Appendix E. Use of Analytic Theory Orbital Data in Conjunction Assessment

The following is an amplification of a statement made by the Conjunction Assessment Technical Advisory Committee in 2016 on the use of analytic theory for conjunction assessment, and in particular the Two-Line Element (TLE) set publicly accessible catalog produced by USSPACECOM.

Due to enormous increases in computational capacity over the last twenty years, the space surveillance industry has been able to transition from analytic to higher-order theory approaches to space catalog maintenance. This transition brings a number of significant advantages to modern space catalogs such as the modeling of additional perturbations that were not typically represented in analytic theories (e.g., solar radiation pressure and solid earth tides), an improved fidelity in modeling the major perturbations (e.g., geopotential and atmospheric drag), and a durable covariance matrix to accompany each state estimate. Some of these features existed with previous analytic and semi-analytic theories, but it is only in recent times that all the additional attributes are routinely produced and available for space surveillance applications. This development has prompted the question of how data from analytic models (such as the Simplified General Perturbations Theory #4 [SGP4], which is used to produce the publicly available TLE catalogs) and higher-order theory models (such as the USSPACECOM Special Perturbations (SP) Space Object Catalog, although higher-order theory data can also be furnished from other sources) should be employed in the conjunction risk assessment process.

Risk assessment, whether it is quotidian consumer risk, the extremely serious discipline of nuclear accident risk, or the present problem of satellite collision risk, must always contain within its calculus some determination of the likelihood of occurrence of the event in question. For CA, this determination is most commonly assessed by calculating the Probability of Collision (Pc; see Appendix I for a development of the calculation theory), and this calculation requires both the state estimates for the primary and secondary objects and the accompanying measures of these estimates' uncertainties, typically through a covariance matrix, at the time of closest approach (TCA).

Most implementations of analytic orbital theories lack the ability to produce a meaningful covariance (this is true of the SGP4 TLEs), so they cannot enable a probabilistic conjunction assessment calculation and thus for conjunction assessment applications will yield only a miss distance at the time of the two objects' TCA. Because no uncertainty information on this miss distance is provided, no probabilistic conclusion can be drawn. A miss-distance threshold can be (arbitrarily) selected to enable a binary "safe / not safe" conjunction evaluation, but there is no ability to determine, for example, how often the real miss distance, for which the nominal miss distance is only an estimated mean value, is actually likely to exceed or fall below such a threshold. No probabilistic assessment of whether a conjunction will be a dangerously close encounter is possible, so the likelihood of occurrence of a collision cannot be determined.

This dynamic is illustrated in Figure E-1, which gives a Cumulative Distribution Function (CDF) plot of miss distance distributions, both for critical (Pc > 1E-04) and non-critical (Pc < 1E-04)

events. The lack of any strong correlation between Pc and miss distance is clear: the span of miss distances for the critical events (less than 100 m to more than 10 km) is extremely broad and overlaps significantly with the results for the non-critical events. For example, to eliminate 95% of the high-Pc events using a miss-distance criterion alone, a 10 km miss-distance threshold would need to be chosen—a value that would also include more than 50% of the non-worrisome events, creating a density of events to mitigate so great that an unsustainable number of such mitigation actions would be required.

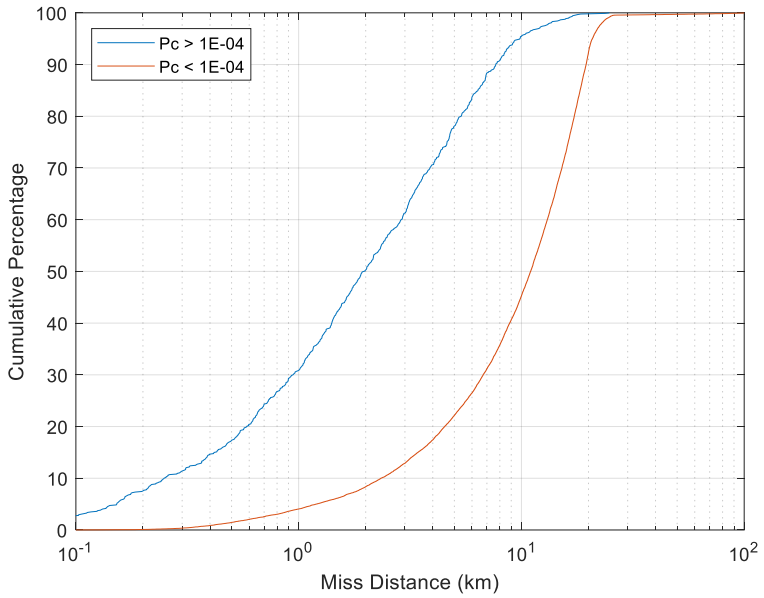


Figure E-1 Miss-Distance Distributions for Critical (Pc > 1E-04) and Non-critical (Pc < 1E-04) Events

Additionally, the inherent theory error in general perturbation approaches makes them ill-suited for even a risk-tolerant, miss-distance-based conjunction assessment. A published study (Hejduk et al. 2013) provided TLE accuracy information for the two methods presently used by USSPACECOM to generate them: traditional SGP4 orbit determination and Extrapolated General Perturbation (eGP) TLE generation, which performs the SGP4 orbit determination not against sensor measurements but against pseudo-observations generated from a higher-order theory trajectory for the object, which includes some forward prediction of the trajectory. Both methods are in use presently to generate the publicly-released TLE catalog, and it is not always easy to determine which method was used to produce any given TLE. Performance data for the orbital regime corresponding to Figure E-1 above (LEO with perigee heights greater than 500 km) are given in Figure E-2. These results show that, while the eGP methodology outperforms traditional SGP4 as expected, both are noisy in the epoch to 72-hour propagation states, which is the time-frame during which most conjunction risk mitigation decisions are made; errors from 600 m to 3 km are observed.

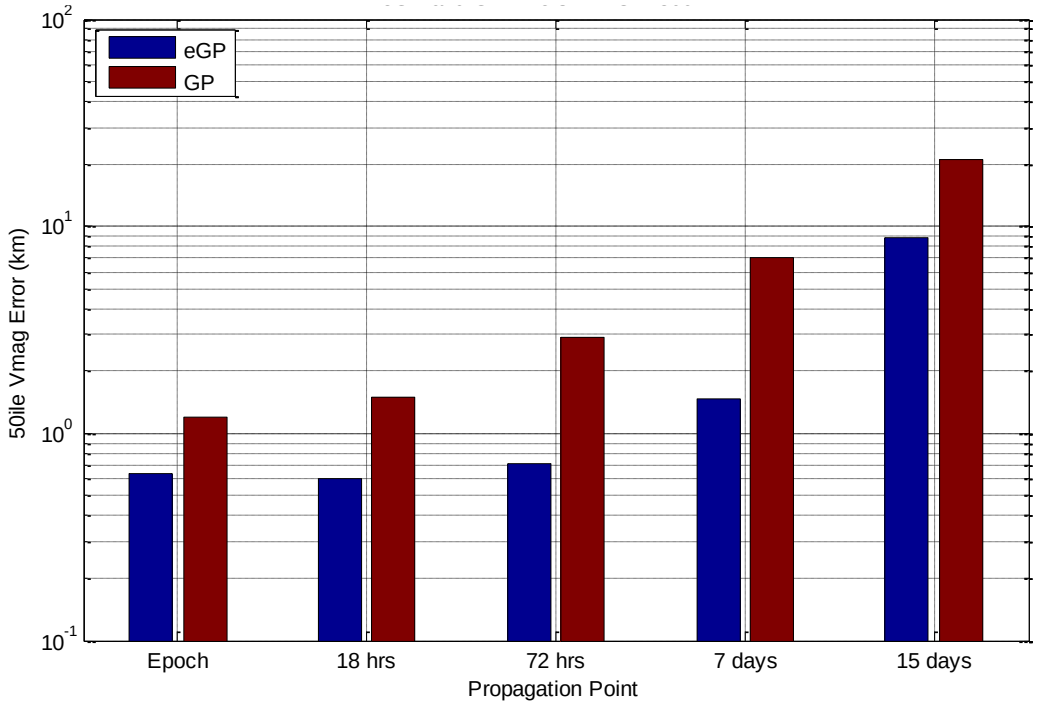


Figure E-2 eGP and Regular SGP4 Errors: LEO > 500 km (log scale)[11]

Figure E-3 below is an expansion of Figure E-1 to include CDF plots of miss-distance distributions for events with both Pc > 1E-03 and, separately, 1E-02. Figure E-3 suggests that a miss-distance criterion of several hundred meters, or perhaps 1-2 km, could be used as a simplified method of performing a risk-tolerant conjunction assessment: remediate any conjunctions with a miss distance less that 1-2 km. This move is ill-advised because it would result in a huge number of unnecessary mitigations (as an example, for events with a miss distance less than 1 km, the number with a Pc less than 1E-03 and thus not serious is six times greater than the number with a Pc greater than 1E-03). In addition, the inherent theory error (600 m to 3 km, from above) is a substantial portion of the miss-distance threshold of 1-2 km that would be imposed—in some cases even greater than the threshold itself. So even under the very limited case described here, the use of general perturbation results for conjunction risk assessment is not tenable. Thus, analytic theory data, such as that contained in the public TLE catalogs, should not, taken alone, serve as a basis for risk assessment decisions and subsequent mitigation actions.

[11] From Hejduk et al. 2015.

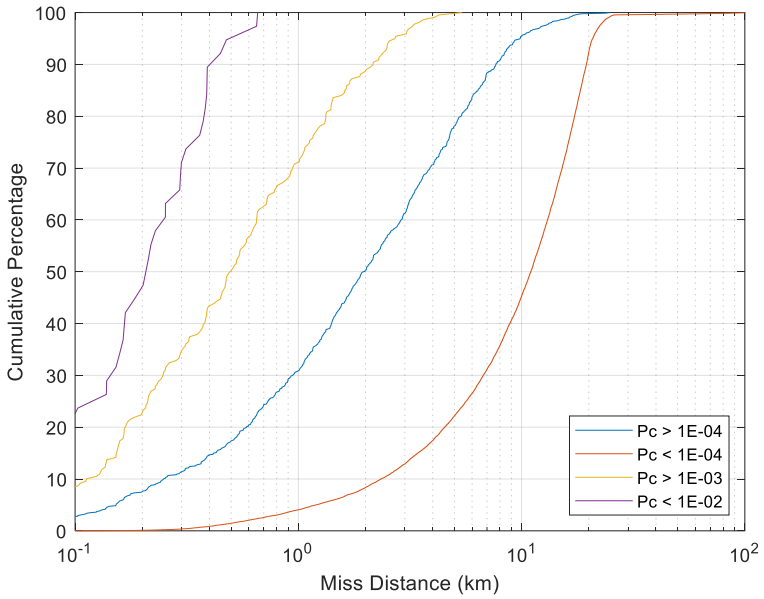


Figure E-3 Expansion of Figure E-1 to Include CDFs for Pc Thresholds of 1E-03 and 1E-02

Nonetheless, analytic theory orbital data, usually furnished as TLEs, does have its uses in conjunction assessment; some of these have included the following:

- Mission planning activities such as the selection of satellite future orbits and final disposal orbits;
- General situational awareness information such as the positions of neighboring objects and the general effects of the basic perturbations (e.g., atmospheric drag);
- Pre-filtering of candidates for conjunction assessment screenings for which simple TLE-based filters can eliminate substantial numbers of candidate pairs from requiring explicit screening runs;
- Assessment of the status of a payload that appears as a conjuncting secondary, as active secondaries require special analysis and alerting when performing conjunction remediation; and
- Consistency checks of the higher-order theory data. While the higher-order theory data are typically more accurate, they themselves are vulnerable to updates with bad or inadequate tracking data, errors in force model and integration control settings, and incorrect maneuver solutions. Comparisons with the analytic data can often identify such problems and, while not enabling their repair, can at least make clear that such cases constitute non-actionable conjunction assessment data.

Although the above uses of analytic theory data can be tangentially helpful to the conjunction assessment enterprise, these uses do not themselves directly support conjunction risk assessment, for which more accurate state estimate data accompanied by uncertainty information is required.

Appendix F. Expected Conjunction Event Rates

The conjunction assessment workload that a mission experiences is a function of:

- The total number of CDMs that it receives,
- The number of associated events with risk levels high enough to require enhanced monitoring, and
- The number of events that have a risk level so high as to require mitigation planning and potential execution.

During mission planning and development, it is helpful to have a general idea of the frequency of such situations to model spacecraft fuel consumption, identify staffing needs, and determine the appropriate degree of conjunction assessment-related tool automation.

The following profiling of conjunction assessment event densities is intended to provide at least some first-order information to answer these questions. These data summarize three years' recent conjunction assessment data history for the payloads that NASA protects in the LEO orbit regime (event densities in the non-LEO region are both low and highly dependent on the specific orbit, so summary information is not given here). In November 2021, Russia conducted a kinetic anti-satellite (ASAT) test that destroyed a derelict Russian satellite at approximately 480km, which generated an additional 1500 cataloged debris fragments in LEO. This increase is reflected in the three-year averages given below, although by 2024 most of these fragments will have decayed.

The graphs presented in Figure F-1 contain several component parts.

The x-axis gives the hard-body radius for the conjunction, which is essentially the combined sizes of the primary and secondary objects. Although different methods are used to determine this size, perhaps the most common is to construct a circumscribing sphere about the primary object and add its radius to a standard radius size for the type of secondary object (i.e., payload, rocket body, or debris) in the conjunction (Mashiku and Hejduk 2019). Hard-body radius (HBR) values of 20 m, 10 m, 5 m, 2 m, 1 m, and 0.1 m are used, the last few values included to give some statement of the expected situation for small satellites. All the CARA events were reprocessed with these different HBR values to produce these varied results.

The y-axis gives the number of expected events per payload per year, shown on a logarithmic scale, with the type of event indicated by the color in the stacked bar charts, which are to be read in the following way:

- The height of the red bar indicates the frequency of red events, which are defined as events containing at least one CDM with a Pc value > 1E-04. These are the most serious events; they require detailed monitoring, mitigation planning, and in a minority of cases, the actual execution of a mitigation action.

- The height of the yellow bar indicates the frequency of yellow events, which are defined as events containing at least one CDM with a Pc value > 1E-07. These events generally merit increased monitoring and attention, including in some situations, mitigation

planning (if they are close to the red threshold). Based on the definition above, all red events are also yellow events.

- The height of the green bar indicates the frequency of events of any type, including those that never manifest a Pc > 1E-07. Based on this definition, all red and yellow events are also green events.

The overall height of each bar thus indicates the CDM generation rate. Note that this number, while in some ways a useful loading figure, is a function of the size of the screening volume that is employed and thus is not that meaningful in itself. For a screening volume similar in size to those used by NASA CARA, the green event generation rates (which include the yellow and red events) are likely to be similar. Because they require relatively small miss distances, the yellow and red generation rates are expected to be largely independent of the screening volume size selected as long as the volume used is reasonably-sized.

Three different graphs are presented representing three different regions of LEO sorted by perigee height (p): a high-drag regime (p < 500 km), a low-drag regime (perigee height between 500 and 750 km) that includes a large concentration of satellites somewhat higher than 700 km, and a second low-drag regime from 750 to 1400 km. There are moderate differences in event frequency among the three different regimes, although some of that variation may be due to the different levels of sampling among the three, given that NASA operates substantially different numbers of payloads in each of these three regions. These results are more notional than precise; nonetheless, they do give some level of orientation to the expected loading for each event type (green, yellow, and red).

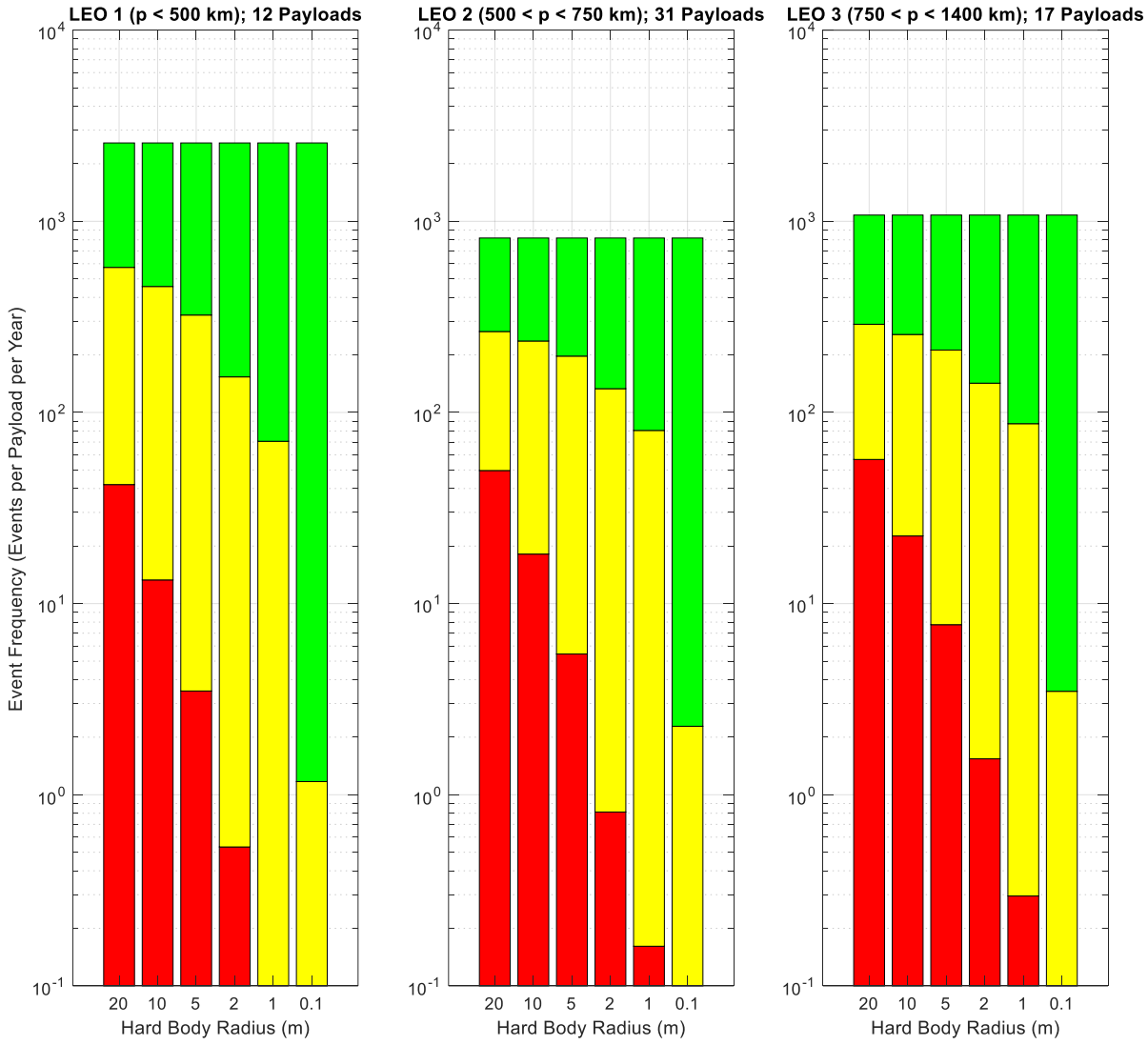


Figure F-1 Conjunction Event Frequencies as a Function of Orbit Regime and Event Severity

Appendix G. Orbital Debris Density

Orbital debris can be generated by a number of processes that have conspired to produce the present debris environment:

- Detritus cast off as part of the launch and injection activities;
- Off-scourings from healthy satellites themselves, such as peeling of Mylar thermal insulation;
- Collisions between satellites;
- Explosions of satellites; and
- Satellite debris-generating events in which (without any clear cause) a satellite is observed to break into more than one piece.

Because of the orbital locations in which major debris-producing events have taken place, and due to the differences in efficiency of natural debris cleansing arising from varying drag at these different altitudes, orbital debris density is not uniform. It is thus helpful to examine the current debris density situation when choosing an operational orbit because the orbital debris density does influence the rate of close approach events as well as the likelihood of satellite damage due to collision with an object too small to be tracked.

The graphs in Figure G-1 are produced from the NASA Orbital Debris Engineering Model (ORDEM) 3.2 release developed by the NASA Orbital Debris Program Office (ODPO). Generated for the 2023 debris environment, these graphs give debris flux (per satellite unit area and time on orbit) as a function of orbital altitude (a circular orbit is presumed) and inclination.

Given that satellites often have many years of orbital lifetime, it could be argued that the graphs should be projected further forward in time and thus reflect debris generation events and natural debris depletion processes over as much as the next decade. However, such a prediction requires a stochastic execution of the model to try to guess at the number, severity, and location of future debris-producing events. Since the purpose here is to compare the relative debris densities of different regimes, the simplicity of a single rather than probabilistic presentation has been selected.

The graphs are arranged by inclination, and while it is observed that the highest inclination band represented here does have a slightly higher debris density than the two lower bands, the differences in both absolute debris flux values and the overall curve morphology among the three inclination bands are not great, indicating that inclination is not a major contributor to debris density. However, variation with orbital altitude is more marked: the density peaks at about 850 km and tails off in either direction from that peak at relatively equal rates. And the reduction is significant: about an order of magnitude in each direction within the bounds of the graphs (400 to 1200 km orbital altitude).

Three lines are shown on each graph:

- The amber line represents objects of sizes greater than 10 cm, which is the size of objects in the present space catalog without any contributions from the new Space Fence SSN radar and is thus related to satellite conjunctions presently tracked and reported.

- The reddish line is for objects of sizes greater than 5 cm, which is the size of objects that can be expected to be included in a catalog that does include tracking data from the new Space Fence radar, which should soon be contributing data to the conjunction assessment enterprise.
- The blue line is for objects of sizes greater than 1 cm, which is the size of objects large enough to be generally considered lethal to a satellite if there is a collision.

The gap between the reddish and blue lines thus represents the flux level of objects that are too small to be tracked yet large enough to potentially be lethally damaging. The gap represents a collision risk that, essentially, cannot be mitigated and so is simply accepted as the cost of operating a satellite in space.

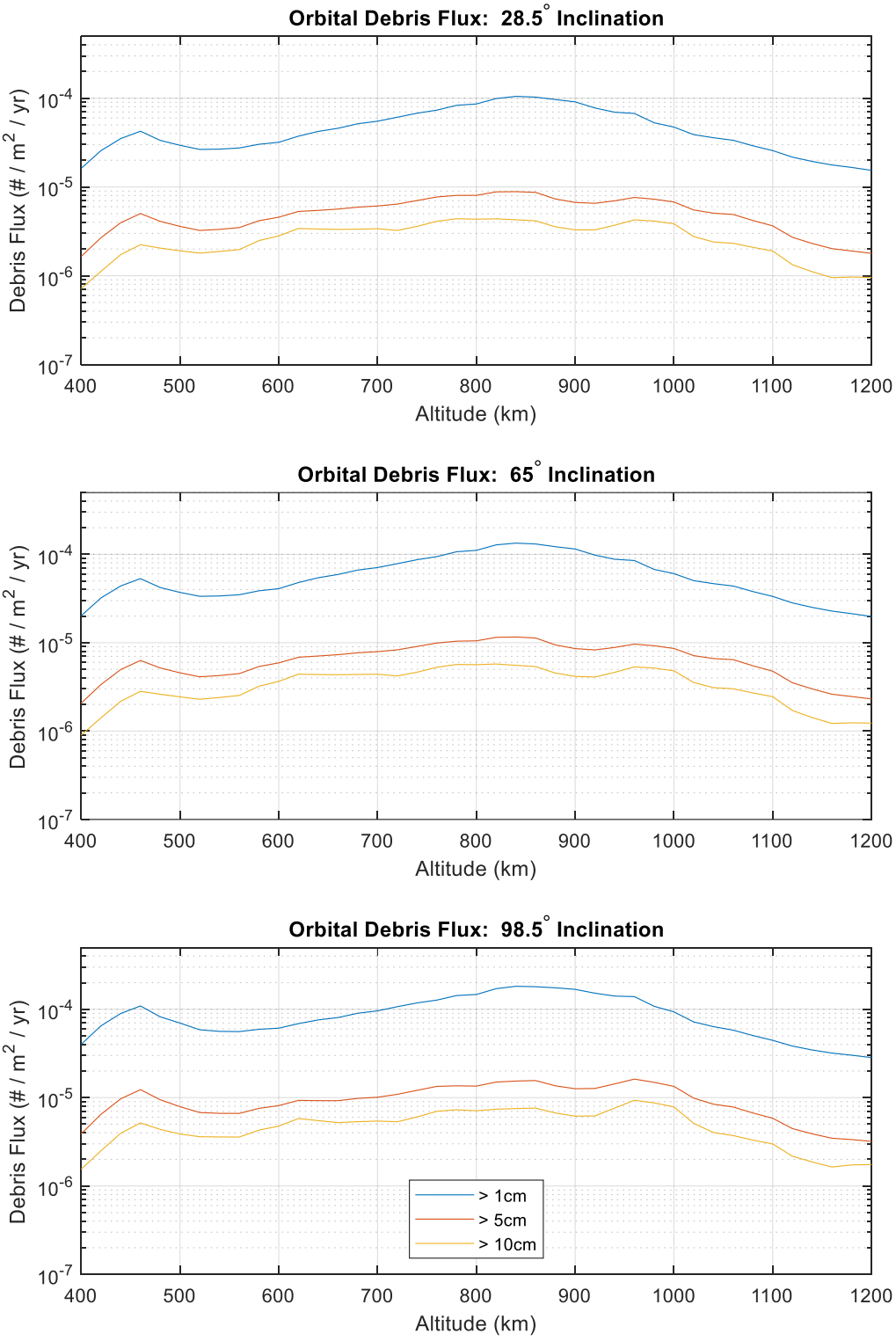


Figure G-1 Debris Flux as a Function of Orbit Inclination, Orbit Altitude, and Object Size

In some cases, the flux densities between the 1 cm and 5 cm bands represent almost an order of magnitude difference. Considering this difference, some might question why conjunction assessment is pursued at all, given that so much of the collision risk cannot be remediated and simply must be accepted. To this question are two distinct responses:

1. First, within reasonable bounds, it makes sense to protect against safety-of-flight hazards that are known and that can, if necessary, be mitigated. NASA CARA's statement of purpose is "to take prudent measures, at reasonable cost, to enhance safety of flight, without placing an undue burden on mission operations." An unjustified over-investment in safety should be avoided, and mission objectives should not be regularly and substantially impaired by conjunction assessment considerations. But since it is not particularly expensive or difficult to identify high-risk conjunctions from the existing space catalog, and since such conjunctions are relatively infrequent, and furthermore, since the need for mitigation action for such conjunctions is less than an order of magnitude less frequent than that, it is good safety practice to conduct conjunction assessment operations according to the CARA mission statement. Space actors thus employ the conjunction assessment enterprise to "prevent the preventable."
2. Second, the satellite conjunctions that occur between objects large enough to be in the space catalog are the ones that have the potential to generate large amounts of space debris. As the NASA ODPO has shown (Johnson et al. 2001), the debris production potential of a collision is a function of the two objects' relative kinetic energy and masses: the ratio of their two masses and their relative velocity. Collisions in which the two objects' masses are quite different will generally result in only the lighter of the two objects being fragmented with the heavier object being merely cratered. Objects in which the two masses are closer to each other will generally result in complete fragmentation of both objects and thus very large amounts of resultant debris. Extremely small secondary objects (i.e., a few centimeters in size and thus below the level of trackability), even at the large relative velocities of LEO conjunctions (~10,000 m/s), are quite unlikely to cause fragmentation of larger active payloads, so the amounts of debris produced by collisions with untracked objects will in almost all cases be relatively small. As stated previously, collisions with untracked objects can still be lethal to a satellite and do constitute an accepted risk simply of operating a satellite in space, but conjunction assessment against tracked and cataloged objects prevents the collisions that will produce large amounts of debris and thus makes a substantial contribution to space sustainability, even though it cannot protect payloads from debilitating conjunctions with very small objects. (Appendix O gives a more detailed treatment of the collision debris production topic.)

Choosing a mission trajectory that occupies a region of space with a lower debris density is thus a favorable selection for two reasons:

- It lowers the risk of a lethal collision with an untracked object; and
- In general, it reduces the rate at which serious conjunctions with cataloged objects will occur.

In addition to achieving the above two goals (as the curves in Figure G-1 demonstrate), choosing orbits with lower orbital altitudes also facilitates compliance with the 25-year disposal guidelines.

While there may be a general correlation, it is important to recognize that there is not in fact a direct relationship between the debris flux for catalog-size objects and the actual rate at which conjunctions, even serious conjunctions, will be identified. The debris flux gives the expected frequency of penetration of the satellite's cross-sectional area by some object in space. According to the curves in Figure G-1, a sun-synchronous object at 700 km, placed at the middle of a 20 m radius sphere (to employ an extremely conservative statement of its exposed area), would over a one-year period expect only 0.006 penetrations of this 20 m sphere. In other words, it should take ~160 years on orbit for a penetration to occur. However, predicted close approaches for such an object are identified daily, yet mitigation actions would be warranted only a few times per year. What is the meaning of this lack of alignment?

When a Pc for a specific event is calculated, it is interpreted as representing the likelihood that the "true" miss distance (which ultimately cannot be known) is smaller than the hard-body radius (HBR) defined about the object (e.g., the 20 m radius of the constructed sphere described in the above paragraph). The "true" Pc corresponding to the "true" miss distance is either 0 or 1; i.e., either the satellites will pass within 20 m of each other or they will not.

However, the orbit determination process does not predict future positions without error, so a Pc value somewhere between 0 and 1 represents a calculation of the likelihood of a penetration given the uncertainties in the input measurement data to the orbit determination, lack of comprehensiveness of the dynamical model, and additional uncertainties expected in predicting the satellites' epoch states to their TCA (such as atmospheric density forecast error).

Generally, when the Pc is greater than 1E-04 (1 in 10,000 likelihood of a penetration), a mitigation action is recommended and executed, This threshold has emerged in the industry as an acceptable balance between safety-of-flight considerations and additional mission burden and encumbrance. Because mitigation actions are recommended when the likelihood of penetration exceeds 1 in 10,000, there is little surprise that the frequency of serious conjunction events exceeds that projected by the debris flux value: in only 1 out of 10,000 cases would the penetration actually occur if the mitigation action were not executed. Fortunately, a conservative approach such as this can be implemented and is still consistent with the mandates of prudent action, reasonable costs, and lack of undue burden on mission operations.

(See Appendix F for information on actual expected rates of serious conjunction events.)

Appendix H. Long-Term Collision Risk and Satellite Colocation Analysis

This section describes an approach developed by NASA CARA that uses long-term collision risk assessment methodologies applied to proposed satellites to predict:

- Long-term average conjunction and collision rates that a new satellite can expect to experience; and
- The potential frequency of repeating conjunctions experienced by a new satellite due to being colocated with another satellite.

These methods provide the means to avoid deploying new satellites into crowded regions of orbital space, and to prevent orbital collocation with one (or more) pre-existing satellites. Although the methods also apply to satellite mid-mission repositioning or end-of-life disposal efforts, they are demonstrated here using an example of a newly-proposed satellite.

With the ongoing deployment of large constellations, such as the Starlink or OneWeb systems, avoiding overly crowded regions of space is becoming an increasingly important consideration. In addition, avoiding orbital colocation prevents problematic repeating conjunctions with other active satellites. To this end, when a new operational orbit is being selected, a long-term collision risk and orbit colocation analysis should be performed. Fortunately, it is often possible to make sufficient modifications to an intended nominal orbit to avoid these unfavorable situations, yet still meet mission objectives.

H.1 Long-Term Risk Assessment Theory and Algorithmic Description

Both semi-analytical and Monte Carlo methods provide the means to estimate statistically-expected long-term collision rates between satellites. Monte Carlo methods also provide estimates of the overall expected frequency of conjunctions, as well as the frequency of repeating conjunctions which can be used to identify potentially colocated satellites.

Analytic Long-term Collision Rate Analysis

The analytic theory developed by Kessler (1981) focuses on estimating collision rates between either spherical objects, or, for this application, circumscribing spheres. It provides estimates of such collision rates by neglecting all orbital perturbations except the orbital precession induced by Earth's J_2 gravitational term (see Vallado 1981 for a description of orbital precession). For each pair of analyzed satellites, the Kessler rate represents an average over six angles (Hall and Matney 2000). More specifically, the method averages collision rates over the three Keplerian orbital element angles for each of the two interacting objects. These three angles are the orbital right ascension of ascending node (RAAN, or Ω), the argument of perigee (ω), and the mean anomaly (M). In other words, Kessler rates represent six-dimensional phase space averages, which are equal to long-term temporal averages in idealized situations for which the ergodic hypothesis is valid. For this reason, Kessler rates are often referred to as "long-term" estimates, even when calculated in more realistic, non-idealized situations to which the ergodic hypothesis does not necessarily strictly apply.

The Kessler (1981) formulation expresses the average collision rate between two satellites as an integral over the volume of their overlapping orbital regions, as follows:

$$\dot{N}_c^K = \sigma_{1,2} \int_{\mathbb{V}} [S_1 S_2 v_{rel}] dV \qquad \text{(H-1)}$$

In this equation, $\mathbb{V}$ indicates the overlapping orbital volume shared by the primary and secondary satellites; $\sigma_{1,2} = \pi(R_1 + R_2)^2$ their combined collisional cross section; S_1 and S_2 their average spatial densities, respectively; and v_{rel} their average relative velocity during conjunction events. (The superscript *K* indicates that the Kessler method has been used to estimate the collision rate, and the dot above the symbol N_c represents the first time derivative.) The Kessler collision rate is zero for orbits that do not cross one another at all over long-term J_2-only propagations, for which the overlap volume $\mathbb{V}$ vanishes. The integrand factors within the square brackets of equation (H-1) depend on first three Keplerian orbital elements of the primary and secondary satellite orbits, i.e., the semi-major axis (SMA), the eccentricity, and the inclination of each. Calculating the integral over volume in equation (H-1) entails two-dimensional numerical integration over geocentric latitude and radial distance.

Monte Carlo Colocation Conjunction Rate Analysis

Monte Carlo methods provide another means to estimate collision and conjunction rates, and allow for more generalities, including:

- Trajectories that account for orbital perturbations other than just those induced by the J_2 gravitational term;
- Both spherical and non-spherical shapes, used to estimate conjunction rates that arise from ellipsoidal or box-shaped screening volumes; and
- The ability to estimate the frequency of repeating conjunctions, used to assess satellite colocation.

The CARA Monte Carlo long-term rate estimation software uses the U.S. Space Force's Computation of Miss Between Orbits (COMBO) astrodynamics standards algorithm to determine close approaches between trajectories propagated using the SGP4 orbital theory. Input consists of a two-line element (TLE) set of mean orbital elements for the primary object plus a file of multiple TLEs for the secondary objects being considered. Each Monte Carlo trial entails propagating trajectories over a duration of several days (typically seven or ten), beginning each propagation at the epoch of the TLE catalog. The initial mean orbital elements for the trial propagations are the same as specified by the input TLEs, except that newly sampled (ω, Ω, M) angles are generated for both objects, each randomly drawn from a uniform distribution spanning 0 to 2π. The Monte Carlo algorithm registers a "hit" for each instance that the trial trajectories approach one another close enough to penetrate a screening volume surrounding the primary object. For sufficiently large screening volumes, multiple hits can potentially occur per trial, so the algorithm also registers the earliest hit that occurs for each primary-secondary pair as a "first-contact" event.

Dividing the total number of hits by the product of the number of trials and the propagation duration yields the Monte Carlo estimate for the rate

$$\dot{N}_c^{MC} = \frac{N_{hit}}{N_{trial}\, T_{prop}} \quad \text{and} \quad \Delta\dot{N}_c^{MC} \approx \frac{\dot{N}_c^{MC}}{\sqrt{N_{hit}}} \tag{H-2}$$

The second equation above provides an approximation of the 1-sigma uncertainty of the estimate due to Monte Carlo counting uncertainties, which decreases as the number of registered hits increases. Rates of first-contact events are similarly calculated and are always less than or equal to the overall rate, i.e., $\dot{N}_{fc}^{MC} \leq \dot{N}_c^{MC}$. As demonstrated later, for colocated orbits that produce mostly repeating conjunctions, the rate of first-contact conjunctions is significantly less than the total rate.

H.2 Long-Term Risk Assessment Methodology Demonstration

To demonstrate long-term risk and satellite colocation assessments, this section introduces a hypothetical proposed satellite called "NewSat" with a nominal initial orbit intentionally designed to be colocated with that of the Hubble Space Telescope (HST).

In this hypothetical scenario, NewSat has a hard-body radius of 3.5 m, and is to be launched and reach its mission orbit on 2022-11-18 00:00:00 UT. The mean orbit of HST at this epoch, propagated from a TLE cataloged shortly before that time, has a mean motion equal to 15.11279168 revs/day, an eccentricity of 0.0002381, and an inclination of 28.4696° — corresponding to an equatorial perigee altitude of about 530.6 km and an apogee altitude of 532.2 km. To create a colocated orbit for this demonstration, the TLE for NewSat's nominal orbit has been made identical to that of HST in all respects except that the inclination has been rounded to the value of 28.5° and the eccentricity increased to 0.001. (These changes prevent the two orbits from being identical but are not sufficiently large to prevent them from being colocated, as will be demonstrated later.) This nominal NewSat deployment orbit corresponds to an equatorial perigee altitude of about 523.7 km and an apogee of 537.5 km, which is 13.8 km higher than perigee. However, for this hypothetical scenario, NewSat is given the flexibility to deploy into an alternate orbit with the same inclination, but with perigee altitudes anywhere from 500 to 600 km, and with an apogee up to 20 km higher than perigee. As discussed below, the long-term risk assessment methodology estimates rates between NewSat and HST over this entire region of orbital parameter space, as well as for all the other satellites contained within the 2022-11-18 TLE catalog, which represent a total of 36,368 potential secondary objects, a population that includes active and retired satellites, spent rocket-bodies, and orbital debris.

Figure H-1shows a plot of Kessler collision rates between NewSat and HST, calculated with equation (H-1), and using hard-body radii of 3.5 m and 10 m for the two satellites, respectively. Specifically, the plot shows the collision rate (on the color axis), as a function of perigee altitude (horizontal axis) and apogee-perigee difference (vertical axis). The white star marks NewSat's nominal deployment orbit. However, as mentioned previously, NewSat could deploy into an alternate orbit at any location within the bounds of the plot. Unsurprisingly, the nominal orbit corresponds to a non-zero long-term collision rate between the NewSat and HST. However, much of the area on the plot corresponds to zero collision rate (shown as the darkest blue color

on the plot); these regions represent candidate orbits for NewSat that do not overlap at all with HST's orbit because they are sufficiently higher or lower in altitude. NewSat could change deployment plans by shifting its orbit into one of these regions, but that would only mitigate the long-term collisional threat due to HST and not necessarily that posed to the other cataloged satellites.

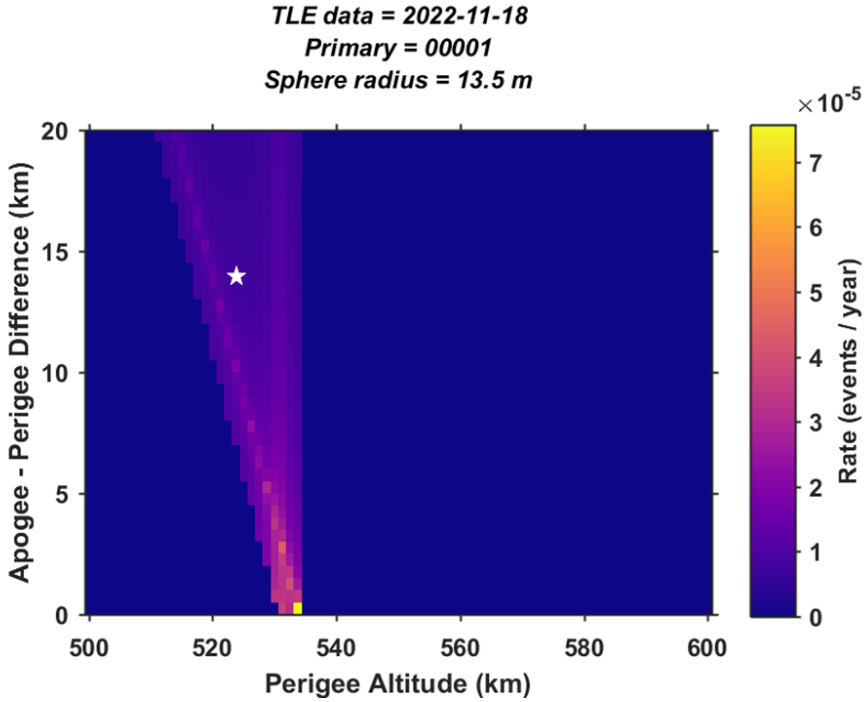


Figure H-1 Kessler Collision Rates between NewSat and HST

Figure H-2 shows collision rates summed over all the satellites contained within the full 2022-11-18 TLE catalog. For the nominal NewSat orbit marked by the white star, a total of 2,183 secondary objects were found to have non-zero Kessler rates among the 36,368 cataloged objects. The calculation uses hard-body radii for these secondary objects taken from a table of published dimensions for known satellites, when available, and based on radar cross section values for unknown objects such as orbital debris (Hall and Baars 2022). Figure H-2 shows that the nominal NewSat orbit is adjacent to (but not within) the region affected by the Starlink constellation satellites, which occupy several orbital shells at higher altitudes (the regions affected by three of these shells are marked in Figure H-2). So, if the NewSat mission were to change its deployment plans, doing so to avoid these already crowded regions would largely eliminate the expected number of Starlink conjunctions and associated risk mitigation actions. For clarity, Figure H-3 shows collision rates for all cataloged objects, but with Starlink constellation satellites excluded. A comparison of the color axis scales on Figures H-2 and H-3 illustrates how relatively large the long-term rates due to the Starlink constellation are in this specific region of orbital space.

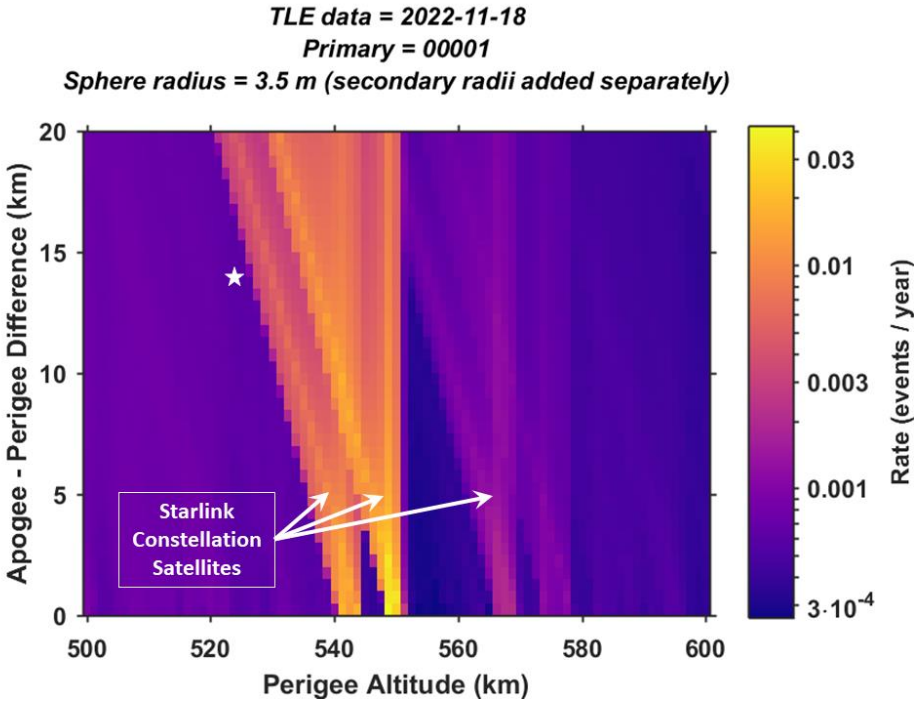


Figure H-2 Total NewSat Collision Rates Among All Cataloged Satellites

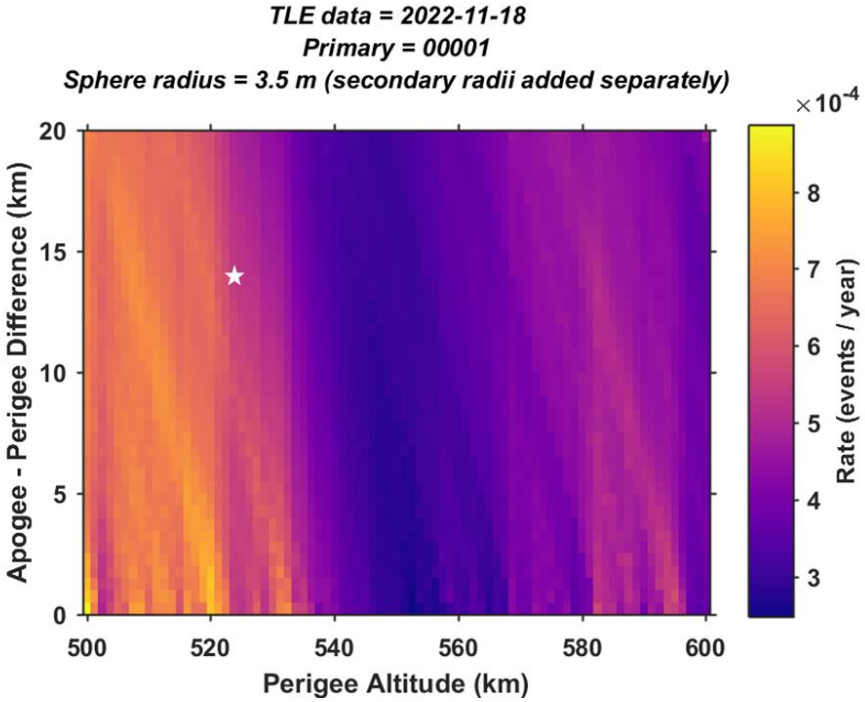


Figure H-3 NewSat Collision Rates Excluding the Starlink Constellation

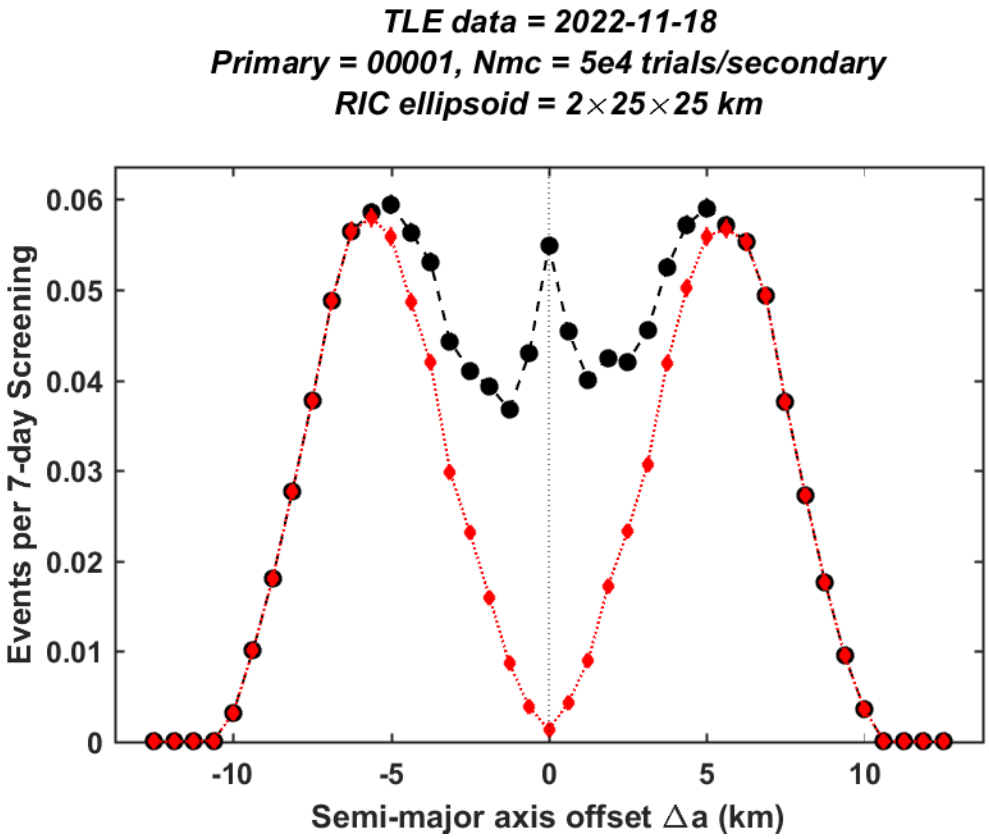
Figures H-1, H-2, and H-3 graphically illustrate how long-term collision rate analyses can be used to avoid deploying satellites into crowded regions, and thereby help minimize the long-term collision risk among tracked objects. For the nominal deployment orbit, Figure H-2 indicates that NewSat's predicted collision rate is ~6×10^{-4} per year, corresponding to a statistically expected collisional "lifetime" (expected time until a collision) of ~1,700 years among cataloged satellites. (As mentioned previously, other physical effects such as active maneuvering, orbital decay, etc., are likely to occur on shorter timescales than this, meaning that the ergodic hypothesis does not strictly hold for this system; nevertheless, long-term collision rates and collisional lifetimes are still used in these situations as indicators of statistical risk.) Notably, shifting NewSat's deployment orbit into one of the regions most affected by the Starlink constellation would lead to a significantly shorter collisional lifetime of 30 to 100 years—in the absence of any risk mitigation maneuvering, which fortunately is being performed by Starlink constellation satellites. However, a careful inspection of Figures H-1, H-2, and H-3 indicates that changing the planned NewSat deployment to an orbit with a mean equatorial perigee altitude of about 558 km and an apogee ≤7 km higher would avoid both the HST and Starlink regions simultaneously and extend NewSat's collisional lifetime to ~3,000 years. Furthermore, an extended analysis that includes all future Starlink constellation deployments (as tabulated by Bassa et al. 2022) indicates that the collision rate predicted for this specific orbital region should remain relatively low, at least among currently cataloged objects and currently planned Starlink satellites.

H.3 Colocation Analysis Methodology

During their on-orbit residence, most spacecraft come into conjunction with other, secondary, objects with some frequency. However, if two active satellites are maintained in similar orbits or in orbits that have consistent intersections, synodic alignment can occur that produces multiple close approaches on successive orbits. If the alignment is very close, such repeating conjunctions can persist over extended periods (e.g., much longer than typical seven- or ten-day conjunction screening durations) and could also recur periodically over a satellite's entire orbital lifetime. Because conjunction management between active payloads is complex, requiring coordination of maneuver plans and joint decisions about near-term courses of action, it is desirable to minimize such situations.

In the example introduced in the previous section, the nominal NewSat orbit was constructed specifically to be highly colocated with HST's orbit. This section demonstrates how Monte Carlo analysis can be used to identify such colocated orbits, which are characterized by a high frequency of repeating conjunctions, or, equivalently, a relatively low frequency of first-contact conjunctions. For this, the Monte Carlo method is used to predict the average rate of conjunctions and the fraction expected to be first-contact events. This specific analysis estimates conjunction rates averaged over a screening period of seven days (which is the same screening duration used for most LEO CARA satellites). Conjunctions occur in the Monte Carlo simulation when an ellipsoidal screening volume centered on the primary satellite is penetrated by the trajectory of a secondary satellite. The ellipsoid used for the present analysis has principal axes that measure 2×25×25 km, aligned with the primary object's radial, in-track and cross-track directions.

The goal of colocation analysis is to determine the orbital conditions that create repeating conjunctions, which can be accomplished by analyzing the frequency of first-contact events. As mentioned previously, first-contact conjunctions correspond to the initial instance that a conjunction screening volume is penetrated. This means that pairs of satellites that experience high frequencies of repeating conjunctions necessarily have low first-contact conjunction rates relative to their total conjunction rates. The colocation analysis graph in Figure H-4 demonstrates this effect by comparing total conjunction rates and first-contact rates, plotted as a function of Δa, which represents the offset of NewSat's orbital semi-major axis (SMA, or a) from its nominal value. (Note, the analysis assumes that NewSat's eccentricity and inclination remain at their nominal values of 0.001 and 24.5°, respectively). The black dashed curve marked with circles in Figure H-4 plots the total conjunction rate, $\dot{N}_c^{MC}$, calculated using equation (H-2). This total includes both first-contact conjunctions and any subsequent conjunctions (i.e., repetitions). The red dotted curve marked with diamonds plots the first-contact rate, $\dot{N}_{fc}^{MC}$, which is always less than or equal to the total rate. The black curve shows that conjunctions of either type are predicted to occur for SMA offsets within about ±10.5 km of the nominal deployment value, i.e., in the range -10.5 km $\leq \Delta a \leq$ +10.5 km. However, the red curve indicates that the rate of first-contact conjunctions is noticeably lower than the total over the range -6.1 km $\leq \Delta a \leq$ +6.1 km. Furthermore, the first-contact rate is less than 50% of the total rate over the more restricted range of -2.2 km $\leq \Delta a \leq$ +2.2 km.



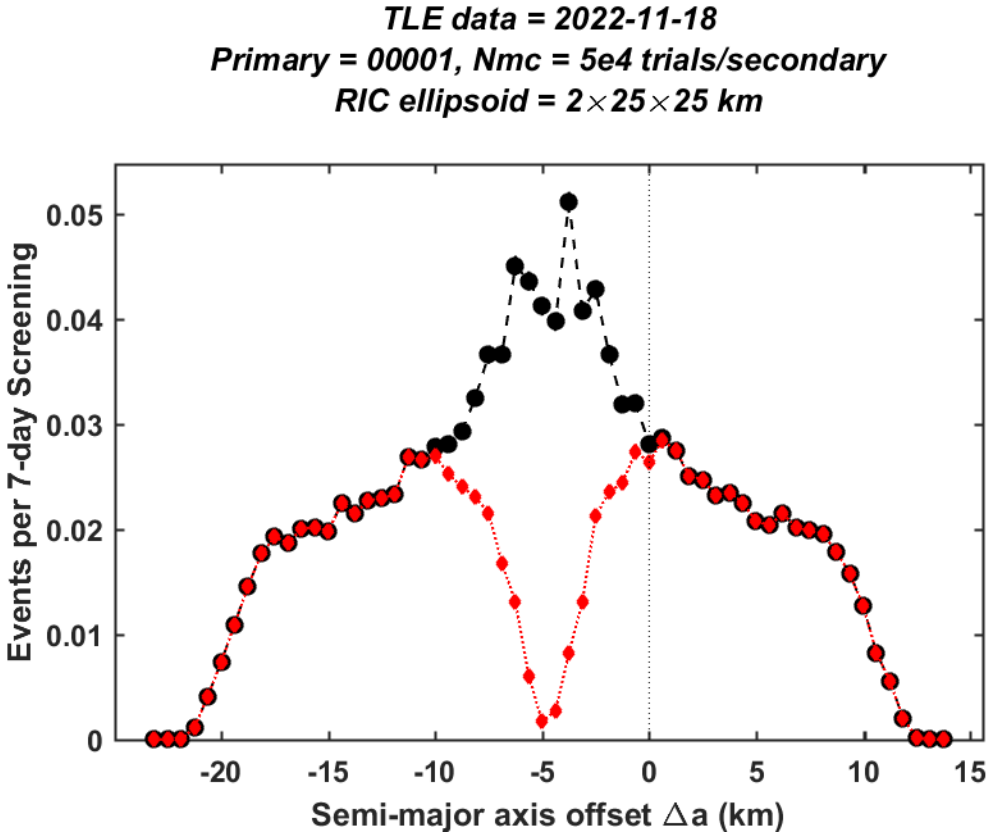
Nmc = The number of Monte Carlo trials RIC = Radial – In-track – Cross-track

Figure H-4 Predicted Rates for NewSat and HST Satellite Conjunctions

This analysis adopts this 50% reduction (i.e., $\dot{N}_{fc}^{MC}/\dot{N}_{c}^{MC} < 0.5$) as a means to identify highly colocated orbits, meaning that, for this specific screening volume and under these conditions, NewSat would need to shift its deployment SMA by at least 2.2 km in either direction from the nominal value to avoid a high level of colocation with HST. Notably, shifting by about 6.1 km would largely eliminate the occurrence of repeating conjunctions, but would not prevent all conjunctions between the two satellites. The one-dimensional analysis shown in Figure H-4 assumes only SMA orbital adjustments, holding the deployment eccentricity and inclination constant at their nominal values. A similar two-dimensional analysis holding only the inclination constant can also be performed to determine how much the SMA and/or eccentricity would need to change to eliminate high levels of colocation (or, equivalently, how much the perigee and/or apogee altitudes would need to change). The generalized method can be extended to even higher dimensions to include all relevant, adjustable orbital elements.

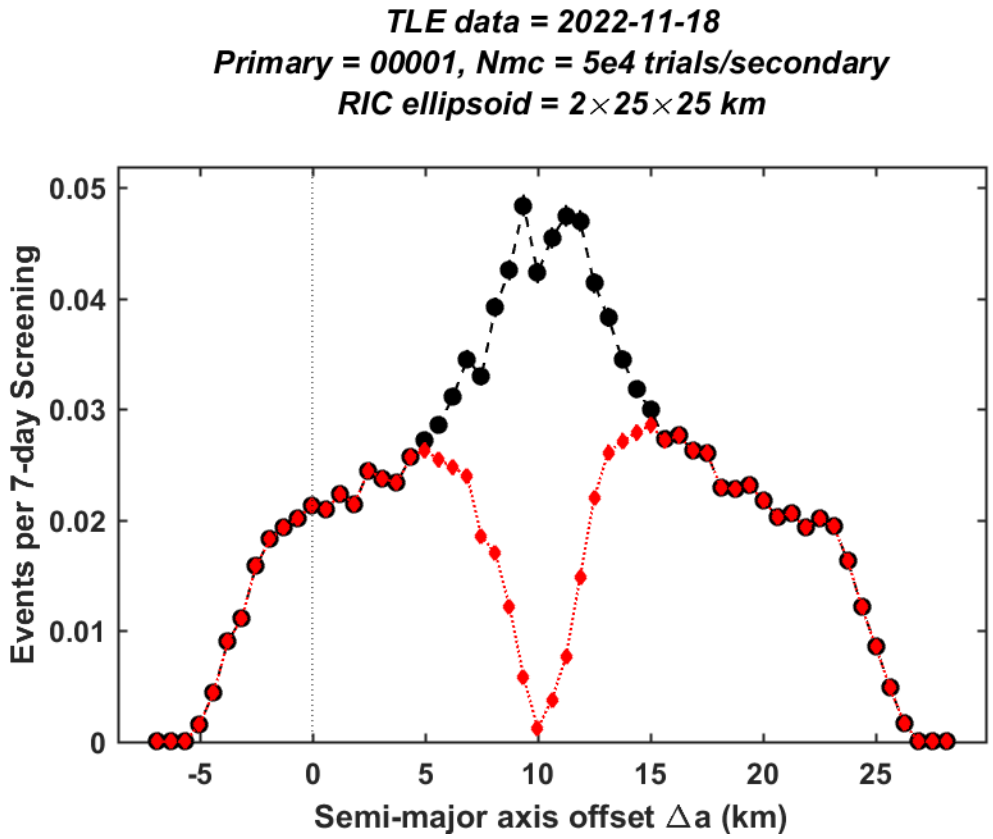
If the NewSat mission were to change the planned deployment orbit to avoid colocation with HST, as described above, then it would be worthwhile to do so in a manner that does not result in it being collocated with any other high value satellites. For example, the Fermi Gamma-ray Space Telescope (FGST) and the Swift satellite, both NASA missions, occupy orbits similar to NewSat's nominal proposed orbit, so it is conceivable that shifting the NewSat deployment orbit

could inadvertently create a high level of colocation with one of these satellites. Figures H-5 and H-6 demonstrate this by showing one-dimensional colocation analysis graphs for FGST and Swift, respectively (each made using the same assumptions as described above for Figure H-4). Figure H-5 shows that FGST's red curve with diamonds is less than half of the black curve with circles for -7.0 km $\leq \Delta a \leq$ -2.5 km, so if NewSat were to shift into this range, then the analysis predicts it would become highly colocated with FGST. Similarly, Figure H-6 indicates that shifting NewSat's orbit into the range +7.8 km $\leq \Delta a \leq$ +12.4 km would result in high colocation with the Swift satellite.



Nmc = The number of Monte Carlo trials RIC = Radial – In-track – Cross-track

Figure H-5 Predicted Rates for NewSat and FGST Satellite Conjunctions



Nmc = The number of Monte Carlo trials RIC = Radial – In-track – Cross-track

Figure H-6 Predicted Rates for NewSat and Swift Satellite Conjunctions

H.4 Long-Term Conjunction Rate and Colocation Analysis Software

The NASA CARA software instantiation of the long-term risk and colocation analysis capability is an ongoing tool development effort, currently being refactored to execute faster (which is especially required for the Monte Carlo analysis modes), and eventually to be more portable (because it currently uses the U.S. Space Force COMBO tool, which has restricted distribution). The main CARA software tool entitled "*LongTermRisk*" uses TLE catalogs as input and allows users to estimate long-term Kessler collision rates as well as Monte Carlo conjunction rates and associated first-contact rates. (The *LongTermRisk* tool was used to create all the figures shown in this appendix.) The tool also has the capability to ingest auxiliary TLE catalogs containing predicted future satellite constellation(s), for instance, or model populations of untracked orbital debris. Because the tool uses the limited-distribution AstroStandards version of the U.S. Space Force COMBO close approach analysis function, and because the associated execution environment is relatively complicated to set up, it is envisioned that the ability to run this tool will remain within the NASA CARA organization.

Appendix I. Satellite Covariance Realism Assessment Procedures

I.1 Introduction

This section outlines standard methods for assessing the realism of a covariance produced as part of the state estimates for a spacecraft. Because probability-based methods for satellite conjunction risk assessment in almost all cases require a covariance for their calculation, realistic covariances for both the primary and secondary objects in a conjunction are important.

The U.S. Space Force has pursued algorithmic improvements over the last twenty years to improve the realism and representativeness of their catalog's covariance matrices, especially in LEO, so covariances that arise from this source have undergone efforts to assess and sustain their trustworthiness. For covariances produced for and contained in satellite O/O ephemerides, the situation is much more uneven. While it is true that in the majority of conjunctions the covariance for the secondary object predominates, in 5% of the conjunctions the primary covariance is larger than the secondary, meaning that it is likely to play a substantial role in collision likelihood assessment in over 10% of conjunctions. This is a large enough subset of a satellite's conjunction history to make covariance realism for the primary object ephemeris an important consideration.

Because the focus of this section is more practical than theoretical, broader questions of covariance formation, methods of covariance propagation, or approaches within the orbit determination process to improve covariance realism are not addressed in detail. Modern orbit determination packages, such as the Ansys Orbit Determination Tool Kit, include features for covariance formation, propagation, and tuning, and their accompanying technical documentation is extensive.

What this section does attempt to do is enumerate the data types needed for covariance realism investigations, introduce the appropriate test statistics, and give some guidance for the evaluation and interpretation of these tests. Most of the information here derives from NASA CARA's practical experience with this problem, as there are few published studies on the subject and no practical guides. The particular focus is on determining the realism of predicted covariance generation for actively maintained spacecraft in the presence of precision position data (e.g., onboard GPS or telemetry tracking information) to use for a realism assessment because this is the particular situation presented to O/Os.

Covariance realism assessment consists of three parts: collection/calculation of position error data, calculation of covariance realism test statistics, and proper assessment of those test statistics. After some introductory discussion, each of the three parts of this process will be addressed in detail.

I.2 General Remarks

It is important to recall that a spacecraft state estimate generated through an orbit determination process is an estimate of a mean state (i.e., mean position and mean velocity), and a covariance is

a stochastic characterization of the expected errors about that mean state. The governing assumption is that these errors in the individual components conform to a Gaussian distribution; this is how a covariance matrix alone is able to represent the error distribution without requiring higher-order tensors. The activity of covariance realism is thus to determine how well actual state errors conform to the Gaussian distribution that the covariance specifies. Such a statement requires additional remarks.

First, the present treatment focuses on the assessment of the realism of the position portion of the covariance, although in most cases the methods advanced here are applicable to a full 6 x 6 or even larger covariance matrix. There certainly are reasonable arguments for evaluating the covariance in element space (usually equinoctial elements) and considering all six elemental representations in any evaluation (Woodburn and Tanygin 2014). However, since the principal purpose of covariance generation at present is for conjunction assessment applications, and for many conjunction risk assessment approaches only the position portion of the covariance is used (typically to calculate the probability of collision employing the two-dimensional simplification of the situation; see Chan 2008), it is acceptable to limit oneself to a Cartesian representation of the covariance and focus only on its position portion. Furthermore, a previous study has demonstrated that potential non-Gaussian covariance behavior brought about by working in a Cartesian rather than curvilinear framework rarely affects the outcome of higher-probability conjunction events, namely those with Pc values greater than 1E-04 (Ghrist and Plakalovic 2012). Given all of this, limiting the analysis to the position portion of the covariance is reasonable.

Second, because a covariance represents an error distribution, its adequacy as a representation can be evaluated only by examining a set of actual error data and determining whether these data conform to the distribution specified by the covariance. In actuality, this is usually not possible in a straightforward way because a particular covariance is propagated to a particular point in time and relevant only at that moment, and at that moment there is only one state estimate and thus only one error point. It is not possible to determine in any definitive way whether a single point conforms to a distribution; one can determine a percentile level for that point (e.g., for a normal distribution with mean 0 and standard deviation 1, absolute values of 3 or greater should occur only 0.13% of the time) and discern that such points will not occur frequently in such a distribution, but they will occur occasionally, and the situation in the above example may well be one such instance. Typically, this problem is addressed by generating a test statistic that compares each error point to its associated covariance, determining what the distribution of such test statistics should be, and evaluating the conformity of the set of test statistics to the expected distribution. More will be said about this in the following sections that directly address the calculation of test statistics and evaluation of their statistical properties.

Third, as stated in the previous paragraph, a covariance is always a covariance propagated to a particular time. If it is not propagated at all, it is an epoch covariance and is purported to reflect the error of the fit;[12] if it is propagated, it is intended to be a reflection of the state estimate error

[12] It is questionable whether the *a priori* covariance emerging from the batch process does indeed do this as it is produced by a calculation that involves only the amount of tracking, times of the tracks, and *a priori* error variances

of that propagated state. Many different covariance propagation methods are available, each of which achieves a different level of fidelity, so one source of covariance irrealism can be the propagation method itself. It is thus important to group realism evaluations by propagation state; for example, all the calculated test statistics for covariances propagated from 1.5 to 2.5 days could be collected into a single pool and evaluated as a group, and the evaluation could be said to be applicable to covariances propagated about two days. It should also be pointed out that covariance irrealism, however it may be evaluated, is unlikely to correlate directly with propagation time in the sense that if the covariance is made realistic for a particular propagation state, it cannot necessarily be expected to be realistic for all propagation states. When pursuing covariance realism remediation, one frequently needs to decide whether some sort of omnibus improvement over all common propagation states is desired or whether it is preferable to focus on a particular propagation state or narrow ranges of states. Many aspects of the U.S. Space Force's orbit determination process are tuned to optimize performance at 3 days' propagation; this is not a terrible propagation state to choose for O/O covariance realism optimization for conjunction assessment applications, although one could also choose a somewhat shorter value if that were to align more closely with maneuver commitment timelines.

Fourth, in the context of a covariance error calculation, the individual state error values will need to be presumed to constitute independent samples, free of correlation between them. Such an assumption is common in most statistical processes, and it is not surprising to see it arise here. While true sampling independence is unlikely to be achieved, there are certain measures that can be taken to promote it. For example, if one has set up a covariance realism assessment scenario by comparing a predicted ephemeris (with covariance) to a definitive ephemeris, it would be best to take only one point of comparison for each propagation state bin that one wishes to evaluate. For example, in evaluating the two-day propagation state, there would be a temptation to take all of the ephemeris points from, say one hour before the two-day propagation point to one hour after that. If the ephemeris were generated with one-minute time steps, then this would produce 120 points instead of just one, which on the surface would seem to give a nice sample set for a statistical evaluation. However, because the propagated and definitive ephemeris points are highly correlated, in fact little additional information is brought to the evaluation by including the entire dataset. Additionally, all of the statistical evaluation techniques that will be deployed to generate a realism evaluation assume sample independence; in such a case as the one described above, they will miscarry and produce what is likely to be an overly optimistic result. In the present scenario, it is best to limit oneself to a single comparison point per propagation state from this pair of ephemerides and seek out a group of such ephemeris pairs rather than use closely grouped data points from them to try to broaden the sample set.

The procedure recommended above eliminates a particular strain of correlation, but it leads naturally to the consideration of a second type: that between successive ephemerides. Definitive ephemerides are generally formed by piecing together sections of ephemeris from essentially the middle of the fit-spans of moving-window batch orbit determinations, and filters process data sequentially but with a forgetting-rate matrix that reduces the influence of older data as a

for each observable. It does not give a statistical summary of the actual correction residuals. For this reason, some practitioners have argued for an alternative formulation of the covariance that considers residual error.

function of data age. But in each case, a fair amount of time between ephemerides needs to pass before the generated points are statistically independent of a given predecessor set. The independence condition could be fulfilled by forcing the generation of products from entirely different datasets, but this is not a practical procedure. In past projects, trying to reduce the data overlap to less than 50%, meaning that subsequent ephemerides to be used in covariance realism evaluations need to be spaced so that they share less than 50% of the generating data with the previous ephemeris, has been considered acceptable in practice, although a lower figure would be desirable. One need not become overly fastidious about this issue, but it is important to take whatever practical steps one can to reduce the influence of sample data correlation on the realism assessment.

I.3 Part I: Error Computation

Satellite state errors are computed by comparing a predicted state estimate to some type of state truth source. The predicted data typically come from a predicted ephemeris with a predicted covariance associated with each ephemeris point. Truth data can come from a variety of sources: onboard GPS precision position data, precision telemetry tracking data, or a precision ephemeris constructed from either of these. Ideally, both the predicted and truth data should align temporally (i.e., their time points should be exactly the same), and they should both possess covariance data at each time point. In actuality, it is rare that both of these conditions are met, and often neither is.

Covariance data are often not available for the truth data source. Precision tracking or GPS data can include covariance information as an output from a tracking or combinatorial filter; when precision ephemerides are constructed, sometimes covariances can be synthesized from the abutment differences observed between the joined “pieces” of ephemeris. However, it is frequently the case that even when this is theoretically possible, the construction software has failed to include this as a feature. If no covariance data are available for the truth source, one must then determine whether the errors in the truth data are so much smaller than the errors in the predicted ephemeris under evaluation that the errors in the truth data can safely be neglected. The degree of accuracy difference that would allow this assumption is a matter of opinion, but most commentators would probably agree that an order of magnitude difference would be an acceptable decrement; and in some cases, even smaller ratios could be tolerated. These differences, however, must be considered at the individual component level. If the error in one component tends to dominate the entire arrangement (as is often encountered with the in-track component due to inadequate drag modelling), then the overall normalized error of all three position points in the truth ephemeris can be much smaller than the overall error in each point in the predicted ephemeris, yet component errors for the cross-track and radial components could be of a similar magnitude in both truth and prediction. In such a case, the omnibus test statistic to be described in the next section would be distorted because it treats the normalized errors in each component essentially equally.

If the time-points between truth and predicted data do not align, then some sort of intermediate-value-determination scheme must be used for one of the two datasets. The highest-fidelity approach would be to use an actual numerical propagator to produce aligning values, but if the points for the source to be adjusted are spaced sufficiently closely, then a number of different

interpolation approaches produce results quite adequate for a covariance realism assessment. While the interpolation of state values is relatively straightforward and can be accomplished satisfactorily with different interpolation approaches (e.g., Lagrange, Hermite), the interpolation of covariances is much more problematic. Different interpolation methods appear to work better in different situations; there is no single accepted method to compare multiple covariances to evaluate their relative fidelity against each other, and there is a danger of certain interpolation methods producing non-positive-definite covariances, which do not make physical sense with the current problem. Given this situation, if both data sources (truth and prediction) can be interpolated, but only one of the two data sources possesses accompanying covariance data (presumably the predicted source), then it is usually preferable to interpolate the data source that lacks the covariance, for this eliminates the question of how to handle covariance interpolation. If both sources possess covariance, then one would in general interpolate the source with the closer point spacing. Alfano (2004) outlines certain covariance interpolation techniques and provides test results. There is also an interest in interpolating not the covariance, but the state transition matrices associated with the straddling ephemeris points. While this approach can in some cases produce slightly less desirable interpolated covariances, it does essentially guarantee that the interpolated result will be positive definite.

Once position and covariance data for the predicted and truth datasets are aligned at the time-point of interest, calculation of the error is straightforward: it is simply the (subtracted) difference between states as long as they both are rendered in the same coordinate system. The subtraction convention does not actually matter given the way the test statistic is computed, but for consistency, one can follow the "observed – expected," or truth – prediction, paradigm. The two covariances for each comparison point (one for the truth position and one for the predicted position) are simply added together, presuming, of course, that they are in the same coordinate system. Thus, for each point of comparison, one will have a position difference (with three components) and a corresponding combined covariance (or just the predicted source's covariance if the truth source's errors are presumed to be extremely small).

While one can conduct an entire covariance realism evaluation in the Earth-Centered Inertial (ECI) reference frame, it is often more meaningful to move it into the Radial – In-track – Cross-track (RIC, which is also called the UVW) frame—a reference frame that is centered on the object itself. If only one of the evaluation data sources has an associated covariance, then it probably makes sense to make that source the center for the RIC coordinate frame. If a covariance is provided for both, then it does not really matter which is selected to ground the RIC frame. The use of the RIC frame is helpful in moving from a finding of irrealism to remediation suggestions.

I.4 Part II: Test Statistic Computation

As stated previously, a position covariance, which in this treatment is the portion of the matrix to be tested for proper error representation, describes a three-dimensional distribution of position errors about the object's nominal estimated state, and the test procedure is to calculate a set of these state errors and determine whether their distribution matches that described for the position covariance matrix. To understand the test procedure, it is best to consider the problem first in one dimension, perhaps the in-track component of the state estimate error. Given a series of state

estimates for a given trajectory and an accompanying truth trajectory, one can calculate a set of in-track error values, here arranged in vector form and given the designation ε, as the differences between the estimated states and the actual true positions. According to the assumptions previously discussed about error distributions, this group of error values should conform to a Gaussian distribution. As such, one can proceed to make this a "standardized" normal distribution, as is taught in most introductory statistics classes, by subtracting the sample mean and dividing by the sample standard deviation:

$$\frac{\varepsilon - \mu}{\sigma} \tag{I-1}$$

This should transform the distribution into a Gaussian distribution with a mean of 0 and a standard deviation of 1, a so-called "z-variable." Since it is presumed from the beginning that the mean of this error distribution is 0 (because the state estimate is an estimate of the mean and is presumed to be unbiased), the subtraction as indicated in the numerator of Equation I-1 should be unnecessary, simplifying the expression to:

$$\frac{\varepsilon}{\sigma} \tag{I-2}$$

It will be recalled that the sum of the squares of *n* standardized Gaussian variables constitutes a chi-squared distribution of *n* degrees of freedom. As such, the square of Equation I-2 should constitute a one-degree-of-freedom chi-squared distribution. This particular approach of testing for normality—evaluating the square of the sum of one or more z-variables—is a convenient approach for the present problem, as all three state components can be evaluated as part of one calculation (ε_u represents the vector of state errors in the radial direction, ε_v the in-track direction, and ε_w the cross-track direction):

$$\frac{\varepsilon_u^2}{\sigma_u^2} + \frac{\varepsilon_v^2}{\sigma_v^2} + \frac{\varepsilon_w^2}{\sigma_w^2} = \chi_{3\,dof}^2 \tag{I-3}$$

One could calculate the standard deviation of the set of errors in each component and use this value to standardize the variable, but it is the covariance matrix that is providing, for each sample, the expected standard deviation of the distribution. Since the intention here is to test whether this covariance-supplied statistical information is correct, the test statistic should use the variances from the covariance matrix rather than a variance calculated from the actual sample of state estimate errors. For the present moment, it is helpful to presume that the errors align themselves such that there is no correlation among the three error components (for any given example, it is always possible to find a coordinate alignment where this is true, so the presumption here is not far-fetched; it is merely allowing that that particular coordinate alignment happens to be the RIC or UVW coordinate frame). In such a situation, the covariance matrix would lack any off-diagonal terms and thus look like the following:

$$C = \begin{bmatrix} \sigma_u^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_w^2 \end{bmatrix} \tag{I-4}$$

and its inverse would be straightforward:

$$C^{-1} = \begin{bmatrix} 1/\sigma_u^2 & 0 & 0 \\ 0 & 1/\sigma_v^2 & 0 \\ 0 & 0 & 1/\sigma_w^2 \end{bmatrix} \tag{I-5}$$

If the errors for a single state are formulated as:

$$\varepsilon = \begin{bmatrix} \varepsilon_u & \varepsilon_v & \varepsilon_w \end{bmatrix} \tag{I-6}$$

then the pre-and post-multiplication of the covariance matrix inverse by the vector of errors (shown in Equations I-3 through I-6) will produce the expected chi-squared result:

$$\varepsilon C^{-1} \varepsilon^T = \begin{bmatrix} \varepsilon_u & \varepsilon_v & \varepsilon_w \end{bmatrix} \begin{bmatrix} 1/\sigma_u^2 & 0 & 0 \\ 0 & 1/\sigma_v^2 & 0 \\ 0 & 0 & 1/\sigma_w^2 \end{bmatrix} \begin{bmatrix} \varepsilon_u \\ \varepsilon_v \\ \varepsilon_w \end{bmatrix} = \frac{\varepsilon_u^2}{\sigma_u^2} + \frac{\varepsilon_v^2}{\sigma_v^2} + \frac{\varepsilon_w^2}{\sigma_w^2} = \chi_{3\,dof}^2 \tag{I-7}$$

What is appealing about this formulation is that, as the covariance becomes more complex and takes on correlation terms, the calculation procedure need not change: the matrix inverse will formulate these terms so as properly to apportion the variances among the *U*, *V*, and *W* directions, and the chi-squared variable can still be computed with the $\varepsilon C^{-1}\varepsilon^T$ formulary. For such a situation in two-dimensions (chosen here for illustrative purposes because the expression is less complex) in which the error quantities are ε_x and ε_y and the correlation coefficient is ρ, the entire equation, with correlation terms included, assumes the form:

$$\varepsilon C^{-1} \varepsilon^T = \frac{1}{(1-\rho)^2} \left(\frac{\varepsilon_x^2}{\sigma_x^2} + \frac{\varepsilon_y^2}{\sigma_y^2} - \frac{2\rho\varepsilon_x\varepsilon_y}{\sigma_x\sigma_y} \right) \tag{I-8}$$

One can observe that if the correlation coefficient is zero, the equation reduces to the two-dimensional equivalent of the form shown in Equation I-7 above. As the correlation coefficient moves from zero to a more substantial value, the test statistic encounters a trade-off between the overall inflating effect of the (1-ρ) multiplier and the subtracted correlation term.

The quantity $\varepsilon C^{-1}\varepsilon^T$ is called the Mahalanobis distance (technically, it is the square of the Mahalanobis distance), named after the mathematician P.C. Mahalanobis (who, by the way, was a personal friend of the 20th-century mathematical genius Ramanujan). This construct is a convenient and useful way to calculate a normalized distance.

I.5 Part III: Test Statistic Evaluation

It is very well that a test statistic can be derived that, if the covariance is realistic, should conform to a known statistical distribution, but of course, there needs to be some method for testing a group of these test statistics to determine if in fact they do conform to the expected distribution. Such a desire leads the investigation to the statistical subdiscipline of "goodness of fit."

Every student of college statistics learns about Analysis of Variance (ANOVA), the particular procedure for determining whether two groups of data can essentially be considered the same or different. More particularly, it is a procedure for determining whether the experimental distribution, produced by the research hypothesis, can be considered to come from the parent distribution represented by the null hypothesis, and the operative statistic arising from the analysis is the *p*-value: the likelihood that the research dataset can be considered a sample drawn from the null hypothesis's parent distribution. If this value becomes small, such as only a few percent, it means that there are only about two or three chances in one hundred that the experimental dataset would have been generated from sampling from the null hypothesis dataset. In this case, the research and the null hypothesis outcomes can be considered to be palpably different, which would be likely to lead to the rejection of the null hypothesis and the embrace of the research hypothesis. This procedure is a specific example of statistical hypothesis testing.

A similar procedure can be applied to evaluate goodness of fit, namely, to evaluate how well a sample distribution corresponds to a hypothesized parent distribution. In this case, the general approach is the reverse of the typical ANOVA situation: it is to posit for the null hypothesis that the sample distribution does indeed conform to the hypothesized parent distribution, with a low *p*-value result counseling the rejection of this hypothesis. This approach does favor the association of the sample and the hypothesized distribution, which is why it is often called "weak-hypothesis testing". Although counterintuitive, this approach is not unreasonable: what is being sought is not necessarily the "true" parent distribution but rather an indication of whether it is reasonable to propose the hypothesized distribution as the parent distribution for the experimental data that one has generated or calculated. In the present case, the question to be posed is whether the observed squared Mahalanobis distance histories, calculated by the procedure in Part II, conform to a 3-degree-of-freedom chi-squared distribution, namely, whether they can be considered to have been drawn from a 3-degree-of-freedom chi-squared distribution as a parent distribution—if they do, then the covariance properly represents the actual observed error distribution.

There are several different mainstream techniques for goodness-of-fit weak-hypothesis testing: moment-based approaches, chi-squared techniques (not in any way linked to the fact that the present application will be testing for conformity to a chi-squared distribution), regression approaches, and Empirical Distribution Function (EDF) methods. The easiest and most direct of these is simply a test of the first moment of the distribution (that is, the mean), which, if normalized by the degrees of freedom of the distribution, should be unity (or, in the present case, should take on an unnormalized value of 3). The square root of this mean, the so-called Mahalanobis distance, is a good estimate of a single-value scale factor describing the covariances departure from reality, and as such, it is a convenient way to compare the results from different

covariance correction techniques as well as estimate a single-value scale factor that could potentially be used to remediate an irrealism situation.

While this test is easy to apply and expeditious for generating comparative results, in comparison to other goodness-of-fit tests it lacks power. Indeed, many different distributions could have the same mean yet be substantially different in the overall behavior or in the tails: a close match of the mean is a necessary but not sufficient condition for matching a distribution. To evaluate the match between entire distributions, the EDF methodology is generally considered to be both the most powerful and most fungible to different applications. For this reason, it does not make sense to apply formal goodness-of-fit tests to first-moment results (i.e., matching of the mean); these should be used merely as a methodology to compare performance for the corrected versus uncorrected cases and intra-correction-methodology performance.

The general EDF approach is to calculate and tabulate the differences between the Cumulative Distribution Function (CDF) of the sample distribution and that of the hypothesized distribution, to calculate a goodness-of-fit statistic from these differences, and to consult a published table of *p*-values for the particular goodness-of-fit statistic to determine a significance level. Specifically, there are two goodness-of-fit statistics in use with EDF techniques: supremum statistics, which draw inferences from the greatest deviation between the empirical and idealized CDF (the Kolmogorov-Smirnov statistics are perhaps the best known of these); and quadratic statistics, which involve a summation of a function of the squares of these deviations (the Cramér – von Mises and Anderson-Darling statistics are the most commonly used). It is claimed that the quadratic statistics are the more powerful approach, especially for samples in which outliers are suspected, so it is this set of goodness-of-fit statistics that are recommended for covariance realism evaluations. The basic formulation for both the Cramér – von Mises and Anderson-Darling approaches is of the form:

$$Q = n\int_{-\infty}^{\infty}\left[F_n(x) - F(x)\right]^2 \psi(x)dx \tag{I-9}$$

The two differ only in the weighting function ψ that is applied. The Cramér – von Mises statistic is the simpler:

$$\psi(x) = 1 \tag{I-10}$$

setting ψ to unity. The Anderson-Darling is the more complex, prescribing a function that weights data in the tails of the distribution more heavily than those nearer the center:

$$\psi(x) = \left\{F(x)\left[1 - F(x)\right]\right\}^{-1} \tag{I-11}$$

The Anderson-Darling construct is thus more sensitive to outliers. Given NASA CARA's experience that outliers frequently creep into covariance realism evaluations and introduce statistical processing issues, it is recommended to choose the somewhat more permissive Cramér – von Mises statistic for covariance realism investigation purposes.

It is a straightforward exercise to calculate the statistic in Equation I-9, discretized for the actual individual points in the CDF for each trajectory (that is, changing the integral into a summation), and for convenience, this quantity is called the "Q-statistic." The step after this is, for each Q-statistic result, to consult a published table of *p*-values (determined by Monte Carlo studies) for this test to determine the *p*-value associated with each Q-statistic (a good source for these tables as well as an excellent treatment of the overall subject is given in D'Agostino and Stephens 1986). The usual procedure is to set a *p*-value threshold (e.g., 5%, 2%, 1%) and then to determine whether the sample distribution produces a *p*-value greater than this threshold (counseling the retention of the null hypothesis: sample distribution conforms to hypothesized distribution) or less than this threshold (counseling rejection of the null hypothesis: sample distribution cannot be said to derive from the hypothesized distribution as a parent). One can also interpolate to determine the precise *p*-value for each test situation. NASA CARA has source code and test cases that perform all the above calculations and reside on the NASA CARA public-facing software repository for free download (See Section 7, Contact Information in this document for the specific URL.)

Finally, it should be noted that, even though there are normalization provisions within the EDF formulation, the results do, to some degree, depend on the size of the sample. In a way this is considered already by accommodation within the *p*-value tables for sample size, but because of the quadratic-sum nature of the test statistic, the procedure can still be overwhelmed by large sample sizes. One approach to mitigating this situation is to pick a standard sample size—perhaps somewhere in the neighborhood of 50 samples—and calculate the test statistic in a (with-replacement) resampled manner, producing a CDF of the *p*-values attained for each sample. As an example, suppose that for a particular evaluation there are 100 error vectors with associated covariance and therefore 100 test statistic ($\varepsilon C^{-1}\varepsilon^{T}$) results. When running the goodness-of-fit test, one might choose 1000 random, 50-point samples (with replacement) from this set of 100 values and test each, producing a CDF chart of the 1000 *p*-values obtained from the resampling investigation.

What is an acceptable level for a *p*-value result—one that would indicate that the error distribution matches that of the covariance? In goodness-of-fit practice, rarely is a significance level greater than 5% required, and levels of 2% or even 1% are often accepted. It probably can be said that values less than 1% cannot allow the conclusion that there is any real conformity to the hypothesized distribution. At the same time, it should be added that the calculation is rather sensitive to outliers and that this should be kept in mind when interpreting results.

If results from different remediation approaches match the hypothesized distribution closely enough, then comparison of different *p*-value levels can serve as a notional indication of the relative performance of these different approaches. However, if the performance is such that the hypothesized distribution is not approached all that closely by any of the correction mechanisms, then a situation can be encountered in which the results fall off the published tables of *p*-values, and it becomes extremely difficult to compare the results of the different methods conclusively. In such a case, it may be possible to draw some broad comparative results from looking at the Q-values rather than the associated *p*-values, but in all likelihood, it will be necessary to revert to simply a comparative-results set such as first-moment tests.

I.6 Part IV: Data Outliers and Conditioning

As remarked in the above section, goodness-of-fit test results can be sensitive to outliers; this is true whether one interprets results visually or uses a formal technique. The deleterious impact on test results is mitigated somewhat using the resampling approach discussed previously, by which the contribution of the outliers to each individual test is lessened. However, the fact remains that bad data do enter the orbit determination process, and the failure to take some account of this reality can produce situations in which the covariance realism assessment problem becomes intractable.

The entire covariance realism assessment process is grounded on the notion that individual component errors are normally distributed, and this situation allows for certain techniques to identify outliers. The usual "x-sigma" filter is a naïve and unscientific method for outlier identification and is especially difficult to justify for the Gaussian distribution, where there is some developed theory for outlier identification. A superior approach is the Grubbs outlier test, which provides a formal statistical test for outliers but works only in situations with a single outlier and cannot be applied recursively (Grubbs 1969). In the case of multiple outliers, the procedures of Rosner (1975, 1977) are applicable but must be applied with an *a priori* guess of the number of potential outliers. That is, one must first inspect the data to assemble a proposed set of outliers and then test this group as outliers for a particular significance level. Because only a single component can be assessed at a time, to use this approach it is probably best to design a tool that can examine a particular error point's behavior in all three components (compared to the rest of the main distribution) to determine which set of points might constitute an outlier set and test that set. Several such attempts may be needed before a set of points can be identified as outliers to a given significance level.

An extended example of the application of the technique described here is available in Zaidi and Hejduk (2016).

Appendix J. CARA Conjunction Risk Assessment Tools Validation

CARA and JSC FOD are chartered to perform conjunction assessment screenings, risk assessment, and mitigation action guidance on behalf of NASA missions, so typically there is no particular need for missions to obtain or develop conjunction risk assessment evaluation tools. However, cases can occasionally arise in which it is not practical or desirable for CARA/JSC FOD to perform these calculations, such as for potential autonomously controlled missions that will perform onboard conjunction assessment. In such cases, the needed calculations or tools will be validated by CARA/JSC FOD before operational use. Methods used for validation include the following:

- **Inspection**. Used to verify that a particular datum or feature is present, that a display or graphic is properly constructed to convey a particular concept, etc.
- **Analysis/Documentation**. Used to ensure that a particular algorithm is theoretically or practically sound, that a needed parameter is properly converted to a different reference frame or units, etc.
- **Formal Test**. Used to ensure that critical calculations are performed correctly by executing a formal test with preconfigured input data and expected results, which are compared to the test article's results and any differences satisfactorily explained.

Formal test is the most frequently used methodology within CARA validation. conjunction risk assessment calculations for which formal test cases with expected results presently exist include the following:

- Two-dimensional Pc calculation with and without covariance correlation correction and non-positive-definite covariance correction
- Two-dimensional Pc calculation when one covariance is missing (Frisbee method)
- Three-dimensional Pc calculation (Coppola-Hall method) with and without covariance correlation correction and non-positive-definite covariance correction
- Pc calculation by Brute Force Monte Carlo (from epoch)
- Pc sensitivity to space weather mismodeling
- Conjunction relative state comparison
- Collision consequence evaluation (anticipated number of resultant debris pieces above a certain size)
- Maneuver trade space calculations for single and multiple conjunctions

The NASA CARA software repository includes test cases for software validation. (See Section 7, Contact Information in this document for the specific URL.)