Flight Test Point Optimization Program for a Self-Protection Application

Oscar T. Klempay[1]

Georgia Institute of Technology, Atlanta, GA, 30332, USA

Planning test points for highly constrained flight tests is a lengthy and iterative process which requires a structured, methodical approach using scientific test and analysis techniques (STAT) including: design theory, multi-objective optimization, and uncertainty analysis. Flight test engineers can spend anywhere from a week to a month determining ideal test points, only to have unforeseen problems arise during the week of testing that can invalidate these points. Genetic algorithms can play a key role in point selection for two common types of tests: model verification and validation (V&V) testing and operational test (OT) design. This paper outlines the methodology behind building a program to quickly identify a set of optimal test points for the trade space. The tool will allow test planners to have confidence in their test point design prior to the test as well as to make on-the-fly adjustments to testing locations during the event based on actual performance. There are a wide variety of parameters captured in the overall evaluation criteria (OEC) that give the planner great flexibility in tailoring the genetic algorithm outcomes for their purpose. The paper will begin by going through the steps behind planning a test and in defining the trade space and underlying uncertainty. Then, it will cover the parameters of the genetic algorithm and future work and recommendations.

I. Nomenclature

SAM	=	Surface-to-Air Missile
RFCM	=	Radio Frequency Countermeasures
MOE	=	Measures of Effectiveness
DOE	=	Design of Experiments
OAR	=	Open Air Range
OT	=	Operational Test
V&V	=	Verification and Validation
CI_{adj}	=	Confidence Interval
n_{adj}	=	Number of trials in binomial experiment
p_{adj}	=	Probability of success on an individual trial
$z_{\alpha/2}$	=	Standard score

II. Introduction

The process of planning test points is an overly constrained, intractable problem that can often lead to suboptimal designs and inefficient, wasteful use of resources without the use of a design framework to scope and optimize the test. For a flight test engineer, the work can be broken up into four main steps: point identification, selection, repetitions and ordering. The primary objective of this research is to investigate the uncertainties of testing and provide clear metrics for the selection of one test point over another. Two types of flight tests will be examined: a model verification and validation (V&V) case and an operational test (OT) design case. In order to analyze each step of the point planning process, the following will be done:

[1] Graduate Research Assistant, Georgia Tech Research Institute, AIAA Student Member.

Table 1 Analysis Techniques and Approaches Used for Each Planning Step

Test Planning Step	Analysis
Point Identification	Bayesian framework applied to a beta distribution to capture uncertainty
Point Selection	Genetic Algorithm with cost function for optimization
Point Repetitions	Adjusted Wald distribution to find confidence intervals
Point Ordering	Scoring combinations of an optimal set of points

III. Background

The test point planning study will be framed using a self-protection scenario. A friendly aircraft is engaged against a ground-based threat utilizing radar and surface-to-air missiles (SAMs) [1]. The aircraft will use radio-frequency counter-measures (RFCM) such as chaff to avoid being hit by the threat. Figure 1 shows an overview of the test. There is a potential threat system being tested against, a self-protection model built from a variety of inputs, and an output of the model showing locations where the aircraft is expected to be hit or missed by the SAM. The trade space is defined as the hit-miss diagram output by the self-protection model that will be used for test planning.

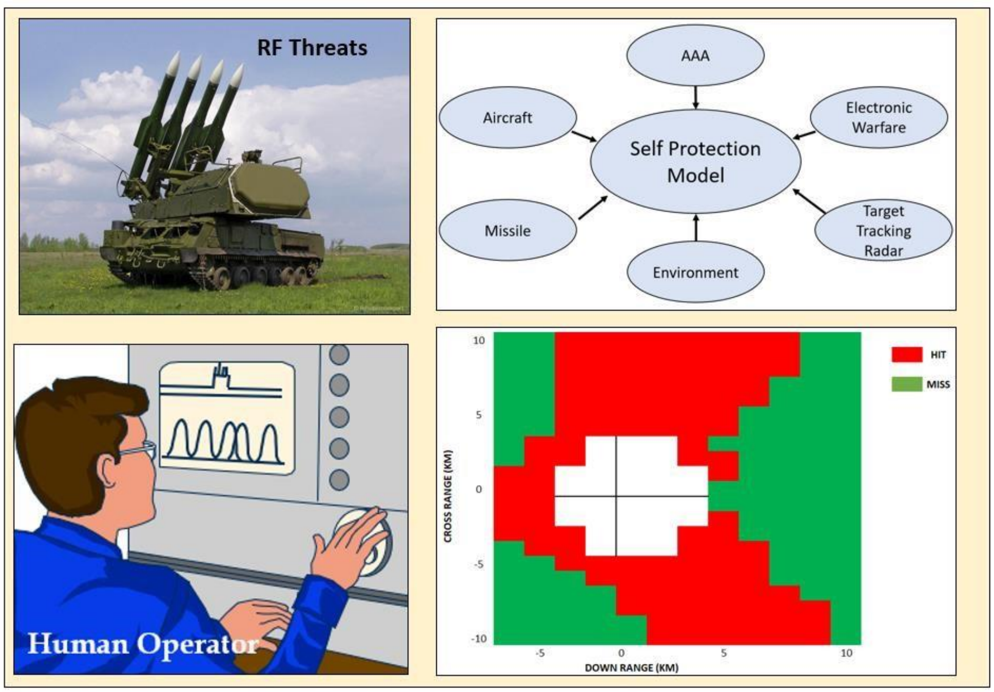


Fig. 1 Overview of Notional RFCM Engagement and Underlying Models

The two types of tests being examined in this paper include V&V and OT design planning. Both tests are related to self-protection. In the V&V case, points are chosen to validate the model's correctness in the hit-miss diagram shown in Figure 1. The goal of the V&V case is simply to validate the model. The OT design case is more involved: a second model would be created that shows the new hit-miss regions when a candidate chaff program is implemented. The difference of the baseline case and candidate case would show how much the hit regions are potentially reduced with the use of the new chaff program. One goal of the OT case is to test at locations where there is a difference in model output between the old and new case to ensure that the locations are correctly identified. Another goal is to measure the effectiveness of the candidate chaff program over the baseline through a variety of parameters like reduction in lethality and chaff bundles expended.

For the V&V test case, the predictive models will go through an iterative cycle to bring them closer to accurately modeling the real world. This helps validate the models such that the overall operational effectiveness is proven: not just for a select few OT points but for the whole model. A study done for the U.S. Army used this iterative approach in creating the Helicopter Mission Survivability Model (HELMS) [2]. An overview is provided in Figure 2.

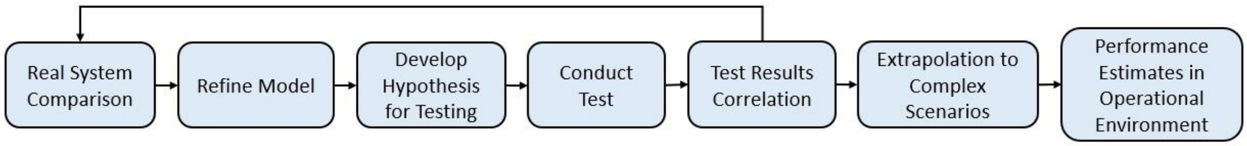


Fig. 2 Cycle of Self-Protection Model Development

IV. Test Planning

Planning a test for high-cost military systems is typically done months before the actual test event. These tests are driven by a need to gather experimental data that a computer model can't provide. Figure 3 provides a rudimentary outline of the steps that go into planning a test [3]. This paper will focus on the final step of the outline: test plan development. These steps can be applied to the RFCM example and two test types mentioned in the background section. The following subsections discuss each step.

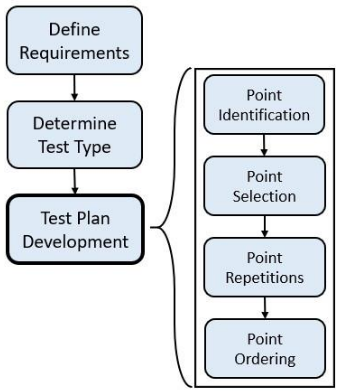


Fig. 3 Test Planning Hierarchy

A. Point Identification

Point identification is a balance between understanding the expected results and knowing the measures of effectiveness (MOEs) going into the test. Test planners need to ensure that selected test points are consistent with the original goals of the test. A large portion of point identification is recognizing which points can and cannot be tested at due to external constraints. Table 2 shows a list of possible constraints that may be imposed on the trade space by planners. These constraints will end up eliminating a large portion of the trade space from being eligible for selection.

Table 2 Trade Space Constraints Limiting Feasible Regions

Trade Space Constraints
Airspace Size
Aircraft Deconfliction
Terrain Avoidance
Protected Habitats/ Environmental Concerns
Climate

B. Point Selection

After suitable regions of the trade space have been identified, individual points within these regions need to be selected. A design of experiments (DOE) approach is used to evaluate points against one another and pick the optimal set within the feasible region. The DOE process needs to manage multiple variables and their possible settings in order to create a draft parameter set table used to build the test design. Table 3 contains an example parameter table:

Table 3 Notional Draft Parameter Table

Altitude (A)	Aircraft Movement (B)	Environment (C)	Range from Threat (D)
Low	Straight and Level	Day	Short Range
Medium	Hot Weave Maneuver	Night	Medium Range
High	Cold Weave Maneuver		Long Range

The distribution of point placement will also vary between the V&V and OT cases. While the goal of the V&V case is simply to validate the model, the OT case has an additional goal of validating the effectiveness of the candidate program being implemented across the entire design space. Table 4 shows different considerations in point locations for the two cases. The relevance of the point patterns will largely influence the optimization problem, which will be discussed later.

Table 4 Differences in Point Placement Considerations Between V&V and OT

Examples of Point Patterns	V&V	OT case
Identifying asymmetrical patterns and opposing locations	X	
Points that *show* a difference between baseline program and candidate program		X
Points that *do not* show a difference between baseline program and candidate program		X
Points on inflection between region of hits and misses	X	X
Points distributed across many tactically relevant aspects (i.e. at least one point in each quadrant)	X	X
Points at consistent range from center	X	X

C. Point Repetitions

Once the optimal set of points have been selected, the test planners need to determine the number of repetitions they will do at each test point. Constraints on repetitions are shown in Table 5.

Table 5 Constraints on Maximum Number of Test Repetitions

Repetition Constraints
Time
Budget
Fuel
Test Expendables
Weather

There is a difference between how repetitions are determined for the two test cases. In the V&V case, planners attempt to reach a target confidence level with the number of repetitions they set. Due to the high expense of flight tests, the sample size for each test point can vary from as few as four to as many as fifteen samples at a point. Because of the limited sample size, it is imperative that the number of repetitions are sufficient to reach the target confidence level. Table 6 compares the inputs and outputs of the two test cases.

Table 6 Inputs and Outputs of V&V and OT cases

Model Type	Inputs	Outputs
Verification and Validation	• Minimum Confidence Level that Results are Accurate at Test Point • # of Needed Test Points	• # of Repetitions **Needed** for Each Point • Expected Test Time • Expected Fuel Amount
Operational Test Design	• Total Test Time Available • Total # of Expendables Available • Total Fuel Amount Available	• # of Repetitions **Possible** for Each Point • # of Possible Test Points • Expected Confidence Level for Each Point

D. Point Ordering

After the set of optimal test points and their repetitions have been found, the sequence which the points are tested are determined for the flight pattern. The largest factors for this step are aircraft deconfliction, terrain avoidance, and the time it will take to fly through all of the points. For RFCM testing of chaff, there is an additional aspect to the time constraint: enough time must be allowed for chaff dispersal before that test point can be flown at again.

E. Summary

It is clear that the test planning process is a complicated and overly constrained problem. There are a very limited number of test points available that are trying to answer too many different questions. The four steps of the process require going through multiple iterations to reach the ideal solution. Unfortunately, all of the planning may be for naught when unforeseen situations arise that can invalidate test points or test locations. The following sections will walk through how uncertainty is being captured in the trade space. This will allow planners to have greater confidence in the points they select before a test and quickly identify alternates should they need to during the test event.

V. Defining the Trade Space

Of equal importance to defining the test planning process is to define the trade space that is being studied. Test events are held at open-air ranges (OARs), which are typically located in restricted airspaces around the continental United States. For the RFCM example, the threat will be set up at specific locations within the OAR. The test planning team will select points to test at around this threat based on the self-protection model. The goal is to understand which areas around a threat are truly safe and how unsafe regions can be reduced.

An example trade space is displayed below to clarify the terminology. Figure 4 shows a notional OAR with a threat at the center. The blue triangles represent a test point and the green triangles represent the direction that the aircraft will fly straight-in through the test point. The red circle represents the limits of the missile’s kinematic envelope outside of which is a sure miss.

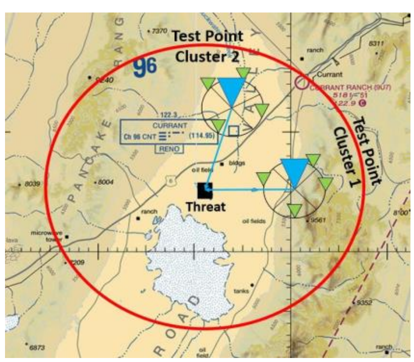
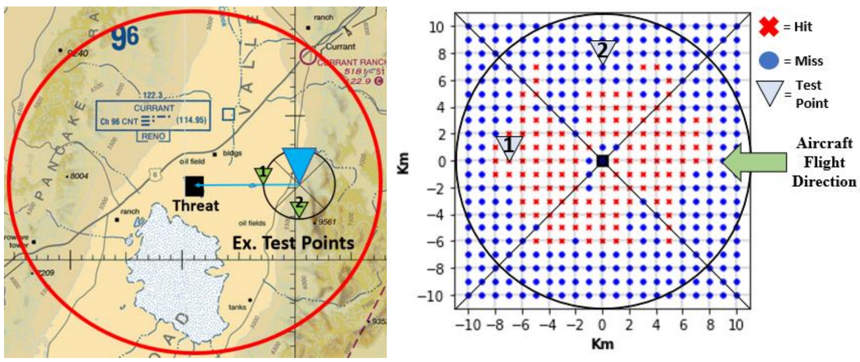


Fig. 4   Notional Trade Space Centered on Threat

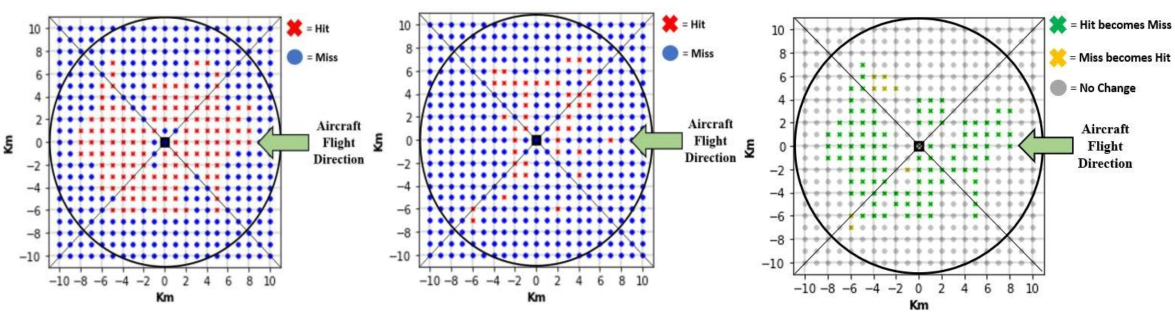
The green triangles are determined by hit-miss diagrams created by the self-protection model. In Figure 5, the blue dots represent miss points and the red dots represent hit points by a SAM. The black box represents the threat. In the hit-miss diagram, the aircraft is assumed to be flying from right to left for reference. The points will then be rotated into the geographic coordinate system as seen in Figure 5. This allows for reduced set-up time in testing at different azimuths against the threat.



(a) Geographic Frame **(b) Model Frame**

Fig. 5   Example of Points in Geographic Reference Frame and Model Reference Frame

These graphs are crucial for allowing planners to go through the point selection and optimization process. When selecting points, it is common to choose locations that are both abeam and fore/aft of the point to fully explore the trade space. As described earlier, an important part of OT testing is understanding how the trade space changes between the baseline and candidate program. The green and orange points in Figure 6 represent a change that occurred in the self-protection model going from the baseline to the candidate case. Note that each of these depictions assumes the same initial ground speed from the fly-in location and a fixed altitude relative to the ground for a given set of environmental and terrain conditions.



(a) Baseline **(b) Candidate** **(c) Difference**

Fig. 6   Notional Model Output Centered on Threat for OT Comparison

A. Uncertainty Analysis and Risk Mitigation

Self-protection models are likely to possess errors and are not a perfect representation of the final results. Thus, test planners must account for this uncertainty in their process. In order to isolate model error, test planners must account for and control test error to the best of their abilities. Figure 7 shows a breakdown of the possible events that may happen when the model determines the outcome of a point.

		Model Prediction	
		Hit	Miss
Experimental	Hit	Correct Decision	Type I Error
	Miss	Type II Error	Correct Decision

Fig. 7 Possible Events that May Occur at each Test Point in Model

Risk mitigation is critical in building confidence in test design and avoiding Type I errors. By controlling test uncertainty through analysis, one additional portion of the experiment is controlled that ensures the data collected will be useable. After the experimentation is done, the model can be validated with test data to uncover errors present and correct them for future tests. This quantified uncertainty will be a useful input for the genetic algorithm. Table 7 provides a brief overview on some of the most prevalent sources of uncertainty in RFCM testing.

Table 7 Inputs and Outputs of Two Test Types

Sources of Uncertainty
Weather
Terrain Clutter
Operator Error (Threat or Aircraft)
Technology
Chaff Dispersal

In order to capture the underlying probability of each point within the trade space, statistical distributions are used to build a "probability grid" that determines which areas of the trade space are more likely to be correct. To construct this probability grid, a beta distribution coupled with Bayes' theorem is applied to the entire trade space. A beta distribution is chosen because the model can self-update over many iterations and move to a more accurate estimate of the true mean with a reduction in variance [4]. Figure 8 shows the flow of the beta distribution. However, this beta distribution is only a placeholder. In future work, a Monte Carlo simulation will be applied to the self-protection model to better estimate the uncertainty distribution of the trade space.

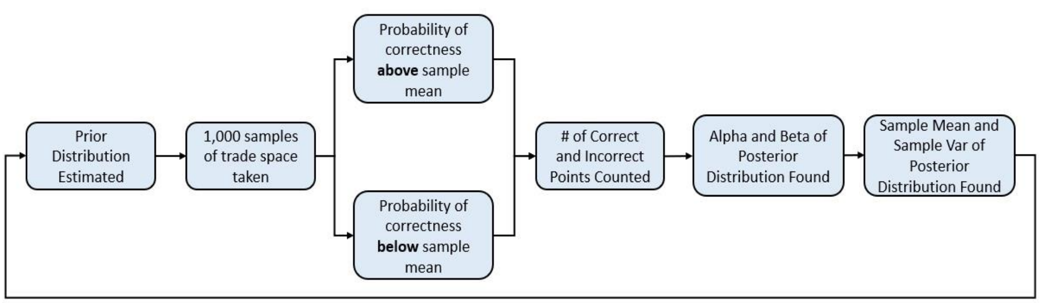


Fig. 8 Flowchart of Beta Distribution with Bayesian Framework

Building an underlying probability grid allows for uncertainty in the trade space to be captured in a succinct and clear manner. It attempts to correct for the fact that the model will not be 100% accurate and that there are some "known unknowns" out there that can only be found through experimentation. Understanding which points are more or less likely to be correct is useful for both the V&V test case and the OT design case, where a test planner would want to know the likelihood that points are correct or incorrectly identified. Figure 9 shows how the probability grid is the foundational piece for having confidence in selecting the right test points.

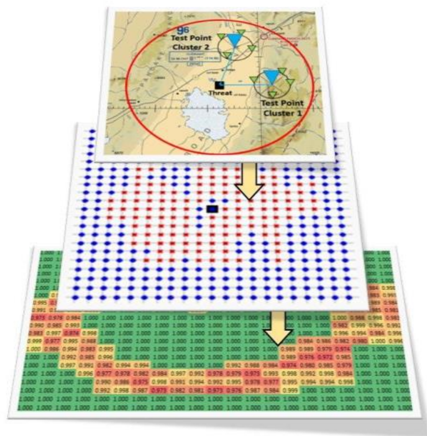


Fig. 9 The Interdependency of Point Selection and the Underlying Probability Grid

VI. Creating the Genetic Algorithm

Point selection is the second step. Several optimization techniques were explored, but the technique that yielded the best result was the genetic algorithm (GA). The GA provided a rapid way to search through the entire design space and avoid getting stuck in local optimum solutions. The following paragraphs will discuss the setup of the GA and the cost function that governs the optimization process.

Figure 10 shows the general outline for how the GA proceeds. Within a particular chromosome, a set of N test points is randomly chosen in a 20x20 trade space and certain metrics (discussed below) captured on the group of N points. From there, the top 10% of chromosomes are selected to reproduce and create the new generation through single point crossover. The top 10% are carried through in each new generation, ensuring that the algorithm keeps improving on point selection. Within each new generation, 0.2% of the generation is allowed to mutate via single point mutation, which prevents the algorithm from getting stuck in a local optimum solution and instead strive for the absolute optimum. The entire process continues to iterate through new generations until the solution converges below a specified limit or the algorithm detects a stagnation in the change in the cost function [5].

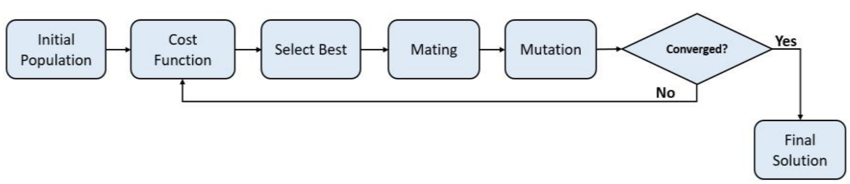


Fig. 10 A Notional Model of Hits/Misses with Underlying Probability Grid

Within a chromosome, each gene is comprised of a single "(x, y)" point. This point represents a value on the trade space. Relevant point patterns for the case under study are identified using the DOE analysis discussed earlier. The number of genes in a chromosome is user-defined. Figure 11 provides an overview on the chromosome setup and the metrics that are captured for a particular chromosome. By gathering data on the genes (i.e. points), the cost function determines which chromosomes are better than others.

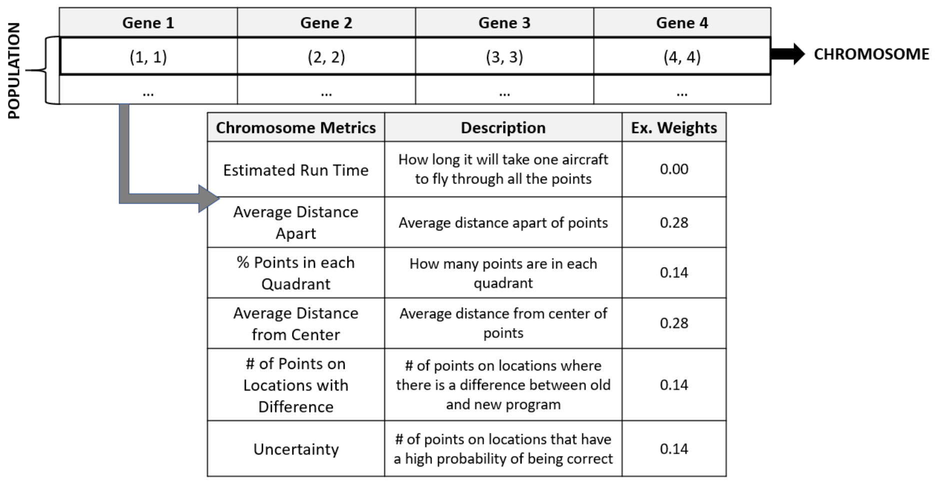


Fig. 11 Breakdown of the Chromosome Structure and Calculated Metrics

The best chromosomes are selected through the cost function using the residual sum of squares difference between a single chromosome score and an optimal set of metrics, which are user-defined. The optimal metrics relate back to the DOE analysis and will differ for the V&V and OT case. A relative weighting scheme, which is also user-defined, is applied to each metric and will greatly affect the outcome of the final solution. When the best chromosomes are selected from the population, single-point crossover is used to generate a new set of offspring that makes the next generation. The best chromosomes are automatically carried through into the next generation. An example of the single point crossover technique is shown in Figure 12:

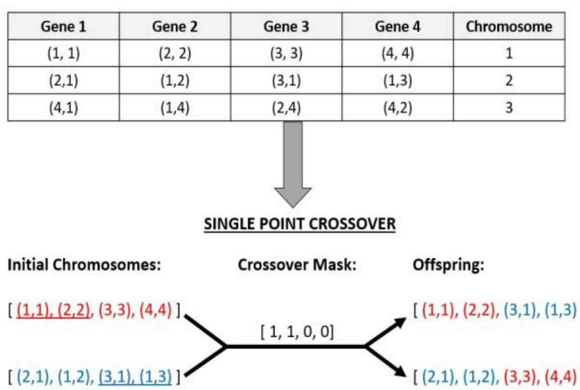


Fig. 12 Using Single Point Crossover to Create Offspring

After the offspring are created, point-mutation is introduced in order to prevent the GA from getting stuck in a local optimum solution. A mutation is considered as changing either the “x” or “y” value of a single gene. Figure 13 shows how the point-mutation is carried out:

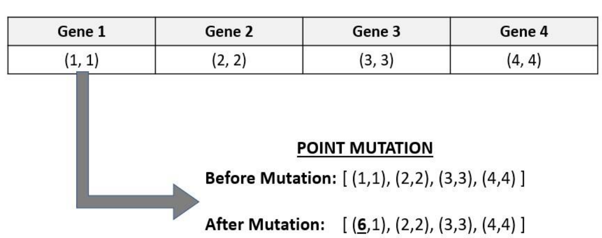


Fig. 13 Using Point Mutations within a Gene

An example output of the convergence behavior of the cost function is shown in Figure 14. More work still needs to be done in weighting each of the error terms and changing their relevancy based on the test case under review. The final output of the GA is a chromosome that contains the N testing points that should be selected.

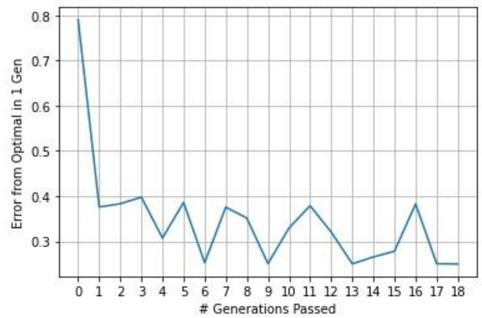


Fig. 14 Number of Generations vs. Cost Function Value of Best Chromosome

A. Determining Point Repetitions

The third step in the test design process involves determining the number of point repetitions. For the case of the V&V test, this repetition number is dictated by a target confidence level. In the OT case, the number of repetitions is instead dictated by the fuel and time available. Despite these differences, the number of repetitions is an important output that is common to both test designs. The following sections will provide an overview on the calculations behind building confidence in test using an adjusted-Wald distribution.

As discussed in the uncertainty analysis section, there is a binomial set of outcomes for whether the point is correct or incorrect in the model. The two most widely known ways to find the binomial confidence interval (CI) are through the Clopper-Pearson Exact method or the Wald method [6]. These are both practical methods when the sample size is fairly small (i.e. less than 30). The adjusted Wald method combines these two to provide the closest CI interval to the true solution and is simple to calculate. Eq. 1 shows how the CI is calculated using the adjusted Wald method.

$$CI_{adjWald} = p_{adj} \pm z_{\alpha/2}\sqrt{\frac{p_{adj}(1-p_{adj})}{n_{adj}}} \tag{1}$$

Table 8 shows the results of setting an 80% confidence level with a historic success rate of 80%. The important metric captured is the CI width and the number successful given a certain number of repetitions. Note that the probability does not directly increase with increasing sample size, but peaks slightly around seven repetitions.

Table 8 Adjusted Wald Calculations for Determining Point Repetitions of a Single Point

Sample Size	Successful	Percent Successful	CI Width
3	2	66.7%	0.290
4	3	75.0%	0.252
5	4	80.0%	0.222
6	5	83.3%	0.198
7	6	85.7%	0.178
8	6	75.0%	0.188
9	7	77.8%	0.173

B. Determining Point Ordering

The final step in the test design process is the point ordering process. Parameters like aircraft deconfliction, total time available, and terrain avoidance are important for figuring out the best ordering. Although further work should be done regarding this phase, the general approach is outlined below.

When the GA identifies an optimal set of test points, it will hand off the set of test points to the ordering algorithm. The ordering algorithm will iterate through the list of test points, setting one test point as the "start point" in each iteration. From this start point, all the combinations of ordering will be scored based on the constraints imposed. If any combination violates a test constraint, the combination will be invalidated. The highest scoring combination among the set of iterations will be deemed the best solution. Note that this ordering is the same regardless of repetition number and assumes a single aircraft test scenario. Figure 15 shows the process.

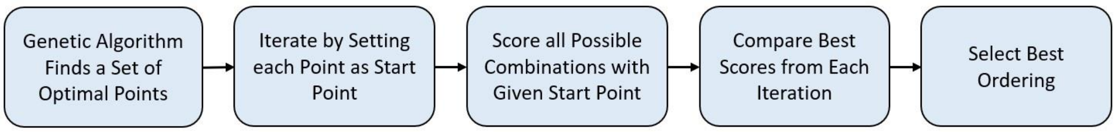


Fig. 15 Flowchart of Point Ordering Algorithm

VII. Future Work and Recommendations

The steps discussed in this paper are intended to be a framework for future research. For one, historical data needs to be gathered to properly validate the results of the GA with real-world data. This will allow better estimates for the weighting schemes in the cost function. Secondly, the "on-the-fly" aspect needs to be addressed. This could allow for adaptive DOE and blocking for each iteration over several days of test. Thirdly, multi aircraft scenarios need to be built into the framework for representation of true flight tests. Implementation of these steps will improve usability, confidence, and adaptability with the test point planning process.

VIII. Conclusion

While the combination of statistics and flight testing is not a new concept, it can be used to improve current metrics in test planning that enable clearer point-to-point comparisons. Design theory and uncertainty analysis are methods that allow for risk mitigation and confidence in point selection. Genetic algorithms can capitalize on the uncertainty analysis to enhance the selection of optimal points. Although there is still work to be done in fleshing out the framework, this application is well on its way to enabling a faster, clearer, and more flexible test design process.

Acknowledgements

The author would like to thank Mrs. Rebecca Douglas for her continued support, guidance, and advisement on the project as well as Mr. Clark Kjorlaug, Mr. Tyler Addison, Mrs. Mahalakshmi Srinivasan, and Mr. Cody Brown for the feedback they provided during development. Thank you to Colonel Anne Clark for assisting with the approval process and working tirelessly to help clear this paper for release. Finally, thank you to the Georgia Tech Research Institute (GTRI) and the Air National Guard Program Office (ANGPO) for whom this research was made possible.

References

[1] StandFair Operations, 2021. *Buk-M2E*. [image] Available at: <https://www.standfairoperations.com/products/air-defence systems/air-defense-systems-and-mounts/buk-m2e/> [Accessed 5 March 2021].

[2] Travis, K., "A Methodology for the Determination of Rotary Wing Aircraft Vulnerabilities in Air-to-Air Combat Simulation", M.S. Thesis, Systems Analysis, Georgia Institute of Technology, Atlanta, GA, 1984.

[3] Reynolds, R., Robinson, G., and Crane, R., "Air Force Flight Test Center Test Plan Preparation Guide", Air Force Flight Test Center, Edwards AFB, CA, Air Force Material Command, USAF, 1999, pp 31-34

[4] Shaw, M. "Use of Bayes' Theorem and the Beta Distribution for Reliability Estimation Purposes", *Reliability Engineering and System Safety*, Vol. 31, 1991, pp. 145-153.

[5] McCall, J. "Genetic Algorithms for Modelling and Optimisation", *Journal of Computational and Applied Mathematics*, Vol. 184, No. 1, 2005, pp. 205-222, DOI: https://doi.org/10.1016/j.cam.2004.07.034 [Retrieved 01 March 2021]

[6] Sauro, J., and Lewis, J., "Estimating Completion Rates From Small Samples Using Binomial Confidence Intervals: Comparisons and Recommendations," *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, 2005.