Identifying On-Orbit Test Targets for Space Fence Operational Testing

Daniel L. Pechkis
Nelson S. Pacheco
Tye W. Botting

Institute for Defense Analyses, Alexandria, VA 22311

ABSTRACT

Space Fence will be an integrated system of two ground-based, S-band (2 to 4 GHz) phased-array radars located in Kwajalein and perhaps western Australia [1]. Space Fence will cooperate with other Space Surveillance Network sensors to provide space object tracking and radar characterization data to support Joint Functional Component Command for Space object catalog maintenance and other space situational awareness needs. We present a rigorous statistical design, using on-orbit objects as radar targets, to test the Space Fence to the letter of the program requirements and to characterize its system performance across the entire operational envelope. The design uses target altitude, size, and inclination as independent factors in statistical tests of dependent variables (e.g., observation accuracy) linked to requirements. Our analysis identifies the type and number of necessary test targets, which includes objects in various orbital and size regimes, objects with high accuracy ephemeris, radar calibration spheres, cubic satellites, and others. Comparing the resulting sample sizes with the number of currently known targets, we determine the areas where modelling and simulation methods are needed. Assuming only a Kwajalein radar coverage and a conservative number of radar passes per object per day, we conclude that tests involving real-world space objects should take no more than 25 days to evaluate all operational requirements requiring on-orbit test targets.

1. INTRODUCTION

Space Fence will consist of two phased-array radar sites with separate transmit and receive antenna arrays, a data processing center for initial orbital determinations, and an on-site local database for managing space object information. The Space Fence radars will cooperate with other Space Surveillance Network (SSN) sensors through the Space Fence Operations Center and provide object tracking and radar characterization data to the Joint Space Operations Center (JSpOC) to support Joint Functional Component Command - Space (JFCC-SPACE) space object catalog maintenance and other space situational awareness needs [2].

The radar system will track space objects, including the growing population of space debris, concentrating on those that pass through the uncued radar fence in low-earth orbit (LEO). Additionally, Space Fence will support cued searches and uncued surveillance at medium-earth orbit and higher [2]. The new system is expected to provide substantial observation accuracy improvements over legacy systems. Capability is also enhanced with respect to catalogued target size (as small as 10 centimeters), timeliness of orbital event information, initial orbital determination, higher cataloging accuracy and completeness, object characterization, and surveillance coverage [2].

We propose selecting on-orbit targets to test the majority of requirements from subsets of all resident space objects (RSO) currently tracked by the SSN and meeting specified orbital limits, as catalogued in the JFCC-SPACE satellite catalog (SATCAT). The SATCAT provides the most comprehensive current set of candidate operational test targets because they are actual objects that Space Fence may be tasked to track. They represent typical orbit types, inclinations, altitudes, sizes, shapes, and rotational motions in their real-world population frequencies. As the Space Fence program matures, the SATCAT will evolve, and the JSpOC may narrow or expand the subset of SATCAT objects tasked to Space Fence. Therefore, test target sets should be reviewed against JSpOC tasking as it becomes better defined. Additionally, the test design will evolve as the Space Fence program matures.

Although the SATCAT is comprehensive, many objects lack position, velocity, and time truth data of sufficient accuracy to test the higher accuracies expected from Space Fence. We propose a two-tier approach in which accuracy requirements are initially tested against a small subset of objects in the SATCAT, where highly accurate information is available, and then against the entire SATCAT. For example, to test observation accuracy at the expected Space Fence precision, Space Fence should be first tested and calibrated against satellites whose positions can be established to a high degree of accuracy (e.g., under one meter for laser ranging). Two such candidate target sets are the International Laser Ranging Service (ILRS) satellites and the High Accuracy Satellite Drag Model (HASDM) satellites [3, 4].[1] Once the radar's operational accuracy is calibrated against those satellites, the positional accuracy estimated by the radar's tracker can be used to validate the radar's performance against all objects within the required orbital regimes.

We present a statistical test design intended to test Space Fence to the letter of the program requirements,[2] and to characterize the system performance across the entire operational envelope. Most of Space Fence's requirements can be expressed in terms of either measurement errors or binomial responses. As such, statistical hypothesis tests on variance and binomial outcomes play a key role in our assessment of overall system performance. We then apply Design of Experiments (DOE) and logistic regression methods to explore the operational envelope by determining if the radar performance varies with the RSO altitude, size, and inclination. We illustrate the appropriate statistical test design techniques in detail by applying them to two common radar requirements, observation accuracy and probability of track. We then apply this same framework to the other radar requirements that are testable with onorbit test targets. Our analysis provides the type and number of necessary targets to test the requirement at specified confidence levels and statistical power. Comparing the resulting sample sizes with the number of currently known targets, we also identify those areas where Modeling and Simulation (M&S) methods are needed.

2. OBSERVATION ACCURACY: STATISTICAL HYPOTHESIS TESTS ON VARIANCE AND DOE

Experimental Design with Accurate Ephemeris Satellites (HASDM and ILRS)

The radar's Time, Elevation, Azimuth, Range, and Range Rate (TEARR) observation accuracy is key to establishing orbital precision and supporting coverage and flexibility of the radar surveillance fence. The sample sizes necessary to determine TEARR accuracy on uncued objects entering the fence can be determined by evaluating measured

---

[1] HASDM satellites are tracked multiple times per day to improve orbital predictions of LEO satellites in high drag regions. Accurate orbital ephemerides could be generated by fusing the data from their frequent tracking, making them useful as calibration targets for Space Fence metric accuracy.

[2] Our analysis does not include Space Fence requirements that do not need on-orbit test targets.

errors in these metrics via a hypothesis test on their variance, assuming a normal distribution. The sample size is determined by a statistical design to test the null hypothesis versus an alternative hypothesis in which thresholds are not met by a given amount of the error, or effect size, between predicted and measured positions. The test statistic for a variance test is (n-1) S_n^2/σ_o^2, where n is the number of metric observations, σ_o is the required standard deviation, and S_n is the sample standard deviation. This statistic follows a chi-square distribution with n-1 degrees of freedom provided the null hypothesis, that σ, equals σ_o, is true. The resulting sample size depends on the effect size, the statistical power (the probability of detecting an effect given that the radar does not meet the required thresholds) desired for the test, and the statistical significance level, α (the probability of falsely detecting an effect given that the radar does meet the required thresholds) [5]. Table 1 lists the statistical power for different sample sizes for a 10 percent effect size and α level of 5 percent.

Table 1. Sample Size verses Power

Sample Size	Power
300	0.7620
400	0.8562
600	0.9507

Table 2. Number of HASDM/ ILRS Satellites in different Altitude Regimes

Altitude (kilometer)	Number HASDM/ILRS
250 - 600	76/4
600-2,000	10/17
2,000-6,000	0/1
> 6,000	0/15
Total	86/37

Since Space Fence will operate in a “target rich” environment in which the number of targets is not, in general, constraining, we choose 95 percent statistical power and conclude that we need approximately 600 data points to meet the desired power for a 10 percent effect size at the 5 percent α level. Of these 600 independent data points, some might come from observations on the same target during different passes over the radar and not necessarily 600 different targets.

To determine target (vice track) needs, we need to assume a reasonable number of test days, estimate an approximate number of orbital passes per day, depending on the orbit, and calculate an approximation to the total number of tracks expected over the test duration. Assuming that only one half of the HASDM/ILRS satellites (Table 2) are available to be used as targets, there would be about 60 such satellites. Assuming a conservative average of two acceptable passes[3] per day through the radar’s field of view (FOV), 600 data points could be obtained in as few as 5 test days. Even if only 25 percent of HASDM/ILRS satellites were available, and/or only one or two passes per day were available, 600 data points could be obtained in 2 weeks or less.

---

[3] By an acceptable pass we mean a pass of sufficient elevation and length of track to allow the radar to gather sufficient data to generate observations.

“All Object” Experimental Design – Spanning the Operational Envelope

Once real-world TEARR accuracy is tested using HASDM/ILRS satellites, accuracy of radar internal error (variance) estimates could be tested using all SATCAT objects crossing the radar FOV. Regardless of the tracking algorithms chosen for Space Fence, the tracker will generate internal estimates of TEARR variances (errors) and covariances on each object for positional prediction, observation association, and track generation.

A single rolled-up “met” or “not met” finding could be made for observation accuracy on the basis of a test on the TEARR variances. While such a finding might be needed to meet the letter of this requirement, it would not completely describe how the radar is performing for subsets of objects that exhibit different physical or orbital properties. It could be that the radar meets (or does not meet) the overall requirement, but different orbital factor and level combinations might affect the measurement error in different regions of the operational envelope in such a way that requirements are not (or are) met in important subsets of the operational envelope.

The power to detect differences between levels of factors such as altitude, size, inclination, and their interactions can be estimated using DOE [6]. Altitude[4] is important because higher altitudes generally make for longer range, and range appears as an inverse fourth power relationship in the radar equation, thus lowering the received radar power's signal to noise ratio (SNR) (although higher altitude is somewhat offset by longer access time windows). Size also affects the received power SNR, making small targets more difficult to track than larger ones. Inclination may not be as direct an influence as altitude and size but it affects the number and duration of passes across the radar’s FOV, indirectly affecting observation accuracy. Additionally, the high object flux density in certain inclination bands might stress the radar’s energy management and/or data processing to the limits.

Consistent with groupings found in the requirements, for altitude, we choose four levels: 250 to 600 kilometers, 600 to 2,000 kilometers, 2,000 to 6,000 kilometers, and 6,000 to 22,000 kilometers. For inclination, we choose three levels: high (80-171 degrees, representing near-polar and retrograde orbits), mid (45-80 degrees, centered on the highly populated mid-60s inclination band), and low (9-45 degrees, capturing low population density and high number of observations per object). For size we choose two levels: 10 centimeters or larger, representing the approximate smallest size trackable by current space radars, and less than 10-centimeters diameter, capturing any sensitivity improvements from Space Fence.

The continuous nature of these variables can be handled statistically either as a multiple regression model, or an Analysis of Variance model in which the continuous variables of altitude, size, and inclination are aggregated into categories. We choose the latter approach because (1) Space Fence requirements are, in general, written with some or all of these factors aggregated into categories, and (2) a phased array radar’s operational modes are restricted by categories of orbital regimes (e.g., high-altitude surveillance mode, low orbit debris fence mode). Hence, our approach is to recognize that the independent factors are continuous, but for statistical design purposes we aggregate their levels to ensure data collection across the entire operational envelope. By doing so we might conclude, for example, that the radar meets metric accuracy requirements in all operational altitude bands, yet has higher accuracy on objects at altitudes between 250 and 600 kilometers than it does for objects between 600 and 2,000 kilometers (or

---

[4] We use altitude instead of range as our factor of choice because altitude is an inherent property of each orbiting object, whereas range depends on the particular pass.

vice versa). From this we may construct a regression model to estimate the loss (or gain) of accuracy as a function of altitude, which could in turn be used to improve the radar's tracking algorithms. This approach preserves the operational significance of the altitude bands stated in the requirements documents, while maintaining the ability to analyze continuous factors.

As with the HASDM/ISLR satellites, we have calculated the data points likely to be available from all catalogued objects crossing the fence. For this calculation, and those that follow (unless explicitly stated otherwise), we assume a conservatively low number of one acceptable pass per day for altitudes less than 600 kilometers, and two acceptable radar passes per day for all targets above 600 kilometers.

It is important to note that the sample size estimate of 600 data points (tracks) represents what would be needed to meet statistical power for an overall estimate of observation accuracy. These 600 data points can be evenly divided across all factor-level combinations to ensure that all representative combinations are adequately included. In Table 3 below there are three levels of inclination, four levels of altitude, and two levels of size, for a total of 24 combinations, so each combination requires 25 data points. This table also contains the number of tracks we would expect to be available from the current SATCAT[5] over a 34-day test period[6] for each factor-level combination compared with the 25 tracks needed. Real tracks from objects in the SATCAT would be available in all inclination, altitude, and size regimes except for objects under 10 centimeter in size at altitudes between 2,000 and 22,000 kilometers. Of the 600 tracks required, 475 could be obtained from real objects, leaving 125 to be met through M&S, should the entire trade space be explored.

---

[5] The publicly available SATCAT as of June 2013, contains 16,845 objects, of which 15,842 are in Earth orbit and have complete data.

[6] Thirty-four days is based on as a test design presented internal Institute for Defense Analyses document, which did not account for the current RSO population. This paper updates that design with the current RSO population, but retains a maximum of 34 days a reasonable cost-effective test period that allows for schedule flexibility.

Table 3. Targets for Testing Observation Accuracy. Targets are organized in terms of inclination, altitude, and size. If 25 tracks can be obtained in less than 34 days, the columns labeled "Real tracks/Min test days" are shaded green, with "25/nn" indicating that 25 tracks can be obtained in minimum of nn days. If at least 25 tracks are not expected to be available over 34 days, the entry is colored red, and the columns labeled "M&S Tracks Needed" represent the number of M&S tracks that would be needed to augment the real tracks to meet the 25-track limit.

Inclination (degrees)	Altitude (kilometer)	SATCAT RSOs of Size (centimeter)		Real Tracks/ Min Test Days		M&S Tracks Needed	
		<10	≥10	<10	≥10	<10	≥10
9-45	250-600	1	32	25/25	25/1	0	0
	600-2,000	4	93	25/4	25/1	0	0
	2,000-6,000	0	6	0	25/3	25	0
	6,000-22,000	0	2	0	25/7	25	0
45-80	250-600	16	85	25/2	25/1	0	0
	600-2,000	534	2,497	25/1	25/1	0	0
	2,000-6,000	0	10	0	25/2	25	0
	6,000-22,000	1	246	25/13	25/1	0	0
80-171	250-600	28	272	25/1	25/1	0	0
	600-2,000	1,372	5,727	25/1	25/1	0	0
	2,000-6,000	0	89	0	25/1	25	0
	6,000-22,000	0	2	0	25/7	25	0
Total		1,956	9,061	175/25	300/7	125	0

The power to differentiate between levels of a factor can be estimated using DOE techniques. Using the JMP program [7] to calculate power for our designed experiment at a 5 percent α level, Table 4 shows the power achieved in differentiating between levels of the main factors of Inclination (I), Altitude (A), and Size (S) or the interactions of I x A, I x S, and A x S, at various levels of statistical SNR,[7] with a 2x3x4 full factorial design using 24 replicates. Based on this table, testers can make appropriate assumptions on statistical SNR to achieve desired power.

Table 4. Power Analysis for Observation Accuracy Design

Factor	Power at Statistical SNR			
	0.25	0.3	0.375	0.5
Inclination (I)	70.4 %	85.0%	96.3%	99.9%
Altitude (A)	62.5%	78.1%	92.8%	99.5%
Size (S)	86.4%	95.6%	99.6%	> 99.9%
Interaction (I x A)	46.0%	60.7%	79.7%	96.1%
Interaction (I x S)	70.4%	85.0%	96.3%	99.9%
Interaction (A x S)	62.5%	78.0%	92.8%	99.5%

---

[7] In JMP, the full factorial design SNR is the ratio of the difference in <u>mean</u> predicted response going from level to level to variation in the response variable due to random events.

3. PROBABILITY OF TRACK: STATISTICAL HYPOTHESIS TESTS ON BINOMIAL OUTCOMES AND LOGISITIC REGRESSION METHODS

Probability of track, p, appears in many Space Fence requirements, and represents the probability of collecting observations on objects passing through the radar's FOV. Assuming independence between tracking attempts, the number of tracks can be modeled as a binomial probability distribution. We develop this approach for p = 0.5 because the associated probability distribution leads to the largest possible sample size of any other value of p.

We define a null hypothesis H_o: $p \geq 0.5$ versus the alternative H_1: $p < 0.5$. The test statistic in this case is $S_n = \sum_{i=1}^{n} \theta_i$, where n is the number of track attempts, and $\theta_i = 1$ if the i-th object is tracked, 0 otherwise, which counts the number of tracks successes per day. S_n has a Binomial distribution with parameters n and p, so that $P(S_n = k) = \binom{n}{k} p^k (1-p)^{(n-k)}$, k =1,...,n. For a given α error size and power, using the p ≥ 0.5 null hypothesis, the sample size is given by that smallest value n* for which: Prob $[(S_{n*} / n^*) \leq 0.5 (1-\delta) \mid p \geq 0.5] \leq \alpha$ and Prob $[(S_{n*} / n^*) \geq 0.5 (1-\delta) \mid p \geq 0.5] \geq$ power, where δ is defined as the effect size.

To minimize approximation errors and calculate efficient sample sizes, we used the logistic regression method as implemented in the R language [8], from which we obtain the sample sizes shown in Table 5. In this table we list sample size for the same 5 percent α and 95 percent power level used to calculate sample size for the observation accuracy requirement, as well as alternative levels.

Table 5. Power and Sample Size for Fence Integrity

Operational Mode	Effect Size	Power	Alpha	Sample Size
p=0.5	10% (p≤0.40)	95%	5%	269
		80%	5%	158
		70%	5%	122
	5%(p≤0.45)	95%	5%	1081

We choose 95 percent statistical power, and conclude that we need approximately 269 data points to meet the desired power for a 10 percent effect size at the 5 percent α level.

The power to detect differences between levels of altitude and inclination cannot be conducted in the same way as the analysis of observation accuracy design as shown in Table 3. The reason for this is that the observation accuracy response variables (e.g., time, elevation, and range errors) are continuous, as contrasted with the binary response variable for probability of track. To analyze this binary response design, our approach is to consider a logistic regression model that uses a function of p, the probability of an object being tracked, instead of θ as a response variable. The response variable p is modeled as a function of the Altitude (A) and Inclination (I) as independent variables.[8] The logistic regression model follows the form: $\ln (p / (1-p)) = \beta_0 + \beta_1 A + \beta_2 I$, where β_i are the regression coefficients. Unlike the original binary response variable of 1 or 0, the transformed response variable $\ln(p / (1-p))$ is continuous and can assume any real value. This regression model was fitted using the R language general linear model "glm" module with a binomial error distribution. Once the regression coefficients are estimated, the probability of an object being tracked for a given altitude and inclination can be calculated as:

---

[8] The effect of object Size (S) was considered in our analysis of the full requirement set and, for the sake of simplicity, was not included in this overview paper.

$p = \exp(\beta_0 + \beta_1 A + \beta_2 I) / (1 + \exp(\beta_0 + \beta_1 A + \beta_2 I))$.

As an example, we ran a Monte Carlo simulation with 1,000 iterations to test whether altitude and inclination are statistically significant factors. The null hypothesis assumed that the factors did not affect the probability, p, of tracking an object and the alternative hypothesis assume that p varied from level to level. A p-value (from the general linear model) less than α for a particular factor indicates that p significantly depends on that factor. The power of the regression, or the probability of identifying a statistically significant effect for altitude or inclination given that there really is such an effect, was calculated on a post-hoc basis by counting the number of iterations where the p-value was less than α and then dividing that sum by the total number of iterations. For example, for the sample size of 170 per factor/level combination in Table 6, a post-hoc power of 90 percent can be obtained to test significance of either inclination or altitude at the 5 percent α level assuming a probability of track of 0.5, 0.45, and 0.4 for the high, mid, and low levels, respectively. A post-hoc power of 90 percent can be achieved in 8 days. This power was verified with the JMP program using a SNR of 0.202 (based on a 10 percent effect), calculated with the logit transformation implemented in the "Binary Response Sample Size Calculator" spreadsheet [9]. Additionally, JMP estimated a 75 percent power for detecting if the probability of track is affected by the altitude and inclination interaction (A x I).

Table 6. SATCAT Targets for Probability of Track Ordered by Inclination

Inclination (degrees)	**Altitude (kilometer)**	**Number**	**Real Tracks/Min Test Days**	**M&S Tracks**
9-45	250-550	22	> 170/8	0
	550-800	52	> 170/4	0
	800-3,000	37	> 170/5	0
45-80	250-550	65	> 170/3	0
	550-800	1,073	> 170/1	0
	800-3,000	1,536	> 170/1	0
80-171	250-550	154	> 170/2	0
	550-800	1,346	> 170/1	0
	800-3,000	4,032	> 170/1	0
Total		8,317	>1,530/8	0

It should be noted that the 5 and 10 percent difference in proportionality is analogous to, but not the same as, the effect sizes used for observation accuracy, and the exact power of the test would not be known until after the data are analyzed (because the size of the error variance would not be known). Given that Space Fence testing will be done in a target rich environment and, if needed, additional data could be readily obtained, our post-hoc approach to power is not overly constraining.

4. TEST DURATION SUMMARY AND MODELLING AND SIMULATION

By applying variance, binomial, and logistic regression methods to other radar operational requirements (surveillance and track coverage, object correlation, radar cross section accuracy, minimum detectible target size and orbital determination), we concluded that Space Fence could be operationally tested with 53 factor/level combinations, using mostly on-orbit targets. M&S would be required for objects less than 10 centimeters at altitudes above 2,000 kilometers, objects less than 10 centimeters in inclinations between 9 and 45 degrees at altitudes between 250 and 3,000 kilometers, and 10 centimeter objects at altitudes above 2,000 kilometers. Over 50 percent of the testing involving on-orbit targets can be completed in 8 days or less, approximately 70 percent can be tested in 2 week or less, and the remainder should take no more than 25 days.

5. SUMMARY

A statistical test design is appropriate to help identify suitable on-orbit targets to test Space Fence requirements and characterize the system performance across the operational envelope. By applying variance, binomial, and logistic regression methods to Space Fence's operational requirements, and identifying orbital regimes in which DOE methods can be used to explore the operational envelope, we estimate Space Fence could be operationally tested in 25 days, with 53 factor/level combinations, using mostly on-orbit targets.

6. REFERENCES

[1] Haines, L. and Phu, P., "Space Fence PDR Concept Development Phase," 2011 Advanced Maui Optical and Space Surveillance Technologies Conference, Maui, Hawaii, 14 September 2011.

[2] "Space Fence Contract Award," Air Force Space Command Website (http://www.afspc.af.mil/news/story.asp?id=123413302), 6 June 2014. (Release Number: 010614).

[3] M.F. Storz, B.R. Bowman, J.I. Branson, S.J. Casali, and W.K. Tobiska, Adv. Space Res. 36, 2497 (2005).

[4] Noll, C. and Pearlman, M., *International Laser Ranging Services 2009-2010 Report*, National Aeronautics and Space Administration TP 2013-217507, June 2012.

[5] Milton, J. S. and Arnold, J.C, "Introduction to Probability and Statistics: Principles and Application for Engineering and the Computing Sciences" Irwin/McGraw-Hill 3rd ed. (1986).

[6] Montgomery, D. C., "Design and Analysis of Experiments" John Wiley & Sons, Inc 7th ed. (2009).

[7] JMP® 11.0.0, 64-Bit Edition; Copyright 2013 SAS Institute Inc.

[8] R version 3.1.0 (2014-04-10) – "Spring Dance" Copyright (C) 2014 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit).

[9] Ortiz, F., "Dealing with Categorical Data Types in a Designed Experiment Part II: Sizing a Designed Experiment When Using a Binary Response," Scientific Test and Analysis Techniques, Test and Evaluation Center of Excellence, 2014.