

I^2SDS
The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2012-12
July 29, 2012

Test and Evaluation Resource Allocation Using Uncertainty Reduction as a
Measure of Test Value

Eileen A. Bjorkman
Department of Engineering Management and Systems Engineering
The George Washington University

Thomas A. Mazzuchi
Department of Engineering Management and Systems Engineering
The George Washington University

Shahram Sarkani
Department of Engineering Management and Systems Engineering
The George Washington University

Test and Evaluation Resource Allocation Using Uncertainty Reduction as a Measure of Test Value

Abstract

Determining the optimum allocation of resources for tests of Department of Defense (DoD) systems is challenging, primarily due to the lack of an accepted and easily obtained value for test results. Past attempts to quantify test value have focused on prioritization schemes or estimates of cost savings postulated to occur by finding and fixing problems as early as possible. These methods have not gained traction, largely due to difficulties in obtaining cost estimates and historical data. In addition, the use of a cost metric does not capture the true value of DoD testing, which is to reduce technical uncertainty and programmatic risk. We propose a methodology to determine test value by estimating the amount of uncertainty reduction a particular test is expected to provide using Shannon's Information Entropy as a basis for the estimate. We apply the methodology to a notional aircraft case study and simulated large test portfolio to allocate resources using a portfolio of tests for a single decision maker involved in the resource allocation. We conclude that using uncertainty reduction to measure test value is easy to apply, produces results that are intuitively appealing, and produces portfolios that outperform those selected using subjective processes.

Managerial Relevance Statement

Using uncertainty reduction as a measure of test value provides managers with a quantifiable and defensible measure for comparing test options and allocating test resources. Using uncertainty reduction as an explicit test objective also provides an objective basis for test planners to use in designing various test options. Finally, uncertainty reduction can be easily related to risk

management and information regarding the uncertainty of the resulting test data can be incorporated into a decision analysis framework if desired.

I. INTRODUCTION

Developmental and operational test and evaluation (T&E) programs for complex DoD systems typically include hundreds or even thousands of individual test points and can be very expensive. Current DoD test planning processes typically allocate T&E resources using rules of thumb and subject matter expert (SME) judgment informed by tests conducted on previous programs; however, there is no evidence that these processes result in an optimum allocation of test resources. Proposed methodologies for improving the resource allocation process have focused on quantifying test value in terms of cost avoidance or early discovery of problems (for example, see [1] and [2]) and have failed to gain traction within the DoD program management, systems engineering, and test communities.

The resource allocation methods just mentioned suffer from difficulties in calculating cost estimates of reworking system problems postulated to exist and in obtaining historical data to use as a baseline. In addition, the use of a cost metric does not really capture the true value of DoD testing, which is to reduce uncertainty and “provide knowledge to assist in managing the risks involved in developing . . . systems and capabilities” [3]. And test data must provide information with high confidence: system operators need confidence in system capabilities and limitations so they can employ the system properly, and program managers need confidence in information used to make acquisition decisions [4]. Thus, test data may be used as an input to a set of decisions made by a program manager and other DoD stakeholders, but test data are also used to provide information to system users so they may properly operate the system. We obviously desire as much confidence as possible in the data collected, within typical program

constraints such as cost and schedule, but current planning techniques do not explicitly account for data confidence or uncertainty during the planning and resource allocation processes.

Although some papers have discussed the value of a test relative to its ability to reduce uncertainty [5], no technique has been found that develops a general methodology for determining this value and using it to allocate test resources.

In this paper, we describe a methodology to determine the value of a test by estimating the amount of uncertainty reduction a particular test can be expected to provide. The methodology is intended to be used to allocate test resources in a way that optimizes the value of a test portfolio to a single decision maker or multiple stakeholders and decision makers within a resource constrained environment. Section 2 first provides additional background on the DoD T&E resource allocation problem. Section 3 then presents the results of a literature review on existing test planning techniques and technical uncertainty characterization. Section 4 presents the overall methodology, describes a technical uncertainty framework derived from the literature review and discusses the use of Shannon's Information Entropy as the basis for our uncertainty reduction measure. Section 5 applies the methodology to a test portfolio consisting of five distinct tests for a notional aircraft and a simulated large portfolio a comparison to SME-selected portfolios is also made. Section 6 discusses results, conclusions and future work.

II. BACKGROUND ON DOD TEST AND EVALUATION RESOURCE ALLOCATION

As mentioned previously, the major purpose of T&E of DoD systems is to reduce risks. Program managers and other stakeholders desire to collect the needed test data as efficiently as possible, within the resource constraints that exist. A typical test portfolio for a major DoD weapon system includes potentially dozens or even hundreds of individual tests that can include hundreds or thousands of individual test points. Typical tests might include: (1) system

performance (for example, speed, range, weapons accuracy), (2) system effectiveness (for example, the ability to conduct a specific type of mission in realistic conditions), (3) human factors evaluations, and (4) reliability and maintainability. For each test, many different types of data may be collected that are used to verify the system design and validate that the system can be used as intended. For example, one test may measure various aspects of system performance and compare those values to a specification. Another test may determine if the system can operate in a specific environment without electromagnetic interference from another system. At the end of each of these tests, the data collected will be used to determine if the system “passes” or “fails” the test. Other tests are conducted simply to collect data to develop training and operations manuals. For these tests, there is no “pass” or “fail” criteria but simply a need to ensure enough data are collected to properly characterize the system. Regardless of test purpose, there will always be some residual uncertainty after the test due to errors associated with the instrumentation and the equations used to transform measurements into parameters of interest; in addition, the phenomenon being measured may not result from a deterministic process.

We demonstrate some of the issues in allocating DoD T&E resources with a simple example. Assume that two tests must be performed to determine whether or not to buy a system and one test must be performed to characterize the system performance so we can write the operator’s manual; we denote these T_1 , T_2 , and T_3 . Test planning personnel have proposed three alternatives for each test with various sample sizes, measurement precision and accuracy, and costs. These test options are summarized in Table 1. If we have \$6.5M available, then there is no issue; we can afford the most expensive option for each test. However, what if only \$3.7M is available? Is it better to do T_{11} , T_{22} , and T_{31} or to do T_{12} , T_{21} , and T_{31} ?

Table 1. Notional DoD T&E Portfolio

| Test | Test Options | Cost (\$K) |
|-------------------------------|-----------------|------------|
| T ₁ (decision) | T ₁₁ | 500 |
| | T ₁₂ | 1000 |
| | T ₁₃ | 1500 |
| T ₂ (decision) | T ₂₁ | 700 |
| | T ₂₂ | 1200 |
| | T ₂₃ | 1800 |
| T ₃ (characterize) | T ₃₁ | 2000 |
| | T ₃₂ | 2400 |
| | T ₃₃ | 3200 |

Decision analysis techniques may help with this situation, but most program decisions are not based on a single piece of information; in many cases, several pieces of information are assembled from multiple tests to determine whether or not to buy a system, and often the data from one test are used to make multiple decisions regarding whether or not to fix something, or to place a note in an operator manual and so forth. In addition, although possible, it is very difficult to place dollar values or military utilities on the potential outcomes of many tests. We believe a better way to allocate test resources is to use a more direct measure of test value.

This research seeks to solve the problem of allocating resources for a test program in a way that optimizes the value of the tests. This research is not about using the data that have been collected to make decisions, but instead is about how to collect optimum sets of test data given constrained resources (we consider cost as the only constraint). Given the purpose of DoD T&E, we believe the optimum test is one that produces the maximum uncertainty reduction associated with various measurements used to the characterize the system and make decisions.

III. REVIEW OF THE LITERATURE

A. Test and Evaluation Planning and Analysis Techniques

Many approaches are used to determine test strategies that optimize the data collected for an individual DoD test. Statistical design of experiments (DOE) approaches and other traditional

approaches (such as hypothesis testing) are common (for example, see [6]); however, these approaches will not work in all situations. Most developmental tests are “analytic” studies (based on a small number of pre-production units) versus the “enumerative” studies required by traditional statistical approaches [7]. In many cases, no statistical DOE approach can be found that results in an adequate statistical model of a system because many of the assumptions of statistical DOE are not valid in engineering experiments [8]. Criticisms of null hypothesis significance testing are abundant in the literature; see [9] for an excellent summary of these criticisms. To overcome some of these problems, Bayesian approaches have been suggested [10], along with other non-traditional methodologies, such as integer programming [1].

The concept of using uncertainty reduction as a measure of value has been applied to product development activities [11]; however, with the exception of a prioritization scheme proposed by Hess and Valerdi [5], none of the techniques described in the literature explicitly account for reducing uncertainty as a test objective. Some techniques do take uncertainty into account indirectly (for example, confidence and power calculations for a statistical DOE) but none use uncertainty reduction as a direct measure of test quality or value. In addition, the focus of most planning techniques is on optimizing individual tests without regard for other tests competing for the same resources. Wong [12] proposed a decision analysis approach to managing a test portfolio using a value structure that allocated test resources based on subsystem performance uncertainty and sensitivity, but this does not solve the problem of testing the system as a coherent whole. Thomke and Bell [13] developed a model for the optimum number of sequential tests during product development, but the formulation seems best suited for software development and small systems that can be designed, assembled, and reworked quickly when design flaws are found versus complex systems such as modern combat aircraft.

Various portfolio optimization techniques and value measures have been applied to research and development or investment problems other than testing. For example, Joshi and Lambert [14] used an entropy measure (among others) to ensure equitable (instead of just efficient) allocation of transportation projects within a geographical region. They also introduced the idea of using a network to depict project allocation; this idea could be useful in test planning, since the system being tested can normally be depicted in a schematic diagram that resembles a network. Although the concept of equitable allocation of tests has merit in a multi-stakeholder environment, the concept is best applied with a baseline value measure. Greiner, et al. [15] used an Analytic Hierarchy Process (AHP) to provide weights for optimizing an investment portfolio; an advantage to AHP is that it can incorporate both subjective and objective value measures. Although these frameworks and concepts developed for optimizing portfolios of projects may be extended to test portfolios, none of the techniques found in the literature can be applied directly due to the difference in the optimization problem (the need to conduct all tests) and the lack of an explicit value measure for tests.

B. Characterization of Technical Uncertainty

Although there are many types of uncertainty associated with a large DoD acquisition program (technical, stakeholder, political, operational, event, safety, cost, and schedule) the focus of our research is on technical uncertainty, since that is the primary type of uncertainty reduced by a test. Uncertainty as used by the DoD T&E community is generally described by the statistics and measurement standards literature as consisting of aleatory, epistemic, and ambiguity components (for example, see [16], [17], and [18]). Aleatory uncertainty is the unknowable uncertainty that will always exist (hence, the other terms for it, random or precision) while epistemic uncertainty is knowable uncertainty that can be reduced by learning more about a

system (hence, the other terms for it, systematic or bias). Although ambiguity is often discussed in the literature as a separate form of uncertainty, it is in fact really epistemic uncertainty – uncertainty related to lack of knowledge. Therefore, we consider ambiguity as epistemic uncertainty instead of treating it as a third component.

If the uncertainty reduction of a test is to be estimated, the sources of the uncertainty related to the test must be understood. Technical uncertainty arises primarily from two closely related sources: information and measurement; for example, measurements may be used directly, or more commonly, measurements may be transformed through models and other processes into information [17]. Specific sources of technical uncertainty include input and output variables[19]; measurement processes [17]; and uncertainties related to modeling, such as model parameter estimates and model structure selection when more than one model is available [20].

Once technical uncertainty sources are understood, there must be a means to estimate the uncertainty. A wide variety of both statistical and subjective techniques for characterizing and estimating uncertainty are available (for example, see [16], [17], and [21]) indicating that there is no single “best” approach to characterizing uncertainty. As an example, although variance reduction may be a useful measure of uncertainty reduction for some tests, variance reduction is not the purpose of all testing. Some tests are conducted specifically to cause system failure; other tests are conducted to measure a specific parameter and compare the estimated value to a specification or the value produced by another system; and so on.

Information theory may provide a consistent measure of uncertainty based on an important information uncertainty measure known as Shannon’s Information Entropy [22]. Although information theory originally studied the theoretical aspects of communication applications, such as data compression, information theory techniques and the use of entropy as a

metric now have wider application (for example, see [23], [24] [25], [26] and [27]). Information theory is not generally concerned with how a measurement is made, but instead with characterizing the quantity and quality of the information obtained. Shannon's Information Entropy depends only on the underlying probability distribution function of the information (which can be estimated prior to a test or from the test data after the test).

C. Literature Review Summary

Although the primary purpose of DoD test and evaluation is to reduce uncertainty and risk, uncertainty quantification and uncertainty reduction are not currently explicitly included in test planning processes. Most test planning processes also focus only on optimizing individual tests versus allocating test resources to optimize the value of a portfolio of tests. Past attempts at using resource allocation techniques have failed due to the lack of a common and easily obtained measure of the value of a test. Our research seeks to overcome these problems by defining the value of a test as its ability to reduce the uncertainty of the data obtained, regardless of how the data will be used. Shannon's Information Entropy is proposed as the basis for the value measurement since it can be applied to uncertainty reduction across a broad spectrum of tests.

IV. EVALUATING UNCERTAINTY REDUCTION FOR DOD T&E PROGRAMS

The next two sub-sections present a planning methodology and technical uncertainty framework for resource allocation based on estimated uncertainty reduction. The final sub-section describes our selection of Shannon's Information Entropy as the value measure basis.

A. Planning Framework

The planning framework used for this research was adapted from [28]. The first step is to define test objectives for each test within the required portfolio of test points. Step 2 is to identify 2-3 alternative test options for each test. More than three options may be developed, but the effort to

do so must be weighed against the benefit. Three test options will usually be sufficient – a “high-end” option expected to maximize uncertainty reduction, a “low-end” option expected to provide minimal uncertainty reduction, and a “mid-range” option.

During Step 3, the problem is further decomposed and the portfolio cost (or other) constraint is determined. The heart of the planning then lies in estimating the technical uncertainty reduction for each test option in the portfolio (discussed further in Sub-Sections 3.2 and 3.3). At the end of Step 3, the problem is ready for resource allocation. All tests must be conducted, so the technique used must select one test option for each test; i.e., unlike a portfolio of stocks, an entire test cannot be eliminated simply because it does not provide as much value as others. The test portfolio optimal resource allocation is a “multiple-choice knapsack problem” for a single constraint [29] or a “multiple-choice multidimensional knapsack problem” [30] if there is more than one constraint. The general multiple-choice knapsack problem is [29]:

$$\min \sum_{k=1}^m \sum_{j \in N_k} v_{kj} T_{kj} \quad (1)$$

$$\sum_{k=1}^m \sum_{j \in N_k} c_{kj} T_{kj} \geq b \quad (2)$$

$$\sum_{j \in N_k} T_{kj} = 1 \quad k = 1, \dots, m \quad (3)$$

$$T_{kj} \geq 0, \text{ integer} \quad j \in N_k, k = 1, \dots, m, \quad (4)$$

where the multiple-choice classes N_k (the test options) are mutually exclusive, $v_{kj} \geq 0$, $c_{kj} \geq 0$, $j \in N_k$, $k = 1, \dots, m$, and $b \geq 0$. In the context of the test portfolio problem, T_{kj} represents the j^{th} option for the k^{th} test, v_{kj} represents the value of the j^{th} option for the k^{th} test, c_{kj} represents the cost of the j^{th} option for the k^{th} test, and b is the cost constraint. Due to constraint (4), we will

assume none of the tests overlaps with another (i.e., no test points are conducted simultaneously; this is a reasonable assumption for many tests). In addition, equation (1) will be maximized instead of minimized (to maximize test value) and the constraint in equation (2) will be less than or equal to (instead of greater than or equal to) the cost constraint b of the total portfolio.

In Step 4, resource allocation is conducted; this can be done either by enumerating all possible portfolios if the number of options is small or by using an optimization algorithm, such as that provided by [29], if the portfolio is large. Sensitivity analysis is conducted in step 5 if desired and then further analysis is conducted in step 6 if needed.

B. Technical Uncertainty Framework

The framework used to describe the technical uncertainty for each test option was derived from the literature review and is depicted in Table 2. The framework provides a common reference point for discussing technical uncertainty within the context of a T&E program.

Table 2. Technical Uncertainty Framework Used for Research

| | Unknowable Uncertainty | Knowable Uncertainty (Ambiguity) |
|---|--|---|
| Essential Elements of Uncertainty: | | |
| Components of Uncertainty | Aleatory | Epistemic |
| Sources of Uncertainty | Measurement(input/output), model structure, model selection, prediction error, inference uncertainty | |
| Application to Test and Evaluation: | | |
| Test Goal | Reduce Uncertainty | Characterize and Reduce Uncertainty |
| Type of Model Available | Physics-based | None or limited |
| | Empirical | |
| Characterization of Uncertainty: | | |
| Uncertainty Reduction | Model Updating and Using: Using test data to reduce or estimate uncertainty and validate/update model | Model Building: Using data to build model and estimate uncertainty |
| Uncertainty Depiction (not an exhaustive list) | Probability Distribution/Summary Statistics Confidence, Prediction or Tolerance Intervals Credible Interval (Bayesian) Akaike Information Criterion | |

| | |
|-----------------------------------|---|
| | Deviance Information Criterion |
| Test Value/Uncertainty Estimation | Measures based on Shannon’s Information Entropy |

Starting from the top of the table, test planners should first determine the nature and sources of the uncertainty that exist for each test in the portfolio. The dominant category of uncertainty surrounding a particular system aspect, the expected sources of uncertainty, and the type of model(s) available, will drive test goals which in turn provide the linkage to how uncertainty will be characterized, estimated, and reduced. When the uncertainty is mostly aleatory, usually a model is already available and the primary purpose of the test is to use test data and one or more models to reduce uncertainty; in other words, we are “model using” [18]. The model is used to make test predictions, establish initial uncertainty, and plan the test; the model is then updated with test data and used to make inferences. However, when the uncertainty is predominantly epistemic, there often is no model, so the test purpose is to both characterize and reduce uncertainty; in other words, the data are used for both “model building” [18] and “model using.” In this case, a subjective estimate of uncertainty may be established prior to the test, test data will be collected to build one or more models, and inferences are then derived from the model(s). Finally, the uncertainty of test results must be depicted. The choice of technique should be based on test objectives, the analysis techniques used, and decision maker preferences; Table 1 suggests some possibilities. We now turn to the last part of the framework, using Shannon’s Information Entropy as a basis for estimating uncertainties.

C. Shannon’s Information Uncertainty as a Basis for Measuring Test Value

We selected Shannon’s Information Entropy to estimate uncertainty and uncertainty reduction for several reasons. First, as seen from the literature review, there is some precedent for doing so. Second, entropy is easy to calculate for a given probability distribution, can achieve both

positive and negative values [25], and provides values for different probability distributions in a common set of units. Finally, Shannon's Information Entropy meets the desirable properties for an uncertainty measurement (concavity and attaining global maximum at the uniform distribution) [27]. Figure 1 illustrates these properties for a Bernoulli random variable.

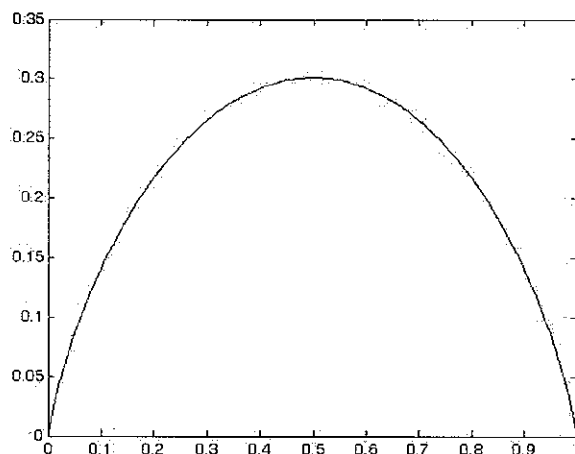


Figure 1. Shannon's Information Entropy for a Bernoulli Random Variable (adapted from Cover and Thomas (2006))

Shannon's Information Entropy is defined for the case of a discrete random variable as:

$$H(X) = - \sum_x p(x) \log p(x) \quad (5)$$

and for the case of a continuous random variable as:

$$h(x) = - \int_S f(x) \log f(x) dx \quad (6)$$

The entropy has a closed form solution for many probability distributions [31]; for the case of a normal distribution, the closed form is:

$$h = \frac{1}{2} \log(2\pi e \sigma^2) \quad (7)$$

Another easily calculated entropy is for the Bernoulli variable depicted in Figure 2:

$$H(x) = -p \cdot \log(p) - (1 - p) \cdot \log(1 - p) \quad (8)$$

We found that equations (5), (7), and (8) are useful for computing uncertainties for a wide variety of test applications, so there is usually no need for complicated calculations. In all cases, the units of the entropy are defined by the base of the logarithm used; a base of 2 results in units of bits (used commonly for communication theory) and a base of e results in units of nats [26]. Base e (nats) will be used for this research for consistency, but the choice of units is entirely arbitrary as long as all uncertainty estimates in the portfolio use the same base.

V. CASE STUDY: THE U-100 UPGRADE PROGRAM

We now apply the proposed methodology to a notional but realistic aircraft (the U-100) upgrade program. Although the U-100 is notional, the data sets are actual military aircraft flight test data sets obtained from the literature or from technical reports in the public domain ([32], [10], [33], [34], and [35]). We also extend the results to a simulated large test portfolio and compare the results of both portfolios to portfolios selected using subject matter experts.

A. Overall Scenario

Several modifications were made to the U-100 that must be tested: upgraded brakes, installation of a radar warning receiver (RWR), upgraded fire control radar, and the addition of an air-to-ground munitions capability. In addition, the operators desire better data on aircraft performance during a condition known as an aerodynamic stall to use to develop a new training syllabus.

B. Application of Planning Process and Technical Uncertainty Framework

Five individual tests in the portfolio are required. The full methodology for one test (the brakes upgrade) is provided in detail here; specific test objectives, uncertainty characterization, and test alternatives for the others can be found in [36].

U-100 Brake Upgrade

The U-100 brakes were upgraded to reduce the landing roll so the aircraft can be operated on shorter runways. The landing roll consists of the phase from touchdown on the runway until the aircraft comes to a complete stop (with brakes applied). Two physics-based models were used to develop initial landing roll distance estimates (see the Appendix for model details).

The overall test objective is to obtain quantitative landing roll data so new landing prediction charts can be incorporated into the U-100 flight manual. Two sub-objectives include: (1) determining the best braking technique (moderate, heavy, or maximum), and (2) determining the maximum landing distance using the best braking technique with a confidence level of 0.99. In updating the physics-based models, the engineers are primarily interested in updating the estimated values for the braking coefficients since they have the highest uncertainty (compared to other parameters such as thrust and drag) and cannot be measured directly during testing; they must be estimated using the physics-based models.

Baseline Uncertainty Estimate

The uncertainty for this test consists of both aleatory and epistemic components. Aleatory uncertainty arises from variability in the landing and braking process and in the precision of the instrumentation available to measure touchdown airspeed, aircraft weight, and roll-out distances. Epistemic uncertainty arises from the use of the two different models since it cannot be determined prior to the test which model will make the better predictions.

A Monte Carlo simulation (9 different test configurations at 10,000 runs each) was used to estimate the baseline uncertainty by using the models to predict landing roll distances under various test conditions (see the Appendix for details). Errors in the landing roll-out data were assumed to be normally distributed, so the entropy was computed using equation (7). Table 3 summarizes the Monte-Carlo simulation results for each model (pooled estimates only). Since

Model 2 has significantly more uncertainty associated with it, the Model 1 uncertainty will be used as the baseline since it will result in a more conservative estimate of uncertainty reduction.

Table 3. Initial U-100 landing uncertainty estimate using Monte-Carlo simulation

| Model 1 σ (feet) | Model 1 entropy, h (nats) | Model 2 σ (feet) | Model 2 entropy, h (nats) |
|----------------------------|---------------------------------|----------------------------|---------------------------------|
| 644.2 | 7.85 | 924.2 | 8.21 |

Initial Test Options

Four test options were initially proposed: Option 1 was a subject matter expert (SME)-designed test (18 test points) using handheld instrumentation to measure the landing roll with a standard deviation of +/- 200 feet; Option 2 was the same test using precise instrumentation with a standard deviation of +/- 50 feet; Option 3 was a full-factorial test (81 points) using handheld instrumentation, and Option 4 was the same full-factorial test using the more precise instrumentation. See Montgomery (2007) for information on full-factorial and other test designs.

Additional Monte Carlo simulations were used to evaluate the desired and expected uncertainty reduction from the tests. First, to establish the desired uncertainty reduction, a Monte Carlo analysis examined the landing distance uncertainty as a function of the possible uncertainty in the braking coefficients. To simplify the analysis, only the full-flap landing condition was evaluated. The average standard deviations from both models (estimated using 10000 simulation runs for each braking coefficient uncertainty) are depicted in Figure 2. At +/- 10 percent braking uncertainty, the heavy and maximum braking landing roll uncertainty is about 300 feet; this is about half the initial uncertainty, so a 10 percent braking uncertainty appears to be a reasonable uncertainty goal for the proposed tests.

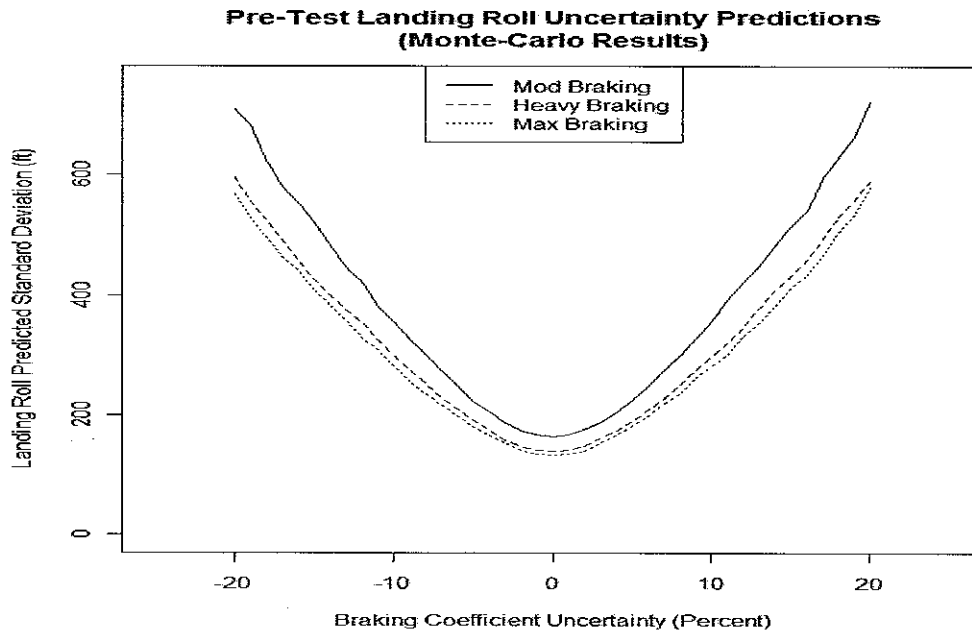
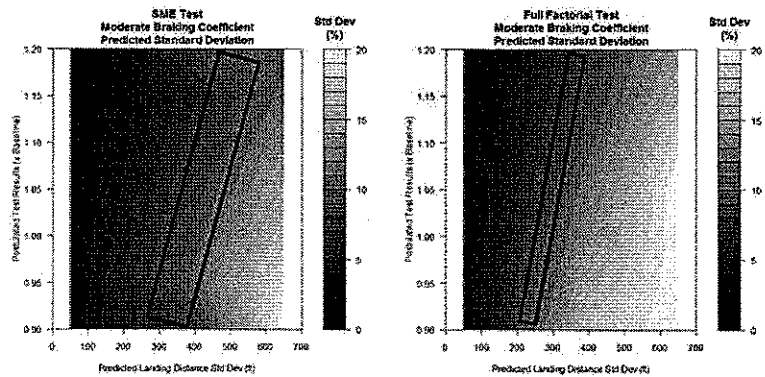


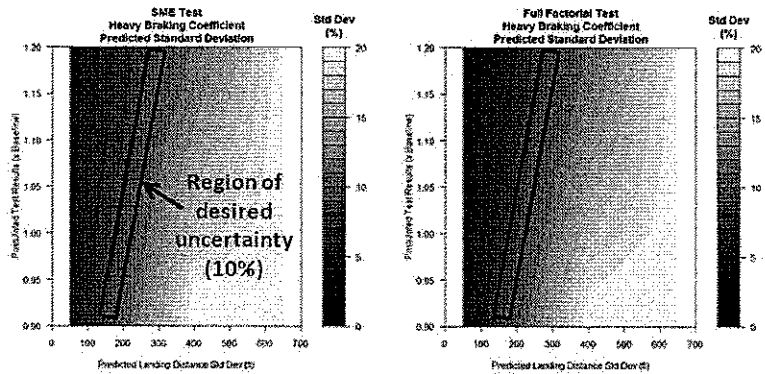
Figure 2. Results of pre-test Monte-Carlo analysis based on braking coefficient uncertainty

The next Monte-Carlo analyses evaluated the ability of each test option to achieve the 10 percent uncertainty desired in the braking coefficients. In addition to the 50 foot and 200 foot test option measurement uncertainties, and the 650 foot baseline uncertainty, a 400 foot landing distance uncertainty was used to capture potential non-linearities. Postulated test results were varied from 90 to 120 percent of the baseline predictions (average predictions from both models) in 5 percent increments. For each landing distance measurement uncertainty, 200 sets of postulated test data were randomly generated and then a maximum likelihood estimation (MLE) technique was used to generate estimated braking coefficients and associated standard errors; model results were averaged for simplicity of depiction.

Test results for moderate and heavy braking, smoothed as contour plots, are shown in Figure 3, with the SME test on the left and the full-factorial test on the right (maximum braking results were similar to heavy so are not included). The x-axis depicts the expected landing distance measurement standard deviation. The y-axis depicts the postulated test results (90-120



(a) Moderate braking uncertainty predictions; heavy rectangles indicate approximate regions of desired uncertainty



(b) Heavy braking uncertainty predictions; heavy rectangles indicate approximate regions of desired uncertainty

Figure 3. Monte-Carlo results comparing SME and full-factorial test designs (percent). The contours depict the predicted standard deviation of the braking coefficient as a percent of the initial braking coefficient; black or near black test results predict braking coefficient uncertainties of about 5 percent or less, gray results predict about 10 percent (the desired uncertainty reduction), and white results predict 20 percent or greater. Overall, the darker the plot, the better the test (the ideal test would have no uncertainty and the entire plot would be black). The SME test, with fewer test points, predicts similar braking coefficient uncertainties compared to the full-factorial test. This is likely due to the non-linear physical models and the MLE technique used. Since the simulation results indicate there is little value in accomplishing the additional full-factorial test points, we eliminate it from further consideration.

Final Test Options

Based on the simulation results in Figure 3, the 200 foot instrumentation will achieve a worst-case braking coefficient uncertainty of 12 percent, which translates to a landing distance uncertainty of about 350 feet (Figure 2, heavy braking), and the 50 foot instrumentation will achieve a worst-case braking coefficient uncertainty of 5 percent, which translates to a landing distance uncertainty of about 200 feet (Figure 2, heavy braking). The uncertainty reduction predictions and estimated cost for each landing test are summarized in Table 4.

Table 4. Summary of initial and predicted uncertainties for landing test options

| Test | Estimated std dev (ft) | Estimated entropy (nats) | Predicted uncertainty reduction (nats) | Predicted relative uncertainty reduction | Cost Estimate |
|-----------------------------|------------------------|--------------------------|--|--|---------------|
| Initial | 650 | 7.91 | N/A | N/A | 0 |
| Option 1 (T ₁₁) | 350 | 7.28 | 0.63 | 0.0796 | \$18,000 |
| Option 2 (T ₁₂) | 200 | 6.72 | 1.19 | 0.1504 | \$28,000 |

Test Portfolio Optimal Resource Allocation

The test portfolio resulting from the five individual tests is summarized in Table 5. The measure used for test value is the uncertainty reduction relative to the initial uncertainty for each test.

While the absolute uncertainty could be used, the relative reduction makes it easier to compare the uncertainty reduction across tests as well as within tests.

To simplify the analysis, we demonstrate the methodology with one resource allocation decision maker with one objective, so the test value was used directly. It is straightforward to extend the case study to multiple decision makers and objectives using techniques such as multi-attribute utility theory or AHP. Using full enumeration, the value and cost of each test combination were computed, and the test combination that generated the highest value was selected for a given cost constraint.

Table 5. U-100 Test Portfolio

| Test | Uncertainty Basis | Estimated Initial Uncertainty (nats) | Test Option | Estimated Uncertainty Reduction (nats) | Test Value: Uncertainty Reduction Relative to Initial | Cost Estimate (\$K) |
|------------------------|--|--------------------------------------|-----------------|--|---|---------------------|
| Brakes | Reducing variance of landing roll | 7.91 | T ₁₁ | 0.63 | 0.0796 | 18.0 |
| | | | T ₁₂ | 1.19 | 0.1504 | 28.0 |
| Radar Warning Receiver | Subject matter expert (SME) estimate from test confidence and power | 16.735 | T ₂₁ | 8.303 | 0.496 | 10.0 |
| | | | T ₂₂ | 8.889 | 0.531 | 18.0 |
| | | | T ₂₃ | 14.627 | 0.874 | 50.0 |
| Fire Control Radar | SME estimate based on p(correct test result) based on predicted pass rates | 0.693 | T ₃₁ | 0.318 | 0.459 | 26.0 |
| | | | T ₃₂ | 0.402 | 0.580 | 36.0 |
| | | | T ₃₃ | 0.422 | 0.609 | 48.0 |
| Stall Performance | Reducing variance in measured stall angle of attack | 2.335 | T ₄₁ | 0.511 | 0.219 | 43.2 |
| | | | T ₄₂ | 1.609 | 0.689 | 71.6 |
| Munitions Performance | Reducing variance in 95% confidence interval of circular error 90 (CE90) | 6.14 | T ₅₁ | 1.03 | 0.168 | 16.0 |
| | | | T ₅₂ | 1.24 | 0.202 | 24.0 |
| | | | T ₅₃ | 1.38 | 0.225 | 32.0 |

Figure 4 shows a plot of overall portfolio value versus cost for all 108 test combinations. As expected, portfolios that cost more generally have more test value; however, for any given cost, there is often a wide range in the portfolio value provided by the test combinations at or near that cost. For example, at a cost of 160, test values range from around 1.65 to about 2.05. Table 6 presents a summary of the optimum (highest value) test combinations selected for cost constraints ranging from \$150K to \$230K in increments of \$10K. To make the table easier to read, the uncertainty reduction listed in the table is a SME judgment of the overall uncertainty reduction provided by the test with respect to operational impact, based on the relative uncertainty reduction, with L=low, M=medium, H=high, and VH=very high. For example, the lower cost landing test option is judged to provide a medium amount of uncertainty reduction because the landing roll distance will likely be known within 350 feet at the end of the test, which is operationally acceptable; however, the higher cost test option is judged to have a high amount of uncertainty reduction (distance known to about 200 feet), which is even better from an

operational perspective. Although the more expensive tests provide the highest overall uncertainty reduction, several mid-priced options also provide good results. For example, with a cost constraint of \$180K, the uncertainty reduction is high for three tests (brakes, radar, and stall) and medium for two tests (RWR and Munitions).

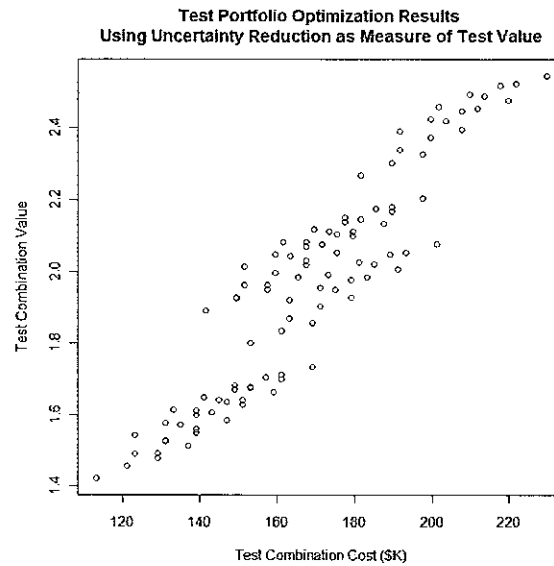


Figure 4. Test Value Versus Cost for All Possible Test Combinations

Table 6. Notional Portfolio Technical Uncertainty Reduction Summary

| Constraint (\$K) | Actual Cost | Test Portfolio Value | Brakes (T ₁) | RWR (T ₂) | Radar (T ₃) | Stall (T ₄) | Munitions (T ₅) |
|------------------|-------------|----------------------|--------------------------|-----------------------|-------------------------|-------------------------|-----------------------------|
| 150 | 149.6 | 1.927 | M | M | M | H | L |
| 160 | 159.6 | 2.048 | M | M | H | H | L |
| 170 | 169.6 | 2.118 | H | M | H | H | L |
| 180 | 177.6 | 2.152 | H | M | H | H | M |
| 190 | 189.6 | 2.304 | M | H | M | H | M |
| 200 | 199.6 | 2.425 | M | H | H | H | M |
| 210 | 209.6 | 2.495 | H | H | H | H | M |
| 220 | 217.6 | 2.518 | H | H | H | H | H |
| 230 | 229.6 | 2.547 | H | H | VH | H | H |

Key: L/M/H/VH represent uncertainty expected from test option selected; L=low, M=medium, H=high, VH=very high

Another way of looking at the uncertainty reduction is to consider the slope of the cost versus uncertainty reduction (see Figure 5). Although the slope of the line varies somewhat, there are two significant “jumps” that occur around \$170K and \$210K. These regions of high or vertical slope indicate that little or no test value is gained for additional investment. So even if the full \$230K is available for testing, the decision maker may want to select the \$210K test combination and use the resources saved for risk reduction in other areas.

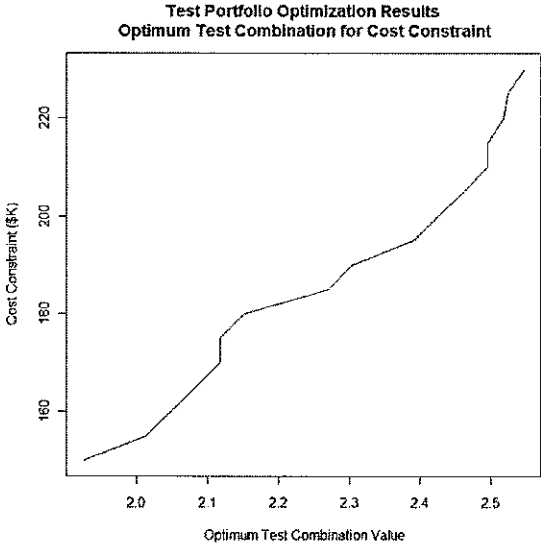


Figure 5. Optimum Test Combination Value Versus Cost Constraint

Scalability

Because large DoD test programs can include dozens or even hundreds of individual tests, this section now demonstrates that the method described above can scale to a large portfolio. Since it was not possible to develop a large test portfolio using a real test program, five large test portfolios with 50 tests each were randomly generated using test values between 0.5 and 0.95 and test costs between \$5K and \$100K (both with uniform distributions); three test options were generated for 22 of the tests and two test options were generated for 28 of the tests.

Figure 6 depicts the results of optimizing the five large portfolios. Each portfolio/constraint combination took less than one second to optimize using a branch-and-bound optimization technique implemented in commercial software. In addition, a global solution was produced in all cases. The optimization methodology was also very efficient in the use of all available resources; the average slack (unused resources) across all five portfolios was 3.1K.

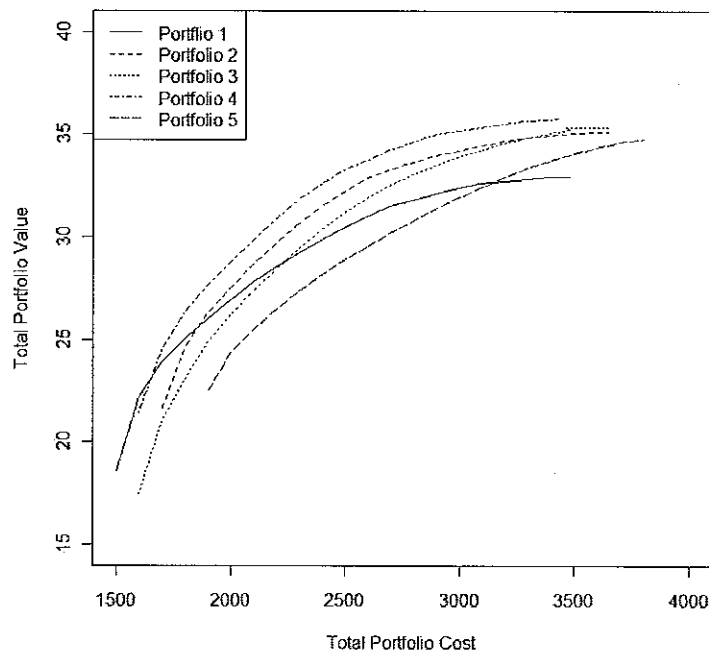


Figure 6. Efficient frontiers for large portfolios

Comparison to Portfolios Selected by Subject Matter Experts (SMEs)

We also compared the results of the optimized portfolios to portfolios that might have reasonably been generated by SMEs (current DoD planning process). Since it was not possible to have actual SMEs select tests, the results were simulated. For the small U-100 case, the SME selected portfolio was simulated by simply using the tests that were actually conducted in the literature. Since the radar test was not actually conducted, three sets of results were examined, one for each potential radar test. Another issue was that the actual stall test conducted had twice as many points as the highest cost test that was in our analysis. This was handled by adjusting

the test cost for the highest cost test to \$93.2K to account for the additional test points; the uncertainty reduction value was left the same, since it was based on the use of the on-board instrumentation, not the number of test points. SME selection results are shown in Tables 7 and 8. Table 7 shows the value and cost of the three SME portfolios given the different radar tests. Table 8 compares the SME results to the optimized portfolio that would have been selected at the same cost of the corresponding SME-selected portfolio. The value of the optimized portfolio ranges from 4.3 to 8.4% higher than the SME-selected portfolio for the same cost. Although this is a small difference, the percentage difference is highest for the lower-valued portfolio, suggesting that optimization is more important in a resourced constrained environment (unless the resources are so low that only the lowest valued tests will be chosen).

Table 7. SME U-100 portfolio selections

| Test | Cost | Value | Cost | Value | Cost | Value |
|-----------|-------|-------|-------|-------|-------|-------|
| Brakes | 18.0 | 0.080 | 18.0 | 0.080 | 18.0 | 0.080 |
| RWR | 50.0 | 0.874 | 50.0 | 0.874 | 50.0 | 0.874 |
| Radar | 26.0 | 0.459 | 36.0 | 0.580 | 48.0 | 0.609 |
| Stall | 93.2 | 0.689 | 93.2 | 0.689 | 93.2 | 0.689 |
| Munitions | 16.0 | 0.168 | 16.0 | 0.168 | 16.0 | 0.168 |
| Total | 203.2 | 2.270 | 213.2 | 2.391 | 225.2 | 2.420 |

Table 8. Comparison of SME and optimized U-100 portfolios

| Portfolio | Radar Low | Radar Med | Radar High |
|-------------------------------------|-----------|-----------|------------|
| Cost | 203.2 | 213.2 | 225.2 |
| Optimized Value | 2.461 | 2.495 | 2.524 |
| SME Value | 2.270 | 2.391 | 2.420 |
| % Difference Optimized to SME Value | +8.4% | +4.3% | +4.3% |

The SME-selected tests for one large portfolio (portfolio 1) used two methods: (1) randomly selecting test options for the 50 tests using a uniform distribution (so expensive, high-value tests were as likely to be selected as inexpensive, low-value tests); and (2) allocating resources to sub-portfolios and then optimizing the individual sub-portfolios. The first method

was not intended to imply that SMEs randomly select tests, but given that they are normally trying to trade off higher priority tests for lower priority tests, in a large portfolio the result is likely to resemble that derived by a random process when completed. The second method resembles the way tests are often allocated in the primary author's experience: a sub-portfolio of tests in each discipline is allocated resources based on past experience and then SMEs try to optimize each sub-portfolio. Since SMEs are more likely to get closer to the true optimum test in a small portfolio, for this method each sub-portfolio was optimized as a knapsack problem.

Two portfolio selections were run for the first method, and a single portfolio was run for the sub-portfolio approach. \$2500K was allocated across 10 sub-portfolios with 5 tests each. The sub-portfolio allocation was close to the mean cost for test options within that sub-portfolio.

The results of the SME-selected portfolios are summarized in Table 9. For the portfolios selected using Method 1, the cost constraint was assumed to be \$2486K, since that was the mean value between the lowest-cost portfolio (\$1493.4) and the highest-cost portfolio (\$3477.7) so was a reasonable "constraint" for the simulated SMEs to obtain. The results show that the optimized portfolios outperform the simulated SME portfolios. For Method 1, the optimized portfolio has about 20% more value than the SME-selected portfolios. In addition, the SME portfolios are inefficient in resource allocation, both with slack values around \$50K. SME Method 2 fares somewhat better, due to the sub-portfolios being optimized separately. However, the overall optimized portfolio is still nearly 7 percent higher in value; in addition, the slack in the SME portfolio is quite high, over \$100K.

Weighted Portfolio

It is straightforward to extend the case study to multiple decision makers and objectives using utility theory. For demonstration purposes, rather than develop a set of utility functions for

postulated decision makers, a simple set of weights was used to simulate a utility function. For the simulation, it was assumed that the landing and RWR tests are more important than the other

Table 9. Comparison of SME and optimized large portfolios

| | | | Optimized Portfolio |
|---------------|--------|--------|---------------------|
| SME Method 1A | Cost | 2435.7 | 2484.9 |
| | Slack | 50.3 | 1.1 |
| | Value | 24.80 | 30.41 |
| | % Diff | +22.6 | |
| SME Method 1B | Cost | 2430.1 | 2384.9 |
| | Slack | 55.90 | 1.1 |
| | Value | 25.23 | 30.41 |
| | % Diff | +20.5 | |
| SME Method 2 | Cost | 2397.1 | 2498.3 |
| | Slack | 102.9 | 1.7 |
| | Value | 28.44 | 30.41 |
| | % Diff | +6.9% | |

Note: % Diff is % difference of optimized portfolio value relative to SME value

tests, since they are related to the safety of the aircraft. Therefore, these two tests were each given a weighting twice that of the other tests. The results are presented in Table 10 for a cost constraint of \$180K. Note that portfolio value has been replaced with portfolio utility, which is somewhat higher than the baseline value since two of the values are now multiplied by 2.

Table 10: Small portfolio results with weighted values, \$180K constraint

| Portfolio | Actual Cost | Portfolio Utility | Brakes (T ₁) | RWR (T ₂) | Radar (T ₃) | Stall (T ₄) | Munitions (T ₅) |
|-----------|-------------|-------------------|--------------------------|-----------------------|-------------------------|-------------------------|-----------------------------|
| Baseline | 177.6 | 2.152 | H | M | H | H | M |
| Weighted | 173.2 | 3.015 | H | H | H | M | L |

Key: L/M/H/VH represent uncertainty expected from test option selected; L=low, M=medium, H=high, VH=very high

At \$180K, weighting does not change the brake test option selected because the high option was also selected for the baseline. On the other hand, weighting the RWR test by a factor of two selects the highest test option. As expected, the weighting comes at a cost to two of the remaining tests, which have lower-valued test options selected. However, the weights ensure that stakeholder preferences are addressed, which may be missed if only the raw values are used.

VI. DISCUSSION AND CONCLUSIONS

Uncertainty reduction, measured using Shannon's Information Entropy, appears to provide a robust measure of test value. In particular, the technique proposed is each to apply and the optimized portfolios can produce between 5 and 20% more total value than SME-selected portfolios. Although this work is very preliminary, it clearly demonstrates the usefulness of uncertainty reduction in a wide variety of test planning situations.

First, as demonstrated in the brake test, using uncertainty reduction as an explicit test objective makes the value of the test options clear so less valuable test options can be eliminated. Eliminating options early in the planning process allows additional resources to be devoted to improving remaining options or in evaluating other options. In addition, the uncertainty contour charts are intuitive and make it clear to even a non-technical decision maker which tests are more valuable than others (or if they are about the same, as in the case study). In any case, decisions to include or eliminate particular test design options from the test portfolio can be defended when uncertainty reduction is used as a basis for comparison.

Second, using uncertainty reduction to measure test value is just the starting point for other potential analyses. We used raw values, but these values can easily be weighted, used as an input to one or more utility functions for use in a decision analysis framework, or used as inputs to other processes. In particular, the network and equity concepts proposed by Joshi and Lambert [14] and the use of a hybrid AHP and portfolio optimization process suggested by Greiner, et al. [15] should be examined as potential extensions to the research.

Third, the uncertainty reduction value can be applied to other activities (such as simulation or laboratory testing) prior to live testing to determine the resources that should be devoted to those activities. The uncertainty reduction value can be used to look for natural

“break points” and determine when additional investment in simulation or laboratory testing has diminishing returns compared to live testing.

Some questions arise in using uncertainty reduction to measure test value. First, how much effort should be invested in computing uncertainty? Not all test programs will have the time, funding, or expertise to conduct a robust uncertainty analysis. However, we suggest that even simple SME estimates to provide a comparative measure of uncertainty reduction across tests can be useful to a decision maker, whether or not those estimates are used to optimize resource allocations. Another issue is what to do if uncertainty goes up at the end of a test? If this occurs, it is likely because of a poor initial uncertainty estimate or improperly conducted test. In either case, an increase in uncertainty at the end of a test should cause a re-examination of all assumptions and prior test results. If the test is correct, then the decision maker must determine if another test is needed, using the test results as the new uncertainty baseline.

Considerable research is still needed. Although the technique was useful for the portfolios examined, additional research is needed to apply the technique to other types of tests and test domains (for example, land, sea, space, and cyber systems). Additional research is also needed regarding the processes used to develop utility functions for the wide variety of stakeholders on a typical DoD acquisition program.

APPENDIX

The first model used in the analysis of braking is described as:

$$S_G = -\left(\frac{WV_{td}^2}{2}\right) / \left(2g(T - D - \mu(W - L))\right) \quad , \quad (9)$$

where

S_G = landing ground roll distance, in ft

W = aircraft weight at touchdown, in pounds (10,800 lbs maximum)

V_{td} = touchdown airspeed, in ft/sec

g = gravity constant = 32.17405 ft/sec²

T = idle thrust = 100 lbs, from prior testing

D = drag, in pounds = $\frac{1}{2} \rho V_{td}^2 S C_D$

ρ = sea level density = 23.77 x 10⁻⁴ slug/ft³

S = reference wing area = 170 ft²

C_D = coefficient of drag, dimensionless, depends on flap setting (see Table 8)

L = lift, in pounds = $\frac{1}{2} \rho V_{td}^2 S C_L$

C_L = coefficient of lift, dimensionless, depends on flap setting (see Table 8)

μ = braking friction coefficient, dimensionless, depends on braking level, to be determined from test (see Table 9 for initial estimates)

Assumptions: dry runway, no slope runway, no wind, data corrected to sea level

The second model uses the same parameters and assumptions as (9) and is described as:

$$S_G = \frac{W}{g \rho S (C_D - \mu C_L)} \ln \left[\frac{T - \mu W}{T - \mu W - \frac{1}{2} \rho V_{td}^2 S (C_D - \mu C_L)} \right], \quad (10)$$

Table 9. U-100 lift and drag coefficients (determined from previous testing)

| Flap Setting (percent) | C_L | C_D |
|------------------------|--------|--------|
| 45 | 0.2840 | 0.0600 |
| 60 | 0.2844 | 0.0725 |
| 100 | 0.2848 | 0.097 |

Table 10. Initial estimates for braking coefficients

| Braking | μ |
|----------|-------|
| Moderate | 0.14 |
| Heavy | 0.20 |

| | |
|---------|------|
| Maximum | 0.22 |
|---------|------|

All uncertainties used normal distributions with mean zero. Touchdown speed standard deviation was 5 ft/sec. All other standard deviations were based on percentages of the baseline parameter values: 1% for weight, lift, and drag; 10% for thrust; and 20% for braking coefficients.

REFERENCES

- [1] L. W. Clarke and L. S. Gardner, "'How much testing is enough?' An iterative, mathematical programming approach," *ITEA Journal*, vol. 16, pp. 26-30, 1995.
- [2] R. Lenz and L. Gardner, "Risk in weapon system acquisition: A decision support approach to the economics of test and evaluation " presented at the ITEA Conference on the Economics of Test and Evaluation, Atlanta, Georgia, 1997.
- [3] DoD, "DoD Instruction 5000.02, Operation of the Defense Acquisition System," ed. Washington DC, 2008.
- [4] J. M. Gilmore, "Rigor and objectivity in T&E," *ITEA Journal*, vol. 31, pp. 161-164, 2010.
- [5] J. T. Hess and R. Valerdi, "Test and evaluation of a SoS using a prescriptive and adaptive testing framework," in *System of Systems Engineering (SoSE), 2010 5th International Conference on*, 2010, pp. 1-6.
- [6] A. A. Tucker and C. H. Dagil, "Design of experiments as a means of lean value delivery to the flight test enterprise," *Systems Engineering*, vol. 12, pp. 201-217, 2009.
- [7] G. J. Hahn and W. Q. Meeker, "Assumptions for statistical inference," *The American Statistician*, vol. 47, pp. 1-11, 1993.

- [8] S. Deaconu and H. W. Coleman, "Limitations of statistical design of experiments approaches in engineering testing," *Journal of Fluids Engineering*, vol. 122, pp. 254-259, 2000.
- [9] R. S. Nickerson, "Null hypothesis significance testing: A review of an old and continuing controversy," *Psychological Methods*, vol. 5, pp. 241-301, 2000.
- [10] A. Hurwitz, W. Kitto, T. Remund, and J. Brownlow, "Estimation of location error for targeting using parametric, Monte Carlo, and Bayesian techniques," *ITEA Journal*, vol. 32, 2011.
- [11] T. R. Browning, J. J. Deyst, S. D. Eppinger, and D. E. Whitney, "Adding value in product development by creating information and reducing risk," *Engineering Management, IEEE Transactions on*, vol. 49, pp. 443-458, 2002.
- [12] P. J. Wong, "Application of decision theory to the testing of large systems," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-7, pp. 379-384, 1971.
- [13] S. Thomke and D. E. Bell, "Sequential Testing in Product Development," *Management Science*, vol. 47, p. 308, 2001.
- [14] N. N. Joshi and J. H. Lambert, "Equity metrics with risk, performance, and cost objectives for the prioritization of transportation projects," *Engineering Management, IEEE Transactions on*, vol. 54, pp. 539-547, 2007.
- [15] M. A. Greiner, J. W. Fowler, D. L. Shunk, W. M. Carlyle, and R. T. McNutt, "A hybrid approach using the analytic hierarchy process and integer programming to screen weapon systems projects," *Engineering Management, IEEE Transactions on*, vol. 50, pp. 192-203, 2003.

- [16] H. W. Coleman and W. G. Steele, *Experimentation, validation, and uncertainty analysis for engineers*. Hoboken, New Jersey: A. John Wiley & Sons, Inc., 2009.
- [17] BIPM, "Guide to the expression of uncertainty in measurement," ed, 2008.
- [18] S. Schrader, W. M. Riggs, and R. P. Smith, "Choice over uncertainty and ambiguity in technical problem solving," *Journal of Engineering and Technology Management*, vol. 10, pp. 73-99, 1993.
- [19] J. R. Wendelberger, "Uncertainty in designed experiments," *Quality Engineering*, vol. 22, pp. 88-102, 2010.
- [20] K. P. Burnham and D. R. Anderson, "Multimodel inference: Understanding AIC and BIC in model selection," *Sociol. Methods Res.*, vol. 33, pp. 261-304, 2004.
- [21] M. T. Pich, C. H. Loch, and A. d. Meyer, "On uncertainty, ambiguity, and complexity in project management," *Management Science*, vol. 48, pp. 1008-1023, 2002.
- [22] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, 1948.
- [23] A. E. Abbas, "Entropy methods for joint distributions in decision analysis," *Engineering Management, IEEE Transactions on*, vol. 53, pp. 146-159, 2006.
- [24] B. Ronen and R. Karp, "An information entropy approach to the small-lot concept," *Engineering Management, IEEE Transactions on*, vol. 41, pp. 89-92, 1994.
- [25] P. S. Abel and N. D. Singpurwalla, "To survive or to fail: That is the question," *The American Statistician*, vol. 48, pp. 18-21, 1994.
- [26] T. M. Cover and J. A. Thomas, *Elements of information theory*, Second ed. Hoboken, New Jersey: John Wiley & Sons, Inc., 2006.

- [27] N. Ebrahimi, E. S. Soofi, and R. Soyer, "Information measures in perspective," *International Statistical Review*, vol. 78, pp. 383-412, 2010.
- [28] R. T. Clemen and T. Reilly, *Making hard decisions with DecisionTools(R)*, Second ed. Pacific Grove, California: Duxbury Thomson Learning, 2001.
- [29] P. Sinha and A. A. Zoltners, "The multiple-choice knapsack problem," *Operations Research*, vol. 27, pp. 503-515, 1979.
- [30] M. Hifi, M. Michrafy, and A. Sbihi, "Heuristic Algorithms for the multiple-choice multidimensional knapsack problem," *The Journal of the Operational Research Society*, vol. 55, pp. 1323-1332, 2004.
- [31] A. V. Lazo and P. Rathie, "On the entropy of continuous probability distributions (Corresp.)," *Information Theory, IEEE Transactions on*, vol. 24, pp. 120-122, 1978.
- [32] J. D. Brownlow, W. Kitto, M. Dorado, and P. E. Kladitis, "F-16 ALR-69A testing at the Benefield Anechoic Facility using design of experiments," presented at the U.S. Air Force T&E Days, 5-7 February 2008, Los Angeles, California, 2008.
- [33] C. J. Naddy, J. P. Hines, G. W. Knowlan, D. G. Riley, Z. T. Schaffer, and T. J. Spaulding, "C-12C parameter estimation: High angle of attack and stall test and evaluation (HAVE HASTE)," USAF, Ed., ed. Edwards AFB, 2010.
- [34] L. F. McNamar and H. C. Gordon, "T-38A Category II Performance Test," Air Force Flight Test Center, Edwards AFB DTIC AD-425650/AFFTC Technical Documentary Report No. 63-27, 1963.
- [35] M. B. Ringel, D. H. Mooney, and W. H. Long, "F-16 pulse doppler radar (AN/APG-66) performance," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-19, pp. 147-158, 1983.

- [36] E. A. Bjorkman, "Test and evaluation resource allocation using uncertainty reduction as a measure of test value (unpublished)," George Washington University, Washington DC, 2012.