IAC-24-A6.7.8

The TraCSS Consolidated pathfinder: Leveraging Commercial Capability in LEO

Matthew Hejduk[a]*, Jess Johnson[b], John Giles[c], Anna Parikka[d], Christine Joseph[e], Sandra Magnus[f], Carson Coursey[g], Miles Lifson[h]

[a] *The Aerospace Corporation, 2011 Crystal Drive, Suite 900 Arlington, Virginia 22202-3780, USA,* matthew.d.hejduk@aero.org
[b] *Lincoln Laboratory, Massachusetts Institute of Technology, 244 Wood Street, Lexington, Massachusetts 02421-6426, USA,* jessica.johnson@ll.mit.edu
[c] *The MITRE Corporation, 7515 Colshire Drive, McLean, Virginia 22102, USA,* jgiles@mitre.org
[d] *NOAA Office of Space Commerce / Global Science & Technology, Inc., 7501 Greenway Center Drive, Suite 1100, Greenbelt, Maryland 20770, USA,* anna.parikka@noaa.gov
[e] *NOAA Office of Space Commerce, USA,* christine.joseph@noaa.gov
[f] *NOAA Office of Space Commerce / The MITRE Corporation, Herbert C. Hoover Building, Room 68015 1401 Constitution Ave., NW Washington, D.C. 20230, USA,* sandra.magnus@noaa.gov
[g] *The Aerospace Corporation, 2011 Crystal Drive, Suite 900 Arlington, Virginia 22202-3780, USA,* carson.d.coursey@aero.org
[h] *The Aerospace Corporation, 2011 Crystal Drive, Suite 900 Arlington, Virginia 22202-3780, USA,* miles.t.lifson@aero.org
* Corresponding Author

Abstract

The Department of Commerce has pursued a pathfinder activity to engage commercial SSA data and service providers to provide augmenting orbital safety products to DoD conjunction assessment (CA) events and to attempt to maintain a LEO space catalogue entirely with commercial data and services. Comprising a preparation period, three months of operational conduct, and a concluding evaluation period, this pathfinder included two commercial SSA data providers and a commercial services provider, and two providers that served as data and product evaluators. The concept of surge tasking for DoD-identified CA events was tested and the quality of the commercially-maintained LEO catalogue was evaluated and compared to the DoD-only analogues for these activities. A number of important conclusions emerged that will inform the Department of Commerce's use of commercial SSA data for operational orbital safety activities.

Keywords: conjunction assessment, space catalogue, mission planning, TraCSS, commercial SSA data

Nomenclature

ε = residual
C = covariance

Acronyms/Abbreviations

CA = conjunction assessment
CDF = cumulative distribution function
CDM = Conjunction Data Message
CVM = Cramér – von Mises
DoD = Department of Defense
DoF = degrees of freedom
EDF = empirical distribution function
GEO = geosynchronous Earth orbit
GNSS = global navigation satellite system
GOF = goodness-of-fit
GSA = U.S. General Services Administration
JCO = The US Space Command's Joint Commercial Office
LEO = low Earth orbit
NASA CARA = NASA Conjunction Assessment Risk Analysis Team
OD = orbit determination
O/O = owner/operator
OSC = Office of Space Commerce
Pc = probability of collision
RIC = radial, in-track, and cross-track (the satellite-centered reference frame)
RMS = root mean square
SGP4 = Simplified General Perturbation Theory #4
SSA = Space Situational Awareness
TCA = time of closest approach
TLE = two-line element
TraCSS = Traffic Coordination System for Space

1. Introduction

The Department of Commerce's Office of Space Commerce (OSC) has been given a mandate to stand up a civilian orbital safety capability, called the Traffic Coordination System for Space (TraCSS). As the Office of Space Commerce advances TraCSS development, it is mindful of the direction to leverage commercial SSA data and services. In order to learn the best way to do this, OSC has identified several pathfinder projects to engage the commercial industry in a number of orbital safety activities in an experimental capacity. Last year a pilot study was conducted that explored commercial orbital safety solution for the GEO regime. This winter and spring (2024) a similar but much larger effort was conducted to explore commercial orbital safety augmentation and catalogue maintenance in LEO. This "LEO pathfinder" employed LeoLabs for LEO radar tracking data, Slingshot Aerospace for LEO optical tracking data, COMSPOC Corporation. for orbit determination and predicted/definitive ephemeris generation, and two additional companies (Kayhan Space and SpaceNav Inc.) to evaluate the data and products from the three main commercial participants. During its operational conduct, the pathfinder Identified a daily list of various-quality CA events based on the DoD catalogue for which additional tracking could potentially be helpful, attempted to perform "surge" tracking on these objects, and produced improved predicted ephemeris and orbital safety products, which were then evaluated against the equivalent DoD products for these events. Additionally, the three pathfinder providers attempted to maintain a LEO space catalogue using only the tracking data from the two data providers. The characteristics of this catalogue were compared to a comparable portion of the unclassified DoD LEO catalogue. While the main objective of the pathfinder effort was to gain experience with this kind of commercial contracting and determine metrics and paradigms for future commercial acquisition activities, much valuable information for shaping future orbital safety solutions that involve commercial industry was obtained by the pathfinder conduct itself.

The active period of the pathfinder consisted of two months of preparatory work, three months of operational conduct, and a final month of data analysis and conclusion extraction. A detailed treatment of the pathfinder construction and organization, with substantial technical explanations of each of its parts, was recently presented at the 2024 AMOS Technical Conference [1], and interested readers are referred to that paper for this full explanation. The purpose of the present paper is to discuss actual performance levels and other conclusions drawn from the three months of pathfinder conduct. The paper is divided into different sections, each of which addresses performance outcomes for one of the major objectives of the pathfinder, and then a conclusions section that summarizes these findings in a single collection of statements.

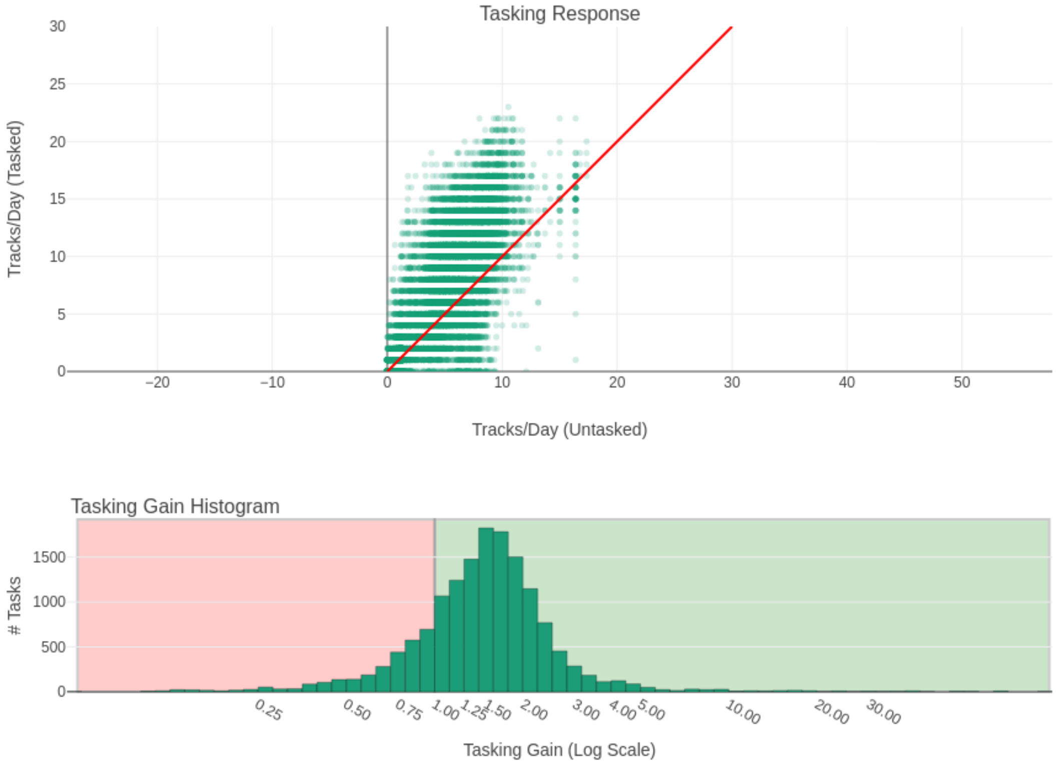
2. Proof-of concept and metrics development

As mentioned above, a recent paper [1] outlines the construction, logistics, and intended purposes of the pathfinder project in detail. While there is substantial interest in how the pathfinder project fared in improving LEO orbital safety and maintaining a LEO catalogue, it is important to establish firmly as the backdrop the principal objective of the pathfinder, which does not include either of the above areas of interest; therefore, the following summary of pathfinder performance is assessed using metrics developed specifically for catalogue-wide operational monitoring. Future analysis may investigate the pathfinder's performance in the domains of LEO orbital safety and catalogue maintenance. The first-tier reasons for OSC's pursuing the pathfinder project include the following:

- To identify the set of metrics that should be used contractually to govern the acquisition of commercial SSA data and services used to improve/conduct the orbital safety mission. The commercial availability of such products and services is a relatively recent development, and there is little government experience with how one should specify the requirements for their production and delivery in a space safety context. The pathfinder project encapsulated all of these activities and metrics types, in a simulated operational context, to allow trade-offs to be explored and a set of durable recommendations assembled to guide future acquisitions. Undoubtedly, further refinement will take place as acquisitions are executed in support of actual TraCSS operations, however, the experience obtained here is already sufficient to construct an informed acquisition framework.
- To experiment with alternative data and services acquisition methods. US government acquisition paradigms often suffer from heavily bureaucratized processes that are inefficient and not easily adaptable to emerging needs, so this pathfinder project experimented with an acquisition platform called Global Data Marketplace. It is an online platform to purchase data and services[*]. The US Space Command's Joint Commercial Office (JCO) has experimented with this acquisition method for some years, but not at the scale of the present pathfinder activity in LEO. Such convenience, of course, does come at a price and with some technical complications; but the procurement of the

---

[*] Global Data Marketplace is maintained by Bluestaq LLC.: https://www.globaldatamarketplace.com



Total	Never Tracked	Lost	Gain < 1	Gain ≥ 1	Found
18023	884 (4.90%)	1848 (10.25%)	2908 (16.13%)	12360 (68.58%)	23 (0.13%)

Fig. 1. Tracking gain for objects placed on mission planning tracking list.

data and services for this pathfinder was vastly simpler and more expeditious than that experienced using conventional methods.

- To determine necessary structures and conditions for commercial competitors working together on a large joint project. While each of the three participating providers has their particular specialty that sets them apart from the others, all three do at some level provide both data and catalogue maintenance services for purchase. It was not known, therefore, whether these companies would recognize the offsetting nature of their strengths and construct a working relationship grounded in this complementarity, but it is pleasing to report that this is in fact what happened. The commercial SSA market is unlikely to produce an outcome in which a single company prevails over all the others, so the ability to compartmentalize competitive pressures to allow cooperative projects is very encouraging.

Thus, while the remainder of this paper will focus on results from other pathfinder objectives, namely conjunction assessment data surges/products and LEO catalogue maintenance, one must always remain mindful that these were subsidiary and not the principal objective of the pathfinder project.

3. Tracking "surge" for conjunction assessment

As described in Magnus et al. [1], a basic feedback loop during pathfinder conduct was established to identify satellite conjunctions of interest using the DoD catalogue. This information was conveyed to the pathfinder participants to allow them to elevate data collection on the secondary objects in these conjunctions and produce orbital safety products, which would then be compared ex post facto to those of the DoD. The overall objective of this set of activities was to examine the viability of this "chase and improve" concept of operations with the expectation that a notable improvement would be wrought in the secondary solution, which would then produce more definitive conjunction risk assessment results.

The CA event selection algorithm, which identified events of interest to present to the pathfinder for improved solutions, did not simply choose the most concerning events emerging from the DoD; instead, it attempted to assemble a broader cross-section that considered orbital altitude, DoD tracking density, limiting of Starlink-related events, and representation of some proportion of maneuverable secondaries. Importantly, relatively few of the ~200 daily CA events selected for surge tracking were poorly tracked by the DoD. A collection of ~200 objects conforming to these selection criteria were identified each day and presented to a mission planning algorithm jointly developed by the two data provider participants. The algorithm would assess the 200 objects for tracking adequacy and then assign a relative priority to each, and this prioritized list would be furnished to both providers' scheduling algorithms to collect additional data. The pathfinder tracking data were then made available to the pathfinder's orbit determination process to produce estimated states and predicted ephemerides, which could then be used to recalculate the DoD event CDMs to produce an alternate CA risk assessment result.

In terms of producing a data augmentation in response to this tasking, in an overall sense the approach was successful. Fig. 1 gives a scatter plot of the quantity of tracks (in tracks/day) for the objects on which this was attempted, during periods where the object was not on the tasking list (untasked) versus when it was (tasked). The histogram shows the tasking gain per object per day, which is a ratio of the tracking rate while on the tasking list relative to the mean tracking rate during untasked periods of the pathfinder. At the median value, one sees a gain of about 1.5 to 2 times as many tracks per day. In CA, often just a few additional tracks can be transformative. The Pc will either drop off precipitously with this additional information, thus allowing the event to be dismissed; or it will remain high, certifying that the situation is high risk and merits a risk mitigation maneuver.

In terms of producing a shrinking in the uncertainty (covariance) for the secondary object, which is the principal way in which the Pc behaviour described above takes place, the results were disappointing. Fig. 2 gives a CDF plot of the RMS uncertainty of the covariance (a technique to measure a covariance's size) of the tasked objects' state estimates both before and after being placed on prioritization. The two curves are nearly identical, indicating that the additional tracking obtained did not translate into an appreciable decrease in the covariance size. Thus, while the pathfinder was able to successfully identify and track additional objects of interest, the resulting data did not measurably improve the information available to the satellite operator for actionability.

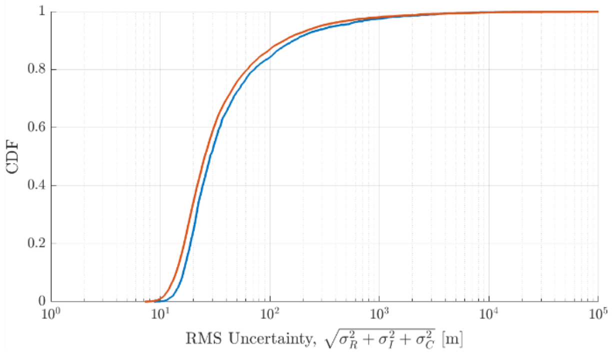


Fig. 2. Covariance size (via RMS uncertainty) before (blue line) and after (red line) tracking prioritization.

A similar conclusion emerges from examining the behaviour of the COMSPOC OD confidence evaluation, a metric that is calculated after each OD to determine how reliable a correction has been obtained. Fig. 3 gives the OD quality statistic both before and after the tracking surge, indicating in this case that the situation actually slightly worsens with additional tracking, although both readings are probably equivalent given the level of precision of the metric.

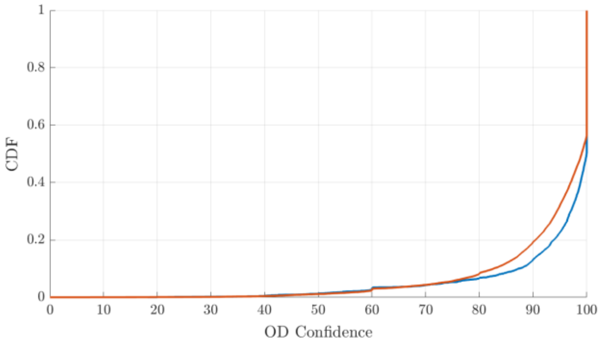


Fig. 3. OD confidence value before (blue) and after (red) tracking prioritization.

The lack of identifiable improvement given the statistically-significant tracking gain appears to be due to a bifurcated situation. If the object is already reasonably tracked, then the pathfinder sensors can acquire it and provide more tracking; but because the covariance size is governed (approximately) by the square root of the number of measurements considered, this incremental increase in data does not result in a significantly different covariance. Conversely, objects that are placed on the CA list because the DoD is struggling with adequate tracking are likely too small or problematic for the current suite of commercial sensors to acquire. These objects are either not tracked at all by the pathfinder or show no tracking gain and thus identical performance pre- and post-prioritization.

This explanation is supported by the outcome of a different usage mode of the mission planning functionality, which was introduced about two months into the pathfinder conduct, namely to additionally consider objects that did not derive from the DoD event histories but rather were identified as objects that the pathfinder catalogue maintenance function was maintaining, but not particularly well. Some number of these underserved objects were placed on the mission planning list each day, and the response was monitored. Figs. 4 and 5 provide the same two graphs as were shown in Figs. 2 and 3 except reporting on performance on these underserved pathfinder objects rather than those from the DoD list.

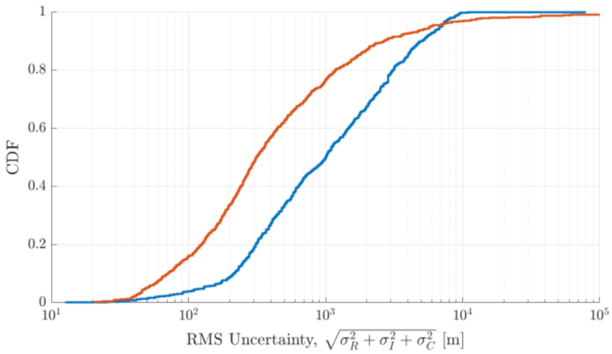


Fig. 4. Underserved pathfinder object covariance size (via RMS uncertainty) before (blue) and after (red) prioritization.

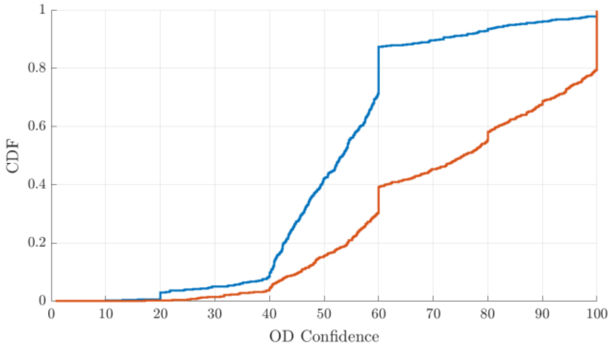


Fig. 5. Underserved pathfinder OD confidence before (blue) and after (red) prioritization.

For objects for which the pathfinder's sensors had likelihood of acquisition, but which had been previously not received adequate tracking, the mission planning paradigm worked well. There is a significant improvement in both covariance size and OD confidence score once such objects get placed on mission planning list and were prioritized for additional tracking.

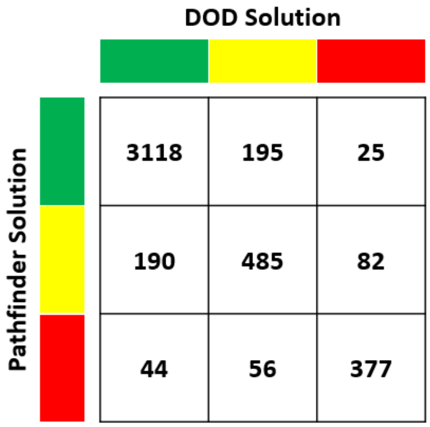
The conclusion from these results is that the "chase and improve" approach to improving data holdings for the secondary objects for in-progress CA events would likely be of questionable efficacy. Events of interest based on the DoD catalogue do not seem to be improved by this concept of operations. Either the objects are already tracked well and therefore do not seem to benefit appreciably from the additional tracking, or they are difficult for the DoD to track and therefore usually below the threshold of detection of the present suite of commercial SSA sensors. However, for states that were based on the catalogue maintained by the pathfinder, the successful use of "chase and improve" showed some ability for the pathfinder sensors to detect and track. Consequently, the process itself is not fundamentally flawed, but its usefulness for TraCSS—namely in response to issues with the DoD states on the secondary objects—it did not present palpable advantages.

Regardless of whether or not surge tasking was in fact employed for any given event (and part of the mission planning algorithm was to determine whether for any particular DoD CA event a tracking increase was in fact advisable), all of the DoD CA events of interest were processed by both the DoD and the pathfinder, and independent solutions for these events exist and can be fruitfully compared. In order to standardize the event processing and reduce the number of free variables, only events with O/O solutions for the primary object were examined comparatively. This forced an essentially identical solution for the primary object, allowing any difference in CDM information to be driven by the differing solutions for the secondary object.

For identical events, comparing the CA results to solutions derived from two different systems is complicated and different approaches are often taken depending on the level of detail desired for each event. In summarizing results for entire groups of events, it has been found to be useful to assign each event a colour, based on the level of the Pc at the "maneuver commitment point" (the point at which a mitigation action decision must be made), and examine instances of difference in assigned colour between the two systems. A good maneuver commitment point to use is 12-24 hours before the time of closest approach (TCA) between the two satellites, a time used by most current operational missions. The colour assignments are those used by NASA CARA as follows:

- Red events are those with a Pc greater than 1E-04, which usually require a risk mitigation action;
- Yellow events are those with a Pc between 1E-07 and 1E-04, which require more engaged monitoring but do not usually result in a risk mitigation action; and
- Green events are those with a Pc less than 1E-07, which are usually not given any further operational regard.

Fig. 6 gives a “confusion matrix” that compares the results for more than 4000 events that were placed on the list that was furnished daily to the pathfinder. If both systems agreed completely on the colour-based results, all of the entries would fall along the diagonal. Off-diagonal terms indicate disagreements, with the most extreme disagreements in the upper right and lower left corners. These represent situations in which one system would have counselled a mitigation action and the other system would have dismissed the event entirely. These extreme situations represent about 12% of the cases in which at least one of the two systems assigned a red status [(25+44)/(25+44+82+56+377)]. This level of divergence is not crushing, as different systems often produce somewhat different solutions; but it is still disturbing that one in ten of the very serious events identified by one of the two systems also manifest the other system’s calling for an outright dismissal.



	🟩	🟨	🟥
🟩	3118	195	25
🟨	190	485	82
🟥	44	56	377

Fig. 6. Confusion matrix comparing both systems’ results for a maneuver commitment points between 12 and 24 hours.

The source of these disagreements and whether the divergence can be traced to issues with one of the two systems is, of course, important. The topic will be treated in more depth in Section 4 as the concepts and results concerning the LEO catalogue maintenance will be deployed to help resolve this divergence in event risk assessments.

4. LEO catalogue maintenance

As was remarked earlier, a tertiary objective of the pathfinder was to maintain a LEO space catalogue entirely with commercial tracking data and orbit determination. For the purposes of the experiment, the LEO catalogue was limited to the public space catalogue of objects with perigee values greater than 300 km and apogee values less than 2500 km. Three main metrics were applied in evaluating the quality of the pathfinder catalogue:

- Catalogue completeness. How many objects in the LEO regime (as restricted for this experiment) was the pathfinder able to maintain well over its three-month operation, and what percentage of the DoD public catalogue did that constitute?
- Catalogue object prediction error. What levels of prediction error did the pathfinder predicted ephemerides manifest, and how did this compare to DoD performance?
- Catalogue predicted covariance realism. How realistic were the covariances that the pathfinder catalogue generated and placed in their predicted ephemerides, and how did this level of covariance realism compare to that of the DoD?

The DoD catalogue performance levels are not publicly releasable, so they cannot be stated explicitly in this paper. However, it is possible to make certain statements about the relative performance that, while not disclosing DoD performance levels, do give insight into pathfinder accomplishments.

4.1 Catalogue completeness

The DoD catalogue is the commonly-applied standard for catalogue completeness. Thus, in examining the pathfinder’s ability to maintain an independent catalogue, it makes sense to compare its catalogue holdings to those of the DoD. One needs first to identify which objects are “well maintained,” both in the DoD public catalogue and the pathfinder catalogue, so that only those objects will be counted and compared. The mere publication of a TLE in the DoD public catalogue is not a testament to the object’s regular maintenance, as some objects have epoch times that are months or years old and are therefore not under regular maintenance. Therefore, it was necessary, to establish a standard for “well maintained” that could be applied to both the DoD public catalogue and the pathfinder catalogue, and the sets of satellites that each catalogue maintained according to this standard could be compared for overlap. The precise standard that was applied cannot be stated here, but it involved evaluating a combination of tracking rate and state update rate.

Table 1 shows the number of objects that appeared in either catalogue, well maintained or not, and the counts of well-maintained objects in different levels of intersection between the two catalogues:

Table 1. Catalogue maintenance comparison between pathfinder and DoD.

Category	Count	%
Appeared in either catalogue	21704	100.0
Well maintained in both catalogues	15119	69.7
Well maintained in DoD catalogue only	2946	13.6
Well maintained in pathfinder catalogue only	178	0.8
Not well-maintained in either catalogue	3461	15.9

The results are overall very good for the pathfinder. If objects that neither DoD nor the pathfinder maintained well are removed (last row in Table 1), the pathfinder maintained 84% of the well-maintained DoD public catalogue. It was expected that some portion of the DoD catalogue would not be maintainable by the pathfinder because the pathfinder sensor suite is known not to be as sensitive as the most capable DoD radars, and indeed the great majority of the ~3000 objects maintained well only by the DoD are less than 20cm in estimated size. The ~200 objects maintained well only by the pathfinder are largely situations in which the DoD misses the "well maintained" standard by a relatively small amount, so they do not constitute some sort of unique DoD deficiency. On the whole, the pathfinder performance was quite good: objects within the detectability limits of the pathfinder sensors were tracked and maintained regularly.

4.2 Prediction error

A catalogue with large prediction errors is not very useful for CA. Even if the associated covariances are realistic, the covariances' appropriate large sizes will depress the Pc values to low levels, potentially masking actual collision risks. It is therefore important that prediction errors be kept reasonably low.

The standard way catalogue prediction error is assessed is after the fact. Predicted ephemerides are produced (at least once daily) and saved. Then, a definitive, "as-flown" ephemeris is constructed from tracking data or from on-board GNSS / telemetry, if available (active payloads. The two ephemerides are then compared at propagation states of interest, such as 1, 2, and 3 days' propagation time. The set of residuals from these comparisons are then summarized statistically, both by individual position component (usually in a satellite-centered frame, such as RIC) and overall vector magnitude of the position difference.

The DoD has deployed software for some years that automatically constructs such a reference orbit for every object. Similarly, the pathfinder OD process was asked to produce a definitive ephemeris for every object they were attempting to maintain. Because the production of definitive ephemerides was not part of any of the participants' usual activities outside of this pathfinder, a series of scripts had to be developed during the experiment period. Consequently, the calculation of definitive ephemerides was not as reliable as one might wish: about 4000 (of the 15119) objects in the pathfinder catalogue lacked reference orbits, and thus, prediction error comparisons were not possible for them. Fortunately, the great majority of orbits were appropriately provisioned, so a broad analysis was still possible.

For orientation, one must examine the reference orbits themselves before using them to evaluate prediction error. The first step was to investigate DoD/pathfinder reference orbits for objects for which there are precision external ephemerides—in both cases, it is these objects that are typically used to calibrate individual sensors. For calibration objects, both the DoD and the pathfinder performed extremely well. The DoD's residual set was slightly smaller, but that is probably due to the availability of a much larger set of tracking data, however the differences between the two residual sets were too small to be of any real operational consequence. The second step was to investigate, by direct comparison, the catalogue-wide sets of reference orbits on all object types. The comparison revealed that occasionally reference orbits get misformed and therefore some outlier exclusion was necessary. After excluding suspect reference orbits, the performance of the DoD and pathfinder catalogues aligns reasonably well, as shown in Fig. 7, grouped by object type. It is still somewhat disappointing that ~25% of the overlaps investigated do show a difference of more than 100 m, but it is also not entirely unexpected. Additionally, overlaps showed smaller differences for the rocket body and unknown objects categories than those for the debris and payload categories, indicating poorer predictions on debris and payload objects.

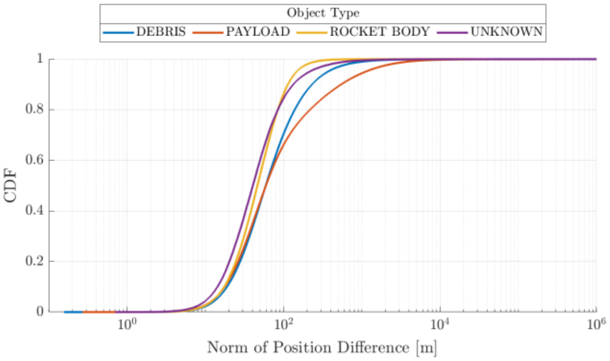


Fig. 7. Position differences between DoD and pathfinder definitive ephemerides, with outliers excluded.

The final activity was to compare DoD and pathfinder predicted ephemerides to their corresponding definitive ephemerides at propagation states of interest. The DoD tool that produces reference ephemerides for all catalogued objects also evaluates every vector that the DoD system produces. The tool does this ex post facto by producing a predicted ephemeris from the vector (using the predicted space weather values valid at the epoch time of the update) and comparing the predicted ephemeris to the reference ephemeris at particular propagation states. The pathfinder evaluation team received these results directly from DoD and therefore did not actually receive the DoD predicted ephemerides. Thus, there was a "mix-and-match" limit on definitive and predicted ephemerides: the pathfinder predicted ephemerides could be evaluated against both the DoD and pathfinder definitive ephemerides, but the DoD predicted ephemerides could be evaluated only against the DoD definitive ephemerides, because this evaluation was performed on the DoD side, using their operational tool. In addition, comparisons between predicted and definitive ephemerides were limited to objects that both the DoD and the pathfinder had under maintenance, and one prediction per object per day was evaluated (results from the first prediction of the Zulu day were chosen from the DoD side). Finally, the evaluation comments were derived only from "in-family" comparisons: pathfinder predicted to pathfinder definitive ephemerides and DoD predicted to DoD definitive ephemerides.

Due to releasability considerations, actual numerical data for prediction error results cannot be shown here. However, what can be said is that pathfinder performance levels for prediction error, for the objects that could be evaluated (recall the ~3000 objects maintained only by DoD and the ~4000 objects for which the pathfinder did not produce definitive ephemerides), were "substantially equivalent" between the pathfinder and the DoD. For well tracked objects at short propagation intervals, the pathfinder's performance was somewhat better. For less well-tracked objects or for objects at longer propagation intervals, the DoD's performance was better. In both cases, however, the differences were not substantial.

4.3 Covariance realism

It is only in the last decade that significant attention has been directed to the quality—that is, the realism—of the covariances associated with precision state estimates. This is partly due to the long legacy of the use of analytic theories such as SGP4, which does not provide a covariance; and the relatively recent advent of more sophisticated orbital safety calculations that require a realistic covariance in order to be meaningful. It is important that orbital safety applications be furnished both a realistic covariance and a low-error predicted state in order to calculate a probability of collision that can truly guide risk assessment. Many CA practitioners would claim that of the two the realistic covariance is the more important.

The first major synthetic publication on uncertainty realism for aerospace applications was by Poore et al. [2], and since that time, multiple publications (*e.g.,* [3, 4]) have documented different approaches both to assess and improve covariance realism, which is the strain of uncertainty realism most immediately transferrable to current OD methods. The general approach used in evaluating the realism of both DoD and pathfinder covariances employs the method of Hejduk et al. [5], which is somewhat enhanced to provide comparison statistics that are more informative. Both the general method and the enhancements will be described here. Only the position portion of the covariances was evaluated, both because only that portion is needed for the two-dimensional Pc calculation and because the evaluation is greatly simplified.

Comparisons between predicted and definitive ephemerides evaluate how well the predicted ephemeris actually models the satellite's actual trajectory; and these comparisons at a particular propagation state create a set of residuals that document these differences. The predicted covariance at that propagation state, if realistic, will statistically characterize the expected residual set arising from the comparison described above. Because a fully-capable OD should solve for all systematic errors, leaving only Gaussian noise, it is expected that the errors in each of the components conform to a Gaussian distribution; and if each residual is normalized by the variance in its associated combined covariance (that is, the sum of the propagated covariance and the formation covariance of the reference orbit), a normalized Gaussian distribution (zero-mean, unity variance) should be produced. The sum of normalized Gaussian variables will constitute a chi-squared distribution, with the number of degrees of freedom dictated by the number of normalized variables summed. The matrix product $\varepsilon * C^{-1} * \varepsilon^T$ for each ordered triple of residuals (ε) and each covariance (C) for a particular propagation state will produce the square of the Mahalanobis distance for that ordered triple, and a set of such Mahalanobis distances for a group of ephemerides evaluated at the same propagation state (for either the same object or for a class of objects) should conform to a 3-DoF chi-squared distribution. A distribution matching technique is used to determine whether the Mahalanobis distance set actually matches the 3-DoF chi squared distribution; the present approach employs the Cramér – von Mises (CVM) quadratic empirical distribution function (EDF) goodness-of-fit (GOF) test [6], nominally at a significance probability of 0.02. For calculated p-values above this significance level, a distribution match is presumed, and the covariance set is considered realistic.

There are additional considerations that have led to emendations/alterations to the above procedure. First, the quadratic EDF test is sensitive to outliers (although not as strongly as supremum versions, such as the Kolmogorov-Smirnov test) and large datasets, and a good way to reduce this vulnerability is through a resampling technique. For example, if a set of 100 ordered triples is available to examine, rather than evaluating the entire set of 100 at once, one might randomly choose 30 ordered triples from the set of 100 and test that set, repeating this procedure (with replacement) 1000 times and then examining the percentage of the 1000 evaluated sample sets that passed the test. This approach keeps a small number of outliers from determining the outcome for the entire residual set. Additionally, comparison between datasets is aided by calculating not just simple pass/fail statistics but additionally with each resampled set, a scale factor by which the covariances could be multiplied in order to produce an evaluation as close as possible to the idealized result, which here means minimizing the Cramér – von Mises test statistic. Thus, instead of just performing the GOF test with each resampled residual set, one calculates a scale factor by which one would multiply the covariance in order to minimize the test statistic; and the viewed results are these scale factors rather than actual pass/fail criteria.

Although the scale factor approach is not strictly an evaluation that can testify to passing and failing (*e.g.,* the minimum CVM value for a residual set might not actually be a passing condition, and a potentially large-scale factor might be required to achieve a minimum for a situation that was already passing at a 0.02 significance level), it is nonetheless helpful and illuminating. Scale factors of unity typically indicate that the furnished covariances are fine as they are—no manipulation is necessary to produce a properly-sized covariance. As the scale factor deviates from unity, one is given a sense of how "unrealistic" the current covariance is, i.e., a further multiplicative distance from unity maps to a more unrealistic situation. The deviation also shows in which direction the error lies, i.e., a greater-than-unity scale factor indicates that the current covariance is too small. This method is the one employed for the evaluation of the DoD and the pathfinder covariances, and it seems well suited to a relative comparison. A more complete explanation of these additional considerations, as well as source code that performs all of the calculations described above, is provided on the NASA CARA GitHub[†].

Just as with prediction error, releasability restrictions prevent a presentation of actual numerical results; but a heuristic statement of the results is possible. Over the past decade, the DoD has made a substantial investment in improving covariance realism for LEO orbits, especially in modelling atmospheric density forecast area in prediction. The DoD covariance realism results were likely superior to those of the pathfinder due to these improvements. Often the median scale factors were comparable between the DoD and the pathfinder, although there were cases in which one was superior to the other. However, in all cases the interquartile range for the scale factor was smaller for the DoD, and usually much smaller. Basically, the DoD's OD and the ability to compensate through consider parameters for atmospheric density forecast error and satellite frontal area uncertainty gives the DoD covariances a level of stability that the pathfinder covariances did not possess.

This result is not entirely surprising, for two reasons: First, the DoD has targeted covariance realism, especially in LEO, as a focus for research and development, whereas the pathfinder OD vendor has not; and second, the pathfinder conduct period was not really long enough to allow a process noise tuning campaign to be conducted to improve covariance realism. The data types processed by the pathfinder, namely a combination of LeoLabs radar and Slingshot optical measurements, have never before been processed together for any large collection of satellites, thus, it is not surprising that the set of OD settings that the COMSPOC brought to the pathfinder were not optimized for this particular data density and mixture. Regardless of the OD settings that were used, however, some modest level of investment in space-weather-connected process noise modulation would be needed. For example, with the availability of an openly available higher-fidelity atmospheric density model, OD improvements that are equivalent to current DoD results are not seen at all as out of reach of the current commercial SSA industry. Of course, one must remain mindful of the ~7000 objects that were not evaluated, due to either the lack of a reference ephemeris or their presence in only one of the two catalogues (DoD or pathfinder) but not both.

5. Individual event performance

Having discussed the levels of performance for the overall LEO catalogue maintenance activity, it is time to circle back to the issue of individual event performance, which was summarized earlier by the confusion matrix (Figure 6). The covariance realism performance of the DoD was somewhat better; thus, one might be inclined to favour the DoD solution for individual events. However, high-level trends do not necessarily translate to individual event comparisons. Two events will be profiled below that show that, with the benefit of ex post facto analysis that allows the prediction error and covariance realism for those particular secondary objects to be evaluated, one can identify the preferred solution for each particular event. The preferred solution can come from either the DoD or the pathfinder catalogue. The particular satellites and dates have been redacted to allow the actual performance data to be presented.

5.1 Event #1

For this event, Fig. 7 shows the time series of Pc values. For this example, one should concentrate on the dark blue solution (DoD CDM based on O/O primary and DoD secondary), which never exceeds the red threshold of 1E-04, versus the light blue solution (O/O primary versus pathfinder secondary), which shows a Pc value above the red threshold and therefore in principle would counsel a mitigation action at the 1-day-to-TCA

[†] https://github.com/nasa/CARA_Analysis_Tools

maneuver commitment point. Thus, there is a difference in operational actionability in this event: one source would counsel a maneuver, and the other very much would not.

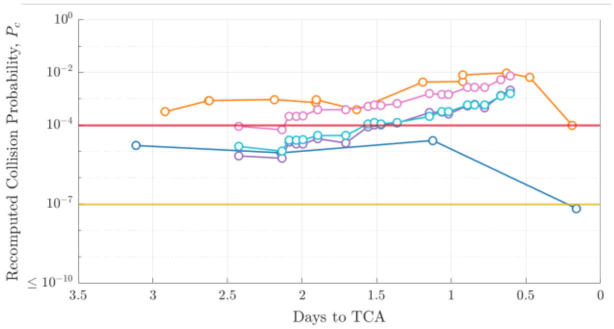


Fig. 8. Pc evolution for different input data arrangements; dark blue is DoD and light blue pathfinder.

Because the focus is on CDMs that are using O/O ephemerides for the primary and catalogue solutions (DoD or pathfinder) for the secondary, one looks to historical performance information on this secondary object in order to choose a favoured solution. Fig. 8 gives a CDF of prediction errors for the secondary, and Fig. 9 a box-and-whiskers summary of covariance realism scale factors for the secondary.

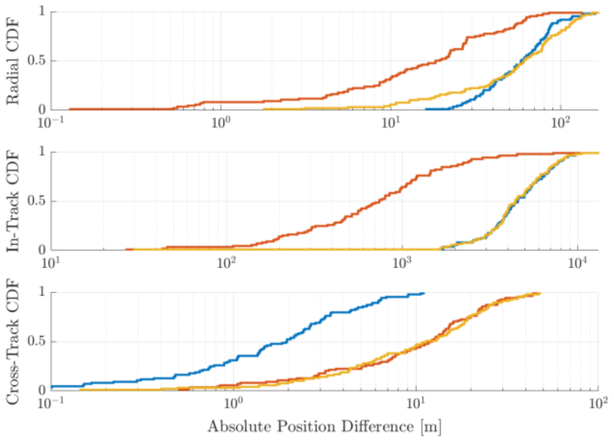


Fig. 9. CDF of prediction errors by component for Event #1. Blue line is pathfinder; red line is DoD.

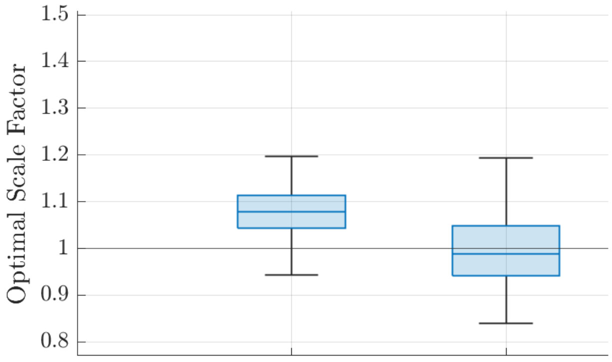


Fig. 10. Box-and-whiskers plot of covariance realism scale factors for Event #1. Left side is pathfinder; right side is DoD.

For this event, the prediction errors are smaller for the DoD solution, especially for the radial component, which tends to be the determinative in the calculation of the Pc. The covariance realism results are quite similar, with the DoD results for this satellite also somewhat better. In an actual adjudication, one would want to investigate the recent tracking density, sensor mix, and OD goodness-of-fit indices in order to understand the integrity of the actual final correction for the secondary for each side. These being comparable, however, historical accuracy and realism information would favour the DoD solution for this event.

5.2 Event #2

For this event, the focus is again on the dark blue (DoD) and light blue (pathfinder) lines (here in Fig. 10); and at a 12-hour maneuver commitment point, the DoD results will counsel a mitigation action as the Pc falls above red line, whereas the pathfinder results will not.

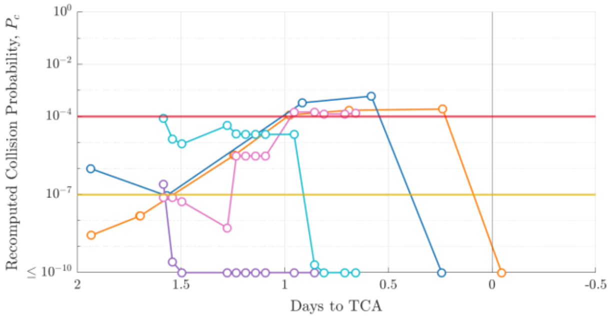


Fig. 11. Pc history graph for Event #1; dark blue is DoD, and light blue is pathfinder.

Results for prediction error and covariance realism, given in Figs. 11 and 12, give a different interpretation here than for Event #1: the pathfinder fared better in prediction for this secondary satellite and covariance realism results were essentially equivalent.

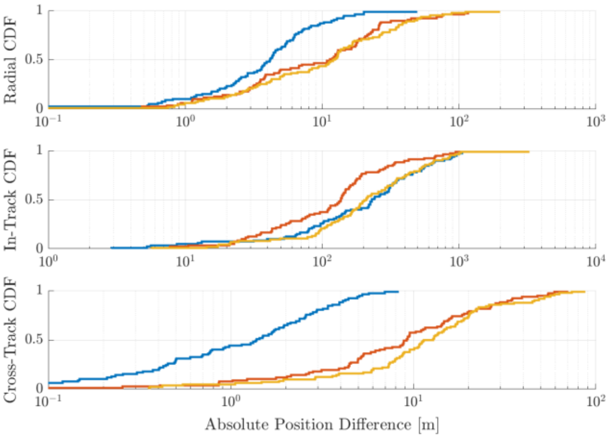


Fig. 12. Prediction error by component for pathfinder (blue) and DoD (red).

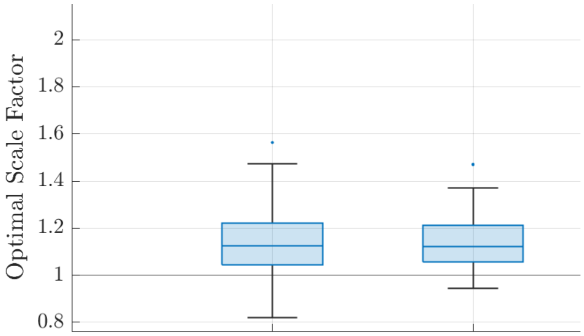


Fig. 13. Covariance realism scale factor distribution for pathfinder (left) and DoD (right).

In this case, in the absence of any other data pertaining to the particular OD fit for the final Pc calculation, or any direct maneuver detection and recovery information, one would embrace the pathfinder results and refrain from a mitigation action.

The purpose of these two short vignettes is to show that general performance data are not necessarily indicative of the performance situation for any particular event. Historical performance information for particular objects is required, and even then, there will be indeterminate situations. This result points to a more general problem with multiple CA solutions, namely that different catalogues maintained with different data will frequently produce operationally significant differences in results without any regularly reliable method for choosing a preferred solution. As the CA solution space continues to federate, with different providers offering SSA products and CA solutions from those products, the operational confusion resulting from multiple solutions will only increase.

6. Conclusions

The Office of Space Commerce has obtained the following information from this pathfinder exercise:

- It is possible to execute an orbital safety effort, for a portion of on-orbit assets that commercial sensors are capable of tracking, using commercial SSA data with "competing" companies working collaboratively under a public-private partnership to perform difficult tasks. Future acquisitions can be managed via contractual metrics, and a reasonable set of such metrics, and good initial values for associated threshold, have been established from this effort.
- The concept of "chase and improve" commercial tracking augmentation for bringing clarity to CA events is under strain. This approach was not entirely successful in the pathfinder. Event objects either were already well tracked and therefore not ameliorated by additional tracking, or they were not well tracked and therefore not acquirable when additional tracking was attempted.
- For objects that the pathfinder data providers could track regularly, catalogue maintenance efforts were successful, producing prediction error results that were "substantially equivalent" to the DoD. Covariance realism is an area in which performance lagged vis-à-vis the DoD and could therefore benefit from commercial investment and attention.
- Different orbital safety systems produce different orbital safety results, and it is often not obvious which solution out of a collection of solutions produced by different entities should be favoured. In the present pathfinder, a non-trivial number of situations emerged in which one system counselled mitigation action and the other system a complete dismissal of the event. Further analysis showed that the source of the favoured result (DoD vs pathfinder) varied case by case. However, no effort has been made to date to quantify at a catalogue-level what percentage of the discrepancies favoured each catalogue. This non-alignment of solutions is a serious problem for orbital safety operations and will only grow as more independent entities issue competing CA solutions.
- It is important to remain mindful that there was a ~3000 object decrement between the pathfinder and DoD unclassified catalogues, that ~4000 satellites on the pathfinder side remained unevaluated due to reference orbit construction issues. In addition, the DoD reference orbits were not available for all objects, and the pathfinder data providers relied on DoD TLEs for detection cueing. These issues prevented a fully definitive evaluation of the pathfinder as an independent data augmentation and catalogue maintenance entity. However, within the limitations stated, the data collected were adequate to draw the general conclusions outlined above.

References

[1] S. Magnus, M. Hejduk, J. Johnson, J. Giles, A. Parikka, C. Joseph, C. Coursey, M. Lifson, A. L. Antunes de Sá, D. Baker, M. Duncan, C. Grey, R. Hall, S. Johnson, P. Mesalles-Ripoll, D. Oltrogge, J. Stauch, D. Strobel, The TraCSS Consolidated pathfinder: Leveraging Commercial Capability in LEO, 25th Advanced Maui Optical and Space Surveillance Technologies Conference, Maui, Hawaii, USA, 2024, 17 – 20 September.

[2] A.B. Poore, J.M. Aristoff, J.T. Horwood, Covariance and Uncertainty Realism in Space Surveillance and Tracking, Report of the Air Force Space Command Astrodynamics Innovation Committee, 2016, 27 June.

[3] A. Cano, A. Pastor, D. Escobar et al., Covariance determination for improving uncertainty realism in orbit determination and propagation, Advances in Space Research, https://doi.org/10.1016/j.asr.2022.08.001.

[4] M. Schubert, C. Kebschull, S. Horstmann, Analysis of different process noise models in typical orbit determination scenarios. 8th European Conference on Space Debris, vol. 8. ESA Space Debris Office, Darmstadt, Germany, 2021.

[5] M. D. Hejduk, D. Plakalovic, L. K. Newman, J. C. Ollivierre, M. E. Hametz, B. A. Beaver, R. C. Thompson, Trajectory Error and Covariance Realism for Launch COLA Operations, AAS 13-355, 23rd AAS/AIAA Spaceflight Mechanics Meeting, Kauai, Hawaii, USA, 2013, 10 – 14 February.

[6] R. D'Agostino, M. Stephens, Goodness-of-fit techniques. Statistics: Textbooks and Monographs, vol. 68. Marcel Dekker Inc, New York, 1986.