true system performance.

The generic flight-test problem may be viewed, as shown in Fig. 1. The experimenter wishes to vary certain controllable inputs in such a way as to deduce which inputs, alone and combined, may affect the output of the process. Furthermore, he would like to vary these factors in such a way as to minimize his chances of drawing an erroneous conclusion, such as finding that an inert factor has an effect or failing to detect an active factor. The presence of background variation (e.g., fuel state, wind, and pilot technique), measurement error, and other noise sources complicate the experimenter's problem. The large number of combinations available to be tested within the test envelope also complicates the problem. The flight-test envelope in Fig. 1 contains an infinite number of combinations of four continuous variables: altitude, Mach number, angle of attack, and angle of sideslip.

A flight-test program planned, executed, and analyzed as a designed experiment would provide several benefits. First, a model

Received 13 August 2008; accepted for publication 15 July 2009.

*C-5/C-17 Experimental Test Pilot, 418th Flight Test Squadron, 35 North Flightline Drive; aaron.tucker@us.af.mil. Senior Member AIAA.
†Wing Operations Analyst, 46th Test Wing, Room 211, Building 1.
‡Professor of Systems Engineering, 229 Engineering Management, 1870 Miner Circle.

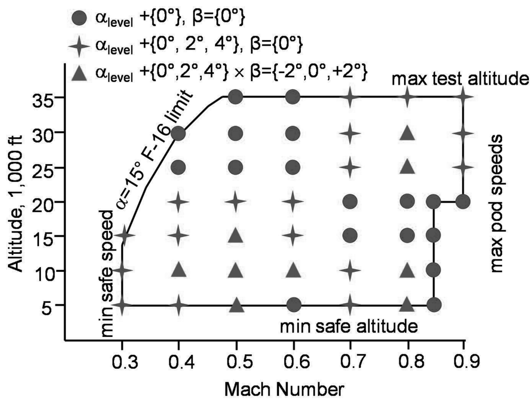

Fig. 1 Baseline experiment.

of the system's behavior would be generated. The model would indicate the probability that those results are representative of the true system behavior (statistical confidence) and how likely the model could detect a response of a given magnitude (statistical power). Deliberate planning decisions can consider the required quality of the system-response model. The resultant design is the most efficient use of very expensive, and often limited, flight-test resources. For example, a relatively low-resolution model would require a given investment of flight-test resources (time, money, aircraft, and people) that results in a particular confidence and power. A given increase in the fidelity of test points in the test space could be purchased for a calculable increase in resources according to the specific needs of the customer. Further, the data set gathered under one flight-test program using a designed experiment could serve as the baseline for the next system modification. Such savings are particularly attractive in the current spiral development programs in defense acquisitions. Finally, the system-response model with its accompanying statistical confidence and power would be provided to decision makers. It would educate their technical, safety, and programmatic decisions with rigorous analyses. The result would enhance efficient stewardship of the nation's flight-test resources.

Description of Designed Experiments

"Experiment is defined as a test or series of tests in which purposeful changes are made to the input variables of a process or system so that we may observe and identify the reasons for changes that may be observed in the output response" [5].

Design of experiments (DOE) allows for planning and conducting experiments with the goal of analyzing the resultant data so that valid and objective conclusions are obtained. It is the application of a proven test methodology coupled with statistics to analyze both the main effects of input variables and interactions of those effects on response (output) variables. The generalized framework has been instrumental in closing the gap between engineering and statistics [6]. The experimental design lends itself to efficient statistical interpretation of experimental data. Two important statistical measures are used to measure the quality of the experimental results:

1) Statistical confidence provides a measure of the experiment's ability to determine the probability that the observed system response is representative of the true behavior.

2) Statistical power gives the probability that a certain sample size was adequate to detect defined changes in the system response under test.

Through each phase of flight testing, DOE is a robust method of planning, executing, and analyzing flight-test data. It is particularly adept at isolating system performance from external sources of variability. Designed experiments enhance traditional flight-test procedures and allow systems engineers and program managers to create objective, verifiable, and traceable empirical models that reveal both main effects and their interactions, if they exist [7]. The advantages of designed experiments applied to flight testing are as follows:

1) DOE provides a structured planning process that can be used to involve stakeholders and synergize individual subject matter knowledge to generate test and analysis plans that are comprehensive and efficient.

2) Sequential testing and analysis leads to immediate system discovery and understanding (accelerated learning).

3) Empirical statistical models can be used for estimation and prediction [1].

This paper will address the experimental design and the properties of such designs, especially their scope, statistical power and confidence, and flexibility. Other topics will be covered, such as the models that can be generated from designed experiments, the flexibility of designed experiments, and the ability to design experimental subsets, which conserve test resources while accepting diminished resolution on variables that have little effect in the investigation. The authors believe that a retrospective redesign of an actual flight test in multiple ways can illustrate the efficacy and efficiency of the application of design of experiments to the challenges of test.

Description of the Case Study

The case study was a flight-test program charged with demonstrating the effect of active-flow control on targeting pod-induced buffet on an F-16B aircraft's ventral fin [4]. A targeting pod shape was equipped with six synthetic jet actuators designed to inject high-velocity jets of air into the flowfield at a specific frequency. The U.S. Air Force Research Laboratory had conducted wind-tunnel and preliminary-flight research on the effects of the synthetic jet actuators [8–10]. These jets of air were intended to actively control the buffet levels experienced by the F-16B's ventral fin approximately 13 ft downstream of the targeting pod.

To test the effect of the synthetic jet actuators on ventral fin buffet, a modified ventral fin was installed with strain gage and static-pressure port instrumentation. The test aircraft was stabilized at a specific test condition with a specified altitude, Mach number, angle of attack, and angle of sideslip. Two 10 s data points were then collected: 10 s with the synthetic jets turned on and 10 s with the synthetic jets turned off. Time-domain data was transformed to the frequency domain in a power-spectral density chart. Ventral fin buffet was characterized by several peaks in the frequency domain: strain (80 Hz), pressure (158 Hz), and pressure (225 Hz). Effects of the synthetic jets were analyzed from this frequency-domain data. The case-study experiment did not detect any significant change in ventral fin buffet level across a wide range of flight conditions in 10 flight-test sorties and 14 flight hours despite good statistical power.

The baseline case study was executed in traditional fashion with no planned controls to account for day-to-day variability. It lacked randomization and blocking in order to protect against unknown sources of variability. Further, the baseline case was not designed to produce a data set supporting a nonlinear model of the behavior of the fin vibration across the envelope of flight-test explored. All the designed experiments considered in this paper are specifically designed to fit an empirical model that describes how the vibration varies across the flight envelope of interest. In the baseline case, a statistically designed experiment would have likely reached the conclusion that the jets were ineffective in only a few sorties, saving the additional test resources to explore other aspects of the flight-test problem.

Application of Designs of Experiment to a Case Study

This paper examines the results of applying various designs of experiment on a baseline experiment's flight-test envelope investigation. The baseline experiment's test plan was originally planned and executed as a classic subject matter expert (SME)-designed survey. These flight data were used as the basis for data sets, which were analyzed in several designed experiments. The results of each analysis are compared for their resultant statistical confidence and power and required number of data points. Operational flight-envelope limitations (α, M, and altitude) were honored during the design of each experiment in order to illustrate the challenge of applying experimental methods to a real flight-test envelope.

Baseline Experiment

The case study applied a typical grid investigation of a Mach number-altitude flight envelope [4]. In Fig. 1, circles indicate a typical survey grid placed at the edges of the flight envelope and at regular intervals within the flight envelope. These data points represent a test condition at the level-flight angle of attack , $\alpha_{\text{level}} + \{0°\}$, and zero angle of sideslip, $\beta = \{0°\}$. The case-study subject matter experts also selected specific areas of the flight envelope in which to further investigate the system response. Therefore, variations in angles of attack and sideslip were applied in these flight regions. Cross stars indicate that data were collected at $\alpha_{\text{level}} + \{0°, 2°, 4°\}$, $\beta = \{0°\}$, and triangles indicate that data were collected at

$$\alpha_{\text{level}} + \{0°, 2°, 4°\} \times \beta = \{-2°, 0°, +2°\}$$

At each test condition, data were recorded with the synthetic jet actuators operating and turned off. The trials consisted of 14 flight hours in an instrumented F-16B flight-test aircraft for a total of 324 test points.

Central Composite Design

Figure 2 shows the central composite design (CCD) in altitude, Mach number, angle of attack, and angle of sideslip. It efficiently fits a second-order polynomial with interactions to describe the effects of the control variables (Mach number, altitude, α, β, and synthetic jet operation) on the vibration response. The CCD requires $2^k + 2k + 3$ points. Four variables ($k = 4$) result in $2^4 + 2 \times 4 + 3 = 27$ points. Multiplied by the two operating states of the synthetic jets (on and off), the total number of test points required is 54, which is 17% of the baseline experiment. Not only does the designed experiment require fewer resources, but it will also better characterize the main part of the flight envelope because each altitude/Mach number condition is investigated through the full ranges of angles of attack and sideslip. While the axial points on the edges of the envelope are remote from the main factorial portion, Fig. 3 illustrates that the center of the envelope appropriately receives much more operational utility than the edges of the envelope [11].

Another benefit to the CCD is the flexibility of the design, particularly its ability to be efficiently augmented in numerous ways to explore interesting or unexpected behavior. Points can be added that enable the analysis of an entirely new orthogonal matrix, a CCD + n design. A CCD + n design is a classic CCD with n orthogonal points added to the axial points to create another factorial centered near the edge of the envelope. For instance, if the test team noted interesting results in the low-altitude, low-Mach-number region of Fig. 2, another two-level factorial could be analyzed by adding a single altitude-Mach test condition ($\alpha_{\text{level}} + \{0°\}$, $\beta = \{0°\}$), the square, resulting in a CCD + 1 design. The resultant system-response model may provide further insight into an area of regular operational utility (Fig. 3). If variations in angle of attack warranted investigation, the $\alpha_{\text{level}} + \{0°, 2°, 4°\}$, $\beta = \{0°\}$ data set could be added to the experimental design with the addition of three test conditions, a CCD + 3 design. Completing the investigation with angles of sideslip

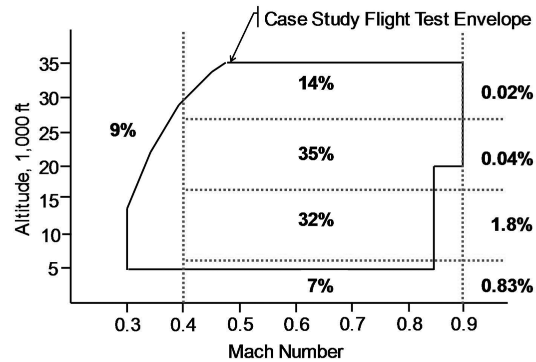

Fig. 3 F-16CG operational employment spectra.

$$\alpha_{\text{level}} + \{0°, 2°, 4°\} \times \beta = \{-2°, 0°, +2°\}$$

would result a total of nine additional test conditions and a CCD + 9 design.

Such flexibility should be anticipated by the flight-test team with test and safety planning that preserves the comprehensive evaluation of risk but allows for flexibility in the placement of specific test points. Consider that similar factorials could be created to further refine the model at the various edges of the envelope, often a necessity before releasing an operational capability to the user.

Face-Centered Central Composite Design

Another design choice would be the face-centered CCD (FCD), where the axial points of the CCD are projected into the face of the factorial (Fig. 4). The FCD was developed to address the problem of physical limitations on the test article that did not allow the axial points to be placed remotely from the main portion of the design [12]. It has the added benefit of efficiently exploring a hyper cubic test region, such as the Mach number/altitude envelope commonly encountered in loads, vibrations, limit cycle oscillation, and environmental characterization testing. As with the CCD, the FCD allows an explicit quadratic model of responses with calculated levels of experimental noise. If desired, smaller adjacent FCD matrices can be constructed, as shown in the low-Mach-number/altitude corner of the envelope (Fig. 4 dashed lines). Similar to the CCD, each FCD consists of 54 test conditions, which is 17% of the case study.

Consider that both matrices could be planned in a flight-test program and undergo the required test and safety planning processes with flexibility in the placement of the points. The major FCD would be executed, and the data could be reduced and analyzed before placing the points for the minor FCD. This allows the test team to learn from the fully modeled system response, better understand the system, and apply that knowledge to the minor FCD at the edge of the flight envelope and realize the benefits in test efficiency, risk management, and confidence in system response.

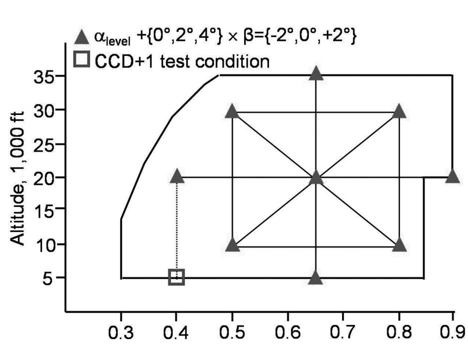

Fig. 2 Central composite design.

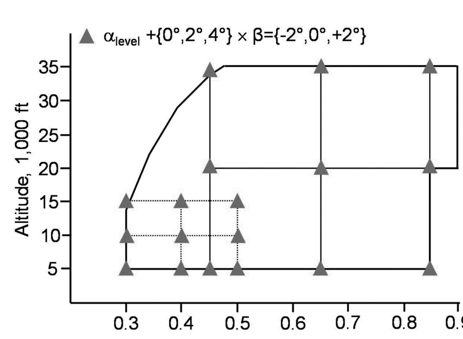

Fig. 4 Face-centered central composite design.

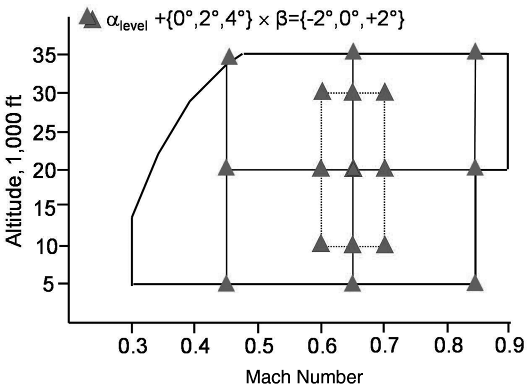

Fig. 5 Embedded FCD.

Embedded FCD

Figure 5 shows a designed experiment consisting of two embedded FCDs: one within the other (triangles) [1]. The resultant composite design allows up to third-order polynomials to be fit through the experimental region at modestly increased cost. The two 54-test-point FCDs can be run together for a total of 108 test points (including three common-center points repeated to estimate experimental error).

Fractional FCD

For efficiency, a further refinement of the CCD and FCD designs is to fractionate factorial portions of the design. A fractional factorial experiment carefully chooses a subset of the original experiment that preserves information about selected variables while sacrificing the ability to resolve other interactions [13]. This technique increases efficiency in execution at the cost of slightly increased complexity in analysis and the possibility of later requiring additional test points to resolve uncertainties in the system-response model. In practice, fractional factorial experiments are commonly employed with five or more test variables. A resolution-IV half-fraction of the four continuous variables yields a design with $2^{(4-1)} + 2 \times 4 + 3$ center points for a total of 19 test conditions, yielding 38 data points for the entire experiment. As before, a quadratic model can be fit through each of the variables and explicitly model uncertainty and reproducibility.

Three-Level Design

If the experimenter believes the variables might interact nonlinearly, a suitable choice might be a three-level design, referred to as the $3k$ class of designs. For these problems, one often decides to fractionate the resultant design because the powers of 3 grow so rapidly with the number of test conditions. For the four variables under consideration, a 3^4 full factorial or 81 test conditions, repeated for both on- and offjet conditions yields a total of 162 data points. This design is half the size of the baseline but with vastly improved modeling possibilities. A fractionated three-level design would typically be a one-third fraction for a total of $3^{(4-1)} \times 2 = 54$ test conditions. Generally, three-level designs are not the best choice for nonlinear modeling but are certainly among those that may be considered.

Box–Behnken Designs

Another response surface choice is a class of designs located on a sphere or circle in the experimental region. Although not particularly appropriate for this largely rectangular test region, the Box–Behnken designs add to the toolbox of the experimenter seeking to characterize nonlinear responses efficiently [14].

Results

To compare the efficiency and efficacy of five experimental methods, the data from the case study were formed into data sets as prescribed by each designed experiment. However, the original case study did not include some test conditions that a designed experiment would have prescribed, particularly in the $\beta = \{-2°, +2°\}$ data set. Therefore, the $\beta = \{0°\}$ data sets for the various designed experiments were completed using interpolation, extrapolation, and hand fairing of system-response curves. The case study's $\beta = \{-2°, +2°\}$ data set, however, did not have enough original data and could not be completed. Analyses of variance were used to analyze only the $\beta = \{0°\}$ data and statistical power was calculated for each design.

Statistical Confidence

As discussed in the Introduction, a tester can make one of two errors in concluding the results of a test program: determining that a factor makes a difference in the response when it does not (measured by statistical confidence) or failing to detect a difference when such a difference actually exists (measured by statistical power). Therefore, the tester would like to have high confidence in his conclusions while preserving high power.

Table 1 compares the experimental designs in terms of the data points required to execute the design and the p value. The number of points represents the efficiency of the experimental design. The tabulated p values allow comparison between the various designs' efficacy and represent statistical confidence $(1 - p)$.

Table 1 includes a p value for the effect of the system under test on each of three system responses: strain (80 Hz), pressure (158 Hz), and pressure (225 Hz). Note that none of the synthetic jet p values reach a level of significance ($p < 0.05$). This is important to the tester because it shows the synthetic jets were unable to affect the system response in the flight envelope of interest. The conclusion is that the synthetic jets did not significantly affect the vibration of the F-16 ventral fin. An analysis of the 324 data points gathered in the baseline experiment reached the same conclusion, with similar confidence to the designed experiments. Importantly, the designed experiments provided the same conclusion with 16 to 33% of the test resources expended. The challenge for the test team is to find the balance between the size of the experiment and the information gained [15].

Statistical Power

Figure 6 compares the statistical power achieved by several types of experiments. The SME-designed baseline experiment has the statistical power to detect very small changes in ventral fin vibration for each response variable. For example, a synthetic jet actuator effect of 0.50 standard deviation, 0.50σ, has a 90% probability of being detected by the baseline experiment. In comparison, a CCD has only a 40% probability of detecting a 0.50σ effect or would take a 1.0σ effect to have a 90% probability of detection. It should be noted that

Table 1 Results of various designed experiments

Experimental design	Number of points in data set (% baseline)	Ventral fin response	p value
Baseline	324 (100%)	Strain (80 Hz)	0.77
		Pressure (158 Hz)	0.45
		Pressure (225 Hz)	0.95
CCD	54 (17%)	Strain (80 Hz)	0.65
		Pressure (158 Hz)	0.98
		Pressure (225 Hz)	0.77
CCD + 3	60 (19%)	Strain (80 Hz)	0.65
		Pressure (158 Hz)	0.94
		Pressure (225 Hz)	0.82
Face-centered CCD	54 (17%)	Strain (80 Hz)	0.66
		Pressure (158 Hz)	0.89
		Pressure (225 Hz)	0.82
Embedded FCD	108 (33%)	Strain (80 Hz)	0.62
		Pressure (158 Hz)	0.88
		Pressure (225 Hz)	0.79
Half-fractional embedded FCD	53 (16%)	Strain (80 Hz)	0.65
		Pressure (158 Hz)	0.88
		Pressure (225 Hz)	0.92

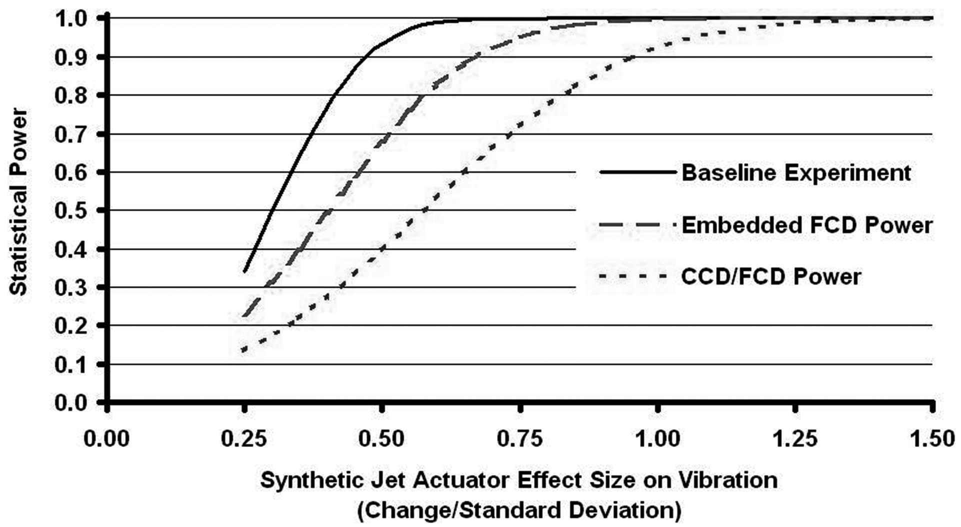

Fig. 6 Experimental statistical power.

such small changes in the response (half the noise level or less) are typically not of much practical interest in a flight-test program. One might take the position that the SME-generated baseline was too powerful for the intended purpose (detecting ventral fin buffet) and therefore wasteful of test resources. Conversely, the test team could have used those test resources to examine an alternate experimental configuration (e.g., without external wing tanks), as recommended by the original test team [4].

Statistical power is a crucial commodity in flight testing. A test design that lacks power will fail to detect the very behavior it was designed to explore, whereas an overpowered test simply wastes resources. Properly designed, the analysis can calculate the power of the design using a selected significance (e.g., $p < 0.05$) to determine the sample size required to give a specified probability of detecting a given effect size. The baseline experiment can detect very small effects with a high probability of detection but comes at a cost of six times the invested resources of the CCD experiment. The test team can select a required statistical power for a given effect size and design and fund a test program with the statistical rigor to successfully investigate the system response. The alternative to a designed experiment is to blindly collect and inspect data with no insight into whether that test could have detected the effect of interest.

Conclusions

The application of the design of experiments to a flight-test investigation is valuable by increasing both test efficiency and the understanding of the data collected. The information gained from the data set is exploited to generate a system-response model with a known statistical confidence and power. The test is efficient because no more data are collected than required to produce that system-response model. Compared to a typical flight-envelope investigation, a flight-test program using a designed experiment would require only 16 to 33% of the data points to fully model the flight envelope. Also, the flexibility in designed experiments allows for test teams to apply safe test practices, investigate interesting results, and continually build on the data set in a structured manner. Designed experiments can assist with safety planning and constructing a buildup approach by producing insight into the system response during the initial data collection. This insight will educate the judgment of the test team and help to mitigate risk through increased confidence in a refined system-response model as the data collection progresses closer to the predicted edge of the operating envelope. Also, the flexibility of the experimental designs allows the test team to analyze a complete data set, note areas of interest, and adjust the test plan to completely investigate those regions. Often, the test team can use part of the existing data set instead of launching an entirely new test program. In a designed experiment, data is collected in such a manner that the effect of an independent variable can be isolated from other variables. Therefore, future test efforts can use previous data as part of the structure of another designed experiment.

In times of enormously expensive flight-test programs, the efficiencies realized through the application of designed experiments to flight testing could mean the difference between the timely delivery of a needed capability to the customer; an overcost, late, underperforming system; or outright cancellation of the system. Design of experiments has the capability to make flight testing safer and more efficient.

Acknowledgments

Aaron Tucker would like to thank George Ka'iliwai III for his generous support and direction in this line of inquiry and Elwood T. Waddell for his statistical expertise and guidance.

References

[1] Landman, D., Simpson, J., Mariani, R., Ortiz, F., and Britcher, C., "A High Performance Aircraft Wind Tunnel Test Using Response Surface Methodologies," AIAA Paper 2005-7602, Dec. 2005.

[2] Luner, J. J., and Healey, M. D., "Modern Design of Experiments Techniques to Optimize a Leading Edge Extension," AIAA Paper 2003-655, Jan. 2003.

[3] Simpson, J. R., and Wisnowski, J. W., "A Systematic Approach to Planning Flight Tests Using Designed Experiments," AIAA Paper 2000-2695, June 2000.

[4] Tucker, A. A., Morgenstern, S. D., Giebner, H. C., Wong, D., Violetti, A., and Burgess, G., "Limited Characterization of Dual Bimorph Synthetic Jet Actuators: Aeroelastic Load Control—Phase I," AIAA Paper 2007-475, Jan. 2007.

[5] Montgomery, D. C., "Introduction," *Design and Analysis of Experiments*, 6th ed., Wiley, New York, 2005, pp. 1–22.

[6] Nair, V. N. (ed.), "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, Vol. 34, No. 2, 1992, pp. 127–161. doi:10.2307/1269231

[7] Simpson, J. R., and Wisnowski, J. W., "Streamlining Flight Test with the Design and Analysis of Experiments," *Journal of Aircraft*, Nov.–Dec. 2001, pp. 1110–1116; also AIAA Paper 0021-8669.

[8] Shaw, L. L., Smith, B. R., and Saddoughi, S., "Active Control of Pod Wake—Mid-Scale Application," AIAA Paper 2005-799, Jan. 2005.

[9] Hamburger, C. L., Smith, B. R., and Shaw, L. L., "Wake Large Eddy Simulation over a Blunt-Base Pod with and Without Base Pulsating Blowing," AIAA Paper 2005-5014, June 2005.

[10] Shaw, L. L., Smith, B. R., and Saddoughi, S., "Full-Scale Flight Demonstration of Active Control of a Pod Wake," AIAA Paper 2006-3185, June 2006.

[11] Wickert, D. P., "Mitigating Risk at the Edge of the Envelope: Revising Basic Aircraft Limit Test Procedures," Society of Experimental Test Pilots Paper 2015, Sept. 2006.

[12] Montgomery, D. C., "Response Surface Methods and Designs," *Design and Analysis of Experiments*, 6th ed., Wiley, New York, 2005, pp. 405–463.
[13] Montgomery, D. C., "Two-Level Fractional Factorial Designs," *Design and Analysis of Experiments*, 6th ed., Wiley, New York, 2005, pp. 282–346.
[14] Box, G. E. P., and Behnken, D. W., "Some New Three-Level Designs for the Study of Quantitative Variables," *Technometrics*, Vol. 2, No. 4, Nov. 1960, pp. 455–475.
doi:10.2307/1266454
[15] Antony, J., "Some Useful Practical Tips for Making Your Industrial Experiments Successful," *Design of Experiments for Engineers and Scientists*, Butterworth-Heinemann, Amsterdam, 2007, pp. 93–104.