JOURNAL OF AIRCRAFT
Vol. 38, No. 6, November–December 2001

Streamlining Flight Test with the Design and Analysis of Experiments

James R. Simpson*
Florida State University, Florida A&M University, Tallahassee, Florida 32310-6046
and
James W. Wisnowski†
United States Air Force Academy, Colorado Springs, Colorado 80840

Traditional flight-test methodology typically results in large test matrices using the approach of changing settings of one input variable at a time, keeping all others at some nominal level. Although this method is functional, another method exists that can dramatically reduce the amount of testing required, increase the quality of the data analysis, and provide additional insight into system performance. We propose using the principles of statistically designed experiments; and we provide a process for planning flight test using these principles. The proposed process complements traditional testing methods and places a greater emphasis on structuring test design to increase the precision and interpretability of the data using statistical analysis. Another enhancement to the test process is recommending small and sequential testing so that information learned from earlier tests is incorporated in subsequent phases and unnecessary tests can be avoided. The proposed methodology has three phases: building the strategic flight-test plan, developing detailed test objectives, and planning flight testing and analysis. This paper discusses the activities conducted in each phase and illustrates the main points with an example from a recent experience with the CV-22 operational test and evaluation. Although our proposal is directed to flight test, it can be easily implemented in most test environments.

I. Introduction

ENGINEERS and scientists have traditionally generated new aircraft system flight-test plans by building a comprehensive set of test matrices in order to assess the aircraft in all conceivable operational flight conditions. These test plans are developed using a combination of past experience and current system knowledge. The intent is to expand gradually the performance envelope and expose the aircraft to actual flight conditions so that the majority of the system flaws can be identified and repaired. Flaws in the system are typically identified through a pass/fail rating system for each flight-test point. Once a flaw is identified, expert knowledge is used to determine additional relevant flight tests used to pinpoint problem areas.

The aircraft engineers who are specialists in the potential problem areas can often quickly identify the root cause. However, sometimes the investigation is lengthy requiring additional test flights and subsequent study. The test engineers typically gather aircraft configuration information starting at the time that problems are first detected. Although this information is certainly helpful, an alternative testing approach that takes advantage of special test matrix structures and statistical analysis methods can provide significantly more evidence regarding potential relationships between problem root causes and aircraft system configurations. In fact, by using a statistical approach to designing flight-test experiments system engineers could use flight-test data to develop empirical models relating aircraft performance to changes in flight system control settings. Using this modeling approach, system engineers can determine which aircraft control parameters significantly impact various aircraft performance characteristics. The models can be further used to predict aircraft performance under specified flight conditions not flown in the test program.

Two management concerns in new aircraft flight-test programs are reducing the test budget and completing the test schedule on time. Test plans are often lengthy, and many of the flight-test points appear redundant. Fortunately, by developing test matrices using an efficient designed experiment approach and by gaining knowledge of the aircraft systems using sequential testing procedures these concerns can be alleviated. The suggested test matrices can be greatly reduced by running only a portion of all possible configurations. This streamlined test plan will still include all configurations deemed essential to demonstrate operational proficiency.

Streamlining the test plans requires modifying the approach to build test matrices. Current test plan methods suggest changing a single input flight-control parameter at a time, while keeping all other parameters fixed. Once all variations of that single factor are tested, the plan switches to focus on the second factor. The second factor is varied across all its levels again keeping all other factors constant at some nominal setting. This approach is called one-factor-at-a-time (OFAT) testing. The OFAT method is traditionally used in many sectors including the flight-test community. Unfortunately, the OFAT method is inefficient in terms of number of tests required and often times does not reveal the true relationship between the inputs and the output performance measures (see Ref. 1, for example). An alternative method is to use formal design of experiments (DOE) techniques that propose a test matrix structure allowing for simultaneous changes in the input parameters. These test matrices are efficient and allow for simple development of empirical statistical models that can be validated as representative of the true aircraft systems. Montgomery[2] shows that these DOE techniques, called factorial designs, outperform the OFAT method as a function of the number of input factors. For example, in Fig. 1 we see that an experiment with five input variables or factors would require one-third (that is, 1/relative efficiency) the number of tests (runs) to obtain the same information. We will show that DOE actually produces more information about the system in significantly fewer runs than OFAT.

Designed experiments methods have been used successfully in industrial settings, primarily in manufacturing (including aerospace) and in chemical process companies. These industries have

Received 17 January 2001; revision received 15 July 2001; accepted for publication 21 August 2001.

*Assistant Professor, Department of Industrial Engineering, 2525 Pottsdamer Street.
†Associate Professor, Department of Mathematical Sciences.

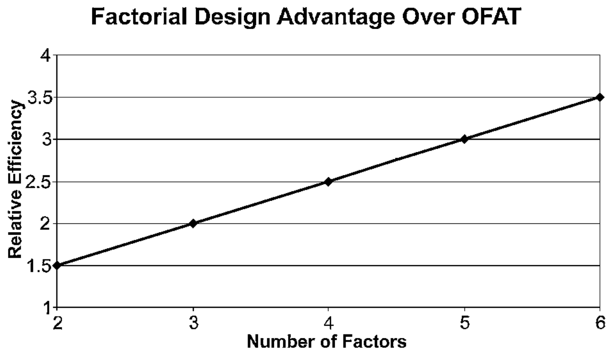

Fig. 1 Test savings from using DOE techniques vs traditional OFAT methods. The relative efficiency is the ratio of tests needed by OFAT to those needed by DOE to statistically estimate the factor effects. For example, twice as many tests are need for a three-factor experiment.

incorporated designed experiments into all phases of the product and process life cycle and as a result have realized significant gains in productivity, performance, quality, and reliability. These techniques, however, have not received as much attention in the flight-test community. Many flight tests still use traditional OFAT methods and limit analysis to graphical studies. Engineers have recently been incorporating DOE methods into ground-based aircraft testing, including wind-tunnel tests.[3–5] The purpose of this paper is to introduce a procedure for using DOE in planning, conducting, and analyzing the results of flight tests.

Specifically, we propose a general flight-test planning approach that incorporates the powerful methods of designed experiments and statistical analysis while maintaining the integrity of traditional flight-test principles. The approach consists of three phases: 1) building the strategic flight-test plan, 2) developing detailed test objectives, and 3) planning sequential flight testing and analysis. Each of these phases contains a procedure that is outlined and then discussed. Included in the discussion is an introduction to design of experiments and appropriate references for further study. The advantages of using this method over traditional methods are described in the conclusion section. For illustration purposes we discuss an actual flight-test planning exercise that used our recommended approach for evaluating a radar system for a new U.S. Air Force aircraft. Results from the planning effort[6] are woven throughout the discussion to demonstrate procedure implementation.

II. Case Study Background

The Bell-Boeing CV-22 Osprey is a tilt-rotor, multipurpose, multimode aircraft. The CV-22 is capable of helicopter mode for vertical takeoff and landing and quick conversion to airplane mode for higher speed, long distance travel. The CV-22 is the U.S. Air Force version of the Marines' and Navy's V-22 and will serve as the mainstay of the U.S. Air Force Special Operations Command. The aircraft will be used for search and rescue as well as for troop exfiltration and infiltration of Special Forces personnel. The U.S. Air Force plans to purchase 50 of these aircraft by 2010.

The CV-22 Bell-Boeing Integrated Test Team decided to incorporate experiment design methods in planning the flight tests for the aircraft's terrain-following/terrain-avoidance (TF/TA) radar system. They formed a TF/TA Tiger Team to investigate the potential application of formal designed experiments that could possibly be helpful in designing the test points and analyzing the data after flight tests. After being introduced to experiment design methods, the Tiger Team decided to proceed with the new approach to planning flight tests. The 20-member TF/TA Tiger Team consisted of U.S. Air Force, Boeing and Bell CV-22 senior leadership, CV-22 avionics engineers, flight-test engineers, members of U.S. Air Force Operational Test and Evaluation Command, U.S. Air Force CV-22 test pilots, and a U.S. Air Force officer expert in design of experiments. The planning exercise involved several meetings over an 18-month period culminating in a test plan[6] for evaluating the TF/TA radar system.

For many of its missions, the CV-22 pilot will fly low over variant terrain to avoid detection by enemy radar. The objective of the TF/TA radar system is to combine the skills of the pilot with advanced aircraft avionics to maintain altitude over often-rugged terrain and keep the aircraft out of harm's way. The ability of the TF/TA radar system (including the pilot) to maintain altitude may be dependent on various flight conditions. For example, maintaining constant altitude above ground level over level terrain, in airplane mode, cruising straight ahead at 200 kn may not be as challenging as turning hard over an isolated peak in helicopter mode at 75 kn.

Incorporating a designed experiment approach to flight test resulted in a process that contained the necessities of traditional plans and also added the efficiencies and analytical insight gained by knowledge of the methods that have been so successful in industry. The process steps will be discussed in detail and examples from the CV-22 study will be used to illustrate actual implementation of the principles.

III. Phase 1: Building a Strategic Flight-Test Plan

The procedure for industrial planning experiments using a designed experiments approach has been proven over decades of implementations. Several authors have provided guidelines for successful planning. For example, Coleman and Montgomery[7] suggest a seven-step process for experimentation including master guide sheets to facilitate the process. These guidelines are certainly helpful for planning flight test, but because of the nature of the mission many other issues must be considered. The following sections present a three-phase process specifically for flight test.

The initial requirement of the first phase (Fig. 2) is to clearly and completely define the strategic objective of the flight test. Although often only a sentence or two, the strategic objective is the reference point for justifying all detailed objectives. Senior members of the flight-test planning team are involved in generating this objective, usually before assembling the entire flight-test planning group. The strategic objective must convey the mission of the study in a clear, concise, and comprehensive manner. For the CV-22 project the objective was to optimize the performance of CV-22 multimode TF/TA radar system.

Determining the relevant members of the planning team is the second step of phase I. The core of the group should already have been involved in stating the problem, and additional members should

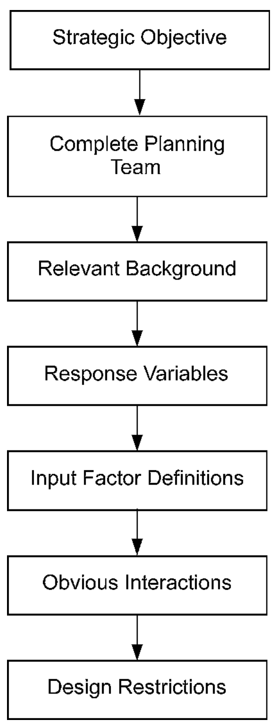

Fig. 2 Sequence for phase I: developing the strategic flight-test plan.

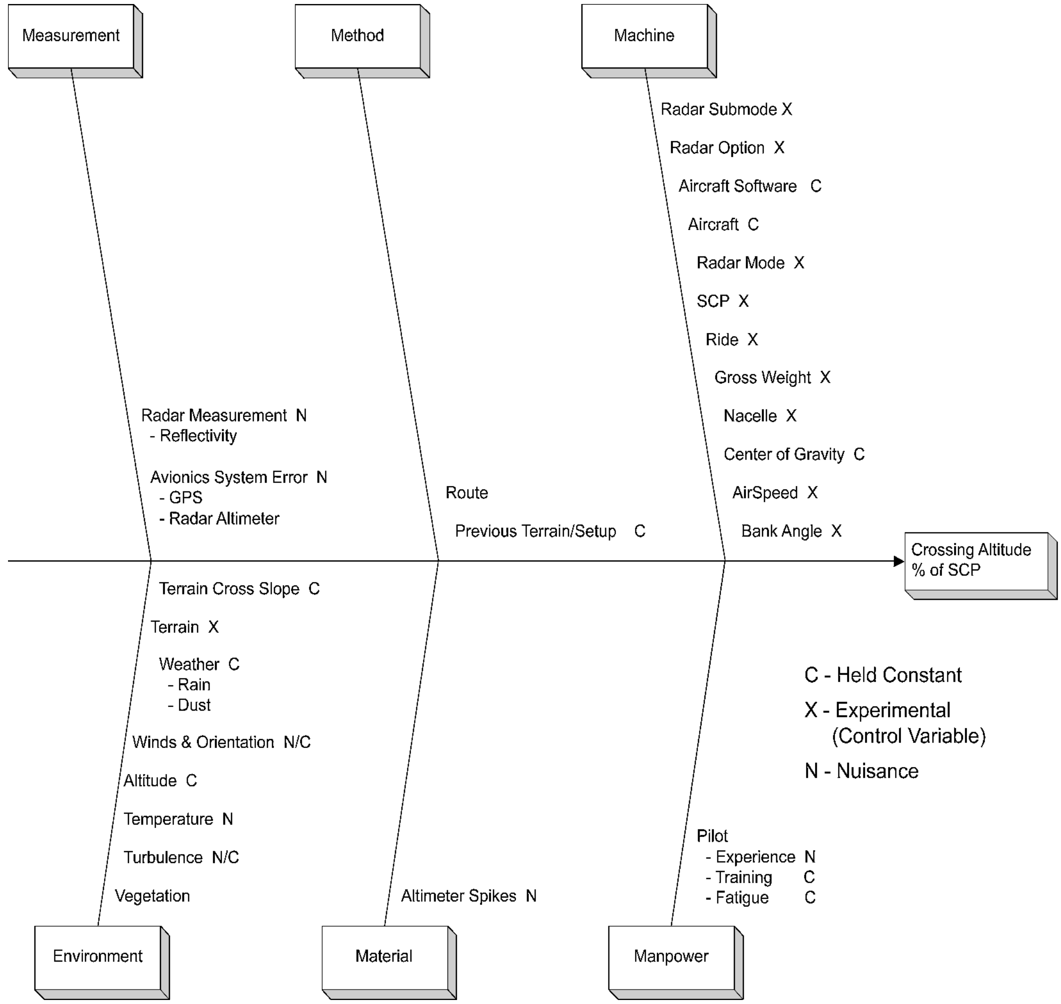

Fig. 3 Cause and effect diagram for the CV-22 TF/TA radar system.

represent missing expertise areas. Senior leadership involvement is also highly recommended. The members of the CV-22 DOE team represented all required expertise areas as well as representation from all interested customers and senior leadership.

The third step is to acquire relevant background information pertinent to the system being studied. Possible relevant information includes estimated system capabilities, requirements, and performances of related systems. The remaining four steps require the team to quantify and define the system inputs and output performance characteristics associated with the strategic objective. The team must first define the output characteristics (or responses) to be studied. The team should attempt to quantify carefully each response and determine the precision associated with measuring these values. Sometimes it is helpful to use scales, ranks, or a scoring system to quantify qualitative responses.

For the CV-22 TF/TA radar system an example quantitative response is the deviation in set clearance plane. The set clearance plane measures the aircraft location relative to the ground. A possible response associated with the goal to evaluate TF/TA radar system performance is the deviation (in percent) between the actual set clearance plane (SCP) and the desired SCP. This measure is referred to as the SCP deviation. To illustrate a qualitative response, suppose the flight-test engineers are interested in knowing the how easily the aircraft responds to pilot input commands in order to maintain a certain SCP. This pilot rating response can be quantified using a scoring system, where each score value is clearly defined.

Once the responses are clearly detailed, the team should focus attention on the input parameters that potentially influence any of the responses. An effective approach for identifying factors is to use a cause and effect (or fishbone) diagram. The team brainstorms all potential factors and groups them according to one of three categories: (X) factors can be controlled in flight test and are of interest in the eXperimental study, (C) factors can be controlled but are not of primary interest and are best held Constant for the study, and (N) factors cannot be controlled in flight test and will be regarded as Nuisance variables during experimentation. The CV-22 radar cause and effect diagram for the deviation in set clearance plane response, showing examples of each factor type, is provided in Fig. 3. Our experience has shown that just going through this process alone with the team significantly improves the flight-test program.

Once all factors are identified and labeled, the team should then provide further detail for each factor. For the hold-constant variables and the nuisance factors determine the measuring technique, level of precision, and any anticipated effects that factor will have on the response. Include a strategy to minimize the effects of nuisance factors. For example, specify criteria for selecting pilots with high experience levels and determine acceptable crosswind velocities to initiate flight test.

More information is needed on the experimental factors because they will be the focus of the test plan. Determine how they will be measured or set and with what precision. Also detail the normal levels and range for each factor and its anticipated effects on the different responses. Finally decide proposed test settings for each experimental variable based on the predicted effects on the response. For qualitative factors list the settings required for study. For quantitative factors determine two settings (called low and high levels) for each X factor, such that the team anticipates a similar and significant change in the response. Table 1 lists the criteria for a set of experimental factors considered in the CV-22 study. Design efficiencies are optimized if all factors have only two levels, but in some instances (including this example) qualitative factors with more than two levels might be necessary.

Table 1a Experimental factors for the CV-22 radar study: qualitative factors

Factor	Normal operating range	Level 1	Level 2	Level 3
Nacelle angle	0–97.5 deg	Airplane 0 deg	Conversion 60 deg	Helicopter 85 deg
Terrain	Flat, rolling, isolated peak	Flat	Rolling	Isolated peak

Table 1b Experimental factors for the CV-22 radar study: quantitative factors

Factor	Normal operating range	Low setting (−)	High setting (+)
SCP	100 ft; 1,000 ft	300 ft	500 ft
Ride setting	Medium, hard	Medium	Hard
Turn rate	0–5.5 deg/s	0 deg/s	5 deg/s
Airspeed	0–300 kn	——	——
Airplane	——	160 kn	230 kn
Conversion	——	80 kn	120 kn
Helicopter	——	40 kn	70 kn
Gross weight	32,000–60,500 lb	47,500 lb	55,000 lb

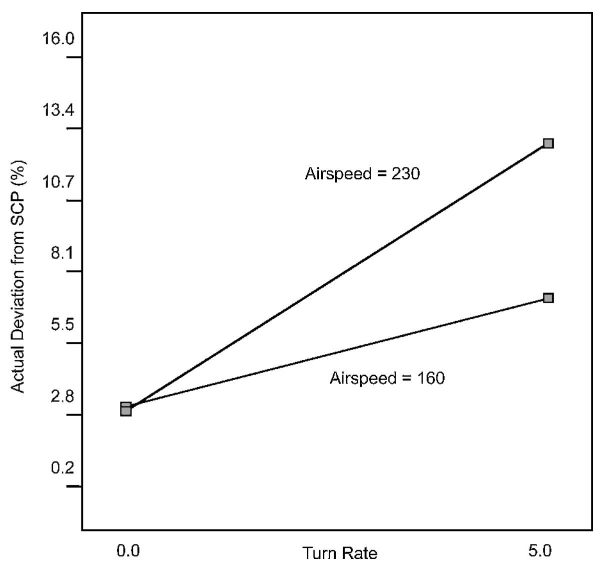

Fig. 4 Interaction graph of turn rate and airspeed vs SCP deviation (%) for CV-22.

Following the specification of each input factor, the group should then discuss anticipated response effects from pairs of inputs. The combining of inputs to affect the response is called a *two-factor interaction* and is very common in actual systems. A two-factor interaction is defined as the failure of one factor to have the same effect on the response at different settings of another factor. Consider the CV-22 radar system evaluation. The SCP deviation response might be affected by an interaction between airspeed and turning rate (Fig. 4). At low airspeed increasing the turning rate might have only a small impact on SCP deviation. However, at high airspeed increasing the turning rate might greatly affect SCP deviation. Identifying potential interactions prior to experimentation can aid determining appropriate input combinations for flight test. However, DOE analysis of flight-test data will reveal which interactions significantly impact performance even if you have not considered them beforehand. In virtually all physical systems two-factor interactions significantly impact the response variable, and they are critical to system understanding. Traditional flight-test analysis and OFAT often overlook these important factors.

The final step in phase I planning is to discuss test conduct procedures and potential flight-test restrictions as a result of safety concerns or physical limitations. Decisions should be made regarding the actual collection of response data on a flight, how often data points should be collected, ease of configuration of the aircraft for each data point, and compatibility between the requirements of the experimental factor settings and the flight-test range natural conditions. Primary safety considerations include sortie length, and flight-control envelope expansion. The proposed flight-test plan will include many smaller groups of test runs. Each group will be sequenced to gradually increase the performance envelope characteristics of the aircraft. For the CV-22 initial tests will primarily be straight and level flight at modest speeds with a high SCP over flat terrain. Later tests will expand the flight envelope to test the aircraft's ability to satisfy operational requirements.

Once the planning procedure has been executed for the strategic objective, the group can concentrate on generating detailed flight objectives and test matrices for each aspect of the program. The following section describes this phase of the planning effort.

IV. Phase II: Developing Detailed Test Objectives

One approach to planning a flight test is to determine all possible operational flight configurations needed to evaluate and generate a single, and usually massive, test matrix. Our recommended alternative approach requires planners to determine specific flight-test objectives associated with studying various aircraft subsystems or certain aspects of operational requirements. This approach allows the team to plan each detailed objective as a separate series of tests and forces them to assess the potential relevant factors and responses. This process also allows for experimenting using small sets of test points, learning progressively about the system so that decisions on future experiments can be influenced by knowledge gained from previous tests. This experimentation approach is often called *sequential testing.*

The first step is to define the relevant specific objectives. Members of the team with significant experience in related flight-test programs are extremely beneficial in this effort. Some of the CV-22 objectives are listed here: 1) quantify the peak crossing altitude over three basic terrain types in maximum turn rate (MTR) radar mode; 2) qualify SCP changes; 3) qualify altitude letdowns into TF flight; 4) expand the envelope of the radar system for high isolated peaks and rough terrain; and 5) qualify smooth TF operation through radar mode changes. The first objective involves assessing the radar system performance as the aircraft passed over isolated peaks in three types of terrain: flat, rolling, and rough.

For each detailed objective the team should follow the iterative procedure outlined in Fig. 5. Determine the appropriate response variables. Then identify the input nuisance, hold constant,

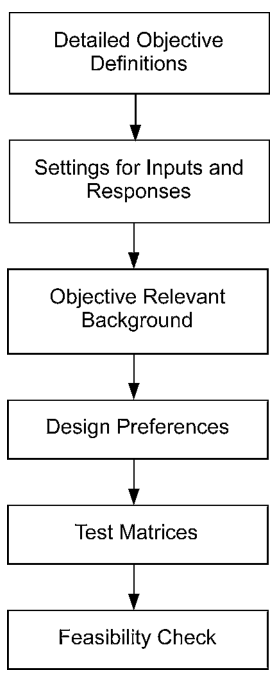

Fig. 5 Sequence for phase II: developing detailed test objectives.

and experimental factors. Finally, set the levels for each of the experimental factors.

All potentially interesting response variables should be listed and then arranged to be collected during flight test. Settings for the experimental factors will depend upon the stage of testing. More conservative levels are typically selected for the earlier objectives when safety risks are higher. The selected levels should be wide enough (between the high and low settings) to see a significant difference in the response in the presence of known system variation. Also consider all potential two-factor interactions and determine the most likely to exist for each response.

Unique characteristics will accompany each objective and must be addressed. This specific relevant background will assist in deciding the appropriate tests to perform. Provide substantiation for performing these tests in support of the strategic objective goal.

Prior to determining individual test points, the group should revisit the objective and decide the purpose of the tests relative to the response variables. Determine whether the goal is to identify the physical relationship between input variables (e.g., flight controls) and performance responses, or whether the intent is to determine the input settings such that performance is optimized. For example, the CV-22 team in some cases determined the aim was to identify which experimental factors affect SCP deviation. In other instances the purpose is to determine which factor settings resulted in the largest SCP deviation, meaning the worst possible performance.

The team should now have all pertinent information to build the test matrix. The purpose of these statistically based designs is to test combinations of experimental factor settings that cover the region of possible factors settings efficiently. The resulting designs will consist of well-suited combinations of all possible test-point settings. The suggested approach is to use full or fractional factorial designs consisting of consecutive tests points with simultaneous changes in experimental factor settings. These designs will enable building empirical models of the relationship between the responses and the experimental factors. The models will include any factors that individually have a significant effect on the response and any two-factor interactions that are also important. Higher-order interactions (three-factor and above) typically do not exist to a significant degree in real-world systems (see, e.g., Montgomery[2]), and so the ability to estimate these terms, although possible, is generally not necessary. This assumption that real-world systems are driven primarily by a subset of main effects and two-factor interactions is called the *sparcity of effects principle.* Models can then be used to estimate or predict response values for given input settings.

Factorial designs typically contain experimental factors set at only two levels: low and high. In some cases the experimental factors are qualitative in nature and cannot be easily set at two levels. For the CV-22 study terrain type with its flat, rolling, and rough levels is a qualitative experimental factor. Other factors such as airspeed and turn rate are quantitative and can be reasonably set at low and high levels.

To determine the actual test points to fly, the team considers the number of experimental factors and associated levels, the specific relevant background information, and the purpose of the test in terms of the desired level of understanding regarding the inputs and response. The resulting design matrix (list of test conditions for each flight) is usually a factorial design structure. A full factorial design contains all possible combinations of experimental factor settings, whereas a fractional factorial design is a specially selected subset of the full factorial. Suppose the team decided to develop a design for five experimental factors, each with two levels. A full factorial consists of 2^5 or 32 test points. Only a subset of these points is required to determine which main effects (individual factors) and which two-factor interactions affect the response. In fact, only 16 test points are needed (half fraction) to estimate all main effects and all two-factor interactions. Fractional factorial designs have effect and interaction aliasing. Aliasing means that the algebraic computations required to estimate the effect of a factor or interaction term are exactly the same computations required for another factor or term. Therefore, it is unknown which factor/interaction is contributing to the change in response variable. Fortunately, shrewd aliasing in the design process and sparcity of effects often suggest which term is significant. In this example design each main effect is aliased (paired) with a four-factor interaction. Each two-factor interaction is aliased with a three-factor interaction. So although terms are aliased, sparcity of effects enables us to determine which term in the group is most likely important, which allows us to reduce the number of runs required. Note that the relative efficiencies given in Fig. 1 are comparisons of OFAT against full factorial designs. Therefore, for this five-factor fractional factorial the relative efficiency is six over an OFAT.

These candidate designs arrange the design points efficiently over the region of all possible points so that each factor is set at each level the same number of times. A similar approach is available for factors with more than two levels. Easy-to-use computer programs (e.g., StatEase Design Expert, MINITAB, SAS JMP) help determine the best design for a given number of design points. In determining the appropriate test matrix it is also important to replicate the same setting of experimental factors in each set of tests. Replication allows for noise estimation, the amount of inherent variability in the system. This allows for a more precise estimation of how the input factors contribute to the response variable.

Example: To illustrate many of the DOE principles mentioned to this point, consider a test for the response variable SCP deviation as a function of three input factors: turn rate, ride mode, and airspeed. The CV-22 radar system allows the pilot to select medium or hard ride mode to indicate the flight-control responsiveness. A full factorial with two replicates (shown in Table 2) is performed; the data for the 16 test points are collected and analyzed. The resulting empirical model is

$$\begin{aligned}\text{SCP deviation}(\%) = {} & +6.51\\ & +3.46 \times \text{turn rate}\\ & -1.08 \times \text{ride mode}\\ & +1.39 \times \text{airspeed}\\ & +1.46 \times \text{turn rate} \times \text{airspeed}\end{aligned}$$

where each variable is coded so that -1 represents the low level and $+1$ represents the high level.

The equation can then be used to predict SCP deviation for any combination of inputs. From the preceding equation we can see that SCP deviation increases as airspeed and turn rate increase. SCP deviation decreases as the pilot changes to a more responsive ride mode (medium to hard). The coefficients represent the change in the response per unit change in the corresponding variable. A change from the low level (-1) to high level ($+1$) is a two-unit change. For example, switching from medium ride (-1) to hard ride ($+1$) decreases SCP deviation by $2 \times 1.08 = 2.16$ %. This type of empirical model would not be possible to generate using the OFAT approach. The turn rate × airspeed interaction term, important for predicting the response, cannot be estimated using an OFAT design.

In addition, the estimates of the effects are more precise using DOE. The factorial designs are balanced vertically, meaning each

Table 2 Design matrix for factor settings (actual units and coded units) for example problem (Note that there are eight distinct combinations of factor settings and two replicates for each one resulting in a total of 16 tests points)

Factor setting	Turn rate	Ride	Airspeed	SCP deviation replicate 1	SCP deviation replicate 2
1	0 deg (-1)	Medium (-1)	160 kn (-1)	——	——
2	5 deg ($+1$)	Medium (-1)	160 kn (-1)	——	——
3	0 deg (-1)	Hard ($+1$)	160 kn (-1)	——	——
4	5 deg ($+1$)	Hard ($+1$)	160 kn (-1)	——	——
5	0 deg (-1)	Medium (-1)	230 kn ($+1$)	——	——
6	5 deg ($+1$)	Medium (-1)	230 kn ($+1$)	——	——
7	0 deg (-1)	Hard ($+1$)	230 kn ($+1$)	——	——
8	5 deg ($+1$)	Hard ($+1$)	230 kn ($+1$)	——	——

factor is run the same number of times at each level. A 16-run two-level experiment consists of eight runs, sometimes called hidden replications, at each level. The corresponding estimates of the main effects are based on all 16 observations, whereas OFAT uses only a subset of the experiments to estimate effects.

Strategies for selecting appropriate designs should also include considerations for randomization, noise factors, and nested factors. Randomization refers to the actual testing order for the different combinations of factor settings. We suggest performing the test points from a test matrix in random order if possible. That is, we would not want to run the 0 deg (−1), medium (−1), 160 kn (−1) setting twice in our above example followed by two runs of 5 deg (+1), medium (−1), 160 kn (−1), etc. Instead, we would run the 16 test points in a random order—even at the expense of testing convenience. Randomizing the run order ensures approximate validity of the test results by averaging out the effects caused by nuisance factors not controlled that might be present in the tests. Randomization also helps satisfy some of the empirical model assumptions. Adding noise factors to a design can sometimes enhance system understanding.

Noise factors are typically present in the system but usually are not controlled during operation. Aircraft gross weight is an example noise factor. Designs involving noise factors can utilize the factorial structure, but the analysis should take into account the random variable nature of these factors. The statistical software programs just mentioned can properly analyze designs with noise factors.

Nested factors can be identified if levels of one factor are similar but not identical for different levels of another factor. For the CV-22 the airspeed factor is nested with nacelle angle. Low and high airspeed levels depend on whether the aircraft is flying in helicopter or airplane mode. With nested factors running separate designs for each combination of nested factors is a reasonable approach. Details on these issues and factorial and fractional factorial designs are available in designed experiments texts including Montgomery,[2] Mason et al.,[8] and Box et al.[9]

Once the appropriate design is selected, each test point must be examined to determine whether that combination of factors is feasible and within safety tolerances for that series of flight tests. If one or more test points are deemed infeasible, alternatives must be considered. One option is to ignore infeasible points and plan to analyze only the remaining data. If a fractional factorial design is used, a better option is to investigate another fraction of the full factorial and see if the number of infeasible points decreases.

V. Phase III: Planning Flight Testing and Analysis

Planning flight testing and analysis are inseparable functions in the third and final phase. Testing will be performed in small batches with analysis immediately following each batch. The prompt analysis will be used to modify as necessary the next series of tests. This process should not imply that the entire test plan is not well known at the outset. However, flexibility is built into the implementation of the plan to allow for a reduction of test points later based on knowledge gained in initial tests. The plan should also accommodate additional tests based on insights not anticipated.

Because testing will be accomplished sequentially, models will be developed after each test matrix is completed, and so analysts can often use subsequent test matrix data to verify and validate the previous test matrix model. Model verification and validation are critical aspects of the process that are often overlooked as a result of the expended test resources. Model verification and validation demonstrate that the empirical model adequately represents the behavior of the true system. The sequential approach to testing clearly supports verification and validation.

An integral aspect of flight-test planning is the ability to integrate successfully ground-based simulators to gain system knowledge at reduced cost and increased safety. Simulators can be used early in testing to assess the level of inherent variability in the responses being measured. Depending on the fidelity of the simulators relative to the actual test environment, the tests can be very informative. Simulators can also be used when system problems arise, and testing can be performed in a safe environment to assess the level of success of various repair efforts. A designed experiment approach using factorial design structures and statistical model building is recommended also for these instances.

Table 3 Test matrix using formal designed experiments for qualifying SCP objective [Separate matrices are built for different settings of terrain type and nacelle angle. Example qualitative factor settings: Isolated peak terrain; airplane mode (0 deg nacelle)]

Test point	SCP, ft	Ride	Turn rate, deg	Airspeed, kn	Gross weight, K lb	SCP deviation, %	Pilot rating (score)
1	300	Medium	0	160	47.5	——	——
2	500	Medium	0	230	47.5	——	——
3	300	Hard	0	230	47.5	——	——
4	500	Hard	0	160	47.5	——	——
5	300	Medium	5	230	47.5	——	——
6	500	Medium	5	230	47.5	——	——
7	300	Hard	5	160	47.5	——	——
8	500	Hard	5	160	47.5	——	——
9	300	Medium	0	160	55.0	——	——
10	500	Medium	0	160	55.0	——	——
11	300	Hard	0	230	55.0	——	——
12	500	Hard	0	230	55.0	——	——
13	300	Medium	5	160	55.0	——	——
14	500	Medium	5	230	55.0	——	——
15	300	Hard	5	230	55.0	——	——
16	500	Hard	5	160	55.0	——	——

Prior to conducting each batch of flight tests, an initial set of flight tests is usually performed to ensure the aircraft and subsystems are functioning properly. These tests can often be performed during the same sortie just before collecting actual test data. Depending on the aircraft and the type of system being evaluated, it may take one sortie or several sorties to collect the information required from a single design matrix (for a single objective). The design matrix is often flown a second or third time in separate sorties if repeatability is a concern. This approach uses a formal designed experiments' strategy called blocking to reduce unwanted sources of system nuisance variability. Blocking is invoked when known factors, not of primary interest, are thought to influence the system. In this example, it is understood that each sortie will be flown under slightly different conditions, but the engineer is not necessarily interested in calculating a sortie effect. As such, each sortie represents a block, and the variability caused by sorties is removed from experimental error used to determine factor significance.

For the CV-22 the design matrices varied in size, but each of the matrices contained 16 or fewer test points. An example test matrix using a formal designed experiment approach for a detailed objective is provided in Table 3. Notice that two of the factors (terrain and nacelle) are varied outside the test matrices, and are essentially analyzed by comparing the model differences across matrices. Sometimes it makes sense to use such an approach. Terrain is not an easy to randomize factor, and so it is kept outside the randomized test matrix. Nacelle is varied outside the test matrix also because conversion to different configurations is time consuming. The inability to estimate the terrain × nacelle interaction is offset by the tremendous savings in flight time. We anticipated easily performing 16 test points per sortie so that in some cases more than one test matrix could be performed on a single flight. The analysis could be performed within minutes after data retrieval so that the engineers would be more informed to make decisions regarding the next flight.

VI. Conclusions

Formal designed experiments methods only enhance traditional flight-test procedures. The intent is to slightly modify traditional planning processes and conduct flight tests so that engineers and program managers can take advantage of objective, verifiable, and traceable empirical models that reveal potential problem area root causes. These models are easily generated when the planning process is modified to facilitate experimental design. This approach leverages detailed objectives and uses factorial designs to generate efficient test plans. Appropriate data are collected during flight test, and real-time analyses are performed during postflight to generate empirical models. This enhanced approach to flight testing has several clear advantages.

First, the structured and systematic planning process bridges the knowledge gap among specialists and managers to generate a flight-test plan that is efficient and instructional. The engineers and system designers provide system capabilities, specifications, and expectations of system behavior under various conditions. Senior leadership provides program direction and guidance on objectives and goals. The aircraft end user specifies the operational requirements. The expert in designed experiments facilitates design construction and provides analysis direction. The entire team develops and shapes the plan.

The DOE approach forces the team to determine system performance characteristics that are relevant, quantifiable, and measurable. Once these response variables are defined and effectively linked to the program objectives, the team can focus on determining the essential input factors to set during test. This process should ensure only the factors of interest are included.

By using sequential testing and analysis with many small test matrices, system discovery and understanding takes place immediately. The knowledge gained is based on objective results from analyzing the key performance characteristics. Insights gained can often lead to anticipating problems before they manifest in flight.

The major benefit in analyzing data from formally designed experiments is that the empirical statistical models, once verified and validated, can be used for estimation and prediction purposes.

These design and analysis methods, as well as the proposed structured planning process, have the potential to significantly improve flight-test operations. As we suggest, only a minor modification to the planning process is required along with having someone familiar with designed experiments provide a test plan and analysis approach that will reveal substantial insight regarding system behavior. The resulting test matrices are more efficient, and the test matrices are specifically tailored to problem source identification.

References

[1] Czitrom, V., "One-Factor-at-a-Time Versus Designed Experiments," *The American Statistician*, Vol. 53, No. 2, 1999, pp. 126–131.

[2] Montgomery, D. C., *Design and Analysis of Experiments*, 5th ed., Wiley, New York, 2001, p. 175.

[3] DeLoach, R., "Applications of Modern Experiment Design to Wind Tunnel Testing at NASA Langley Research Center," AIAA Paper 0713, Jan. 1998.

[4] DeLoach, R., "A Factorial Data-Rate and Dwell-Time Experiment in the National Transonic Facility," AIAA Paper 2000-0828, Jan. 2000.

[5] Possolo, A., Booker, A., and Witkowski, D., "Statistical Models and Error Analysis for Aerodynamic Variables," AIAA Paper 2000-2696, June 2000.

[6] Simpson, J. R., and Koewler, D. A., "CV-22 TF/TA Flight Test Design Plan," U.S. Air Force Academy, TR-01-1, Colorado Springs, CO, April 1998.

[7] Coleman, D. E., and Montgomery, D. C., "A Systematic Approach to Planning for a Designed Industrial Experiment," *Technometrics*, Vol. 35, No. 1, 1993, pp. 1–12.

[8] Mason, R. L., Gunst, R. F., and Hess, J. L., *Statistical Design and Analysis of Experiments*, Wiley, New York, 1989.

[9] Box, G. E. P., Hunter, W. G., and Hunter, J. S., *Statistics for Experimenters*, Wiley, New York, 1978.