Statistical Approach to the Operational Testing of Space Fence

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

The next-generation space surveillance system, known as Space Fence, uses radar to track space debris and operational satellites in low and medium Earth orbit that may threaten U.S. space assets. The enhanced capability provided by Space Fence is expected to increase the number of routinely tracked orbiting objects from approximately 17,000 to more than 100,000 (U.S. Air Force 2014; NASA 2012). Given that existing sensors cannot quickly validate the new radar's complete set of observations, the question becomes how to test successfully this new space surveillance system.

Our approach quantifies the probabilities of meeting requirements; determines how performance varies as a function of an object's altitude, inclination, or size; estimates a 25-day test duration; and determines that modeling and simulation methods may be needed to represent 125 additional satellites.

Difficulty of Testing Space Fence

Space Fence is a new ground-based, space-directed radar system the United States is acquiring to detect, track, and catalog orbiting space objects, including the growing population of space debris. The system will consist of two S-band (2–4 GHz) phased-array radar sites from which it will perform autonomous cued and uncued surveillance and cued searches for objects in low and medium Earth orbits. Space Fence will provide tracking and radar characterization data on orbiting objects to the U.S. Air Force Joint Space Operations Center to support maintenance of the Satellite Catalog (SATCAT) and support other space situational awareness needs (U.S. Air Force 2014).

Space surveillance radar performance has traditionally been tested by comparing radar observations against truth data on position, velocity, and time for a small number of well-understood objects with known positions measured to within 1 meter using laser ranging or onboard beacons (Mochan and Stophel 1968; Noll and Pearlman 2011; Joint Range Instrumentation Accuracy Improvement Group 1995; Martin et al. 2011). However, truth data for the majority of the approximately 17,000 objects tracked by existing radar systems and optical telescopes may not be sufficiently accurate to validate the data on the much larger number of satellites expected to be tracked by the highly accurate Space Fence. Further, truth data on the relatively small number of objects may not extrapolate to an operationally representative population of space objects of different types, inclinations, altitudes, sizes, shapes, and rotational motions. Because Space Fence will have a larger field of view and higher sensitivity than existing radar systems, it is expected to be able to routinely track more than 100,000 orbiting objects (U.S. Air Force 2014; NASA 2012).

Testing a space surveillance system whose complete set of observations cannot be validated in a timely manner by existing radar systems and optical telescopes presents some challenges.

- How can testers know Space Fence is capturing all the objects it is intended to observe?
- Are the radar measurements on all the objects observed by Space Fence of sufficient accuracy and precision to both meet its requirements and support orbital prediction and SATCAT maintenance?
- Can adequate testing of an operationally representative sample population, covering all intended object sizes, altitudes, and inclinations, be performed in a timely manner?

Proposed Testing Method

To address the issue of Space Fence performance across the full operational space, we propose extending initial calibration tests into broader rigorous statistical test designs, using on-orbit test targets that span the orbital limits of Space Fence's operational requirements. Through this approach, we characterize Space Fence performance by using a relatively small subset of the publicly available SATCAT (~1,500 out of ~17,000 objects),[1] grouped by altitude, inclination, and size (Pechkis, Pacheco, and Botting 2014, 2016).

Building on recent experimental design work for assessment of naval surface radar performance (Cortes and Bergstrom 2012), we used the target altitude, size, and inclination as predictor variables (or factors) in statistical tests for measures of radar performance requirements (e.g., range accuracy) as dependent (or response) variables. This approach quantified the probabilities of Space Fence meeting its performance requirements, determined whether and how satisfaction of individual requirements depends on an object's orbit and size, and estimated the sample sizes needed for statistical confidence in this evaluation. Comparing the resulting sample sizes with the number of currently known targets, we determined the areas where augmentation with modeling and simulation (M&S) may be needed because of an insufficient number of targets. Finally, we estimated the necessary test duration by assuming a radar coverage solely for the first radar site (located in Kwajalein in the Marshall Islands) and a conservative number of radar tracks per object per day.

Evaluating Space Fence in Terms of Operational Requirements

We chose four Space Fence operational requirements—metric accuracy, probability of track, object association, and data latency—as the response variables to illustrate the different statistical test design methodologies needed to support Space Fence operational test and evaluation.

[1] Analyses are based on data from the entire publicly available SATCAT on space-track.org as of June 2013.

Metric Accuracy

The metric accuracy of Space Fence is stated in terms of measurement errors (error variance) for each of five radar observation components: time, elevation, azimuth, range, and range rate. Metric accuracy is key to establishing orbital precision and for supporting the coverage and flexibility of radar surveillance.

We determined the sample sizes necessary for computing time, elevation, azimuth, range, and range rate accuracy of uncued objects entering the observation field of view by evaluating measured errors in these metrics with a hypothesis test for their variance, assuming a normal distribution. The hypothesis test results were agnostic to specific requirement thresholds; instead, they depended on the effect size (the amount a parameter exceeds its threshold), the desired statistical power (the probability of correctly determining that the requirement is met), and the significance level, referred to as α error (the probability of incorrectly determining that the requirement is not met).

Accuracy requirements were initially tested against a subset of SATCAT objects with highly accurate information available and then against the entire SATCAT inventory. For the initial testing, we selected two subsets of satellites known to contain accurate position, velocity, and time data—the International Laser Ranging Service (ILRS) satellites (Noll and Pearlman 2011) and the High Accuracy Satellite Drag Model (HASDM) satellites (Storz et al. 2005).

Six hundred object tracks were necessary to achieve a statistical power level of 95 percent for an effect size of 10 percent and an α error of 5 percent. We chose a 10-percent effect size because it would be sufficient to detect meaningful improvement or shortfall between Space Fence and legacy systems. Assuming that half of the satellites in the two subsets (60 satellites) are available and each had a conservative number of two *acceptable tracks* per day over a Kwajalein-based radar,[2] we calculated that 600 tracks could be obtained in as few as 5 test days:

$$5\text{ days} = 600\text{ tracks} \times \left(\frac{1}{60} \times \frac{1}{satellites}\right) \times \left(\frac{1}{2} \times \frac{days}{(tracks/satellites)}\right)$$

We used the analysis of variance (ANOVA) method to determine the probability of detecting whether or not any factor, or a combination of factors, affects the metric accuracy measurements of Space Fence. The factors we considered are altitude, inclination, and size, and we chose levels for each factor consistent with Space Fence requirements (Table 1). To implement this ANOVA approach, we first searched through the SATCAT for satellites likely to be observable from a Kwajalein-based Space Fence and estimated the average number of tracks per day.

[2] An acceptable track is a radar track of an object passing through the radar's field of view at a sufficient elevation and for a sufficient distance to allow the radar to gather enough data to generate observations.

Table 1. Number of Available SATCAT Objects by Inclination, Altitude, and Size

Inclination (Degrees)	Altitude (Kilometers)	Number by Size (Centimeters)*					
		SATCAT Objects		Real Tracks/ Minimum Test Days†		M&S Tracks Needed‡	
		< 10	≥ 10	< 10	≥ 10	< 10	≥ 10
9 ≤ I ≤ 45	250–600	1	32	25/25	25/1	0	0
	600–2,000	4	101	25/4	25/1	0	0
	2,000–6,000	0	6	—	25/3	25	0
	6,000–22,000	0	2	—	25/7	25	0
45 < I ≤ 80 (centered on the highly populated band in the mid-60s)	250–600	16	85	25/2	25/1	0	0
	600–2,000	534	2,498	25/1	25/1	0	0
	2,000–6,000	0	10	—	25/2	25	0
	6,000–22,000	1	246	25/13	25/1	0	0
80 < I ≤ 171 (representing near-polar and retrograde orbits)	250–600	28	276	25/1	25/1	0	0
	600–2,000	1,372	5,728	25/1	25/1	0	0
	2,000–6,000	0	89	—	25/1	25	0
	6,000–22,000	0	2	—	25/7	25	0
Total	—	**1,956**	**9,075**	**175/25**	**300/7**	**125**	0

* Objects <10 cm are included to capture sensitivity improvements from Space Fence; objects ≥10 cm sizes are tracked by current radars.

† The notation 25/*n* indicates that 25 tracks can be obtained in a minimum of *n* days.

‡ The number of M&S tracks that would be needed to augment the real track to meet the 25-track limit.

For this calculation, we assumed a conservatively low number of one acceptable track per day for altitudes less than 600 kilometers, and two acceptable tracks per day for all targets above 600 kilometers. The ANOVA design evenly divides the 600 tracks needed to test the radar calibration across all factor-level combinations to ensure that all combinations are tested.

As shown in Table 1, there are a total of 4 × 3 × 2 = 24 combinations of object altitude, inclination, and size levels, so each combination requires 600 ÷ 24 = 25 data points. Table 1 also contains the number of objects expected to be available from the SATCAT over an approximate one-month test period for each factor-level combination, compared with the 25 tracks needed. (As of June 2013, the publicly available SATCAT contained 16,845 objects, of which 15,842 were in Earth orbit and had complete data.) A one-month test period allows for schedule flexibility and is consistent with historical cost-effective operational test periods.

Tracks from objects in the SATCAT are available in all inclination, altitude, and size regimes, except for objects smaller than 10 centimeters at altitudes between 2,000 and 22,000 kilometers, for which M&S would be needed.

Probability of Track and Object Association

Space Fence has probability requirements for tracking objects that pass through its field of view (*probability of track*[3]) and for associating those tracks with objects in the catalog (*object association*[4]). Unlike metric accuracy requirements, which are expressed as continuous responses, probability of track and object association requirements are stated in terms of binary responses (tracked or not tracked, associated or not associated). As such, we propose statistical hypothesis tests on binomial distributions to assess the system against documented system requirements. We then apply a logistic regression/Monte Carlo method to determine if system performance varies with an object's altitude and inclination.

Unlike in the metric accuracy analysis, the sample sizes necessary to demonstrate Space Fence can meet its probability of track and object association requirements depend on the specific requirement threshold values. For illustrative purposes, we chose threshold requirements of 50-percent probability of track and 97-percent object association. Sample sizes of 268 and 81 tracks, respectively, can demonstrate the radar's probability to meet these threshold requirements for 10-percent effect size, 5-percent α error, and 95-percent power. Using logistic regression/Monte Carlo methods, we determined that the effects of altitude and inclination on probability of track and object association can be tested with 1,530 and 540 tracks, respectively, at 10-percent effect size, 5-percent α error, and at least 90-percent statistical power in 8 days. Table 2 shows the required number of data points for $3 \times 3 = 9$ combinations of altitude and inclination levels for both probability of track and object association. Each combination requires $1{,}530 \div 9 = 170$ and $540 \div 9 = 60$ data points, respectively.

Data Latency

Our final response variable is data latency—the time from when the sensor has finished collecting the data to when the U.S. Air Force Joint Space Operations Center has received the data. For Space Fence, we assumed data latency will be no more than 2 minutes 99 percent of the time and used tolerance intervals to determine the number of data points necessary to evaluate the latency requirement. Data latency is not typically sensitive to the characteristics of the orbiting objects, so we did not account for factor analyses.

A sample of 856 data transmissions can achieve a 90-percent power level for a 10-percent effect size and a 5-percent α error. A 10-percent effect size for latency corresponds to a 12-second delay in the 2-minute latency threshold. Although this may seem like a short delay, it

3 Probability of track is the probability of keeping track of the position and velocity of a given object that penetrates the radar's field of view.

4 Object association is the probability of associating detected objects with known SATCAT objects (to determine if a detected object is already known or newly discovered).

Table 2. Number of Available SATCAT Objects to Test Probability of Track and Object Association, Ordered by Inclination and Altitude

Inclination (Degrees)	Altitude (Kilometers)	Quantity	Real Tracks/ Minimum Test Days* Probability of Track	Real Tracks/ Minimum Test Days* Object Association
9 ≤ I ≤ 45	250–550	22	> 170/8	> 60/3
	550–800	60	> 170/2	> 60/1
	800–3,000	37	> 170/3	> 60/2
45 < I ≤ 80	250–550	67	> 170/3	> 60/1
	550–800	1,094	> 170/1	> 60/1
	800–3,000	1,536	> 170/1	> 60/1
80 < I ≤ 171	250–550	156	> 170/2	> 60/1
	550–800	1,356	> 170/1	> 60/1
	800–3,000	4,039	> 170/1	> 60/1
Total	—	**8,367**	**> 1,530/8**	**> 540/3**

* Indicates the minimum number of days needed to obtain 170 tracks.

could prove significant for certain conjunction alerts and consequent collision avoidance maneuvers.[5] For the International Space Station, for example, with a collision-avoidance-maneuver velocity of 0.5–1 millisecond (Hutchinson 2013), a 12-second delay would mean being 6 to 12 meters closer to a potential conjunction.

Summary

Space Fence will be a ground-based radar designed to perform surveillance on Earth-orbiting objects. Its capabilities will increase the number of objects tracked in the current SATCAT from approximately 17,000 to over 100,000. Testing a system whose complete set of observations cannot be validated in a timely manner by existing systems presents challenges for gathering detection and accuracy truth data while ensuring a reasonable test duration. We proposed a rigorous statistical test design with candidate on-orbit test targets that span orbital limits defined by Space Fence operational requirements. We characterized Space Fence performance across the entire operational envelope by using relatively small subsets (containing no more than 1,530 satellites) of the public SATCAT grouped by altitude, inclination, and size. We identified the type and number of on-orbit test targets needed for evaluating metric accuracy, probability of track, object association, and data latency. Our approach quantifies the probabilities of meeting requirements; determines how performance varies as a function of an object's altitude, inclination, or

[5] A conjunction alert occurs when the predicted time and location at which two or more objects in space will cross orbital paths, creating the potential for a collision. Satellite operators use these alerts to assess the need for collision avoidance maneuvers.

size; estimates a 25-day test duration; and determines that modeling and simulation methods may be needed to represent 125 additional satellites.

These results provide testers and users with a statistical basis of evaluation for Space Fence operational deployment decisions.

References

Cortes, L. A., and D. Bergstrom. 2012. "Using Design of Experiments for the Assessment of Surface Radar Detection Performance and Compliance with Critical Technical Parameters." 80th MORS Symposium. Colorado Springs: Military Operations Research Society.

Hutchinson, L. 2013. "How NASA Steers the International Space Station around Space Junk." *Ars Technica* (July 4). https://arstechnica.com/science/2013/07/how-nasa-steers-the-international-space-station-around-space-junk/.

Joint Range Instrumentation Accuracy Improvement Group. 1995. *Using Satellites for Radar Performance Monitoring and Calibration*. Document, U.S. Army White Sands Missile Range: Secretariat Range Commanders Council.

Martin, L. K., N. Fisher, W. Jones, J. Furumo, J. A. Heong Jr., M. Umeda, and W. Shiroma. 2011. "Ho'opono pono: A Radar Calibration CubeSat." *Proceedings of the AIAA/USU Conference on Small Satellites*. Technical Session VI, Small but Mighty, SSC11-VI-7. http://digitalcommons.usu.edu/smallsat/2011/all2011/42/.

Mochan, J., and R. A. Stophel. 1968. "Dynamic Calibration of Space Object Tracking." *The Space Congress Proceedings*. Paper 1. http://commons.erau.edu/space-congress-proceedings/proceedings-1968-5th/session-1/1/.

NASA. 2011. "Orbital Debris Frequently Asked Questions." Last updated September 2, 2011. http://orbitaldebris.jsc.nasa.gov/faqs.html.

Noll, C., and M. Pearlman. 2011. *International Laser Ranging Services 2009–2010 Report*. Recurring Report, Greenbelt: NASA Goddard Space Flight Center.

Pechkis, D. L., N. S. Pacheco, and T. W. Botting. 2014. "Identifying On-Orbit Test Targets for Space Fence Operational Testing." In 15th Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS 2014). https://amostech.com/TechnicalPapers/2014/Poster/PECHKIS.pdf.

———. 2016. "Statistical Approach to the Operational Testing of Space Fence." *IEEE A&E Systems Magazine* 31, no. 11 (November): 30–39. https://doi.org/10.1109/MAES.2016.150176.

Storz, M. F., B. R. Bowman, J. I. Branson, S. J. Casali, and K. W. Tobiska. 2005. "High Accuracy Satellite Drag Model (HASDM)." *Advances in Space Research* 36, no. 12: 2497–2505. https://doi.org/10.1016/j.asr.2004.02.020.

U.S. Air Force. 2014. "Space Fence Contract Awarded." Published June 5, 2014. http://www.af.mil/News/Article-Display/Article/485271/space-fence-contract-awarded/.

Dan Pechkis (right), a Research Staff Member in the Operational Evaluation Division (OED) of IDA's Systems and Analyses Center, holds a doctorate in physics from the College of William and Mary.

Tye Botting (left), also a Research Staff Member in OED, holds a doctorate in physical and nuclear chemistry from Texas A&M University.

Nelson Pacheco, an OED consultant, holds a doctorate in mathematical statistics from Colorado State University.

Acknowledgments: The authors are thankful for the insights provided by Kelly Avery and Kassandra Fronczyk of OED and Michael Tuley of the Science and Technology Division of IDA's Systems and Analyses Center.

The original article "Statistical Approach to Operational Testing of Space Fence" by D. L. Pechkis, N. S. Pacheco, and T. W. Botting was published in I*EEE Aerospace & Electronics Systems Magazine* 31, no. 1: 30–39. https://doi.org/10.1109/MAES.2016.150176.