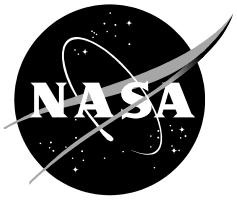
NASA/TP—2015–000000



Launch COLA Operations: an Examination of Data Products, Procedures, and Thresholds
Revision A (12 JAN 2014)

M.D. Hejduk
a.i. solutions, Inc., Maryland

D. Plakalovic
a.i. solutions, Inc., Colorado

M.E. Hametz
a.i. solutions, Inc., Florida

L.K.Newman
Goddard Space Flight Center, Maryland

J.C. Ollivierre
Kennedy Space Center, Florida

B.A. Beaver
Kennedy Space Center, Florida

R.C. Thompson
The Aerospace Corporation, Virginia

March 2015

NASA STI Program ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

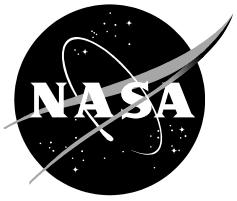
- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TP—2015–000000



Launch COLA Operations: an Examination of Data Products, Procedures, and Thresholds
Revision A (12 JAN 2014)

M.D. Hejduk
a.i. solutions, Inc., Maryland

D. Plakalovic
a.i. solutions, Inc., Colorado

M.E. Hametz
a.i. solutions, Inc., Florida

L.K.Newman
Goddard Space Flight Center, Maryland

J.C. Ollivierre
Kennedy Space Center, Florida

B.A. Beaver
Kennedy Space Center, Florida

R.C. Thompson
The Aerospace Corporation, Virginia

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, MD 20771-0001

March 2015

Acknowledgments

The authors would like to thank Mr. Carson Sparks of NASA JSC for his assistance with the review process of this study and report. Additionally, the authors wish to acknowledge the multiple launch vendors who graciously granted permission to use in this study and report their predicted launch trajectories from historical launch events; without these data, none of the parts of this study would have been possible. Finally, a special acknowledgement is due to Mr. Ryan Frigm, who originally postulated that a cumulative probability of collision calculation might be more fitting to LCOLA calculations and thus encouraged that particular area of enquiry.

This report is available in electronic form at

http://

If you need a multiple-column layout, or need to switch to multiple columns at some point in the report, follow these steps (also shown in the instructions file):

1. Ensure that the cursor is at the beginning of a line
2. Select the Insert menu, Break…, and then choose the Continuous section break type
3. Select the Format menu, Columns…, select a preset column layout, and click OK
4. If you want to change back to one column, follow these steps again, selecting the one column layout

Shown below are the text styles available for your report. These styles are also shown in the instructions file, if you want to delete them here, but still be able to see their format.

"Heading 1 STI" – 14 pt. Arial Bold, Centered

This is the "Report Text" style for the body of your report. 11 pt. Arial, flush left, 0.25 first line indent, widow/orphan control on, line spacing single to accommodate inline equations.

"Heading 2 STI" – 12 pt. Arial Bold, Centered

"Heading 3 STI" – 11 pt. Arial Bold Italic, Flush Left

"Heading 4 STI" – Indented bold italic heading with paragraph. Highlight the rest of the paragraph, including the dash, and turn off the bold and italic.

"Heading 5 STI" – Indented italic heading with paragraph. Highlight the rest of the paragraph, including the dash, and turn off the italic.

- This is "Report Text Bullet"

Executive Summary

NASA GSFC and KSC, acting in response to headquarters NASA direction, performed a year-long study of launch collision avoidance (LCOLA) operations in order to determine and recommend best risk assessment and mitigation practices. The following condenses the findings and recommendations of the study into one short summary, a more expanded version of which appears as Section 10.

1) Predicted launch trajectories are less accurate than both the general perturbations (two-line elsets, called GP) and precision (called SP) space catalogues—more than an order of magnitude less accurate than the current GP space catalogue. There is thus no compelling reason, from a space catalogue perspective, to use the precision space catalogue for LCOLA screenings; the GP catalogue is sufficient.

2) The covariances associated with the predicted launch trajectories, though large, are appropriately sized and thus a realistic statement of trajectory error. This means that probability of collision (Pc) conjunction calculations are possible; and thus this parameter (Pc) should be used rather than miss distance for conducting LCOLA operations.

3) When the same conjunction events are screened with both GP and SP approaches, 60% of the time they yield essentially equivalent Pc results. In nearly all of the remaining 40% of the cases, the GP Pc value is larger (more conservative) than the SP value. GP-based screenings are thus a reasonable conservative proxy for SP screenings—one more reason why use of the precision space catalogue is not necessary for LCOLA operations.

4) If one wishes 80% of a launch window to be "open," GP screenings can be conducted at a Pc level of 2E-06 and SP screenings at 8E-07. If the "open-percentage" is allowed to decrease to something closer to the 60% value, then levels of 1E-06 and 4E-07 can be used. The evenness of excluded times throughout the windows indicate that decreasing to openness levels of 60% or so is less encumbering than was previously believed.

5) Using a Pc calculation approach that encapsulates the combined risk of all of the close approaches, rather than using the single highest Pc value from a screening, is methodologically easy and offers certain advantages. It is recommended that LCOLA agencies migrate to this approach whenever their LCOLA software programs are undergoing repair/enhancement for other reasons.

6) It is possible to determine miss-distance values that emulate the screening results from a particular Pc threshold, to a stated confidence level. However, the window-closure results are cripplingly heavy-handed, and thus this procedure is not recommended.

7) Screening only against "defended assets," such as active satellites, greatly improves the launch window closure results. However, it ignores the risk of collision loss of the launching payload itself and the deleterious effects of the debris-producing event.

8) From a collision risk standpoint, screening at or above a Pc threshold of 1E-05 (GP) or 5E-06 (SP) provides no appreciable benefit over performing no LCOLA screenings at all. Screening at a level to provide a half-order-of-magnitude improvement over the "do nothing" risk posture results in a 50% launch window closure level for both GP and SP. Screening to provide a full order-of-magnitude improvement results in closure levels of 75% in GP and 90% in SP.

Table of Contents

Section 1: Introduction

The increasing number of orbiting objects poses a significant threat to spacecraft health and safety. The threat of collision with objects affects missions not only on-orbit but also during launch and ascent. The current estimate of the number of 'tracked' objects that are larger than 10 cm is greater than 17,000, with the number of objects increasing by several hundred per year. Most of these tracked objects are characterized as orbital debris. Recent events such as the Iridium–Cosmos collision (2009), China's anti-satellite (ASAT) test (2007), and other break-ups such as the Breeze-M rocket explosion (2007) have led to an even greater awareness and concern in the satellite community.

While Launch Collision Avoidance (Launch COLA or LCOLA) has been a standard launch range practice at some level for more than a decade, at present there is neither a requirement nor standardized process for LCOLA across the Agency. NASA's Launch Services Program (LSP) for Expendable Launch Vehicles (ELVs) at Kennedy Space Center provides LCOLA analysis to NASA's robotic missions if either requested or required by the spacecraft customer. LSP works with the Aerospace Corporation (Aerospace) to perform these functions, as Aerospace also performs similar functions for Air Force missions. The NASA Human Spaceflight Mission Control Center (MCC) team has also provided limited pre-launch conjunction screening for the Shuttle. At GSFC, the Conjunction Assessment Risk Analysis (CARA) Team has significant experience in conjunction risk assessment of orbital debris for on-orbit robotic missions. Leveraging the experience and expertise of these organizations, this report documents the results of a joint study to investigate and make recommendations for a standard practice for NASA Launch COLAs.

This Launch COLA analysis and recommendations report documents a five-part study. The first part (Section 1) is a survey of current industry practices and techniques. These will be presented in a trade space of advantages and shortcomings, the purpose of which is to identify LCOLA approaches to advance or modify as a recommendation is formulated. The second part (Section 2) investigates actual launch and injection dispersions from typical launch vendors. Included in this part is a determination of the ability for launch vehicles to quantify expected dispersions, derived from a survey of historical launch accuracies, as well as a comparison to space object catalogue accuracies in order to determine what levels of catalogue accuracies would be appropriate for launch COLA operations. This investigation helped to define requirements for the third part (Section 3), which is an analysis to determine the ability of vendor-supplied covariances realistically to represent the actual trajectory error—a requirement for a meaningful probability of collision (Pc) calculation. Part four consists of a survey of an extended experiment of launch COLA screening results in which five pre-launch trajectories each had their launch times modified in one-minute intervals from -15 to +15 days from the nominal launch time and was each subsequently screened (43,200 screenings each, for a total of 216,000). Examination of the results of these screenings gave insight into the effects of certain launch COLA screening thresholds and differences between TLE-based general perturbations (GP) and special perturbations (SP) results (Section 4); determined the difference between maximum P_c and cumulative P_c approaches to evaluate a single screening (Section 5); examined the relationships among Pc / miss distance, launch window length, and percentage of windows open (Section 6); explored a probabilistic link between Pc and miss distance for given Pc thresholds, as well as the window closure penalties incurred if one wishes to use this approach (Section 7); and considered the effects of conducting screenings against only defended assets rather than the entire space catalogue (Section 8). Finally, part five considers the entire LCOLA

enterprise in terms of improved overall risk posture (Section 9) and then outlines conclusions from all of these study data and articulates launch window closure thresholds and recommendations (Section 10).

Background

Launch Collision Avoidance is the general term that describes the process of actively screening for potential collisions between a launch vehicle and known, tracked on-orbit objects from liftoff through the end of the launch phase, and subsequently taking action to avoid any unacceptable close approaches, or conjunctions. During the LCOLA screening, a powered flight launch trajectory is compared against a catalog of on-orbit objects prior to launch to identify conjunctions. For a conjunction that exceeds a threshold, the corresponding launch time is removed from the launch window. There are two main analysis techniques used to perform LCOLA assessments.

Analysis Techniques

Initially, LCOLA screenings were performed using solely a miss distance approach, and in many cases this technique is still used today. The miss distance method is a comparison of the nominal trajectory of the launched object(s) versus each object in the resident space object catalogue. At each integration step in the trajectory, the point-to-point distance between the objects is computed; and if the minimum distance between a launched object and an on-orbit object is within the evaluation criteria volume, then the vehicle is unable to launch at that evaluated launch opportunity. These screenings are run for each launch opportunity across the launch window and are typically evaluated from first reaching orbit (>150 km) until payload separation + 100 minutes. For different object types, such as manned assets, active payloads, inactive payloads, rocket bodies, and debris, different miss distance criteria may be used to determine if there is a conjunction violation. The advantages of the miss distance technique are the relatively simple required data inputs and the speed of the screening runs: only the nominal launch trajectories of the vehicle and separated objects are required (*e.g.,* no covariance data are needed), and the calculation at each integration step is a simple comparison rather than a complex numerical integration. This speed and convenience, however, comes with limitations. Since the nominal trajectory is all that is considered, the inherent problem is that a vehicle will never fly exactly to the nominal trajectory. To compensate for the lack of a quantification of trajectory error (usually represented as a covariance), it is necessary to set the allowable miss-distance thresholds to larger values, which often results in a proliferation of conjunction violations and the excessive closure of launch windows. This drawback is dramatically highlighted when comparing liquid-fueled to solid-fueled engines. For a launch vehicle that consists of all solid propulsion motors, all engine propellants must be consumed, necessitating the use of energy scrubbing techniques to dissipate excess energy in order to minimize injection errors. Energy scrubbing is well known to result in much higher trajectory variations in flight as compared to liquid propulsion systems, which can be shut down when the desired orbit targets have been achieved. However, this sort of system-dependent variation in capabilities is not captured in miss distance evaluations or thresholds.

To account for these expected variations during flight, a Probability of Collision (P_c) method has become more widely utilized, as it allows for a higher fidelity launch vehicle representation and a more accurate measure of risk. The probability of collision associated

with a given on-orbit conjunction is a function of three parameters: the effective collision area (which quantifies the sizes of the objects), the nominal closest approach distance, and the orientation and size of the uncertainties in the positions of the objects. In this computation, the P_c is determined for each conjunction using the covariances of the objects (as an expression of position uncertainty) to determine the likelihood that both objects will be at the same location at the time of closest approach. More inputs are required for this calculation than for the miss-distance method, and it is more computationally intensive; but the yield is an actual probabilistic statement of collision risk, which requires no hedging or inflation of thresholds in order to compensate for unexpressed errors.

The launch vehicle covariances have been a source of significant concern within the LCOLA community. As the relative size of the launch vehicle one-sigma position ellipsoid will typically dwarf the variability ellipsoid of the on-orbit objects, it is the accuracy of the launch vehicle covariance that most influences the overall accuracy of the LCOLA P_c calculations. There has been concern that the covariances for the launch vehicle are overly conservative and thus do not represent actual vehicle performance. This concern is explicitly addressed in Section 3 of this study, as actual flight histories are compared to predictions in order to validate the covariances produced by the launch service contractors.

LCOLA Policies

Prior to February 2010, the launching Range would screen each launch attempt against only manned/mannable objects, such as the International Space Station (ISS), the Space Shuttle, and Soyuz spacecraft. Historically, this screening has been referred to as the "Range Safety COLA" process and resulted in mandatory closures of launch opportunities when conjunctions exceeded safety thresholds. Mission owner/operators were responsible for all COLA screens against other Resident Space Objects (RSO) and closures resulting from these analyses are usually called Mission Assurance COLAs (MA COLA). It was quite common for MA COLAS to have different threshold criteria than those used for Range Safety COLAS. In February 2010, United States Air Force (USAF) instruction AFI 91-217 was enacted, which mandated operational launch conjunction assessment/collision avoidance screenings by Air Force-controlled ranges. This AFI continues to require Range Safety COLAs but extends the screening to include all RSOs in the Range Safety analyses. Range Safety criteria have relatively high thresholds and continue to be mandatory closures, but the higher thresholds result in a very small number of closed opportunities. Many owner/operators continue to request MA COLAs at lower thresholds because they are willing to accept more closed launch opportunities in exchange for a lower risk of collision during launch and early orbit operations.

For NASA ELV launches, the LSP performed an analysis in July 2001 (ELVL-2001-0025354) to determine the overall probability that a collision could occur during the launch phase. Various launch vehicles were examined in different orbit regimes, and the results indicated that the chance of hitting an on-orbit object during the relatively short launch phase is so small that Mission Assurance COLA measures have essentially no effect on overall mission success (It should be noted that at the time of the LSP analysis, the object catalogue had significantly fewer tracked objects than the current catalogue). Following this study, LSP released a Program Directive (PD) on Collision Avoidance (LSP-PD-120.04), stating that Mission Assurance COLAs would not be routinely conducted for NASA LSP missions (this directive does not have any bearing on COLAs conducted by the appropriate

launch range for every LSP mission). However, if LSP's spacecraft customer requires a MA COLA analysis be performed, LSP must execute it as a formal mission requirement.

Under this LSP directive, NASA Jet Propulsion Laboratory (JPL) generally does not opt to perform any LCOLA screenings beyond the Range requirements. For many JPL missions, the launch window is either a very short duration or instantaneous. Additionally, many planetary or other deep-space missions have annual or even less frequent launch opportunities. These two factors combine to magnify the operational impact of standing down due to a low-risk conjunction.

In contrast, LSP has been conducting MA COLAs for every Goddard Space Flight Center (GSFC) mission since February 2004. In Goddard Policy Requirement (GPR) 8000.1, GSFC set forth the policy that each mission shall perform a MA COLA assessment against all on-orbit active satellites; no screening is required against on-orbit debris or spent upper stages. The objective of the GSFC policy is to protect orbiting assets, not necessarily the payload being launched.

Table 1-1 below summarizes the screening criteria used for the Air Force Ranges (per AFI 91-217) and the GSFC Center Policy. Both policies include provisions for P_c and miss distance criteria.

Table 1. Summary of the Range and GSFC LCOLA Screening Criteria

	Manned - Range	Payloads - Range	Debris - Range	Payloads - GSFC
Probability of Collision	$1x10^{-6}$	$10x10^{-6}$	$10x10^{-6}$	$1x10^{-6}$
Miss Distance (Ellipsoidal)	200 x 50 x 50 km	Not used	Not used	Not used
Miss Distance (Spherical)	200 km	25 km	2.5 km	5 km
NASA Waiver	Not allowed	NA	NA	$3x10^{-5}$ cumulative (prior to launch day)
LDA Waiver	Not allowed	100 x 10^{-6} / 2.5 km	100 x 10^{-6} / 0.25 km	Not allowed
NAF/CC Waiver	Not allowed	1,000 x 10^{-6} / **	1,000 x 10^{-6} / **	Not allowed
MAJCOM/CC Waiver	Not allowed	10,000 x 10^{-6} / **	10, 000 x 10^{-6} / **	Not allowed

Launch COLA Practices

For LCOLA screens performed by the Range, the data inputs are sent to the Range by the launch service contractor (nominal trajectories, launch window, object areas, and flight covariances). The launching Range provides all the necessary data to the JSpOC (who performs the computer runs), along with the details of the launch support required, via Air Force Form 22. Typically, the JSpOC will perform a test run approximately one week before launch followed by preliminary screens at L-48 and L-24 hours for both the primary and backup launch dates. At L-4 hours, the final screen is performed, which will be used for determining the window cutouts. The screens are typically run until payload separation + 100 minutes. The results of the screens are sent from the JSpOC to the requesting Range,

who converts the raw output into a formatted report with the window green and red times. This report is then sent to the appropriate launch decision authority(ies).

The JSpOC uses the program SuperCOMBO/CALIPER to screen against all objects in the SP catalog. The program output includes both P_c and miss distance information, which the Range then uses to determine the window cutouts for each object. Conjunctions with the ISS are determined by miss distance using a spherical 200 km stand-off distance. This is a mandatory criterion (no waiver possible). For objects and debris, window cutouts are determined using P_c with the screening criteria of $10x10^{-6}$, per Table 1. When using the SP catalog, the JSpOC may encounter on-orbit objects for which a covariance was not determined. In this situation, a P_c cannot be computed; and therefore only miss distance information can be provided and evaluated per Table 1-1. While conjunctions with payloads and debris do have waiver criteria to less restrictive limits, to date no waiver has been requested.

For LCOLA screens performed for GSFC missions, LSP works with The Aerospace Corporation, who use the program CollisionVision to perform the LCOLA screens. The required data inputs are identical to those sent to the Range by the launch service contractor (nominal trajectories, launch window, object areas, and flight covariances). LSP provides all the necessary data to Aerospace and coordinates the details of the launch support required. As with the JSpOC, Aerospace typically performs a test run approximately one week before launch followed by preliminary screens at L-48 and L-24 hours for both the primary and backup launch dates. The final screen is performed at L-8 hours, the results of which will be used for determining any window cutouts. LSP performs the final run prior to the operational support staff arriving on console in case COLAs close the full window, causing a scrub for that date. The screens are also run until payload separation + 100 minutes. The results of the screenings also are sent to LSP, who evaluates the results of the screen against the full catalog to determine which conjunctions, if any, are violations of the Goddard COLA policy. The launch window cutouts from the GSFC screen are combined with any cutouts from the Range screen, and the superset is provided to both the NASA and launch service contractor Launch Directors.

The Aerospace program uses the GP catalog to screen against all objects. The program output provides both P_c and miss distance information. However, when evaluating the results to determine the window cutouts, LSP uses the P_c results exclusively. Because the GP catalogue does not include covariance information, Aerospace computes an estimated covariance from GP element set error growth information, grouped by the orbit classification of the object. While conjunctions with satellites are not waivable on launch day, per GSFC Policy a waiver may be issued to perform the launch day screens using a cumulative screening criterion over the full launch opportunity. This option was exercised for the CALIPSO/CloudSat launch, as each launch date had only an instantaneous launch window.

Launch COLA Issues

As the various Launch COLA policies and implementation methods illustrate, there are no commonly accepted best practices among the agencies. The USAF policy serves to limit orbital debris, uses the SP catalog, and limits the use of the P_c method due to their concern about unrealistic launch vehicle covariances. The GSFC policy strives to be a good steward and protect on-orbit assets, uses the GP catalog, and exclusively uses the P_c method in response to the inherent limitations of the miss distance method.

Furthermore, in cases when similar methods are used, different criteria are employed with little or no rationale for how the criteria have been established. Additionally, an LSP analysis has shown there is no equivalent miss distance criteria that can be established for a given P_c criteria (ELVL-2008-0040593). LSP evaluated hundreds of conjunctions identified for LSP missions since 2001 and after examining the P_c and miss distance data, found very little correlation between the methods and further concluded that miss distance is ineffective as a screening criterion. While the AFI does state that miss distance evaluations are believed to be a more conservative approach, the LSP case study demonstrated that this is not always the case. These concerns over the difference in screening methods continually arise when the JSpOC screenings are unable to compute a P_c due to lack of covariance data for on-orbit objects. More information on the use of miss distance as a proxy for P_c is given in Section 7.

Although not addressed in this study, another important issue for LCOLA that should be mentioned is the duration of the screening process. While the GSFC policy states no duration, the AFI states that the LCOLA will screen until the objects are cataloged by the JSpOC and become part of the orbital CA process. However, it can take one or more days before launched objects can be folded into existing on-orbit screening efforts; and existing launch COLA methodologies cannot be reasonably used beyond a few orbits due to large launch vehicle dispersions. The result is a gap in COLA screening capability between the end of the effective LCOLA screening period (typically spacecraft separation + 100 minutes) and the commencement of on-orbit screening two days or more after launch. Consequently, the USAF and NASA currently perform geometric-based “COLA gap” analyses between the newly launched objects and the ISS if the object orbits intersect during the COLA gap period.[1]

Section 2: Launch COLA Trajectory Errors

Launch conjunction assessment, as stated in the previous section, is presently only a lightly-standardized activity in which different indices and launch window closure criteria are used at different launch agencies and centers. However, regardless of the particular calculation and window closure approach, the so-called "miss distance," or the vector magnitude of the position vector difference between the predicted launch trajectory and a secondary object, figures prominently in the computation; and in some cases this miss distance itself is the parameter of interest. As such, it is important to gain some understanding of the overall errors in the predicted launch trajectories, as these errors affect how a miss distance value should be interpreted. Additionally, the size of these errors will indicate whether analytic orbit models, such as the Simplified General Perturbations Theory #4 (SGP4), are adequate to the calculations for the Launch COLA process or whether high-precision satellite catalogues should be employed.

The launch pre-flight predicted trajectories against which collision avoidance screenings are run consist of state estimates (position and velocity information) and covariance data at regular time intervals (typically three-second intervals), with the data beginning from three to thirty seconds after launch and continuing until approximately two hours after launch, although the actual durations are mission-specific. The state estimates are provided in the Earth-Fixed Greenwich (EFG) coordinate reference frame, primarily for the ease of trajectory adjustment when modifications to the nominal launch time are required. The furnished covariance information is the lower-triangle of the position covariance, usually presented in the *UVW* (radial, in-track, cross-track) reference frame. This covariance information is derived from an uncertainty apportionment model that begins with the expected uncertainty in each of the physical components that govern spacecraft position (thrusters, gyros, &c.) as measured in the laboratory and combines all of these uncertainty data to synthesize an overall anticipated trajectory uncertainty at each time point. In some sense these covariance data are provided as a courtesy, as there appears to be no actual "realism" requirement levied on these error representations. The actual examination of the realism of these error representations is the topic of Section 3 of this report.

The statement of trajectory "truth" against which these pre-flight trajectories can be evaluated is the actual flight trajectory telemetry data collected during each launch event. These groups of telemetry data for the missions under analysis were obtained by KSC and converted into a format convenient for analysis: individual position and velocity measurements of the spacecraft (from on-board Global Positioning System [GPS] instruments), rendered in the Earth-Centered Inertial (ECI) True-of-Date reference frame, captured at typically one-second intervals. While this increased data density (increased, that is, over the three-second intervals in the predicted trajectory data) is certainly welcome, it is tempered somewhat by frequent data drop-outs. These drop-outs are due to lack of continuous ground-station coverage and other communications-related difficulties; and while they are more frequent and severe before Tracking and Data Relay Satellite System (TDRSS)-enabled communication became more widespread, they remain common throughout the five-year period from which trajectories were drawn for analysis.

In comparing the pre-flight trajectories to the associated telemetry in order to determine trajectory errors, a number of issues arise. First, the telemetry data lack any sort of accompanying error statement, either point-by-point or a single overall value (such as a

standard error). In order to use the data, one is required essentially to presume that the telemetry data represent error-free truth. One hopes, of course, that the errors in the telemetry data may be substantially smaller than those of the pre-flight trajectories so that the failure to account for the telemetry errors will not affect results. One will be better able to evaluate the rectitude of this assumption when the actual sizes of the trajectory errors are determined, as one can form some general opinions about the expected magnitude of GPS-derived position errors inherent in the flight telemetry.

Second, the time points of the pre-flight and telemetry data do not align precisely; so in order to compare the two datasets some sort of interpolation must be performed. The interpolation of telemetry points suggests itself as the better choice, as the spacing of the telemetry data is more frequent and because this approach would eliminate the need to interpolate covariance information. Nonetheless, the interpolation of the telemetry still needs to be performed with caution due to the aforementioned telemetry drop-outs: a simple-minded interpolation scheme that errantly attempts to interpolate over non-trivial data-gaps will produce extremely inaccurate results in such situations. The telemetry interpolation methodology must be robust enough to recognize such data gaps and avoid them, and the approach adopted here is to require that each of the two interpolation boundaries be within five seconds of the predicted point. Distances outside of this range are considered to constitute a data gap and thus are not included in the analysis.

Third, pre-flight telemetry and associated covariance information comes with a particular predicted launch time. In the majority of the cases the predicted launch time differs from the actual launch time; the difference between the two launch times is usually small—less than one second—but it can occasionally be more than this. The question is whether the actual or the predicted launch time should be used for accuracy analysis. The actual launch time scenario represents the trajectory situation as was actually flown, whereas the predicted launch time represents the situation as was actually screened. The decision was made to use the predicted launch time, as this situation represents the actual errors that made their way into the screening process.

As mentioned in the introduction, data were made available and analyzed for thirty-six pre-launch trajectories, comprising twenty-three Delta II, eleven Atlas V, and two Pegasus boosters. In some cases partial information was available for other missions, but it is only these thirty-six that had sufficient, consistent data to merit direct use in the analysis. The launches represented a considerable variety of orbit types, although both pre-launch trajectory and telemetry data exist for only the first two to three hours of each event; so all analysis remained within the near-Earth orbit regime (that region of orbital space for which orbital periods are less than 225 minutes). A visualization was generated for each trajectory analyzed, showing the actual flight path, the predicted path, and the associated error covariance ellipsoid as an additional check of the rectitude of the coordinate conversions &c.

Position comparison was made for each point where comparisons were possible; drop-out periods were simply not analyzed. While it is possible that data dropouts could have systematically excluded certain levels of trajectory error, the lack of a pattern in the drop-outs' temporal location and duration makes this unlikely and suggests the permissibility of straightforward exclusion of the drop-out periods. Displaying the position difference data as an empirical cumulative distribution function (CDF) curve by individual mission allows the full distribution of the trajectory errors to be observed and thus more meaningful comparison among different trajectories. Additionally, the data were filtered by an altitude threshold of 400km, as there is very little threat of close approach with space objects at altitudes lower

than this (less than 5% of the space catalogue exists at such low altitudes, and apparently spurious trajectory error data appear at altitudes lower than this). Figures 2-1 through 2-3 represent the CDF plots of position errors for all three booster types (Atlas V, Delta II, and Pegasus). The legends spell out the particular mission flights used by payload name, and amplifying information for each mission is given in Table 2-1 at the end of this section.

Figure 2-1 gives the position error results for all Atlas V launches. At the 50th percentile, nearly all Atlas V launches have errors greater than 10 km; and at the 95th percentile a little under half of them exceed 100 km.

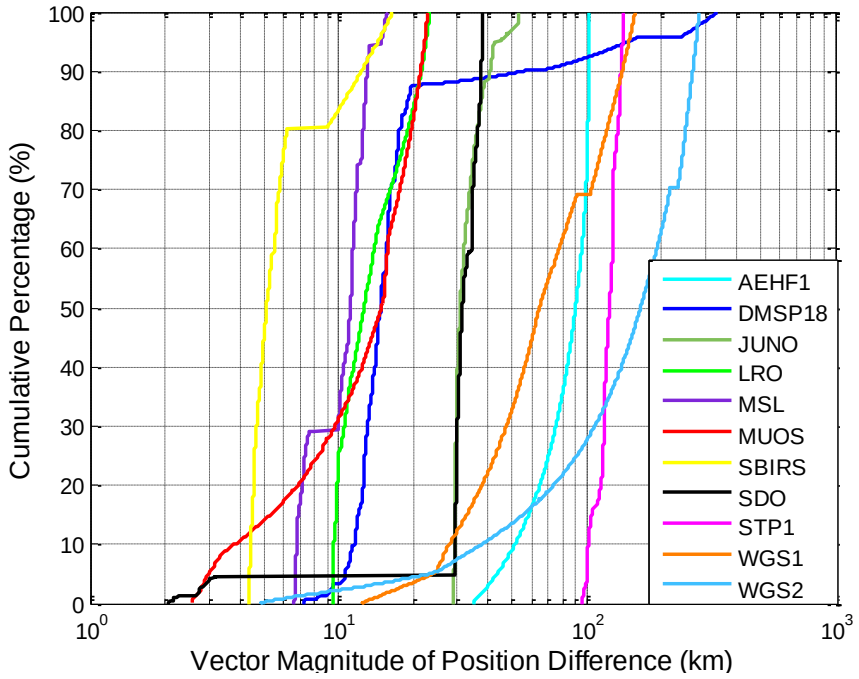
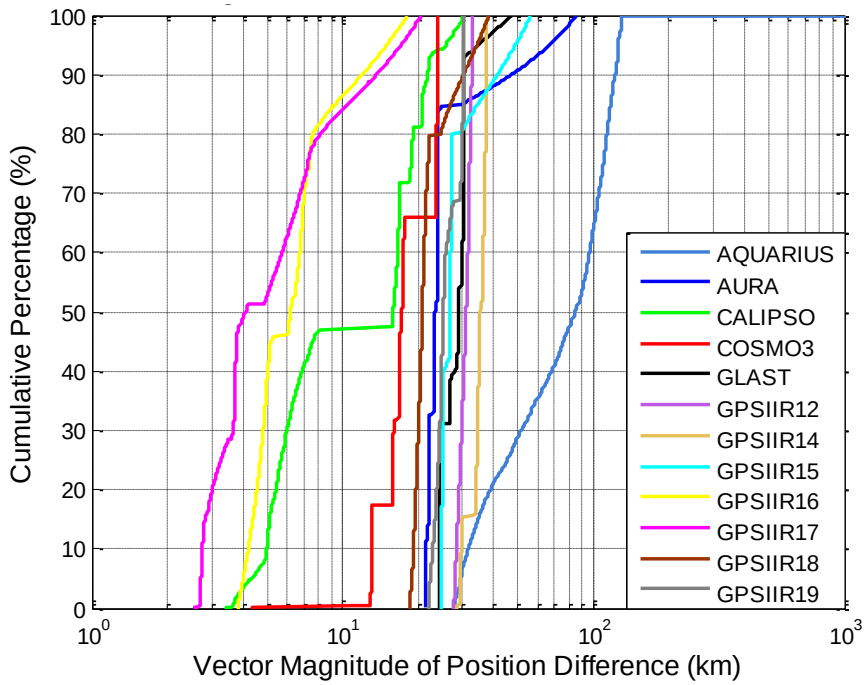
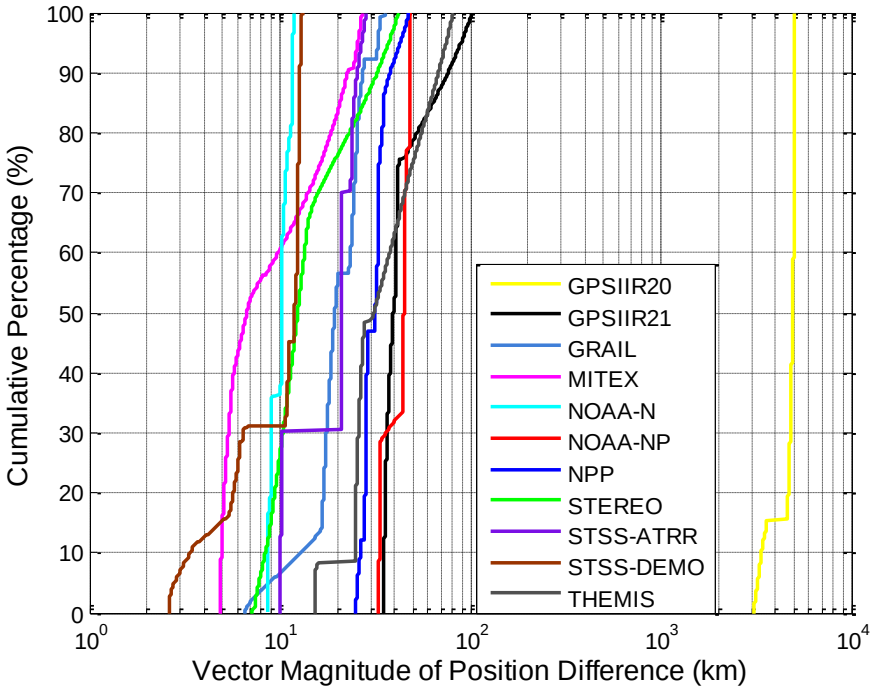


Figure 2-1: CDF plot of trajectory errors for all Atlas V launches





Figures 2-2a and 2-2b: CDF of trajectory errors for Delta II launches

Figures 2-2a and 2-2b give CDF representations of the position errors for the Delta II launches. On the whole, these show an error behavior that is somewhat more bounded than that for the Atlas V, although the majority of trajectories show errors above 10km at the

50th percentile. One trajectory, that of GPSIIR20, shows errors two orders of magnitude greater than all the others; and further examination was not able to determine any particular cause (*e.g.,* coordinate transformation problems). Because of its very different error character, this particular trajectory was judged to be anomalous and thus not considered in the remainder of the analysis.

Figure 2-3 represents the CDF plot of the position difference accuracy for Pegasus launches. Paired pre-launch trajectory and telemetry data were available for only three Pegasus missions; and all three of these were exceedingly short, with only a few minutes of total data in each. This is too little data from which to draw any durable conclusions, and as such the Pegasus trajectories are not used in the analyses of the subsequent sections; but as there is some interest in determining whether these solid-fuel motors seem to exhibit the same general error behavior as their liquid-fuel colleagues, the available data are presented here. The data for one of these trajectories never exceeded the 400km altitude threshold and thus are not shown in the figure below; for the two others, all available data are plotted. If this abbreviated sample is truly representative of the entire error distribution, then one can conclude that the Pegasus error magnitudes and distribution are roughly similar to those for the Atlas V and Delta II launches.

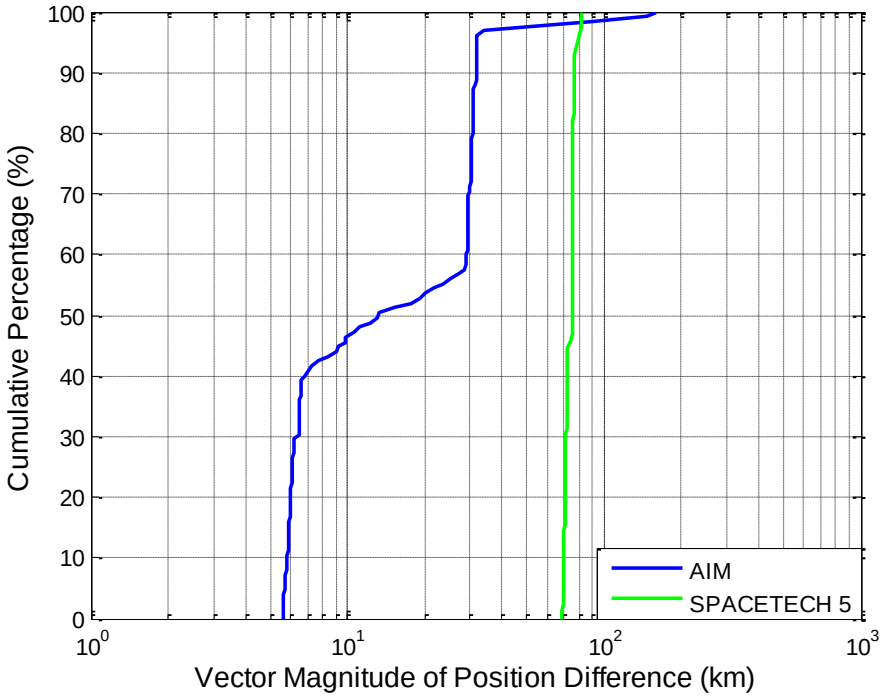


Figure 2-3: CDF of trajectory errors for Pegasus launches

Figures 2-1 – 2-3 give useful presentations of absolute error values and their distributions among the three different boosters analyzed, but it remains to put these error values in context; and the natural *comparandum* is the set of state estimate errors for orbiting satellites. Such a comparison will also assist in determining the level of catalogue precision appropriate to the orbital safety evaluations of these trajectories.

The United States Strategic Command (USSTRATCOM) space catalogue comprises three distinct sub-catalogues of orbital information. The first is the Special Perturbations (SP) catalogue, which uses a higher-order orbital theory to produce a precision set of state estimates; these data are used internally for JSpOC processes and orbital safety calculations for satellites but are not generally circulated to external users. The second is the traditional General Perturbations (GP) catalogue, maintained with SGP4 orbit determination (OD), circulated to external users, and in recent times posted on the SpaceTrack website (www.space-track.org) for anyone to download. There is, however, a new third option that is in beta testing now and is expected to transition to maintaining 75% or more of the GP catalogue: the so-called "extrapolation DC," or in some circles "eGP." This approach, described in detail in Cappellucci (2005),[2] takes a higher-order ephemeris (such as those from the SP catalogue) propagated into the future, creates synthesized observational data (called "pseudo-obs") from this ephemeris, and executes a GP OD from these synthesized data. Because the OD is fitting data derived from high-accuracy future predictions, the quality of the "predictions" using this approach is much better than the traditional, sensor-observations-based OD. Since the generally-available GP catalogue will be composed of these two types of GP OD approaches, a comparison of their errors to those of the pre-launch trajectories can serve both to ground the trajectory errors against some sort of benchmark and at the same time indicate whether the GP catalogue serves as an adequate basis for launch COLA calculations.

Figures 2-4 – 2-6 compare the errors in the launch trajectories to those of the publically-available portion of the satellite catalogue. In each of the figures each CDF line is of a particular accuracy percentile value: in Figure 2-4, it is the 50th percentile, in Figure 2-5 the 68th percentile, and in Figure 2-6 the 95th percentile (these percentile values were chosen to remain analogous to the familiar percentile values for the mean, one-sigma, and two-sigma levels from a Gaussian distribution, even though vector error distributions do not in fact follow a Gaussian distribution). Because the construction of the graphic may be somewhat confusing, the following is a step-by-step account of the features in Figure 2-4, the 50th percentile accuracy CDF collection. For the Atlas V case, the 50th percentile error value from each of the eleven trajectories was calculated and a CDF curve constructed of these eleven values (red line); the same was done for the twenty-three Delta II (green line) and two Pegasus (pink line) trajectories. For the satellite catalogue accuracy information, the data are derived from an analysis of all GP and eGP data for May 2012; and the 50th percentile values of a six-hour prediction (roughly equivalent to the amount of propagation used in LCOLA screenings) errors for each satellite were calculated and also summarized in CDFs. The precise method of calculating the satellite state estimate accuracies makes use of the reference orbits calculated for each satellite be the SuperCODAC accuracy estimation program that is presently part of the ASW operational system; this methodology is described at length in Hejduk, Casali and Ericson (2005)[3] and Hejduk (2008).[4]

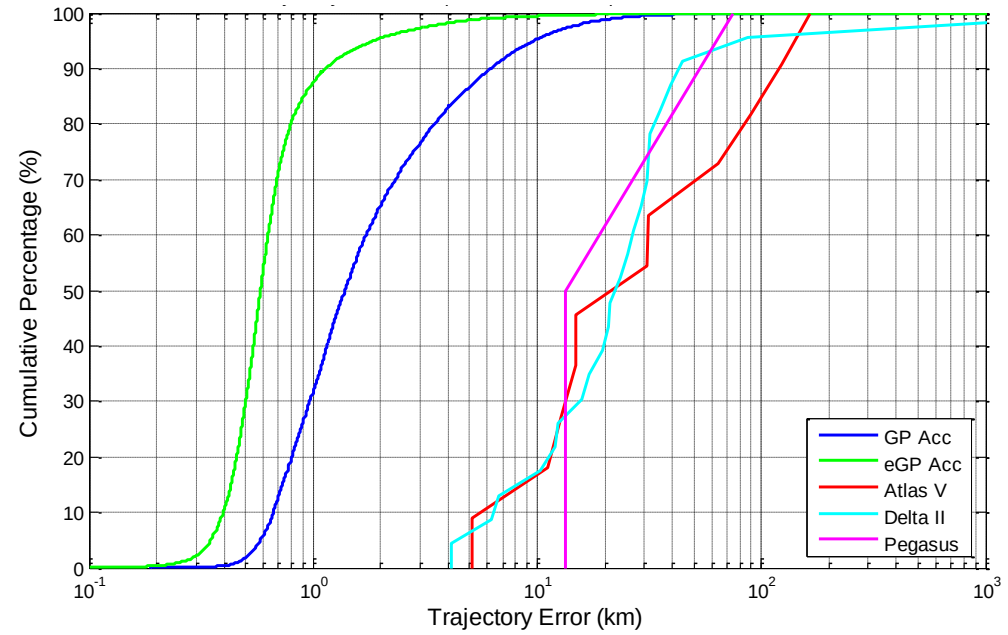


Figure 2-4: Launch / Satellite Accuracy Comparisons – 50th Percentile

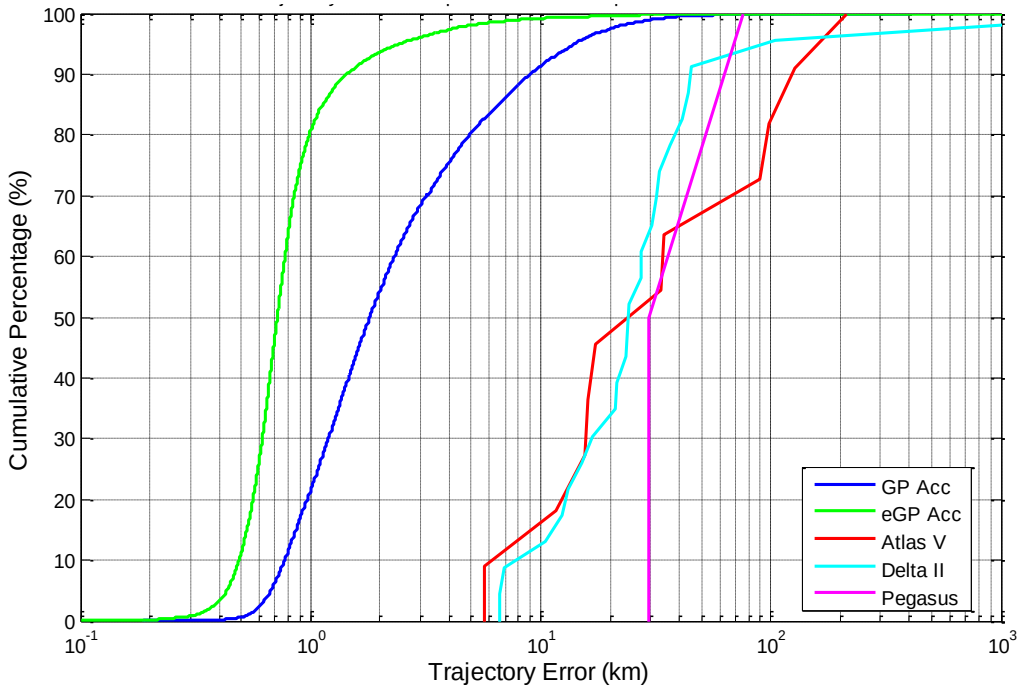


Figure 2-5: Launch / Satellite Accuracy Comparisons – 68th Percentile

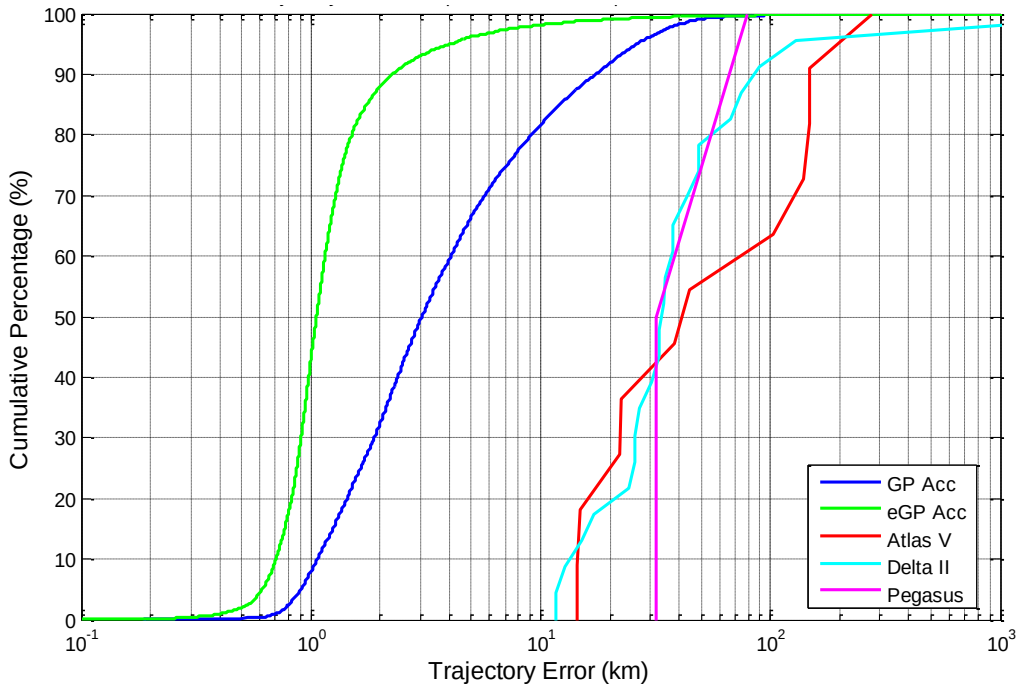


Figure 2-6: Launch / Satellite Accuracy Comparisons – 95th Percentile

In the 50th percentile graph (Figure 4), the pre-launch trajectory errors for most of the cumulative percentage span remain about an order of magnitude larger than the GP catalogue errors and about one and one-half orders of magnitude larger than the eGP errors. In the 68th and 95th percentile graphs, the difference is somewhat smaller but still hovers about an order of magnitude for much of the cumulative percentage span; this is especially so for the eGP data.

There are two conclusions one can draw from these data. First, because the satellite GP catalogue errors are consistently and substantially smaller than the pre-launch trajectory errors, it is clear that there will be little to no appreciable benefit from using the precision (SP) rather than the GP satellite catalogue for LCOLA screenings and risk assessment. It is true that in support of calculating probabilities of collision, actual OD covariance data generated and preserved with the precision catalogue can be used, whereas covariances must be estimated when using the GP catalogue; and the degree of difference this actually makes is addressed in Section 4 of this report. But in terms of accuracy levels alone, the inherent errors in the pre-launch trajectories are so much larger than the GP catalogue errors that if the precision catalogue is not conveniently available, there is little motivation to take on the difficulties of obtaining and using the precision catalogue for LCOLA.

One could also remark that the trajectory errors are so large in an absolute sense that orbital safety calculations will never be reliable enough to be actionable. This question will be taken up repeatedly and explicitly in the following sections of this report, but one should point out immediately that it is not the absolute values of the trajectory errors *per se* but rather the harmony between them and their error characterization (covariance) that truly determines their utility. If the associated covariances adequately model the actual trajectory errors, then the initial gate is cleared: credible orbital safety calculations, especially that of

the probability of collision, become possible. The next section addresses the reliability of launch covariance data to determine whether this initial condition can be met.

Table 2-1: Launch trajectory errors at common percentile values

#	Mission Name	Launch Date	Booster Type	50th Percentile	68th Percentile	95th Percentile
1	AEHF – 1	Aug-14-2010	Atlas V	89.95	98.59	101.94
2	DMSP 18	Oct-18-2009	Atlas V	14.80	16.05	147.47
3	JUNO	Aug-05-2011	Atlas V	30.89	33.48	44.71
4	LRO	Jun-18-2009	Atlas V	12.69	15.63	22.49
5	MSL	Nov-26-2011	Atlas V	11.13	11.72	14.87
6	MUOS	Feb-24-2012	Atlas V	14.87	17.24	22.23
7	SBIRS	May-07-2011	Atlas V	5.13	5.67	14.41
8	SDO	Feb-11-2010	Atlas V	31.53	34.53	38.00
9	STP – 1	Mar-08-2007	Atlas V	123.21	127.32	139.14
10	WGS – F1	Oct-11-2007	Atlas V	64.47	89.38	147.96
11	WGS – F2	Apr-04-2009	Atlas V	164.60	211.89	275.66
12	GLAST	Jun-11-2008	Delta II	28.97	30.21	35.06
13	GPS – IIR12	Jun-23-2004	Delta II	31.07	31.88	32.88
14	GPS – IIR14	Sep-26-2005	Delta II	35.36	36.71	37.54
15	GPS – IIR15	Sep-25-2006	Delta II	26.91	27.13	48.92
16	GPS – IIR16	Nov-17-2006	Delta II	6.21	6.97	15.03
17	GPS – IIR17	Oct-17-2007	Delta II	4.11	6.58	17.14
18	GPS – IIR18	Dec-20-2007	Delta II	20.76	21.44	34.35
19	GPS – IIR19	Mar-15-2008	Delta II	25.22	27.51	30.52
20	GPS – IIR20	Mar-24-2009	Delta II	4915.92	5003.52	5067.46
21	GPS – IIR21	Aug-17-2009	Delta II	39.36	41.36	89.52
22	GRAIL	Sep-10-2011	Delta II	19.41	24.30	32.76
23	MITEX	Jun-21-2006	Delta II	6.72	13.20	25.67
24	STEREO	Oct-26-2006	Delta II	12.39	14.98	37.33
25	STSS – DEMO	Sep-25-2009	Delta II	11.94	12.42	12.82
26	THEMIS	Feb-17-2007	Delta II	31.14	43.72	74.71
27	AQUARIUS	Jun-09-2011	Delta II	86.44	104.81	128.65
28	AURA	Jun-15-2004	Delta II	23.39	23.87	67.13
29	CALIPSO	Apr-28-2006	Delta II	15.90	16.86	25.82
30	COSMO – 3	Oct-25-2008	Delta II	17.15	23.61	24.11
31	NOAA – N	May-20-2005	Delta II	10.25	10.54	11.69
32	NOAA – N’	Feb-06-2009	Delta II	44.66	45.35	48.47
33	NPP	Oct-28-2011	Delta II	31.74	32.90	42.89
34	STSS – ATRR	May-05-2009	Delta II	21.05	21.19	26.87
35	AIM	Apr-25-2007	Pegasus	13.28	29.56	31.80
36	SPACETECH5	Mar-22-2006	Pegasus	74.83	75.22	79.39

Section 3: Launch Trajectory Covariance Realism

The previous section has shown that pre-launch trajectories do have large errors, certainly when compared to the state estimate errors of catalogued satellites. To some degree this level of performance is disappointing, as smaller errors do very much assist in separating real from fallacious orbital safety events. However, the actual enabling factor for orbital safety calculations is not the accuracy of the predicted trajectory itself but rather the precision of the accompanying covariance, which is the characterization of this trajectory error: if the covariance is realistic, then a meaningful probability of collision (P_c) can be calculated for an event and decisions made accordingly. Some can object that predicted trajectories with very large errors and accompanying appropriately large covariances are not useful to the orbital safety mission because the large covariance "dilutes" the P_c to the point that the probabilities are so small that they are not actionable; and indeed groups, even sizeable ones, of low-probability events do leave decision-makers in an indeterminate position *vis-à-vis* risk assessment. But it must be remembered that if the covariance is realistic, the P_c is the actual probability of collision; and this outcome—the presence of a real P_c —is the necessary beginning for a meaningful orbital safety risk assessment. Said differently, if the covariance is not realistic, then there can be no meaningful P_c calculation and risk assessment; whereas if the covariance does accurately represent the error in the estimated trajectory, then there is at least the possibility of performing meaningful calculations and from them drawing risk conclusions. It is thus necessary to investigate the launch trajectory covariances to assess their success in statistically characterizing the trajectory errors.

Before beginning such an investigation, it must be recognized that the formulation of the covariance used for the launch trajectory errors (and, for that matter, with satellite state estimates as well) imposes certain limits to its serving as an error statement. The covariance provides only the second moment (the variance) for each of the state component errors (covariance diagonal) and the terms indicating the degree of correlation among these variances (covariance off-diagonal terms). Embedded in this somewhat laconic presentation are the presumptions that the mean error in each component is zero, so no statement of error bias is needed; and that the error distribution is such that it can be characterized by the second moment alone. Although there are a number of distributions that can be characterized by just two parameters, implicit here is the presumption that the errors in each component will follow a normal (Gaussian) distribution. The examination of this presumption for satellite state estimation is an expanding research area with a considerable and growing collection of critical literature; many question the propriety of this presumption in propagation, but even in some cases at epoch (a good survey of this literature can be found in Ghrist and Plakalovic [2012]).[5] Because the form and framing of the covariance requires a Gaussian assumption, one will for this analysis simply accept it and not attempt to re-evaluate it here; but the tests for covariance realism do implicitly make this assumption. If the tests pass, then it is verified both that the error distribution is at least approximately Gaussian and that the covariance properly represents this distribution; if they fail, one is left with a somewhat indeterminate situation in that either the underlying error distribution is not Gaussian, the distribution is not properly represented by the covariance, or both.

A position covariance, which 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. The test procedure is to calculate a set of these state errors and determine whether their distribution matches that indicated by the position covariance matrix. To understand the particular 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 could calculate a set of in-track error values, here 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 vector 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 mean and dividing by the standard deviation:

$$\frac{\varepsilon - \mu}{\sigma} \ . \tag{3-1}$$

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, the subtraction as indicated in the numerator of Eq. 3-1 is unnecessary, simplifying the expression to

$$\frac{\varepsilon}{\sigma} \ . \tag{3-2}$$

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 Eq. 3-2 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 (*u* represents the vector of state errors in the radial direction, *v* the in-track direction, and *w* the cross-track direction):

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

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; and since the intention here is to test whether this covariance-supplied statistical information is correct, the test

* The velocity portion of the covariance, as well as the elements for the other solved-for terms in the orbit determination (such as drag and solar radiation pressure), are not explicitly tested as part of this analysis because it is only the position portion of the covariance that is used for the P_c calculation. The covariance matrices tested here are provided at each point along the trajectory and therefore do not need to be propagated by the user, so in the present case it is not necessary to consider testing the non-conservative force covariance terms. Some would argue that a testing of the velocity portion should be conducted because these terms affect the coordinate transformation of the covariance into the conjunction plane, a necessary activity in order to calculate the P_c through the typical method; but the usual approach of testing simply the position portion has been followed here.

statistic should use the variances from the covariance matrix rather than a variance calculated from the actual sample of state estimate errors. For the 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 *UVW* coordinate frame). In such a situation, the covariance matrix would look like the following:

$$C = \begin{bmatrix} \sigma_u^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_w^2 \end{bmatrix}, \tag{3-4}$$

and its inverse is straightforward

$$C^{-1} = \begin{bmatrix} 1/\sigma_u^2 & 0 & 0 \\ 0 & 1/\sigma_v^2 & 0 \\ 0 & 0 & 1/\sigma_w^2 \end{bmatrix}. \tag{3-5}$$

If the state errors are formulated as

$$\varepsilon = \begin{bmatrix} \varepsilon_u & \varepsilon_v & \varepsilon_w \end{bmatrix}, \tag{3-6}$$

then the pre-and post-multiplication of the covariance matrix inverse by the vector of errors (shown in Eq. 3-6) will produced the desired chi-squared result:

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

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.

This procedure is easily applied to the dataset in possession: the pre-launch data comprise a state estimate and covariance for each point in the trajectory, and the flight telemetry data provide a truth criterion; so the calculation in Eq. 3-7 can be computed for every pre-launch data point and that entire dataset examined for conformity to a three-DoF chi-squared distribution. Such an approach, however, would have multiple drawbacks: it would fall victim to the obvious correlation between closely-spaced data points, it would not account for the fact that different trajectories have different lengths (and thus more heavily weight the longer trajectories), and it would not take cognizance of the fact that different phases of a launch may have different error characteristics and thus different covariance realism results. To mitigate these difficulties, the following procedure was assembled.

First, the Atlas V and Delta II trajectories were analyzed separately (there were not enough Pegasus data to allow an investigation of their trajectories' covariance realism). Second, an interval of five minutes between analyzed points was maintained in order to attenuate correlation between points; this means that, for example, all of the 5-minute points for the Atlas V trajectories were evaluated together, all of the 5-minute points for the Delta II's, all of the 10-minute points for the Atlas V's, all of the 10-minute points for the Delta II's, &c. Finally, even though there are 11 Atlas V and 23 Delta II trajectories in the dataset, because of drop-outs there are not necessarily that number of data points at each 5-minute boundary point; so for certain statistical analyses a minimum number of data points was imposed in order to consider a particular trajectory / time-point pair.

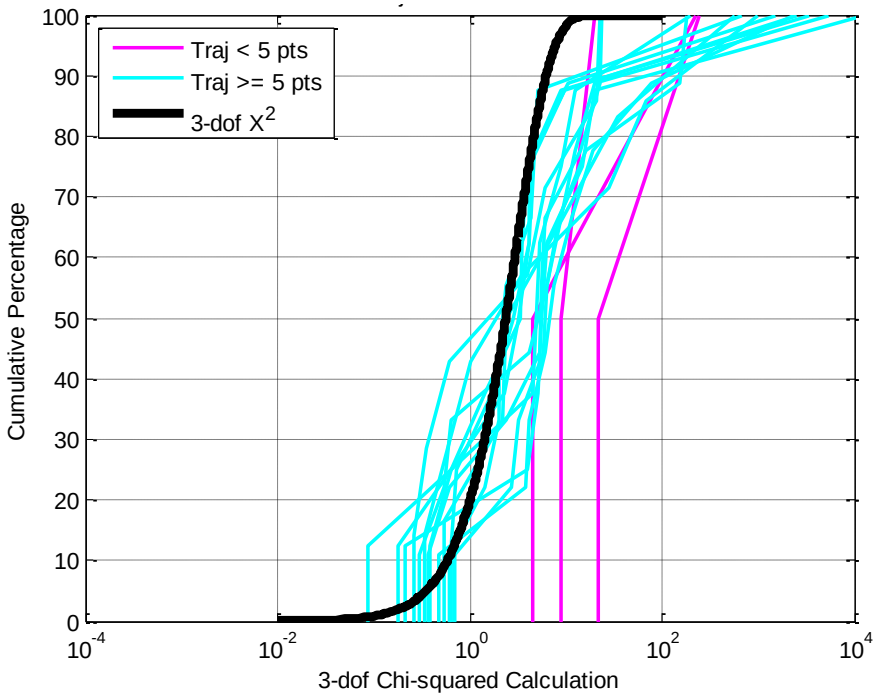


Figure 3-1: Atlas V trajectories: 15-90 minutes, at 5-minute intervals

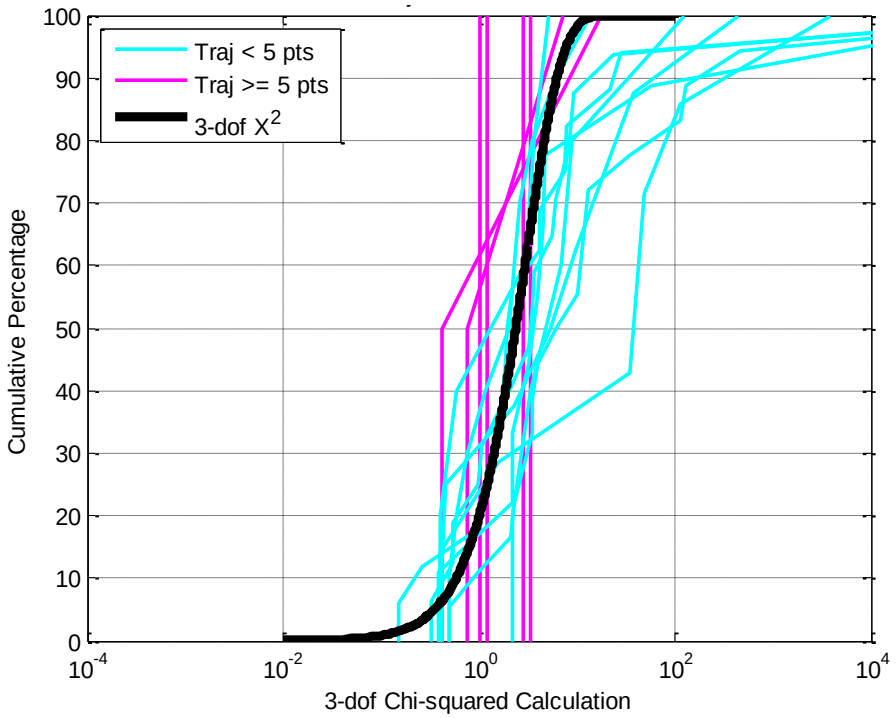


Figure 3-2: Delta II trajectories: 15-90 minutes, at 5-minute intervals

Figures 3-1 and 3-2 show the results, for the Atlas V and Delta II trajectories, of CDF-plotting of the chi-squared results from each trajectory time point; here, five-minute intervals from 15 to 90 minutes were used, resulting in a total of sixteen trajectories for each. Those trajectory time-points that had fewer than five points, and thus are more difficult to interpret in terms of their CDF behavior, are shown in magenta; those with five points or greater are shown in aqua; and the ideal three-degree-of-freedom chi-squared CDF is given as a thick black line. One can see that for both trajectories, the CDF curves, especially the aqua ones, are reasonably similar to the ideal black line. Each booster group included one trajectory with especially large chi-squared values (probably resulting from unusually large trajectory errors that were much larger than the accompanying covariances would suggest); if these two trajectories were excluded from the analysis, the performance would improve considerably in the upper tails of the CDFs, which is where the most divergence is observed. But rather than exclude these based on their poor performance alone, it is best to let them remain in the analysis for the present and remove them only if a reasonable case can be made for such exclusion later.

It is encouraging, certainly, that there is a considerable degree of visual alignment between the ideal and empirical distributions, but this sort of “ocular regression” is hardly conclusive; what is needed is a rigorous statistical test to determine whether these empirical distributions can be considered associated with a chi-squared parent distribution. Such a desire leads the investigation to the statistical sub-discipline of “goodness of fit” (GOF).

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 precisely, 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 hypothesis is 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, say, two or three chances in one hundred that the differences between the two samples (null and research) can be explained by sampling error alone, which in this case 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 GOF, namely, to evaluate of 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"; but that is not necessarily an unreasonable method: 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. Such a view is appropriate to the present purpose, namely whether the behavior of the CDFs for individual time-points in the trajectories of certain booster types can be reasonably ascribed to a 3-DoF chi-squared parent 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. Of all of these, the EDF methodology is generally considered to be both the most powerful and most fungible to different applications, so it is the one selected for use here. The general EDF approach is to calculate and tabulate the differences between the CDF of the sample distribution and that of the hypothesized distribution, to calculate a GOF statistic from these differences, and to consult a published table of *p*-values for the particular GOF statistic to determine a significance level. Specifically, there are two GOF 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 believed that the quadratic statistics are the more powerful approach, especially for samples in which outliers are suspected; so it is this set of GOF statistics that were employed for the present analysis. The basic formulation for both the Cramér – von Mises and Anderson-Darling approaches is of the form

$$Q = n \int_{-\infty}^{\infty} \left[F_n(x) - F(x) \right]^2 \psi(x) dx \ ; \tag{3-8}$$

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

$$\psi(x) = 1 \tag{3-9}$$

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:

$$\psi(x) = \{F(x)[1 - F(x)]\}^{-1} \tag{3-10}$$

Because it is already suspected that each booster type has a single trajectory that is an outlier in the upper tail, it is appropriate to choose the Cramér – von Mises statistic for this investigation.

It is a straightforward exercise to calculate the statistic in Eq. 3-8, discretized for the actual individual points in the CDF for each trajectory (that is, changing the integral into a summation). This calculates the Cramér – von Mises statistic, from this point on called the "Q-statistic," as suggested in Eq. 3-8. 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.[6] 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). In the present case, the *p*-value table instead will be interpolated to determine the *p*-value level associated with each tested time-point, and the *p*-value results can then themselves be plotted as a CDF.

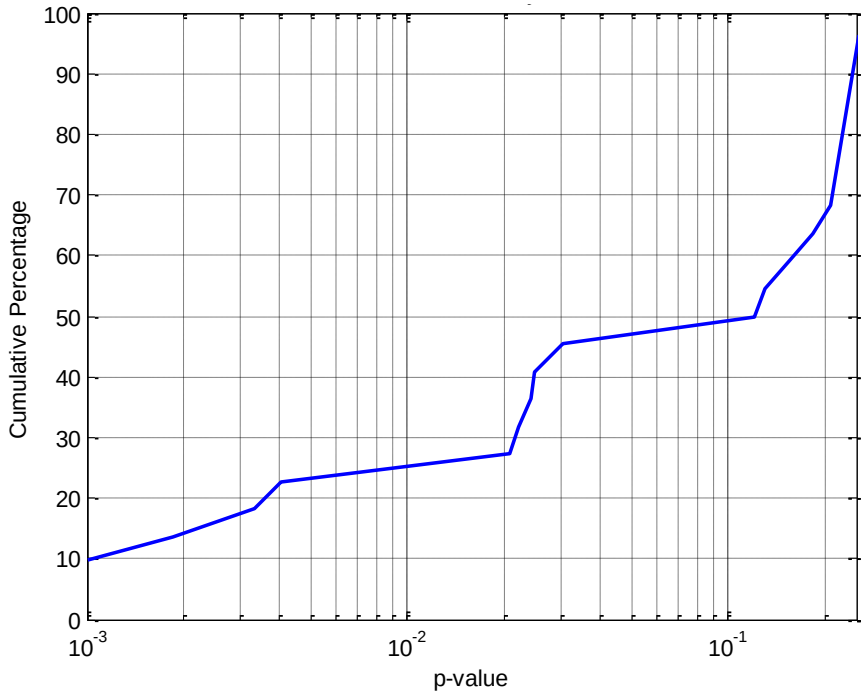


Figure 3-3: p-value results for all trajectories, time points 15-90 minutes, > 5 samples

Figure 3-3 gives the *p*-value results, with outcomes from all trajectory time points (with at least five data points) all mixed together. The boundaries of the graph are set at the boundaries of the published *p*-value tables (0.1% to 25%), so intersections with the vertical axes indicate percentages of the dataset that were not directly evaluatable but yet can still

be interpreted (either bad enough to constitute unhesitating rejection of the null hypothesis or at a level at which there would be uniform acceptance of the null hypothesis).

In understanding these results, it is helpful to keep in mind that a *p*-value range of 2% to 5% is typical for retaining the null hypothesis in a goodness-of-fit evaluation, meaning that values greater than this would indicate that the sample could be considered to derive from the hypothesized parent distribution, in this case a 3-DoF chi-squared distribution. From Figure 3-3 one can observe that a little more than 50% of the cases exceed a p-value of 5% (0.05) and about 75% exceed a p-value of 2% (0.02). This is a very encouraging result: 75% of the cases examined pass outright, without any outlier exclusion or manipulation, a GOF text for the hypothesized distribution. If one were to be more permissive and allow a result of 0.1%, which is the boundary for the published tables, then 90% of the results would pass the GOF test.

It would be unreasonable to expect that all of the results of such an examination, especially given the vicissitudes of actual operational data production, would conform to the idealized case; so the fact that 75% of such cases pass the GOF test at a commonly-chosen *p*-value threshold is quite significant, and that this can be extended to 90% through a somewhat more lenient application of the test. Furthermore, since it is not certain that in all cases individual component errors would follow a Gaussian distribution, these results are all the more impressive. It can be said that the covariances supplied with the launch trajectories do indeed accurately represent the observed trajectory errors.

If the trajectories' associated covariances are realistic, then the calculation of a believable and durable LCOLA Pc is possible; and one can thus proceed to the next phase of the analysis, namely to examine what these LCOLA Pc values look like with different calculation approaches. In order to do this, a large dataset of LCOLA Pc calculations, executed with a variety of different paradigms, is necessary. A screening analysis experiment to produce such a dataset is described, and its results analyzed, in the next section.

Section 4: "Grand Experiment" and Initial Characterization of Pc Results

The previous sections have established that the pre-launch predicted trajectories, while significantly more error-infused than a typical GP satellite ephemeris, nonetheless have their errors properly represented by their accompanying covariances and thus meet the precision criteria for use in orbital safety calculations—as it is not the error values themselves but the ability properly to characterize those errors that is important. It was also shown that, because trajectory errors are about an order of magnitude greater than GP satellite catalogue error levels, orbital safety calculations can legitimately be performed with both a GP and SP paradigm. Since it has been determined that LCOLA can be properly performed with these data, the question thus turns from the "whether" to the more practical considerations of "how" and "how much."

Concerning the "how," one is interested in the differences between the results of GP and SP screenings, especially with regard to the use of the less-precise GP approach as a proxy for SP. The differences between using the maximum Pc value from all the conjunctions in a particular screening as the representation of the conjunction risk for that trajectory, or a cumulative Pc that amalgamates all of the individual Pc values for each conjunction in the screening, should also be investigated. It is clear that a per-screening cumulative Pc value more comprehensively characterizes the overall risk, but it must be determined whether it differs significantly from the corresponding per-screening maximum Pc value, and if so under what conditions, in order to justify a move away from the simpler and more paradigmatic maximum Pc approach.

Once the "how" is settled, the "how much" becomes relevant, namely particular ranges of Pc values, in both GP and SP, that can serve as appropriate launch window closure thresholds. The purpose of this study is not to advance hard values (or value ranges) for these thresholds; rather, it is to give trade-space presentations, such as window closure percentages as a function of window size and Pc screening threshold, so that the practical effects of choosing different window closure (*i.e.,* screening) thresholds can be understood. Some may see this as corrupting a disinterested determination of risk tolerance, which should perhaps be made without considering the effects of the choice on the ease of launch window closure; but it is important to take a holistic view of the problem, as the selection of a desirable screening threshold is likely to be a balance between risk reduction and ability to conduct launch operations without undue encumbrance. Once an understanding of the effects of a threshold choice is available, it is helpful to try to characterize the overall risk reduction secured by the LCOLA process; this can be done by comparing the level of risk assumed between doing nothing at all to mitigate risk versus screening at a given Pc value and closing windows appropriately.

Finally, there are two adjuncts to the "how much" question, and they both center on using miss distance as a substitute for Pc-based screening. While the miss distance that gives an equivalent result (at a certain percentile level) to a Pc screening can be calculated, it is also necessary to take cognizance of the effect of this admittedly simplified procedure on launch window closures, as miss-distance-based screening is a heavy-handed approach that exacts a considerable toll on launch availability. As a solution to this problem, it has been frequently suggested that the screening activity be limited to a certain list of defended assets only, essentially setting aside the collision risk posed by tracked space debris.

Leaving unaddressed the latent policy question of whether one should prescind from considering space debris collision risk, one should at least determine whether limiting screenings to defended assets only succeeds in providing desirable levels of launch window access.

To attempt to answer all of these questions, a large screening experiment was designed and performed to generate a dataset of screening results that could be characterized and mined. Five trajectories from the set established in Section 2 were chosen, representing different orbits and booster types; these trajectories are given and described in Table 4-1. Each of these trajectories was screened against a large number of configurations: the launch time was backed up fifteen days from the nominal launch epoch, and screenings were run at one-minute intervals until fifteen days after the nominal epoch. This resulted in 43,200 screenings per trajectory, for a grand total of 216,000 screenings from all five trajectories.

Table 4-1: Launch Trajectories Selected for Screening Experiment

Trajectory #	Satellite	Launch Vehicle	Target Orbit	Range
1	*NOAA-N'*	Delta II	LEO	Western
2	*GPS II-R21*	Delta II	LEO (Parking)	Eastern
3	*STSS Demo*	Delta II	LEO	Eastern
4	*DMSP-18*	Atlas V	LEO	Western
5	*SDO*	Atlas V	GEO	Eastern

Each of the screenings was run in two different ways. First, the screenings were run at the Aerospace facility against the GP catalogue using their "Collision Vision" GP screening toolset. Because the GP space catalogue lacks covariance information, this utility uses component-based quadratic error growth curves for the appropriate satellite class to estimate component error as a function of time and thus synthesize a GP covariance at the time of closest approach (TCA) between the two objects.[7] The GP satellite file of the day of the trajectory launch was used; this scenario is slightly operationally unrealistic in that it results in back-propagation of element sets in some cases, but this is not expected to alter the outcome notably. The screening is first conducted for a given keep-out volume around the trajectory (100km box); for objects that penetrate this volume, a minimum miss distance and Pc are calculated at TCA. The typical value used for the hard-body radius, which indicates the approximate combined size of the spacecraft, is 10m; and that value was used for the present screening set. The toolset imposes a reporting threshold of 1E-12, so conjunctions with Pc values smaller than this are not reported. Occasionally because of this thresholding a screening run might not report any conjunctions, but this outcome is extremely rare in GP.

The second method, run at the a.i. solutions facility, ran the same trajectories and number and time-phasing of screenings but used the Air Force Space Command, Studies and Analysis Division's COMBO (Computation of Miss Between Orbits) astro standard

against the GP catalogue, with a large screening volume selected (100km box), to identify all of the conjunctions; this volume should be large enough to capture all the conjunctions that would emerge from an SP screening, which typically would be more modest in the size of its screening volume. Once identified, this list of conjunctions was reprocessed using the SP catalogue: in each case the appropriate Vector Covariance Message (VCM, which is the method of encapsulating all information about the SP state estimate; and here the VCM with the epoch time closest but prior to TCA) was retrieved and a covariance-enabled ephemeris was produced in the vicinity of the GP TCA. This ephemeris was compared to the trajectory using an interpolation technique to find the TCA and thus for that time calculate the SP miss distance and Pc. The standard two-dimensional Pc approximation was used, following the Foster calculation approach.[8] As with the Aerospace GP screenings, a 10m hard-body radius was employed.

There is a mismatch that arises between these two datasets, for a number of reasons; and this fact does govern the conduct of some of the subsequent parts of this study. First, the JSpOC SP catalogue attempts to maintain a certain standard level of quality and as part of doing this requires a minimum amount of observational data in order to perform an SP update. Many objects do not receive this minimum and thus do not have an SP vector at all or only one that is too old to be usable; but as there is no such similar restriction on the GP catalogue, element sets do exist for these objects. During the periods of time investigated, there were about 1000 such objects—those that are present in the GP but not the SP catalogue—at any given time, so one would expect conjunctions from a GP screening that are not seen in an SP screening. Second, the Aerospace screening software differs from the GP portion of the approach used for the runs at the a.i. solutions facility. The use of the Aerospace software for the GP screening portion of the experiment was the appropriate choice, given that it is the standard toolset used by KSC for LCOLA. However, it is a different implementation from the COMBO astro standard, it does not use the same version of the SGP4 propagator, and, most significantly, it left-censors the results at 1E-12, whereas the SP results do not introduce any left-censoring (and thus produce a much larger return dataset). Given these factors, each results set (GP and SP) contains conjunctions not found in the other.

This non-alignment of results is addressed in the following way. For investigations that wish to compare SP and GP results directly to determine the degree of difference between the two methods, results for the subset of conjunctions that are found in both results sets will be used. For all other investigations (which is most of the analyses of the remainder of this report), reported GP and SP results are each derived from the complete dataset that emerged from that particular approach (either the Aerospace GP screening set or the a.i. solutions GP-SP screening set). This allows the fullness of each screening approach's results to feed the analyses, restricting this only when it is necessary to have a direct comparison of the GP and SP results.

The next analysis to be presented is such a situation: it seeks to compare the SP and GP Pc and miss distance results arising from the same conjunction events. SP is the more precise theory, and it natively produces a covariance (in contrast to the synthetic GP covariance produced by the Aerospace toolset); so it is typically and reasonably presumed to be the better and more believable result. The GP calculation, however, is a much faster computation and easier to obtain and work with (since the GP catalogue is publicly available). Furthermore, it is not necessarily a worse result. There is suspicion that the SP catalogue's covariances are often too small, and in fact there is at present an Aerospace analytical effort to investigate this;[9] if true, then the SP Pc values will at times be too large

and at others too small, depending on the ratio of the miss distance to the covariance size. The GP synthetic covariance is not a true *a priori* covariance, *viz.* the inverse of the normal matrix; but because it is derived from error growth curves themselves constructed from accuracy information from a large number of element sets, it represents a durable error characterization representative of actual state estimate errors. So it is important to see how different the results are from both approaches, both in a direct comparison and in general behavior over a range of different Pc values.

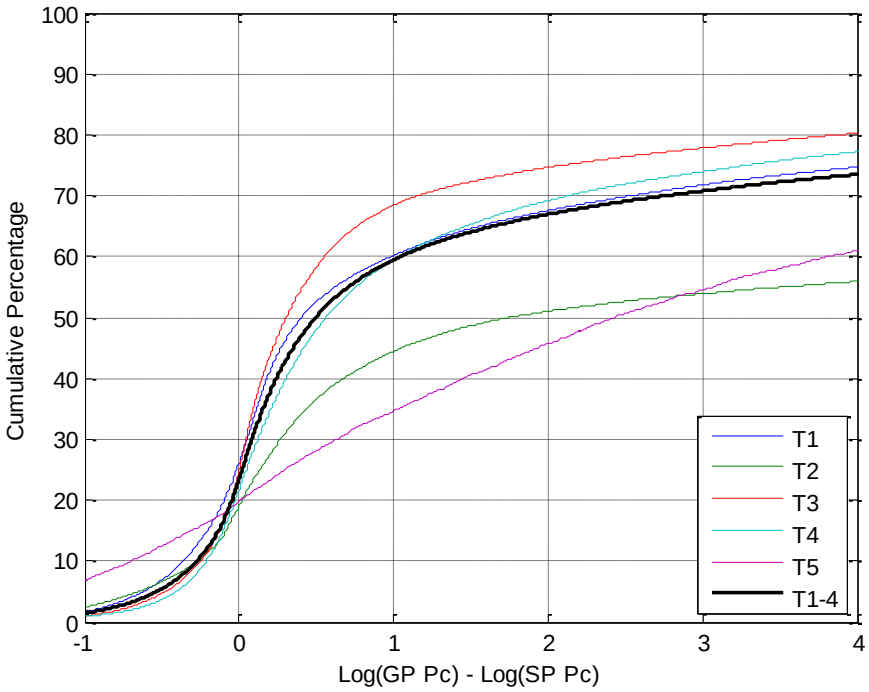


Figure 4-1: Comparison of GP and SP Pc Values

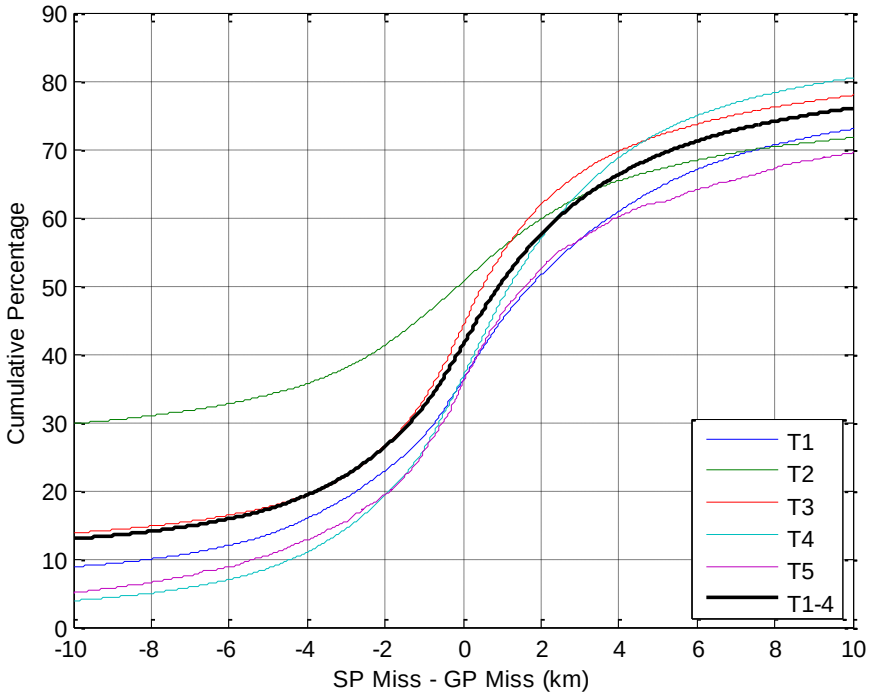


Figure 4-2: Comparison of GP and SP Miss Distances

Figure 4-1 provides a CDF comparison of the Pc values for the experiment's event set that was common to both GP and SP runs; because the graph is dense and may require some orientation to read, an interpretive explanation is given here. One notes immediately the differences among the five trajectories. Three of the trajectories (T1, T3, and T4) clump within a fairly narrow range of behavior; T2's response gives the same general shape but in a manner well apart from the other three trajectories. T5, the one deep-space trajectory, exhibits a very different behavior and also represents a much smaller response set—only between 3% and 5% of the number of conjunctions reported for each of the other four trajectories. The thick black line represents the CDF results for trajectories 1-4 taken together.

The particular construction of the x-axis of this CDF plot arises from the manner in which one determines whether a change in the Pc value is significant. The rule of thumb in the industry, arising from JSC experience with human space flight orbital safety work and subsequently broadly adopted, is that changes in Pc less than an order of magnitude typically do not result in a substantially altered risk posture and thus are not considered significant. This same standard is a reasonable threshold for defining a significant difference between GP and SP Pc values for the same conjunction event: differences greater than an order of magnitude constitute a significant departure. The actual data plotted are the difference between the base-10 logarithm of the GP Pc and the base-10 logarithm of the SP Pc. A value of zero represents no difference between the two Pc values, values falling within the region of -1 to +1 are within the single order-of-magnitude difference range and represent Pc values that are essentially the same, and values falling outside of this region do constitute significant differences between the two Pc values. A

positive value means that the GP Pc is larger (*i.e.,* GP results indicate a conjunction is more likely) and a negative value the opposite (SP Pc represents a riskier scenario).

In interpreting these results, one notes first that 60% of the cases fall within the single-order-of-magnitude range, so a supermajority of cases essentially manifest equivalence between the GP and SP calculations. Of the remaining 40% that do not, nearly all of these cases fall to the right of the +1 boundary of the order-of-magnitude region, meaning that the GP Pc is larger. There is not ideal convergence of the CDF lines in this region—T3 shows 80% of cases to be smaller than four orders of magnitude, whereas T2's results show only 55% of cases to be better than four orders of magnitude; but on the left boundary (at -1) there is a surprising convergence: only about 2% of the cases, for all the main four trajectories and their average, produce an SP Pc that is more worrisome than a GP Pc. These results allow the significant conclusion that GP is a very good LCOLA proxy for SP: GP and SP Pc results do not differ significantly the majority of the time; and when they do, in nearly all of the cases the GP Pc is larger and thus the more conservative value. The only real downside to using the GP Pc is that, because it is conservative, it is likely to result in unnecessary launch window cut-outs.

Figure 4-2 provides an accompanying diagram of miss distance differences. The x-axis for this graph is the simple difference between the SP- and GP-computed miss distances for the paired conjunction events. A positive value indicates that the GP miss distance is smaller and thus a greater collision threat, making the GP calculation more conservative. Here one observes a fairly even distribution of the miss distance differences: while in general the GP miss distance is smaller (60% of the cases) and thus more conservative, it is not at all the lopsided result observed for Pc differences, in which 80% of the cases GP is more conservative and in only 2% of the cases is SP more conservative in a significant way. This miss distance result is consistent with previous studies that have documented the relative lack of correlation of miss distance with Pc and thus with conjunction risk.[10]

Having established that GP Pc is a reasonable proxy for SP Pc, from this point forward all analyses will consider GP and SP results separately, drawing from the two different executions of the screening experiment, so that behavior and thresholding for these different calculation types can be independently established. Even though the results are generally comparable and GP results are conservative, the absolute values of the calculated Pc's are both different and may have different behaviors in a cumulative probability sense; so it makes sense to treat both datasets separately yet display their results in parallel. The first opportunity for this is the analysis that follows here, which is a general investigation of Pc behavior.

In addition to treating the SP and GP datasets separately, the manner in which the screening results are used is also modified. For the previous figures, every individual conjunction was considered in mapping the relationship between GP and SP; this is because one is interested in the general relationship between the two calculation methods for all conjunctions, not just the ones that emerge as the most worrisome from individual screening runs. For most of the analyses that follow, however, the focus will be on the set of single-value summaries of the overall conjunction risk for each screening run. This means that there will be 43,200 such values for each trajectory, one for each screening run (left-censoring reduces this by a little for the four main trajectories and for the fifth trajectory by quite a bit). The important issue is thus how one chooses to represent the overall risk associated with a particular screening run.

Typically, when a screening run is performed, the *maximum* Pc is reported, meaning the largest Pc value arising from the entire set of conjunctions that the screening identified. But since the trajectory usually encounters close approaches with a number of objects and each of these close approaches has a Pc, it can be argued that it makes more sense to calculate and use the *cumulative* probability of the spacecraft colliding with any of these close-approach objects. After all, one is not primarily interested in the possibility of a collision with just the riskiest potential conjunction but with all the possible conjunctions along the trajectory. In this LCOLA differs from on-orbit orbital safety because their methods of remediation are different. For the on-orbit CA case, the mitigation methodology is (almost always) to maneuver the primary satellite; and because maneuvers can really only be designed to eliminate a single potential conjunction, it is appropriate at any given moment to focus on the riskiest conjunction, which is the one with the maximum Pc. For the LCOLA case, the remediation methodology for a risky screening result is to cut that particular launch opportunity out of the launch window; and this can be done as easily working from a cumulative Pc that considers all of the potential conjunctions that a particular launch opportunity presents as with the Pc from the riskiest single conjunction.

The calculation of the cumulative Pc is straightforward because it is reasonable to consider all of the possible conjunctions for a particular trajectory to be independent events. Thus, one needs to calculate the probability that at least one of *n* independent events, each with its own probability, will occur. In point of fact, of course, no more than one collision would actually occur during the flying of the trajectory; but since such fictitious situations (*i.e.,* two or more collisions occurring) already represent the catastrophe of a collision, it is reasonable to include their eventuality in the calculation. The easiest way to perform the calculation is to calculate the cumulative Pc as the compliment of no collision occurring. Let a particular screening run produce *n* close approaches, each with its associated Pc: P_{c1}, P_{c2}, … P_{cn}. For any single conjunction, the probability of no collision is the compliment of the Pc: 1- P_{c1}, 1- P_{c2}, &c. The cumulative probability of no collision is thus the product of all of these probabilities:

$$(1- P_{c1})(1- P_{c2}) \ldots (1- P_{cn}) , \tag{4-1}$$

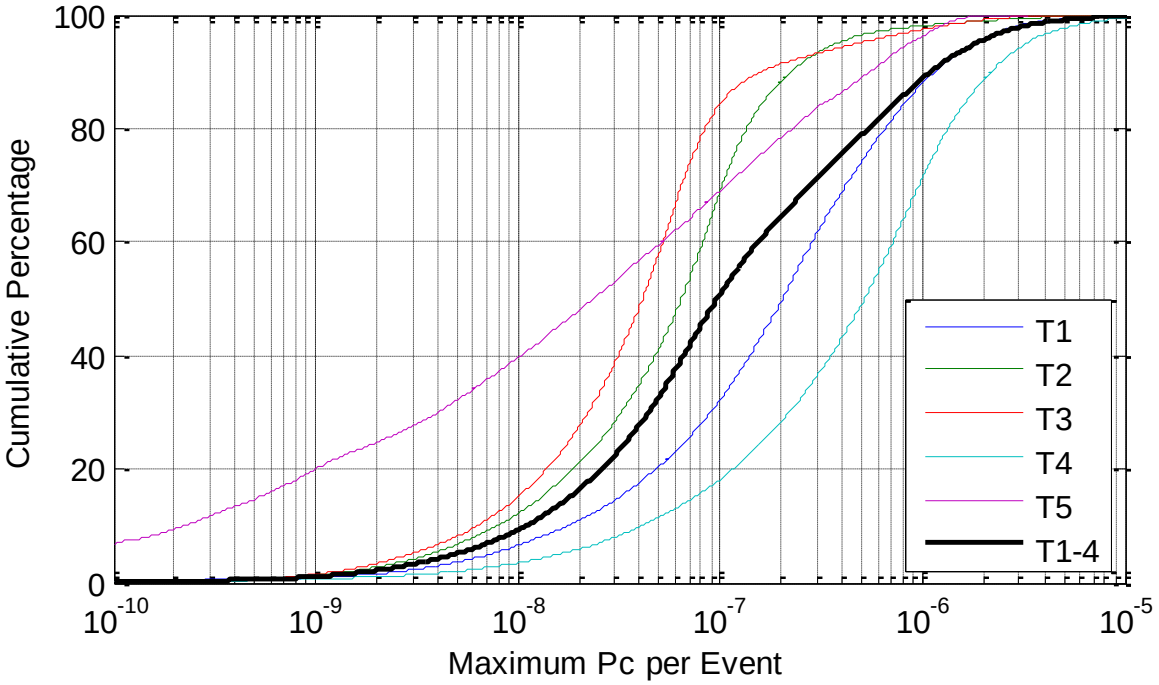
so the cumulative probability of at least one collision is the compliment of this product:

$$P_{ccum} = 1 – [(1- P_{c1})(1- P_{c2}) \ldots (1- P_{cn})] . \tag{4-2}$$

Figure 4-3 begins the discussion of the general behavior of Pc, with a CDF plot of the maximum and cumulative Pc values for each screening event, both by individual trajectory and as the results for trajectories 1-4 taken together. It is helpful to remember that the GP results are left-censored at a Pc level of 1E-12; this fact will become especially important in contrasting these results to those for SP. One can see in both graphs that the behavior of the results for trajectory 5, which represents the deep-space launch, is substantially different from the other four near-earth trajectories; and as such it will receive limited consideration here. Focusing at first on the maximum Pc graph, one observes first that the responses have died out substantially in significance by 1E-08 (only about 8% of the screened trajectories, on average, have a maximum Pc smaller than 1E-08), and by 1E-09 there are virtually none. On the right-hand side of this graph one sees a much greater spread of responses among the trajectories, but the principal interest is at what Pc threshold, if at all, do these curves change their characteristic shape, as such a point may suggest itself as a natural threshold. For trajectories 2 and 3 there is a notable bend of the CDF curve at around 2E-07 and 1E-07, respectively; but for the other curves the behavior is much

smoother and does not present so obvious an inflection point. Perhaps one could suggest the neighborhood of 1E-06 for T1 and 3E-06 for T4, but even this is tenuous. A value of 1E-06 would identify only a few percent of the conjunctions as hazardous, about 10% for T1, and 25% for T4; one could suggest this as a right-side limit, although not in any definitive way. So the neighborhood of 1E-06 to 1E-08 seems to be a reasonable bounding area for a Pc screening threshold, as it represents the region where the behavior of the CDF curves change. But given the lack of a striking change of behavior for at least some of the trajectories, this particular conclusion can be only a starting point for further investigation.

The cumulative Pc graph gives the same general presentation and behavior as that for the maximum Pc graph except that it is shifted somewhat to the right, as one would expect; for the cumulative Pc will always be greater than or equal to the maximum Pc. The shift to the right varies somewhat by trajectory but is a few ordinal levels higher; one might keep the 1E-08 left cut-off but perhaps shift the right boundary of the neighborhood to 2-3E-06.



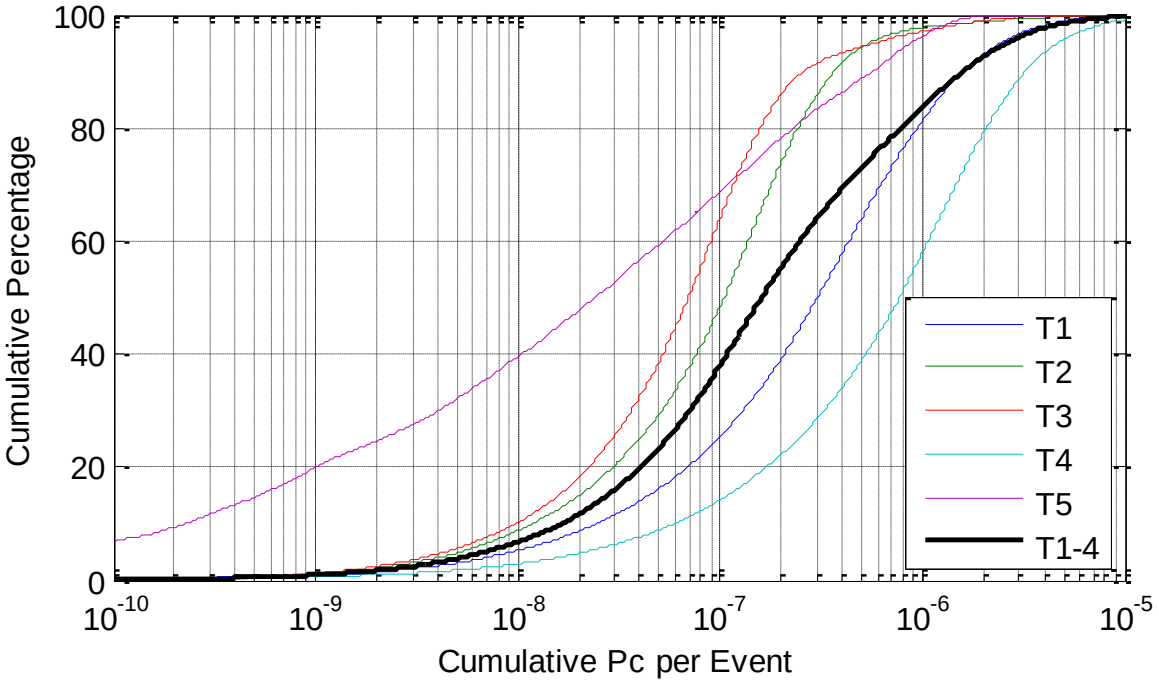


Figure 4-3: GP maximum and cumulative Pc per screening event

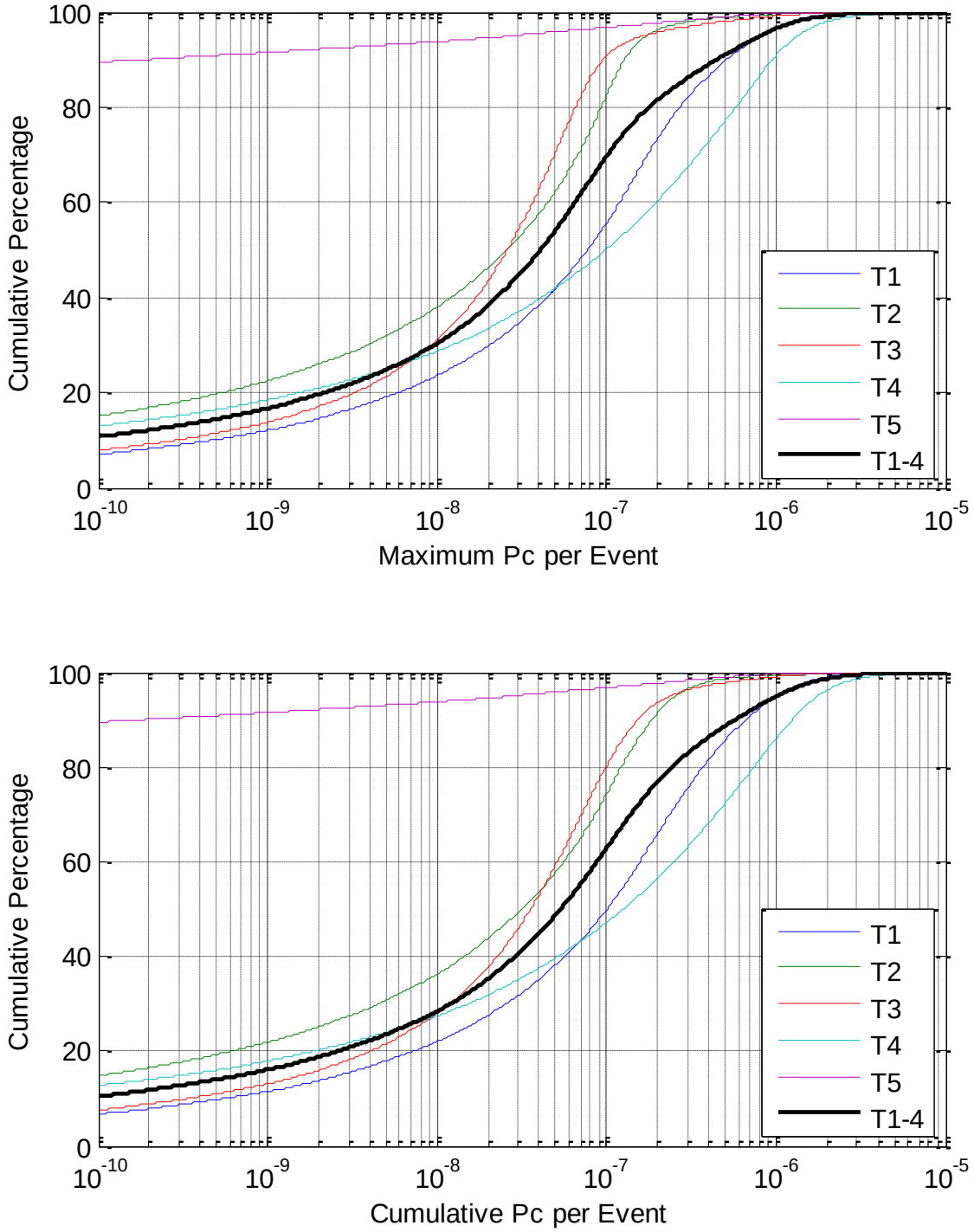


Figure 4-4: SP maximum and cumulative Pc per screening event

The SP graphs, while again giving the same "look and feel" as their GP counterparts, do exhibit some different features, namely in the failure of the curves to decay to a negligible level by 1E-10. The GP Pc values nearly all of the time were higher than the SP calculations for the same events, often by orders of magnitude; so this general behavior is

not surprising. This issue is brought to an extreme with trajectory 5, whose behavior appears significantly different, although the magnitude of this difference is to some degree an artifice of the method of data accumulation. When the left-censoring of the GP screening calculations resulted in no conjunctions at all for a particular screening, that screening was simply eliminated from the data pool. For trajectories 1-4, there were very few such cases, so the CDF results are not significantly affected by this elimination. For trajectory 5, however, the great majority of the screening results produced no conjunctions with Pc values higher than the left-censoring threshold; so because these trajectory results were eliminated rather than represented by a zero value, the percentile performance appears very different than for the SP data, where no left-censoring was used. In any event, one now sees a shift to the right by a few digits in the ordinal: one might begin the viable neighborhood at 7E-07 and perhaps continue it to 5E-09; in the cumulative case, one might use the GP max Pc value of 1E-06 but match the SP maximum Pc case in running the left boundary to 5E-09.

These bounding regions for screening thresholds for the GP and SP cases direct the subsequent investigations of this study by providing ranges of the Pc value for which these additional investigations need to be performed. The most important of these is the effect on launch window “openness” or availability, which will be treated in Section 6. But first it is necessary to take a short detour to explore the differences between maximum and cumulative Pc in order to try to establish the preferred method for assessing the risk embedded in any given set of screening results.

Section 5: Maximum versus Cumulative Pc

In the previous section, the concept of the cumulative Pc, which encapsulates the cumulative risk posed by all of the close approaches experienced by a particular trajectory rather than that of the single most dangerous close approach, was introduced; and it was further remarked that for the purposes of LCOLA this might be a preferable method for performing risk assessment, as it aligns naturally with the way LCOLA risk is mitigated (*i.e.,* elimination of the particular proposed launch time rather than trying to modify the trajectory). To move to this approach would, however, constitute a change of paradigm, requiring software and procedure modification as well as potentially imposing an additional computational burden; so it is important to quantify the degree of difference and/or improvement that might be expected from such a modification before determining whether such a change is warranted. The two sub-parts to this section address the two portions of this question. The first determines the frequency with which the cumulative Pc might indicate the violation of a given Pc threshold when the simple use of the maximum Pc would not. The second investigates the possibility of using, instead of the "main" Pc threshold that actually represents the desired level of risk mitigation, a proxy screening threshold somewhat smaller that the main threshold. The hope here would be to capture, by screening the maximum Pc at a slightly lower (more demanding) threshold, most of the cases in which the cumulative Pc exceeded the main threshold but the original maximum Pc did not. If this latter approach is viable, it could be a way to achieve essentially the same result as that of using a cumulative Pc screening but without requiring a change of software or paradigm.

There are several ways one might look at the differences between the maximum and cumulative Pc values produced by a large number of screenings. A CDF plot of their simple differences (which will be a one-sided difference, since the cumulative Pc for a given screening will always be greater than or equal to the maximum Pc for that screening), for example, would show that in most cases this difference is small yet in a non-discountable number of cases it is large. What such a presentation does not do is give any insight into the significance of those differences *vis-à-vis* the actual purpose of the screening, namely the determination of whether its presented risk exceeds a stated threshold. What is needed is an examination of the frequency of the cumulative Pc's exceeding any given threshold with the maximum Pc remaining below that threshold, as these are the actual cases in which screening by the cumulative rather than maximum Pc would have operational benefit. The results of this particular calculation are given in Figure 5-1. The x-axis represents the Pc screening threshold, and the y-axis represents the percentage of the cases for which the maximum Pc is below the screening threshold and the cumulative Pc above the threshold. Results for the GP dataset are shown in blue and those for the SP dataset in green. It is interesting that for both datasets the percentage of cases peaks around the 1E-07 point, although the peak is nearly twice as high for the GP dataset as for the SP. In both cases, there are Pc threshold values at which a significant number of cases manifest a cumulative Pc that exceeds the threshold but a maximum Pc that does not; in the GP case, the frequency of such cases can be as high as 13%. So clearly there is value in considering the issue of cumulative Pc in greater depth, with the degree of value dependent on the particular Pc threshold selected.

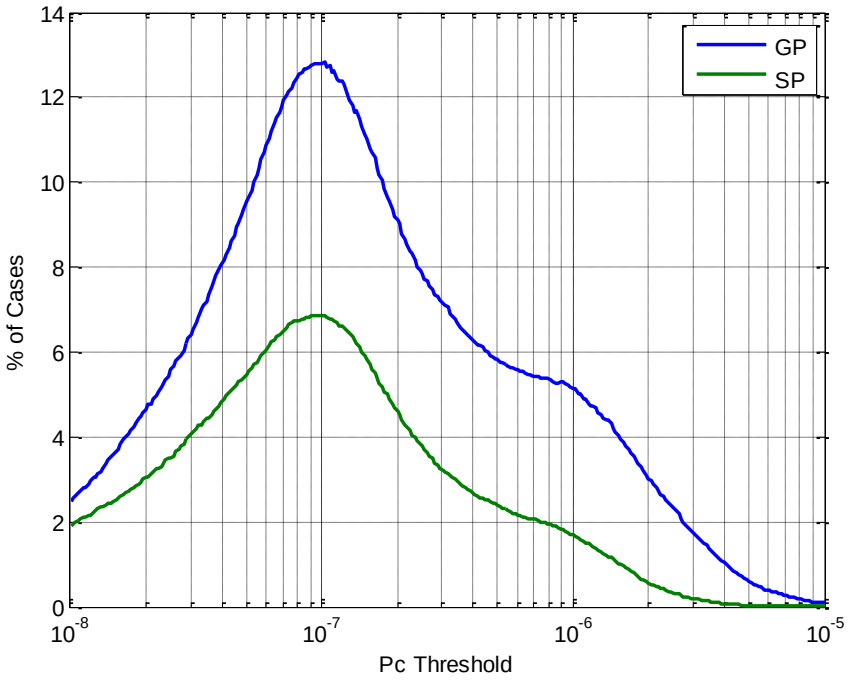
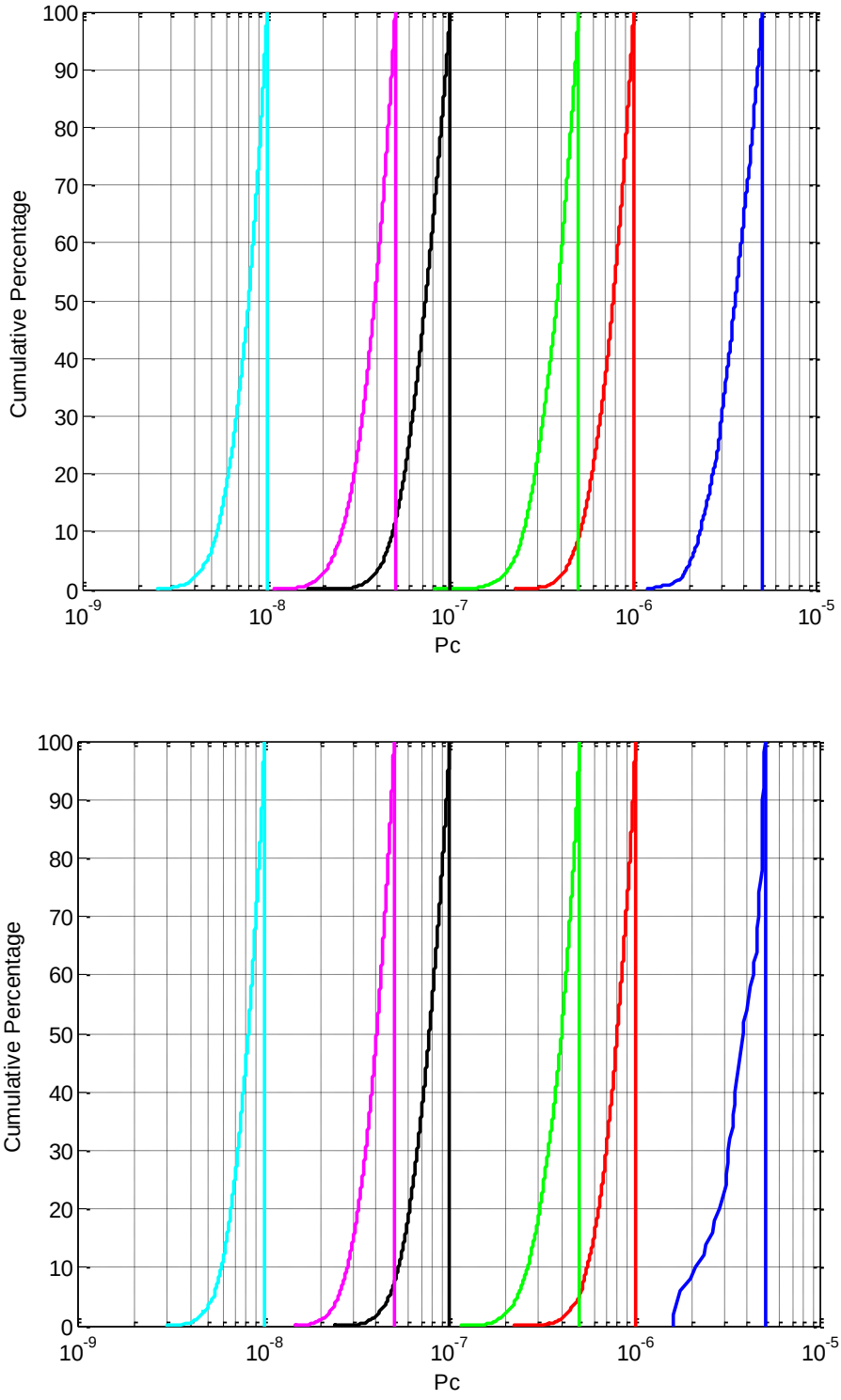


Figure 5-1: Percent of cases in which cumulative Pc exceeds screening threshold but maximum Pc does not

The next level of "depth" would be to determine, for the cases shown in the graph above, how often the differences between cumulative and maximum Pc are significant as opposed to trivial. If in most such cases the difference is very small—representing an increase from, say, 9E-07 to 1E-06—then a change to a cumulative Pc calculation would not be very compelling; or in any case the same effect could probably be accomplished with a very modest modification of the Pc threshold to a slightly smaller value. If, however, the differences are often significant, then clearly there are a number of screenings whose results give the appearance of safety when evaluated by their maximum Pc but are actually quite risky.



Figures 5-2 and 5-3: GP (above) and SP (below) maximum Pc distributions for cases in which cumulative Pc exceeds the screening threshold but maximum Pc does not

Figure 5-2 is an attempt, for the GP case, to try to characterize the distribution of these differences at a number of different Pc screening threshold values in order to get a sense of their significance. Since the graph is somewhat confusing as presented, it is best to take a descriptive tour of one of the colored curves and naturally extrapolate to the others; to do this, the red curve will be selected. This curve plots results for a Pc screening threshold of 1E-06, and this is indicated by the vertical-line portion of the graph, which is placed directly over the 1E-06 point on the x-axis. The curved red portion is a CDF plot of the maximum Pc values for those cases where the maximum Pc is below the screening threshold but the cumulative Pc is above the Pc screening threshold; this presentation allows the distribution of the differences between the maximum Pc and the Pc screening threshold to be observed. One can see, for example, that the 60% line of the CDF plot intersects the red CDF curve at approximately 8E-07, meaning that in 40% of the cases for which the maximum Pc and cumulative Pc straddle the screening threshold, the maximum Pc is between 8E-07 and 1E-06—not an overwhelming difference. One also observes, however, that the 10% CDF line intersects the red CDF curve at approximately 5E-07; ten percent of the cases thus manifest a difference of 5E-07 between maximum Pc and screening threshold, a factor of two difference. Similar behavior is present for all five of the screening thresholds explored in the figure, although the log-plot scaling obscures this and forces the reader to recognize that the relationship between maximum Pc and screening threshold is here geometric rather than linear. For example, at the 10% line was observed for the red curve a maximum Pc to screening threshold difference of 5E-07 to 1E-06, whereas for the green curve the two values are 2.5E-07 and 5E-07; both pairs of values have different linear differences but each differ by a factor of two. Very similar behavior is observed for the SP case (Figure 5-3).

It is necessary to decide whether the differences shown in the figure should be considered significant. At the 50th percentile line, the approximate ratio for most of the curves is 75% (3/4); so even for the screening threshold that produced the greatest number of “straddled” cases between maximum and cumulative Pc (around 1E-07 for GP), this would mean that only 6% of the total cases would have a ratio larger than this relatively small one. As the differences between the maximum and cumulative Pc increase, the number of cases only decreases, as Figures 5-2 and 5-3 show. To expand the example given previously, for the 1E-07 screening the number of cases that have the somewhat large factor-of-two difference total only 10% of the maximum-cumulative straddle cases, or about 1.2% of the total number of cases at that screening threshold. One would need to decide whether this would be considered significant, and there is good reason to suppose that it would not be: an approach that is adequate to 99 out of 100 cases could well be judged acceptable, and the factor-of-two difference is considerably smaller than the order-of-magnitude rule-of-thumb used to assign significance to a change in Pc.

Such an assessment of significance is not, however, adjudicated in a vacuum—the decision is not just an unalloyed assessment of risk tolerance but also includes consideration of the ease of remediation. If, for example, a small adjustment in the Pc screening threshold would allow a maximum-Pc-based screening to capture many, perhaps even most, of the situations in which the cumulative Pc exceeds the true desired screening threshold when the maximum Pc does not, then it may be desirable in fact to perform this remediation because it is so easily accomplished. The desirability of such an approach, of course, is not simply a question of the ease of calculation but also a consideration of efficacy: how often does this alternative screening technique—making the Pc screening threshold somewhat smaller (more demanding) in order to capture the straddle situations—actually successfully perform this capture and how often does it generate false alarms?

Table 5-1: Outcome Matrix for "Adjusted Pc" situations

Row		Pc < Adjusted Threshold	Adjusted Pc Threshold (*e.g.,* 5E-07)	Pc Between Thresholds	Main Pc Threshold (*e.g.,* 1E-06)	Pc > Main Threshold		Result
1		Max Cum						Same
2		Max		Cum				Same
3		Max				Cum		Type 2 Error
4				Max Cum				Type 1 Error
5				Max		Cum		New Finding
6						Max Cum		Same

In order to explore this, all the different combinations of where the maximum Pc and cumulative Pc may fall in relation to a main Pc threshold (the level of risk at which one actually wants to screen) and an "adjusted" Pc threshold (a somewhat more restrictive threshold to catch the situations in which the cumulative Pc exceeds the main threshold) need to be addressed. The situation is complicated, so one must walk through it slowly with the help of a visual aid, provided here in Table 5-1 above. In this table, "max" is the single largest Pc resulting from a screening, and "cum" is the cumulative Pc calculation for that same screening. The column in which these designators ("max" and "cum") fall represents their values relative to the adjusted and main Pc thresholds; in row 1, for example, both the max and cum Pc calculations are smaller than the "Adjusted Pc Threshold" and thus appear in the column to the left of the "Adjusted Pc Threshold" column. The result of such an arrangement of screening results is given in the "Result" column: "same" means that the same result is realized by screening with the max Pc at the adjusted threshold and the cumulative Pc at the main threshold, and type 1 and type 2 errors have their usual meanings. Each row of the table is explained in the paragraph below.

In row 1, both the max and cum results are smaller than the adjusted threshold, so the same result is obtained if a max screening is done against the adjusted threshold (result below threshold) and a cum screening against the main threshold (result below threshold). In row 2, the max result is below the adjusted threshold and the cum above the adjusted threshold but below the main threshold; since both results are below their proper thresholds (max below adjusted and cum below main), the result is the same for both (result below threshold). In row 3, the max is below the adjusted threshold but the cum above the main threshold; this scenario, a type 2 error, is the most worrisome in that it is a real violation (cum above main threshold) but not identified as such by the proxy screening (max below adjusted threshold). In row 4, both max and cum are above the adjusted threshold but below the main threshold; this is a false alarm (type 1 error) because the max result is above the adjusted threshold and thus throws the alarm, but the cum result is not actually above the main threshold. In row 5, the max result is between the thresholds and the cum result above the main threshold; this is the situation that the proxy screening hopes to identify: screening the max value at the adjusted threshold identifies a case in which the cum result is above the main threshold. Finally, in row 6, both max and cum are above the main threshold, so both screening approaches will yield the same result.

In assessing the efficacy of the proxy screening at the adjusted threshold, scenarios that yield the same result (flagged as "same" in Table 5-1) need not be analyzed or reported on explicitly. Rather, one is interested in the percentage of cases that yield type 1 errors (false alarms), type 2 errors (missed conjunction events by using the adjusted Pc threshold), and

situations in which the proxy screening actually identifies a conjunction that fell below the threshold when screened with the max Pc approach but exceeded it when screened with the cum Pc approach ("new finding"). Clearly, one wishes a situation in which the type 1 and type 2 error rates are as low as possible and "new finding" rates are as high as possible. These results will, of course, be a function of the actual values selected for the two thresholds and the degree of their difference from each other; so the best way to get a sense of the situation is to construct sensitivity plots that show these results as a function of these two variables (main and adjusted Pc screening levels). One should also be attentive to the total percentage of cases reflected by each of these values: if a given result, regardless of whether it be good or bad, occurs in a very, very small percentage of the screenings analyzed, it should not serve as a significant decision driver.

The charts in Figures 5-4 and 5-5 present these results. Again, the chart presentation is not immediately intuitive and requires some orientation. Each of the sub-plots is for a given main Pc screening threshold; for the upper-left graph in Figure 5-4, which will serve as the model for this explanation, the main Pc screening threshold is 5E-06. The x-axis represents different values of the adjusted Pc screening threshold. At the far right side of the axis, it is equal to the main Pc value, and it decreases as one proceeds left until it is one-tenth of the main Pc screening threshold. The y-axis reports the percent of cases; the range is from 0.01% to 100% on a logarithmic scale. The percent of cases with type 1 errors is given by the blue line; as expected, it is greatest when the adjusted and main Pc screening thresholds differ the most and decreases as these two values approach each other. The percent of cases with type 2 errors is given by the green line; it is lowest when the adjusted Pc threshold is the most demanding (*i.e.,* furthest from the main Pc threshold) and highest as the two thresholds are brought together. The percent of cases that exhibit "successful" behavior in the sense that they properly identify, via a max Pc screening at the adjusted threshold, situations in which the cum Pc exceeded the main threshold is given by the red line; these values are smallest when the adjusted and main Pc thresholds are close together, and as they spread apart the number of "success" cases increases until it reaches an asymptote. This asymptotic behavior makes sense in that, once the two thresholds reach a certain degree of separation, the likelihood of finding additional successful cases trails off substantially.

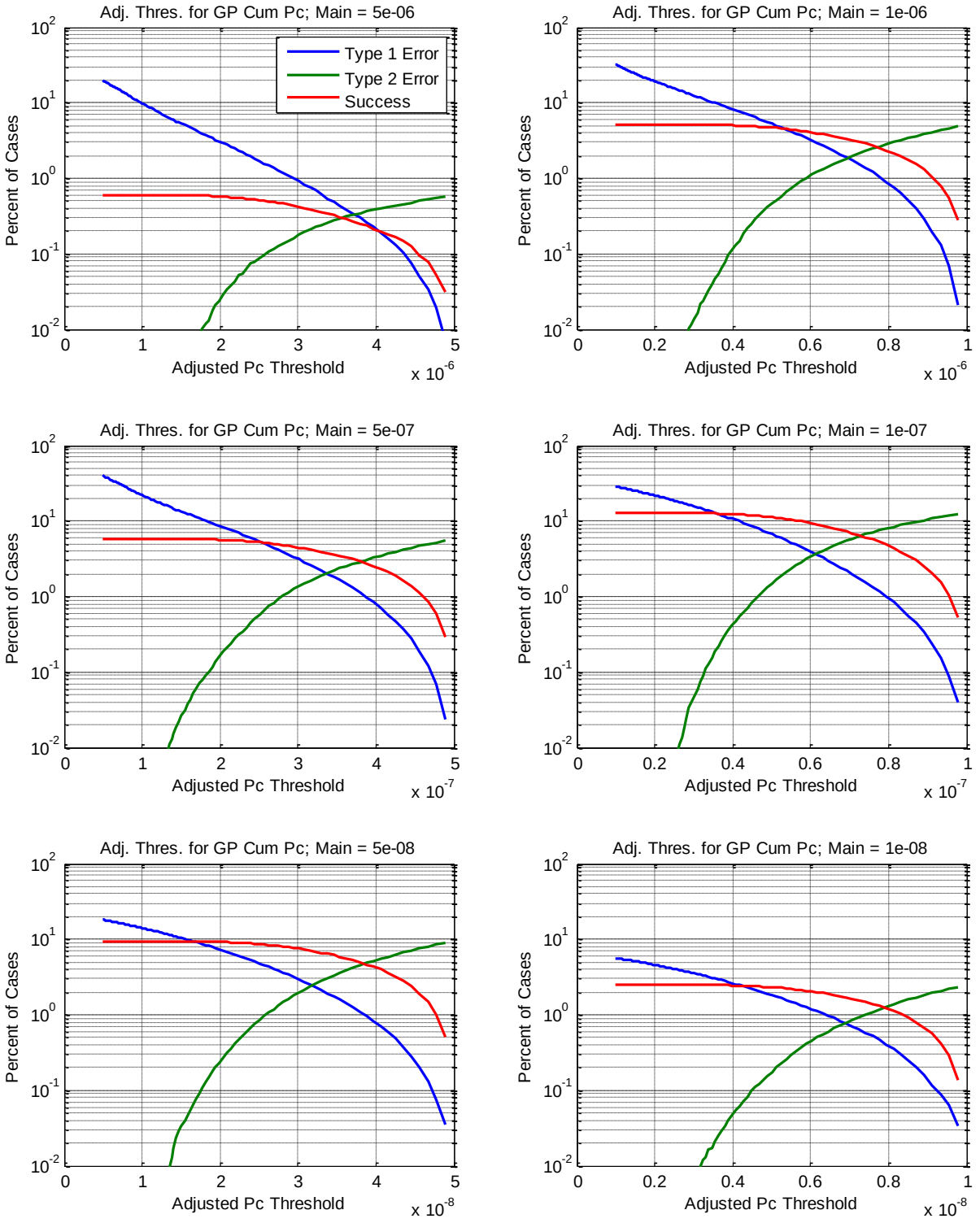


Figure 5-4: Successes and errors for use of adjusted GP Pc screening thresholds

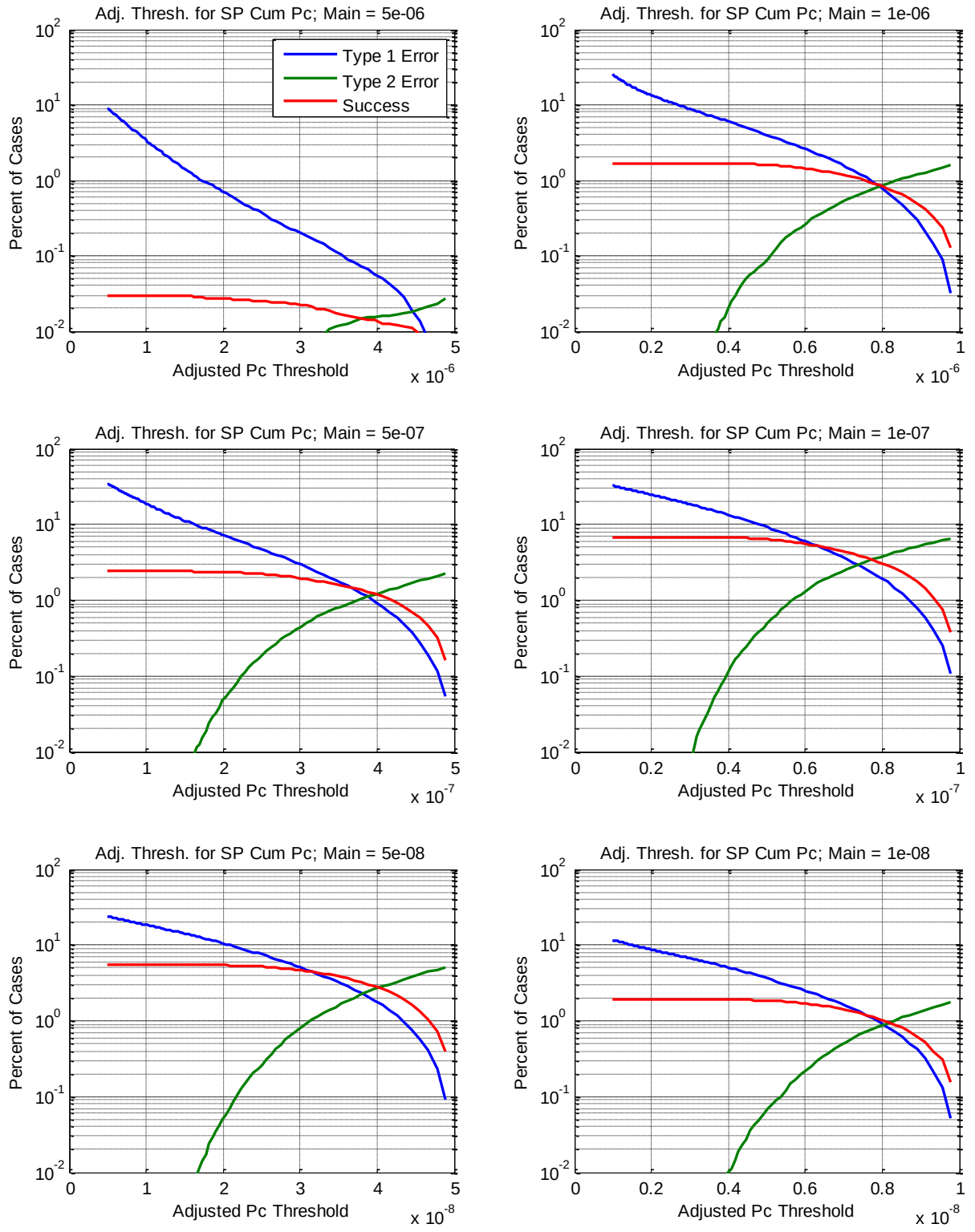


Figure 5-5: Successes and errors for use of adjusted SP Pc screening thresholds

To aid in interpreting these charts, two subplots in the GP results set (Figure 5-4) will be examined in detail. The first of these is the top left subplot, for which the main Pc threshold is 5.0E-06. For most of the adjusted Pc range (x-axis), the false alarm rate (blue line) is much greater than the success rate (red line); and by the time these two curves approach parity, the type 2 error rate (green line) is greater than either. For this particular main Pc threshold (5E-06), there is little to no advantage in using any of the adjusted Pc thresholds. The second case is the middle right subplot, for which the main Pc threshold is 1.0E-07. Here, the situation is more sanguine. At the 0.6 x-axis value, which denotes an adjusted Pc value of 6.0E-08, the success rate is 10% of the cases, with a false-alarm rate and a type 2 error rate each of 3%. The success rate here significantly exceeds the rates for the two detractor cases, but still some might object that the type 2 error rate is too high. Backing the x-axis value up to 0.4 (4.0E-08), one observes that now the false alarm and success rates are equal at about 10% but the type 2 error rate has dropped to 0.5%. This situation might be found to be satisfactory: there are more false alarms to endure; but 10% of the overlooked higher-cum-Pc cases are identified by the adjusted Pc value, and miscarriage (type 2 error) is observed in only one out of 200 cases. The SP graphs (Figure 5-5), for their part, have a similar look and feel to the GP graphs, but in each corresponding case (*e.g.,* same main and adjusted Pc threshold) the SP results are even less encouraging than those for GP. On the whole, the adjusted Pc proxy approach to try to obtain a similar effect to a cumulative Pc screening with the simplicity of a maximum Pc approach has not yielded particularly encouraging results.

However, the calculation of a cumulative Pc is not a particularly burdensome calculation, even though it would require the modification of the GP and SP screening software presently in use. The present analysis has identified a non-negligible number of situations in which the cumulative Pc result is notably larger than that maximum Pc result—not enough to exceed the order-of-magnitude difference criterion that has served as the rule of thumb throughout this study but still enough to present a worrisome discrepancy. Because the cumulative Pc calculation is so easy to accomplish, it does seem warranted to migrate to this paradigm whenever the screening software is naturally opened for repair or upgrade.

One should conclude by pointing out that a wholesale switch to a cumulative Pc calculation, with an accompanying upward adjustment of Pc thresholds, could create an occasional situation in which a relatively solitary maximum Pc event, which would have exceeded the more stringent threshold for maximum Pc events, will not have enough accompanying low-Pc events to produce a cumulative Pc greater than the upwardly-adjusted threshold. A straightforward remediation to this problem is to perform screenings of both types (maximum and cumulative Pc) and close a certain launch opportunity if either threshold is violated. Since by dint of computing the cumulative Pc the maximum Pc will have been computed and is available, it should be extremely easy to screen each launch opportunity against both thresholds.

Section 6: Launch Window Closures

Satellite launch execution is managed through periods of time called *launch windows*. A launch window is a continuous period of time during which the payload, if launched, can obtain the desired orbit and is thus secondarily the time during which all launch support systems are fully prepared for a launch. For some final orbits, the launch window can be exceedingly long, as there are few if any time-based constraints that would prevent the vehicle from reaching the intended orbit. For other launches, especially space probe launches with a trajectory design that requires the gravity-assist of one or more additional luminaries, the launch opportunity may be a single instant in a given day, which, if missed, forces a delay to the next day and then, if the trend continues, ultimately to the end of the viability window. The LCOLA task is to determine which portions of a launch window are safe for launch activities and which are not—the latter of which are called "cut-outs" because they are the periods of time cut out of the continuous window.

The LCOLA approach differs in important respects from the on-orbit CA paradigm. When a potentially worrisome conjunction is identified between two on-orbit objects, a multi-day process begins to acquire additional tracking data, refine the orbit(s), assess the risk through a multiplicity of different amplifying investigations, decide whether an avoidance maneuver is needed, plan the maneuver, and, updating the information as close as possible to the maneuver commit time, decide whether to execute the maneuver. When an LCOLA screening for a particular launch time indicates a potentially dangerous situation, there is no follow-up investigation to perform additional risk analysis; that launch time is simply removed as one of the viable launch times within the launch window. Typically, only certain time intervals within a launch window are proposed as candidate launch times. The majority of launches today, especially for the somewhat longer launch windows, are "launch on minute," meaning that within the launch window only the even minutes (*i.e.,* once every sixty seconds) are actual candidate launch times; this simplifies the LCOLA screening process, as only those particular times need to be screened. Some launches are "launch on 30 seconds," which gives twice as many launch opportunities per unit launch window size; and a few are "launch on second," which is substantially more demanding for the screening process (per unit launch window time). In all cases, the initial evaluation criterion for the desirability of a launch window is the percentage of the launch opportunities that are viable.

A useful contributor to the choosing of a Pc screening threshold is its expected effect on the portions of the launch window that will need to be closed or can be left open. This is not to say, of course, that the selection of a screening threshold should be merely for the convenience of launch operations—choosing an easy-to-meet threshold so that candidate launch times within the launch windows will be mostly available and thus that launches can be conducted with as little encumbrance as possible. At the same time, it only makes sense as part of the threshold selection process to consider the effect of certain threshold choices on the ability to conduct launch operations. As remarked previously, risk mitigation is not a threshold entity but a continuum: as one makes the threshold more demanding, risk is reduced and operational encumbrance increases; but there is rarely a fixed reason to make any particular choice along the continuum. In the end, the selection of a threshold will be a subjective assessment of the desirable balance of risk mitigation and operational fluidity.

To this end, an investigation that attempts to establish the effect of choosing different screening thresholds on launch window closure is a helpful contribution to the overall threshold selection process; and the current screening experiment has produced data that

are quite useful to such an investigation. Four of the five trajectories screened in the experiment are on LEO objects and are reasonably similar in behavior; thus, if one concatenates them into a single dataset, one will thus have 172,800 screening results at one-minute intervals, with only four seams in the dataset (at the points of transition from one trajectory to the next). Each such screening result will have a maximum Pc, a cumulative Pc, and a miss distance.

One can thus construct launch windows of any desired size and, through "moving window" shifting of the window through the entire dataset, examine statistically the portion of such a window that is open or closed. For example, suppose one were to choose a launch window of twenty minutes' length. The first twenty records (records 1-20) of this dataset could be used as a possible set of screening results for this window, and the percentage of these twenty records with a Pc smaller than a chosen threshold would represent the percent of the launch window that was "open" (for the "launch-on-minute case). One could then shift this window down one record (examining records 2-21) and calculate this percentage again, and so on until the entire set of 173,000 records has been examined. One can then take percentile points from the result and thus report at, say, the 50^{th}, 68^{th} (1-sigma analogous equivalent),* 95^{th} (2-sigma analogous equivalent), and 99^{th} percentiles the "percent open" of the twenty-minute launch window.

If this is performed for a large number of different window sizes and Pc thresholds, contour plots can be developed that give the percentage of launch window open as a function of launch window size and Pc threshold. This type of presentation can communicate in a more comprehensive way the impacts to launch mission conduct of choosing different Pc thresholds, and Figures 6-1 through 6-12 use this approach, arranged in quad chart format. For each of the quads, the x-axis is the Pc threshold (maximum or cumulative, depending on the figure); the y-axis is the length of the launch window, in minutes (ranging from one minute to three hours); and the color indicates the percentage of the launch window that is open (blue indicates 0% open and dark red 100% open). Each quad is a different percentile level, representing the four percentile levels given in the previous paragraph. Different figures are given for maximum and cumulative Pc thresholds, for GP and SP results, and for different densities of viable launch points within a window (launch on minute, launch on 30 seconds, and launch on second).

This latter item requires additional commentary because it involves approximations. The screenings in the experiment were run at one-minute intervals, and as such they support only that launch point density natively. Results for this density are thus definitive and properly represent correlation of events between screenings, to the extent that any such correlations may be present. However, these results can be made to approximate those for a greater launch point density by remasting. For example, the twenty-minute moving launch window described earlier, for which groups of twenty records would be systematically evaluated as the window was moved record by record through the dataset, could be made to approximate a "launch-on-thirty-seconds" situation by simply doubling the number of records: the group of records to be examined would be forty records in size, even though this would be evaluated as representing only a twenty-minute window. Of course, this approach ignores, and therefore misrepresents, any screening-to-screening correlation

* Because the range of values is bounded (0% to 100%) and the extremes of this range are encountered, the distribution is unlikely to have Gaussian properties; and even without this there is no real reason to suppose that the dataset would behave in a Gaussian manner. However, using percentile points that emulate those encountered for the 1- and 2-sigma cases allow easier comparison with confidence levels typically encountered in statistical inference.

effects that are in play, as the screening results were not actually run at thirty-second intervals. Most conjunction durations, however, are much shorter than this; so the thirty-second results should be fairly representative. This same approach was also used to produce an approximate result for the "launch-on-second" case by increasing the moving window size by a factor of sixty. In this situation, the result can be nothing more than a first-order approximation, as event correlation is much more likely at a one-second separation interval. So results for the launch-on-second case need to be understood and interpreted with this in mind.

In examining the plots, one observes a discernible "chatter" along parts of certain contour lines that results in localized instances of non-monotonic behavior. This is due to the taking of percentile points against short(er) launch windows which themselves are composed of discrete screening results that are described in a binary fashion of being either open or closed. While this chatter is not a major impediment to interpreting the graphs, it still merits a more thorough explanation. The following example helps to illustrate the origin of the problem.

One begins by postulating a launch window of one minute's duration (for this entire discussion a "launch-on-minute" paradigm will be further presumed). Because only one screening event is incorporated in such a window, there are only two possibilities: it is either 0% open or 100% open. A launch window of two minutes' duration will offer three possibilities: 0% open, 50% open, or 100% open. A three-minute launch window will offer four: 0%, 33%, 67%, and 100%. As the size of the launch window increases, the set of possibilities will become more continuous; but for the smaller launch window sizes, there is a relatively small number of discrete levels for the "percent open" evaluation.

Now one further supposes a set of 100 screening results, 80% open (indicated by a 1) and 20% closed (indicated by a 0) when evaluated at a particular Pc threshold and arranged in the following way, essentially a 1-1-0 repeating sequence of open and closed opportunities until record 85, after which point the window is entirely open:

Table 6-1: Hypothetical set of 100 screening results

1	1	11	1	21	0	31	1	41	1	51	0	61	1	71	1	81	0	91	1
2	1	12	0	22	1	32	1	42	0	52	1	62	1	72	0	82	1	92	1
3	0	13	1	23	1	33	0	43	1	53	1	63	0	73	1	83	1	93	1
4	1	14	1	24	0	34	1	44	1	54	0	64	1	74	1	84	0	94	1
5	1	15	0	25	1	35	1	45	0	55	1	65	1	75	0	85	1	95	1
6	0	16	1	26	1	36	0	46	1	56	1	66	0	76	1	86	1	96	1
7	1	17	1	27	0	37	1	47	1	57	0	67	1	77	1	87	1	97	1
8	1	18	0	28	1	38	1	48	0	58	1	68	1	78	0	88	1	98	1
9	0	19	1	29	1	39	0	49	1	59	1	69	0	79	1	89	1	99	1
10	1	20	1	30	0	40	1	50	1	60	0	70	1	80	1	90	1	100	1

For this results set, moving windows of 1, 2, and 3 minutes' duration were run, producing groups of 100, 99, and 98 results, respectively. In summarizing these results with the percentile points suggested earlier (50^{th}, 68^{th}, 95^{th}, 99^{th}) for window sizes of 1, 2, and 3 minutes, the following summary values are observed:

Table 6-2: Results summary for hypothetical screening results

Window Size	50ile	68ile	95ile	99ile
1 minute	100%	100%	0%	0%
2 minutes	100%	50%	50%	50%
3 minutes	67%	67%	67%	67%

At different percentile points, one sees very different percentage results and different trends as the launch window is lengthened: at the 50th percentile, the results are monotonically decreasing; at the 95th and 99th percentiles they are monotonically increasing; and at the 68th percentile they follow a non-monotonic behavior. As the window size is increased, the impact of this behavior is lessened (and further smoothed by the graphics generation routines in the figures below); but the basic problem of chattery results brought about by percentage evaluation of relatively small binary sets, coupled with percentile-point summary of larger groups of these percentile results, remains a feature of the overall dataset. As stated previously, this is not really an impediment to interpretation if one wishes general information by region, which is all that should appropriately be extracted from the graphs in any case.

Because the charts can be somewhat confusing to read, the following is given as a primer, starting with the upper-left quad of Figure 6-1. If one chooses, say, 1E-06 as the Pc value (x-axis) and a 70-minute launch window (y-axis), the corresponding color for that locus is red but not dark red, indicating that with those parameters *ca.* 90% of the launch window is open 50% of the time (because the upper-left quad is the 50th percentile graph). At the 95th percentile (bottom-left quad), for these same parameters perhaps about 55% of the window is open. Thus it can be said that, for a Pc threshold of 1E-06, half of the time 90% of a 70-minute window is open and nineteen out of twenty times 55% of it is open. Moving left to right along the 70-minute (horizontal) line, the percentage grows from 0% to 100% open, which makes sense in that as the Pc threshold is larger (*i.e.,* more lax), more of the launch window is open. Moving vertically on the 1E-06 line, one sees relatively little variation in the percentage open value—no more than ten percentage points or so. The results of the screenings seem to be more or less uniform this way, especially at the higher percentile points: once a launch window is of a reasonable size, the sampling appears to be large enough that the percentage closed does not vary as the launch window size is further increased.

The “U-shaping” of the contour plots, most visible for the launch-on-minute case but detectable in others, is an expected result. In the upper-left quad of Figure 6-1, the dark blue and dark red actually meet along the launch window size = 1 line (right at the x-axis), without intermediate colors. Because with a one-minute size in a launch-on-minute situation the window must be either entirely open or entirely closed, it makes sense that only two colors appear for it, with the transition taking place at a Pc threshold of about 1.5E-07. As the window size is increased, there is more room for different percentage values (rather than just 0% or 100%); and the U-shape moves quickly to the largely parallel contours that begin at launch window sizes of about 10 minutes.

The progression from a launch-on-minute (Figure 6-1) to a launch-on-30-seconds (Figure 6-2) to finally a launch-on-second (Figure 6-3) situation does not result in significant

changes to the observed response. The response "stretches out" horizontally somewhat in that, due to the longer groupings of discrete-response data, there are somewhat fewer situations in which windows are 0% and 100% open; and the response lines for the same reason (*i.e.,* larger groupings) show less chatter. It is important to recall, however, that the launch-opportunity rates smaller than one minute were synthetically assembled by concatenating launch-on-minute data to assemble these larger data groups; so the results for these densities, because they ignore event correlation, are not fully trustworthy.

The results for cumulative Pc show a slight shift to the right but otherwise the same look-and-feel to their maximum Pc counterparts. This is consistent with what was observed in Section 5, in which the maximum and cumulative Pc issue was examined in detail. The results for SP in comparison to GP show a more substantial shift, this time to the left—10 to 25 percentage points or so, depending on the specific part of the response surface. The overall look-and-feel is, however, still the same as that for GP.

It is very well to discuss verbally the response that is visible from the graphs themselves, but more helpful is an attempt at interpretation; and for this some guidance is needed on what would constitute acceptable operational performance with regard to launch window size and the portion of such windows that are open. The length of the launch window is governed by the orbit trajectory (and thus indirectly by destination) and by deployment issues specific to the spacecraft, such as trying to avoid earth-shadow or other unfavorable environmental conditions during component deployment. For some space probe launches, an "instantaneous" launch is required, meaning there is only one acceptable launch time, and for such situations there is no real issue of a launch window; but for more typical NASA launches windows of tens of minutes are possible, and launch operators very much prefer an hour, encountering the end of their comfort zone at about twenty minutes. For USAF launches, the windows are longer, typically on the order of one to two hours; and it would be nice to preserve that length if possible. Fortunately, for most Pc thresholds, Figures 6-1 through 6-12 show that the response as a function of launch window size has more or less leveled out into straight contours; so choices that are appropriate for a twenty-minute window will probably be more or less appropriate for a 120-minute window and vice versa.

In terms of percentage of window open, it is typical to conduct operations with windows that are mostly open—80% or so; below this level operators begin to feel cramped, and this feeling becomes acute when the percentage decreases to the 50% level. Approximate Pc values, taken from Figures 6-1 and 6-7, for three different open-window percentages are given in Table 6-3 below. These values are for the 95th percentile and represent the situation over all but the smallest launch window sizes.

Table 6-3: Pc Values Corresponding to Open Window Percentages (approximate)

Percent of Window Open	GP Launch-on-Minute Pc, 95th Percentile	SP Launch-on-Minute Pc, 95th Percentile
80%	2E-06	8E-07
60%	1E-06	4E-07
40%	5E-07	1.5E-07

If one wishes to remain within the more conservative levels for window closure, then screening thresholds that preserve an 80% "open-window" level seem appropriate. As one

can see, such levels are already notably more demanding than the current guidance of 1E-05 (from Table 1-1). The next level of increase shown is to move all the way to the 50% level, a level at which one would expect to encounter some operational resistance. However, in some respects more important than the overall open/closed status of a particular launch window is the ability to find, in any smallish sector of the window, some reasonable number of viable launch opportunities. For example, typically the initial launch opportunity is at the beginning of the launch window. If that opportunity is waved off due to a launch exigency, then perhaps an opportunity will be sought around the middle of the window; and if for some reason that opportunity also cannot be used, then one towards the end of the window will be selected. Since the contours in Figures 6-1 – 6-12 are reasonably straight, the "percent open" value for a given Pc contour (straight vertical line) is more or less constant until the extremely small windows are encountered; but even there the deviation is not substantial. This means that sub-parts of any given launch window are likely to have more or less the same "percent open" value as the overall window. For example, consulting Table 6-3, one can see that for the GP case the Pc threshold corresponding to an 80% availability figure is approximately 2E-06; this value can be taken from the bottom-left quad of Figure 6-1 at the 90-minute launch window point. However, one can additionally see that in descending to the 20-minute window size the openness level drops by perhaps only five percentage points: if the 90-minute window were to be divided into thirds, each of twenty minutes' duration, there would be about a 75% launch opportunity availability in each of these thirty-minute sectors.

If this is the case generally, then it may be possible to entertain more demanding Pc screening thresholds than this analysis shows *prima facie*. If the true concern is to ensure that a reasonable launch opportunity can be found in each of the sub-sections of a launch window, and it appears that availability within a sub-section is only slightly lower than that for the entire window, then a Pc that maps to an overall window-availability figure that is somewhat below what historically would be considered acceptable can perhaps be embraced. Given this dynamic, the Pc values that map to the 80% overall figure should certainly be acceptable without reservation; and it would seem that Pc values as low as those associated with 50% availability may be acceptable. Synthesizing these two values, one could propose a GP max-PC screening threshold of 1E-06 and an equivalent SP value of 5E-07, or perhaps even a little lower. After operational experimentation with these values, it may be found possible to lower them even somewhat more without adverse effects on launch operations.

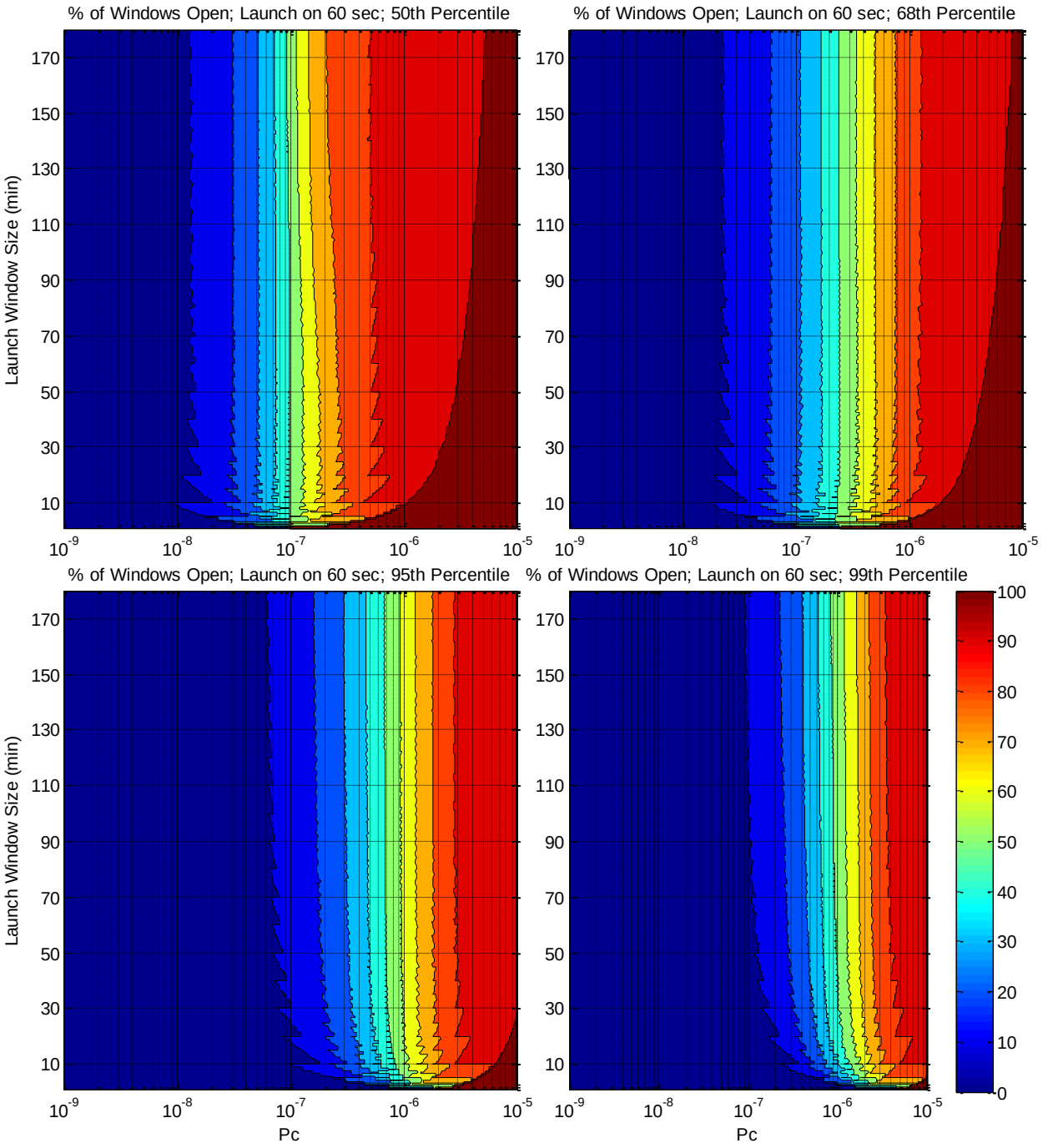


Figure 6-1: Window Open/Closed Percentage vs Launch Window Size vs Pc; GP Screenings, 60-sec Interval ("Launch on Minute"), Maximum Pc

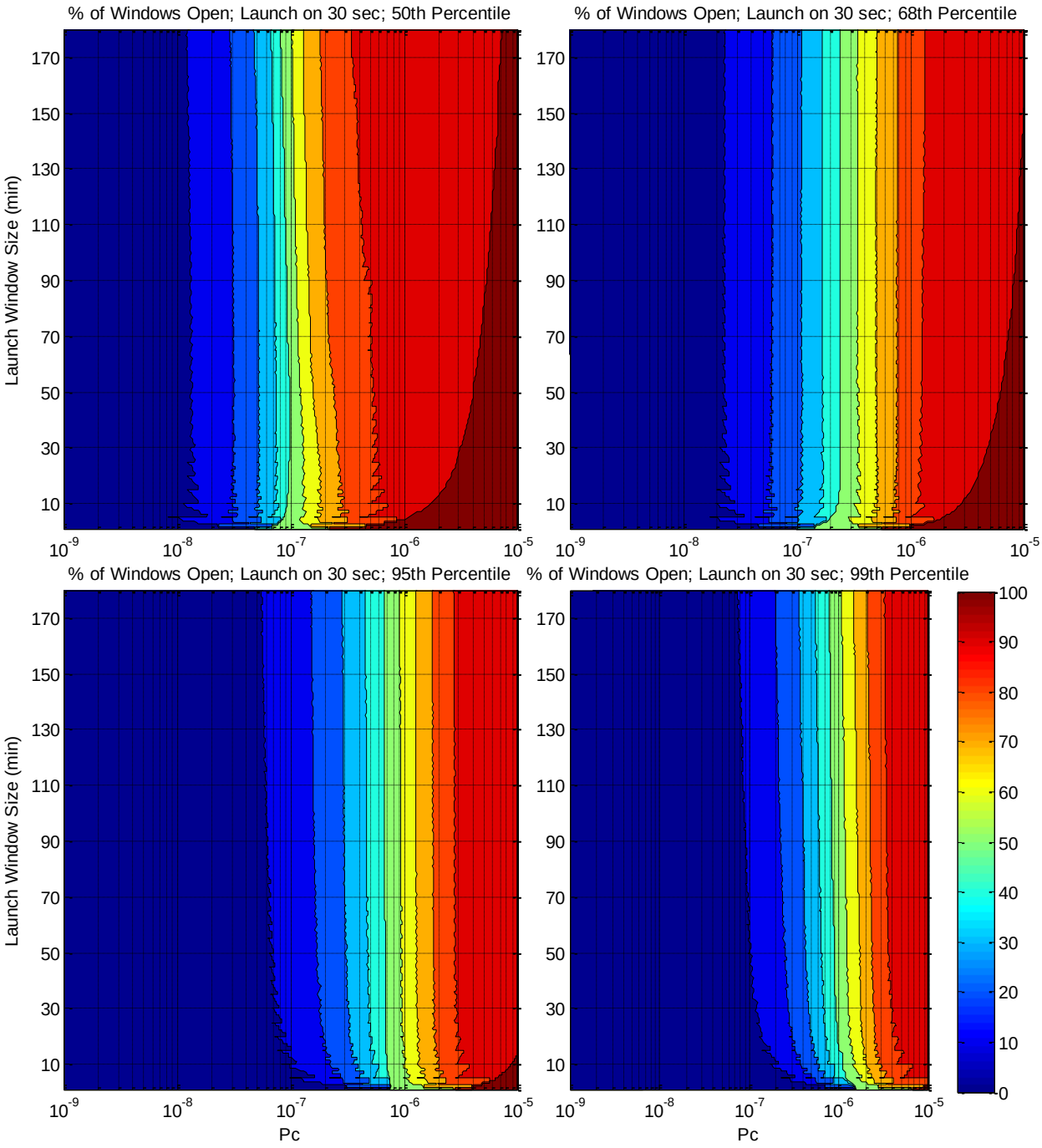


Figure 6-2: Window Open/Closed Percentage vs Launch Window Size vs Pc; GP Screenings, 30-sec Interval ("Launch on Half-Minute"), Maximum Pc

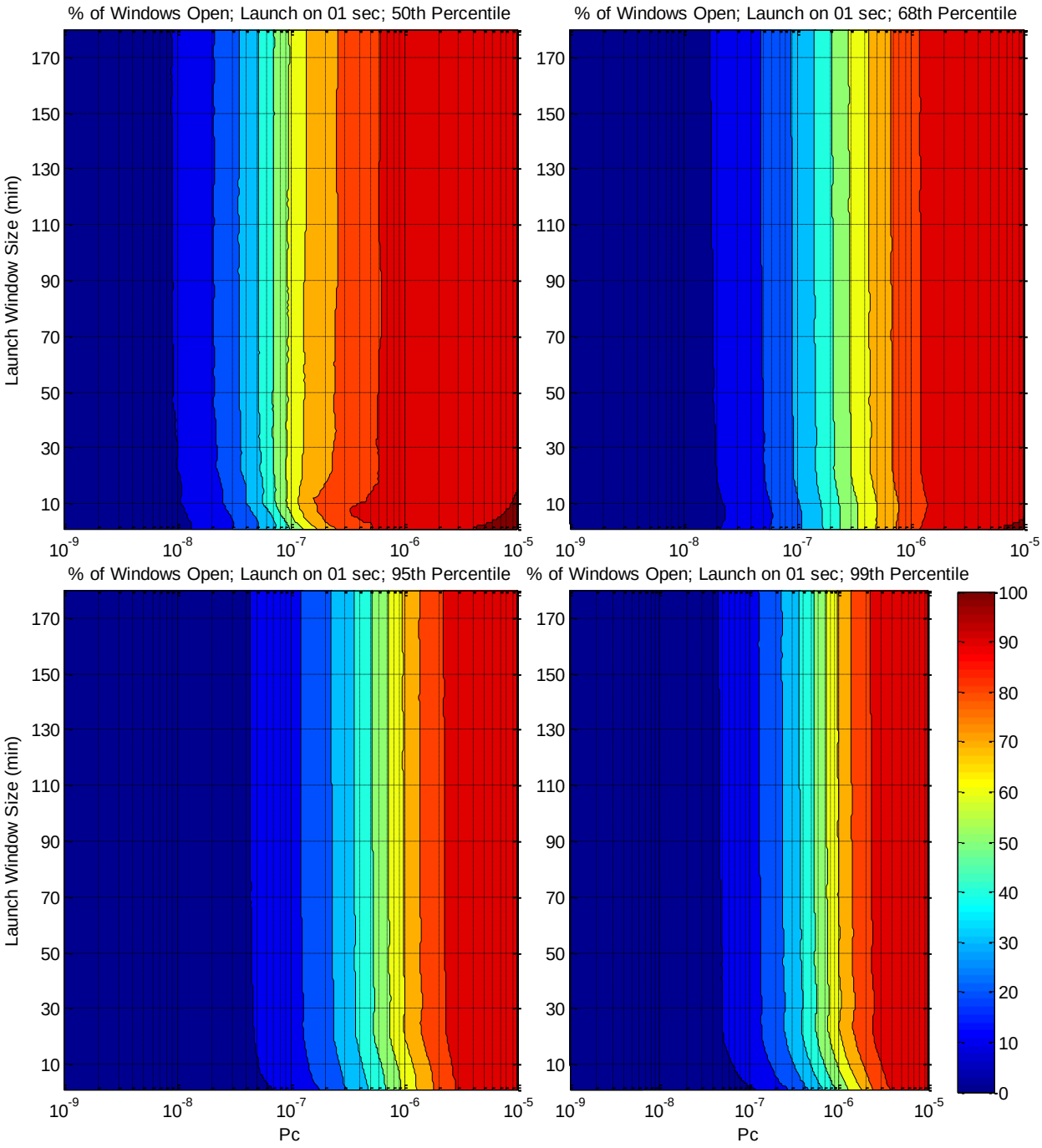


Figure 6-3: Window Open/Closed Percentage vs Launch Window Size vs Pc; GP Screenings, 1-sec Interval ("Launch on Second"), Maximum Pc

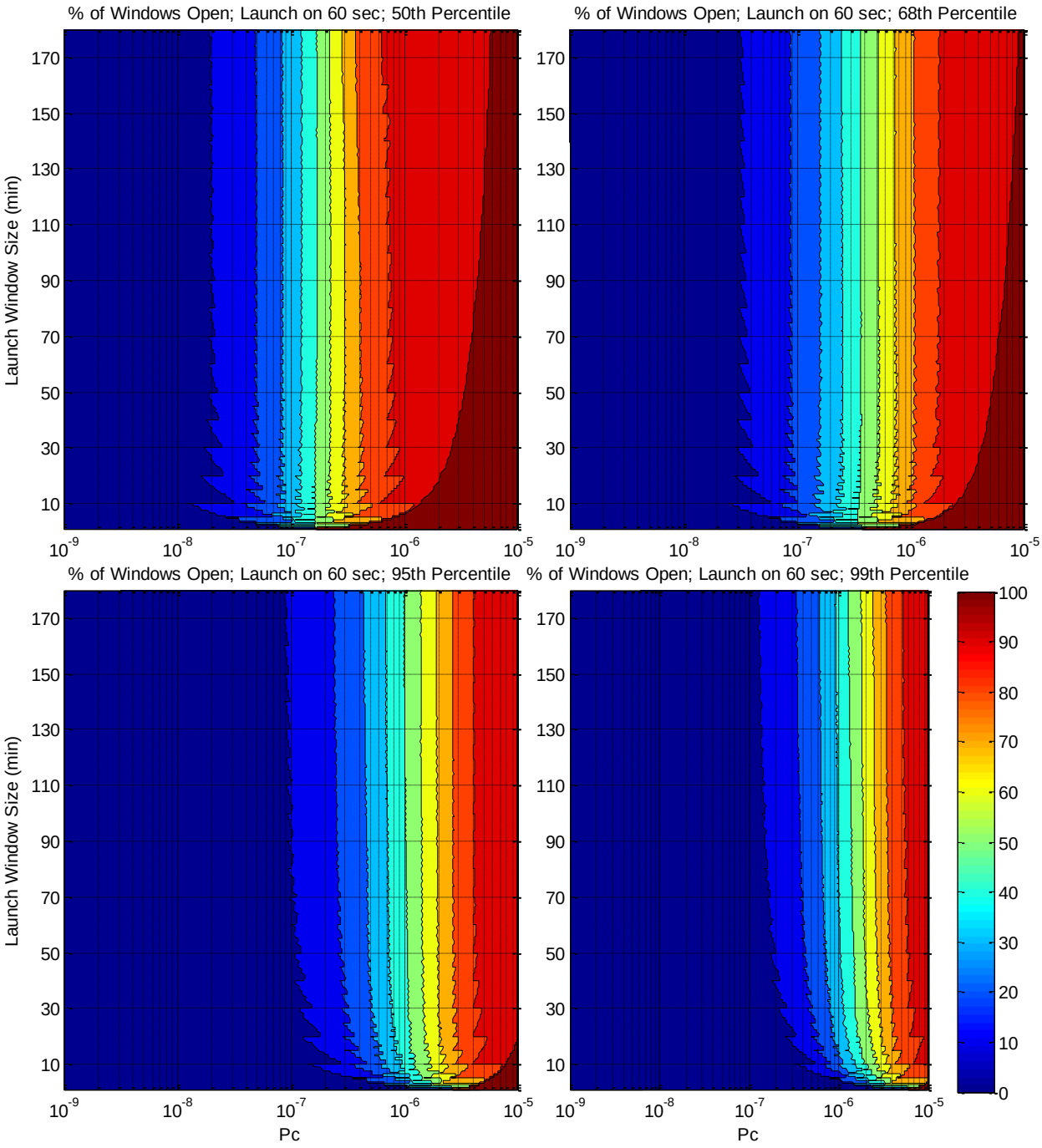


Figure 6-4: Window Open/Closed Percentage vs Launch Window Size vs Pc; GP Screenings, 60-sec Interval ("Launch on Minute"), Cumulative Pc

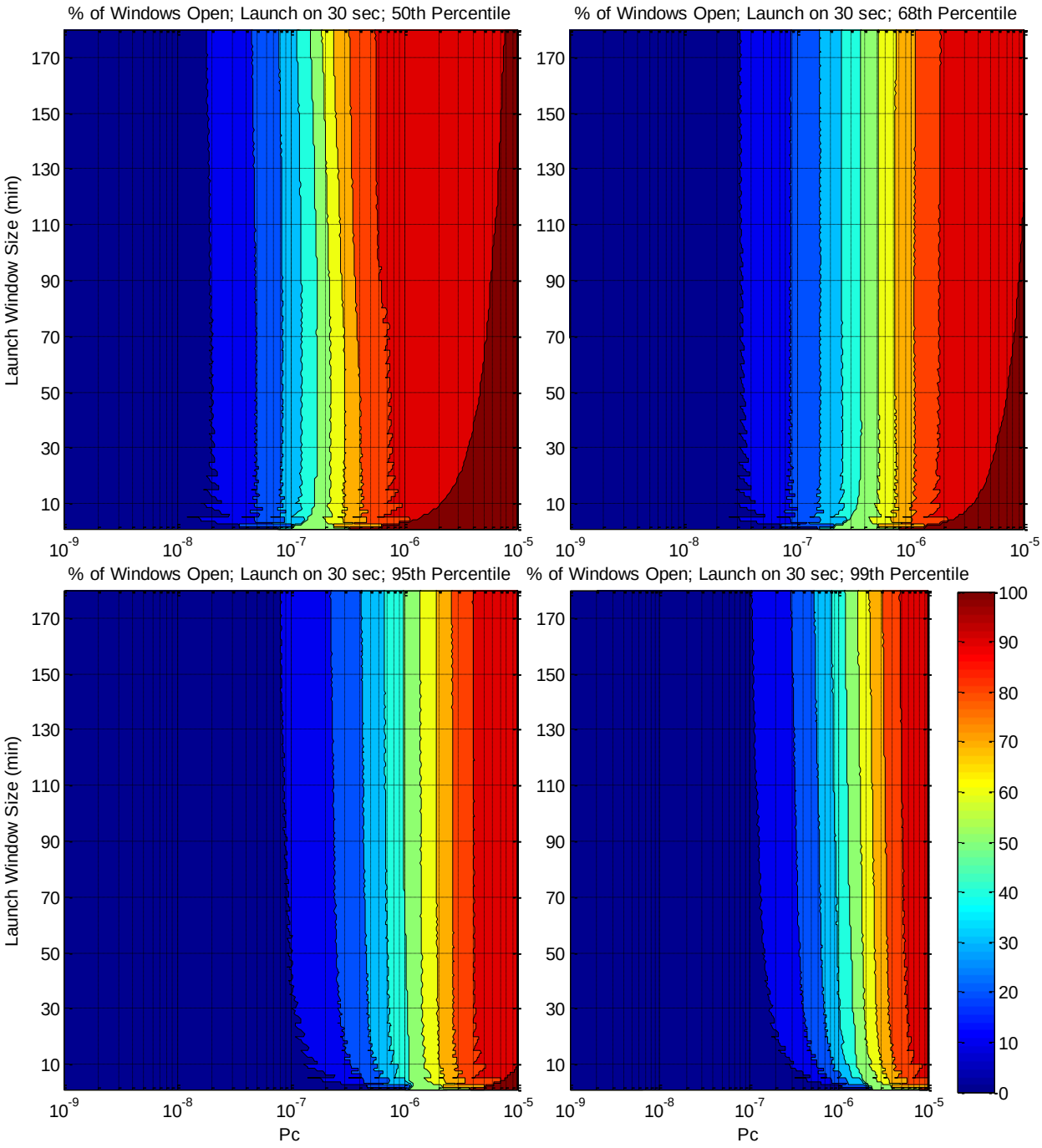


Figure 6-5: Window Open/Closed Percentage vs Launch Window Size vs Pc; GP Screenings, 30-sec Interval ("Launch on Half-Minute"), Cumulative Pc

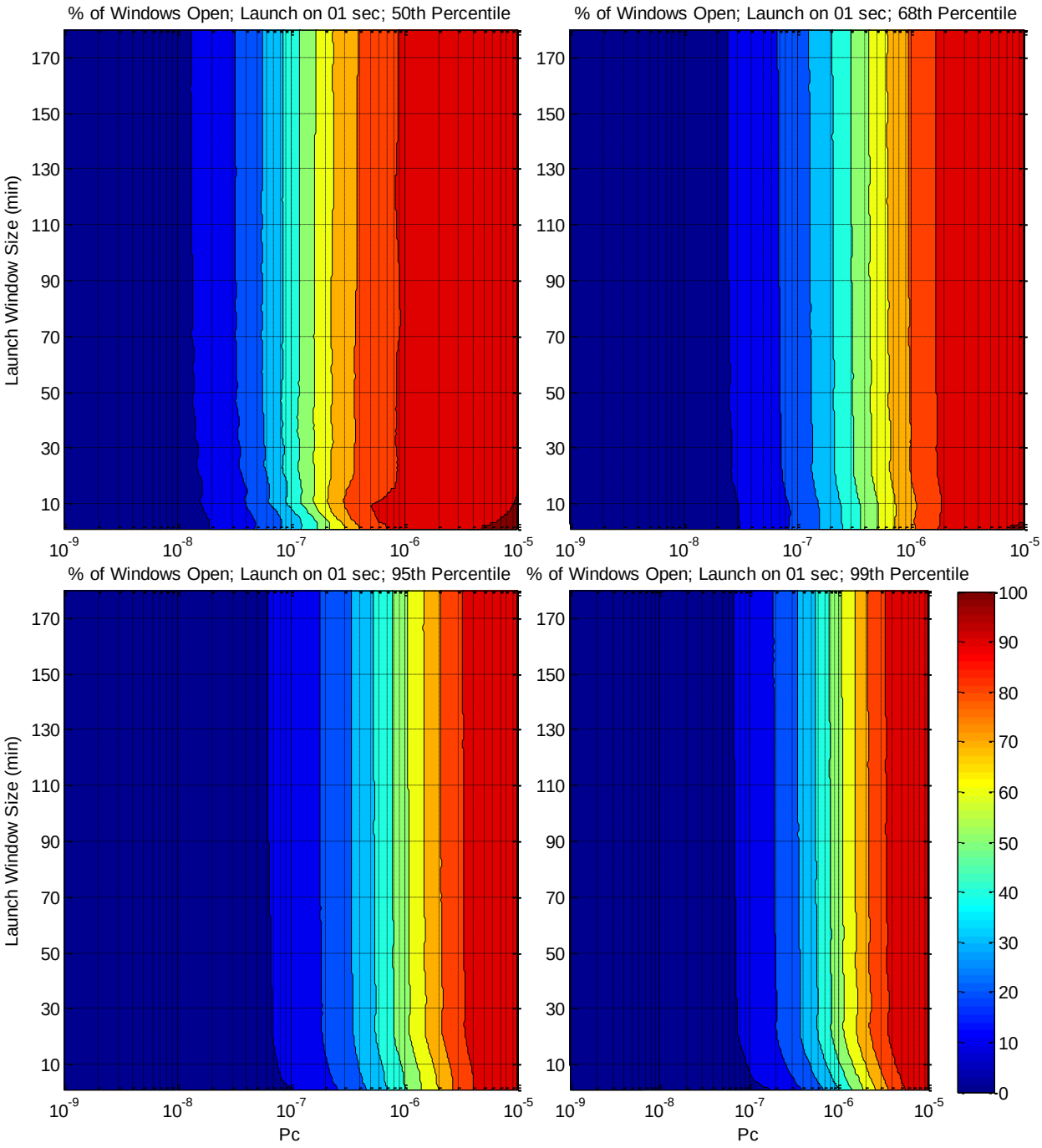


Figure 6-6: Window Open/Closed Percentage vs Launch Window Size vs Pc; GP Screenings, 1-sec Interval ("Launch on Second"), Cumulative Pc

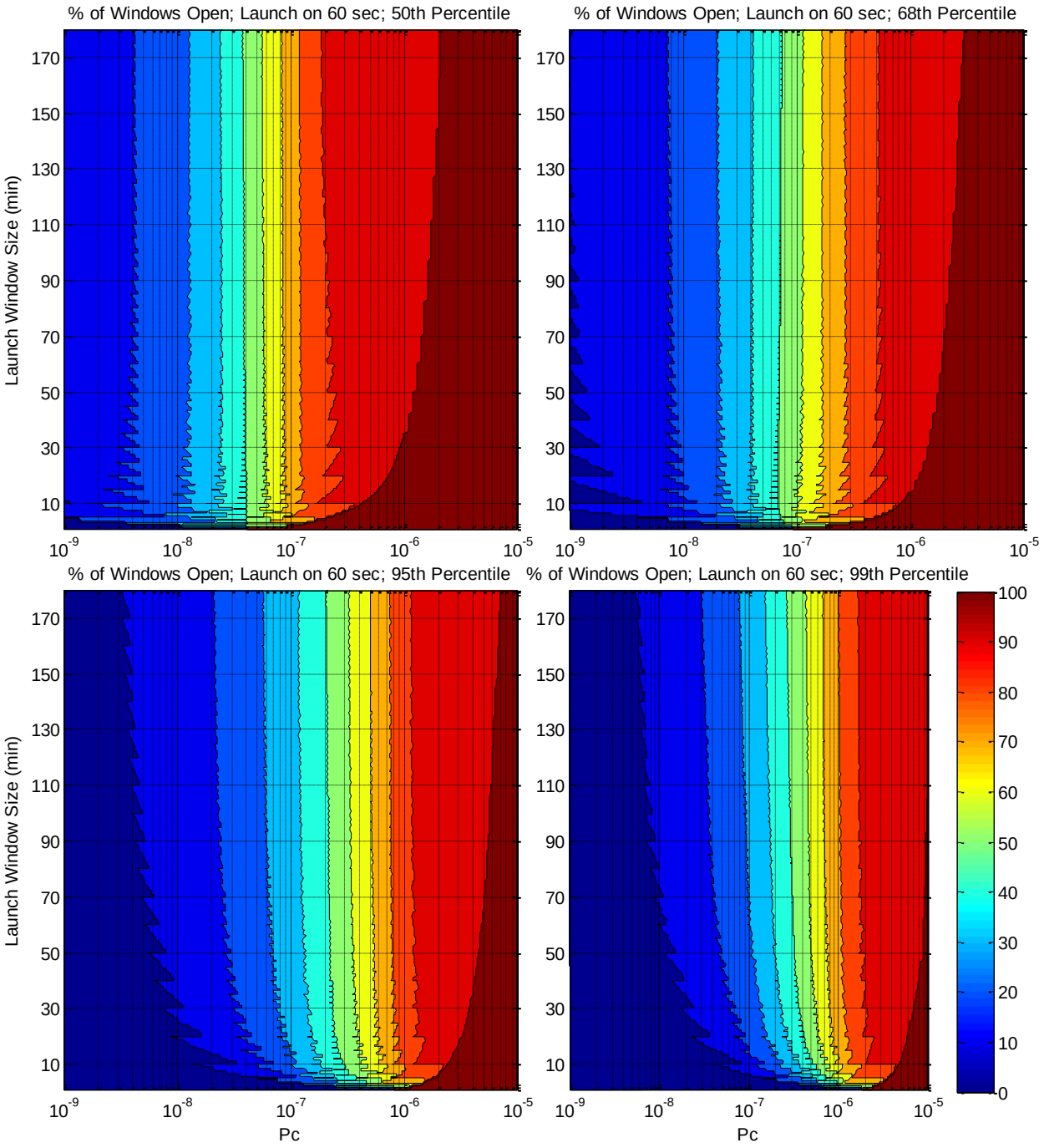


Figure 6-7: Window Open/Closed Percentage vs Launch Window Size vs Pc; SP Screenings, 60-sec Interval ("Launch on Minute"), Maximum Pc

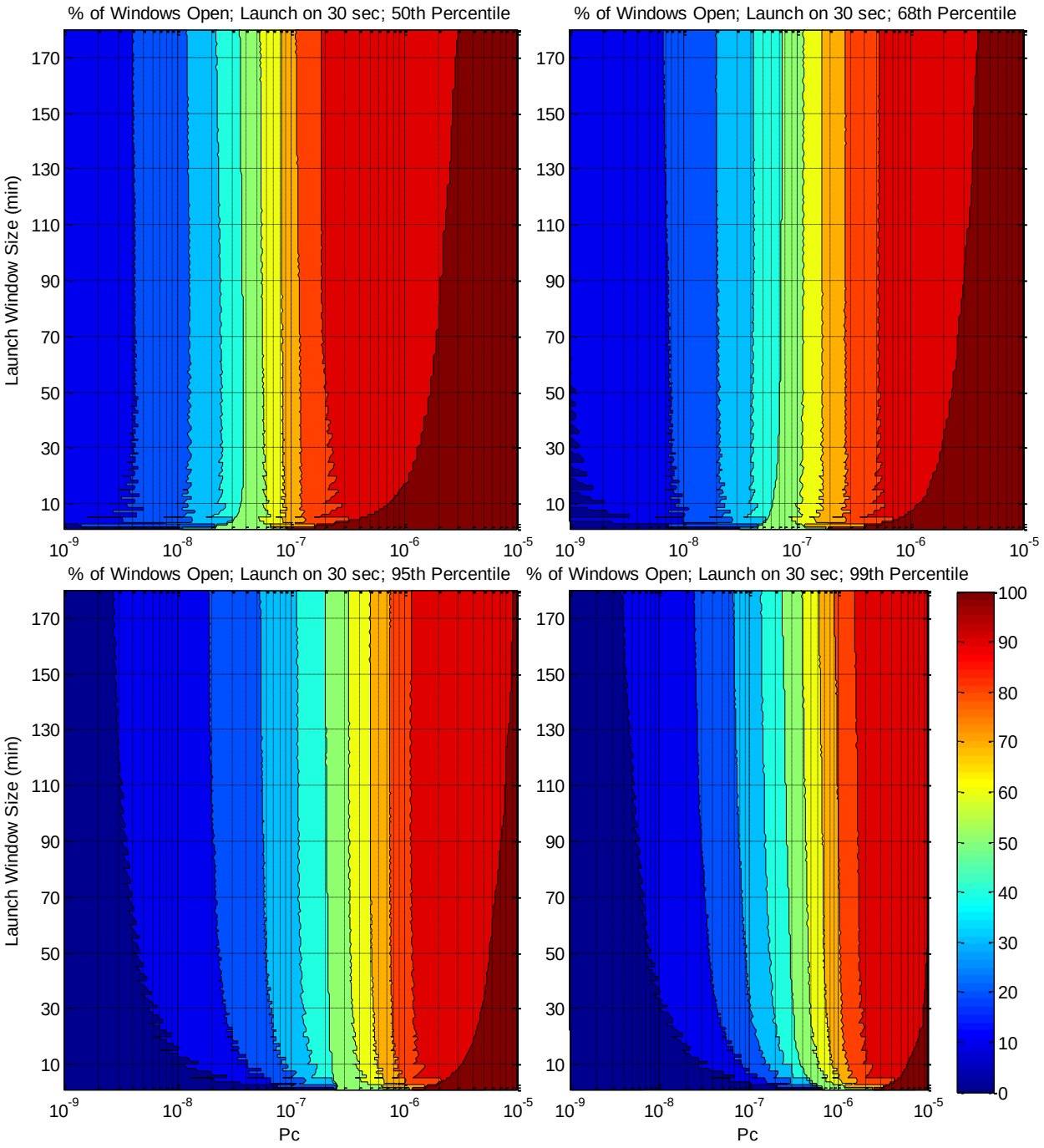


Figure 6-8: Window Open/Closed Percentage vs Launch Window Size vs Pc; SP Screenings, 30-sec Interval ("Launch on Half-Minute"), Maximum Pc

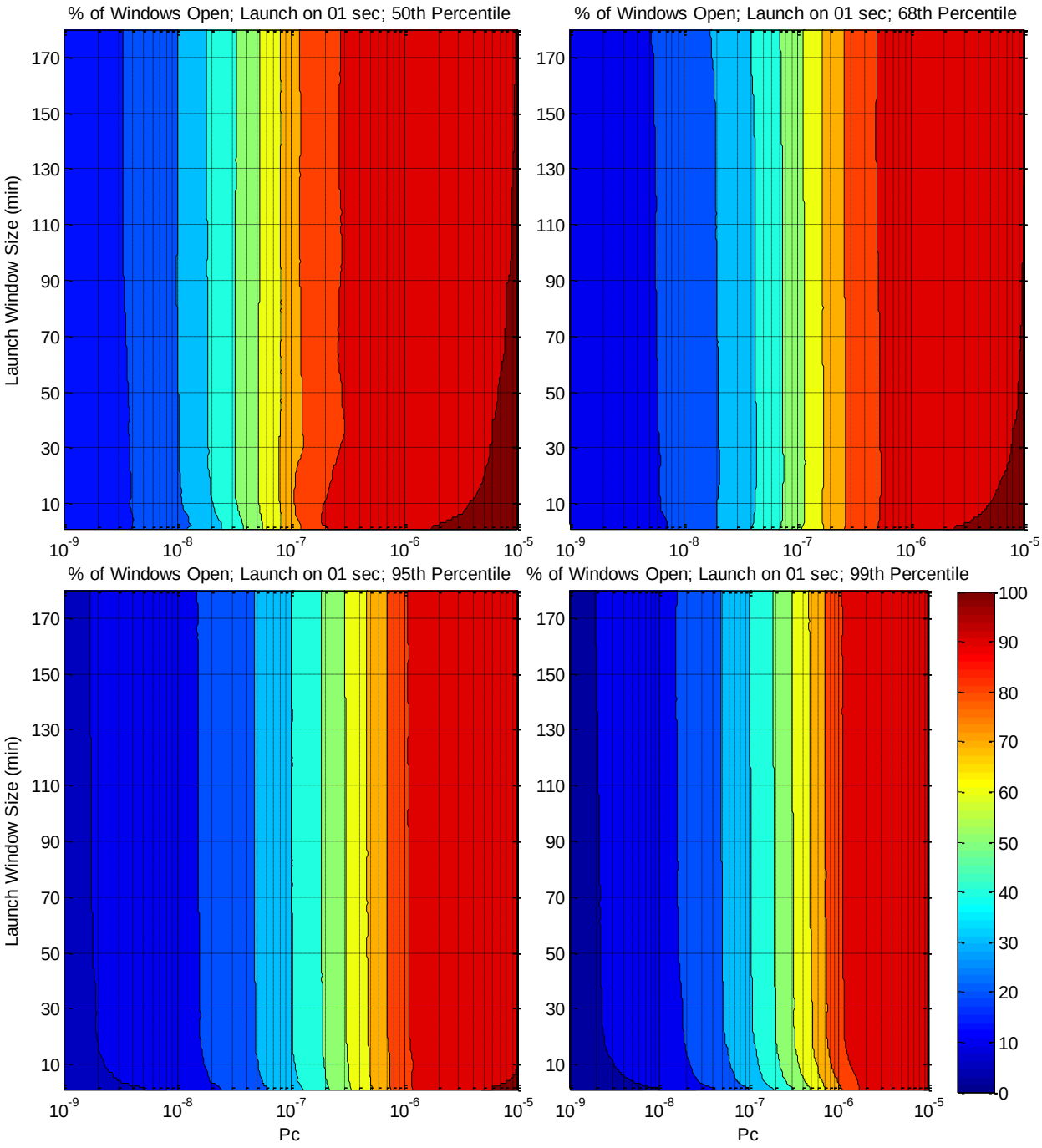


Figure 6-9: Window Open/Closed Percentage vs Launch Window Size vs Pc; SP Screenings, 1-sec Interval ("Launch on Second"), Maximum Pc

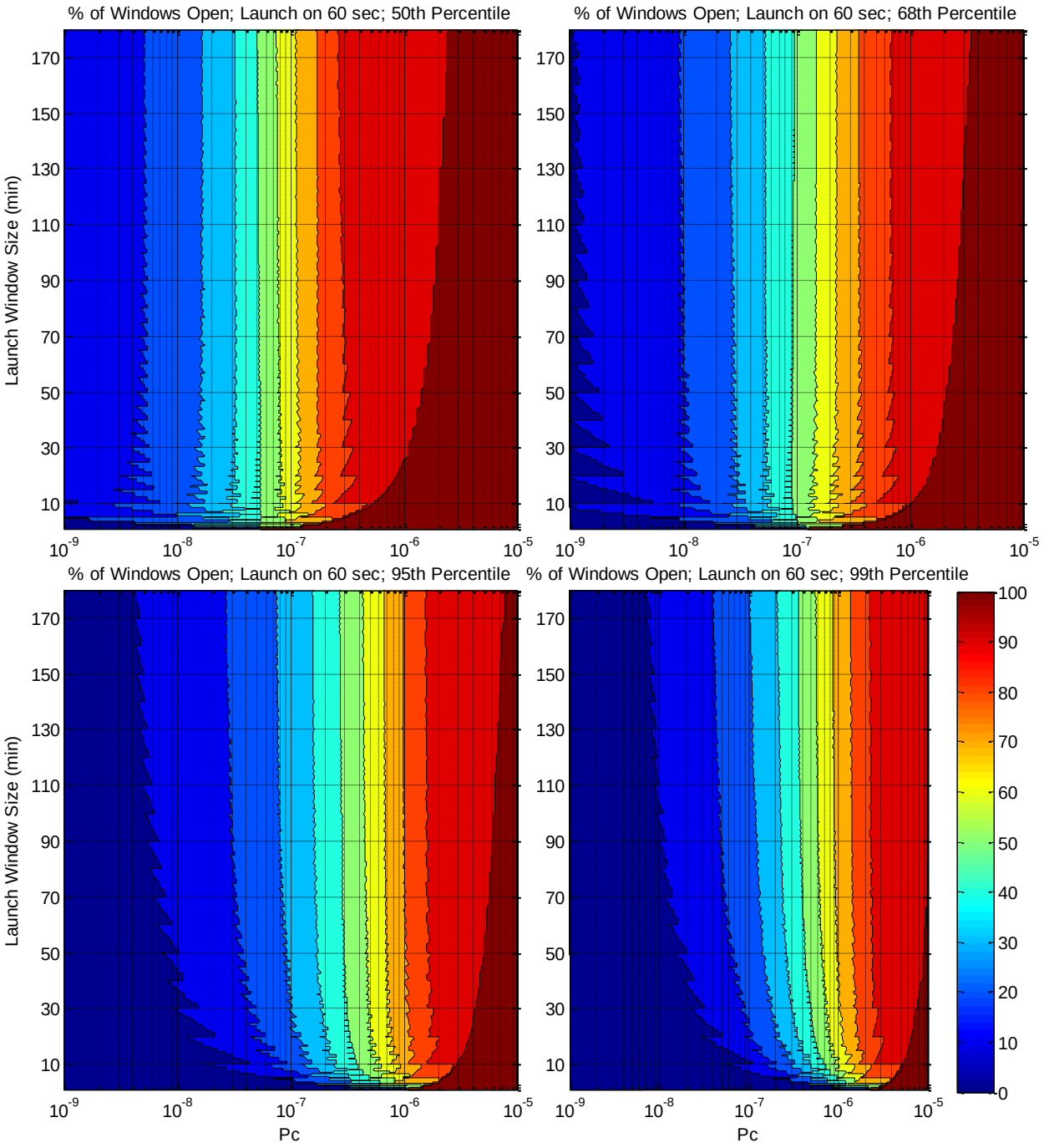


Figure 6-10: Window Open/Closed Percentage vs Launch Window Size vs Pc; SP Screenings, 60-sec Interval ("Launch on Minute"), Cumulative Pc

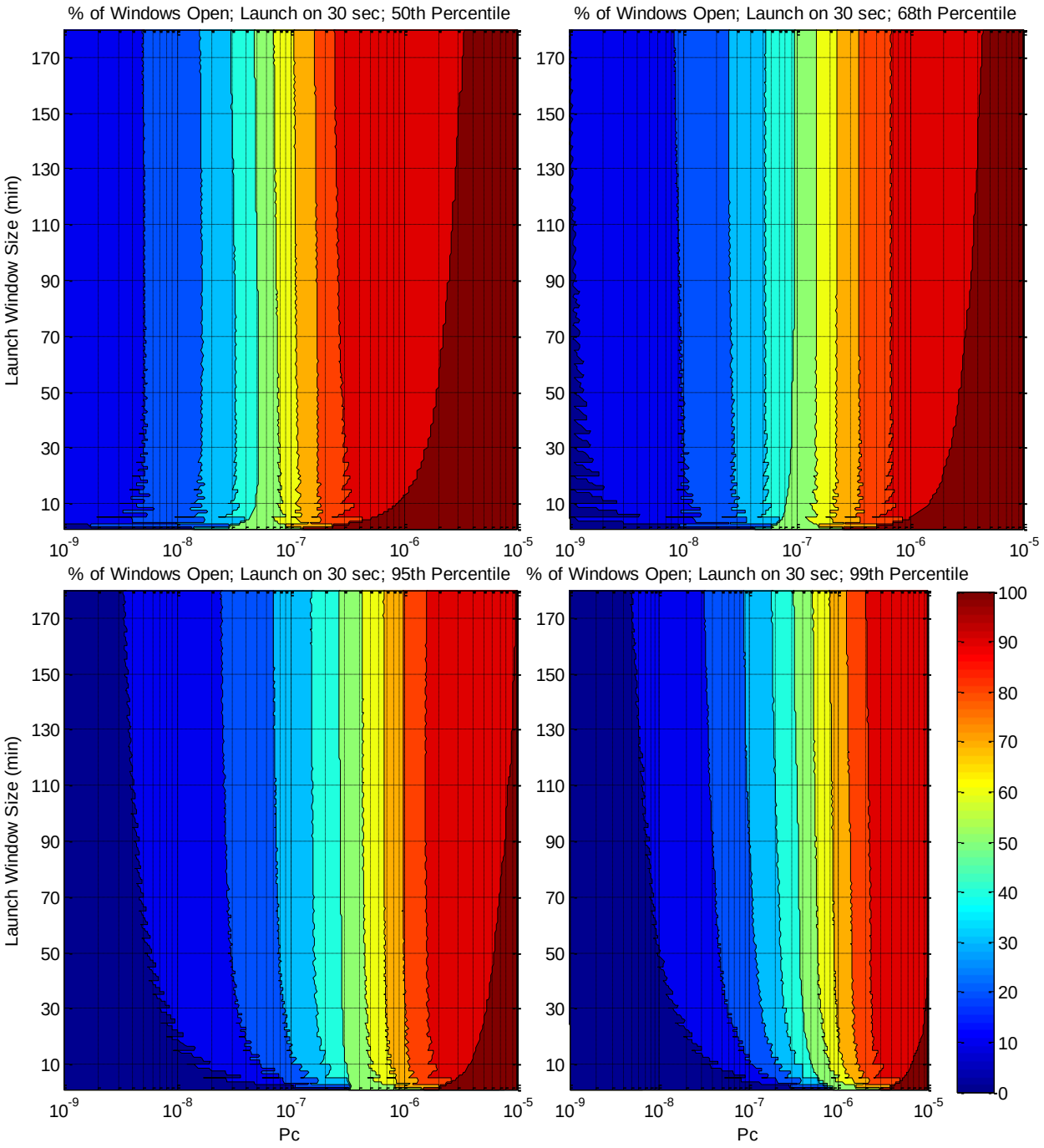


Figure 6-11: Window Open/Closed Percentage vs Launch Window Size vs Pc; SP Screenings, 30-sec Interval ("Launch on Half-Minute"), Cumulative Pc

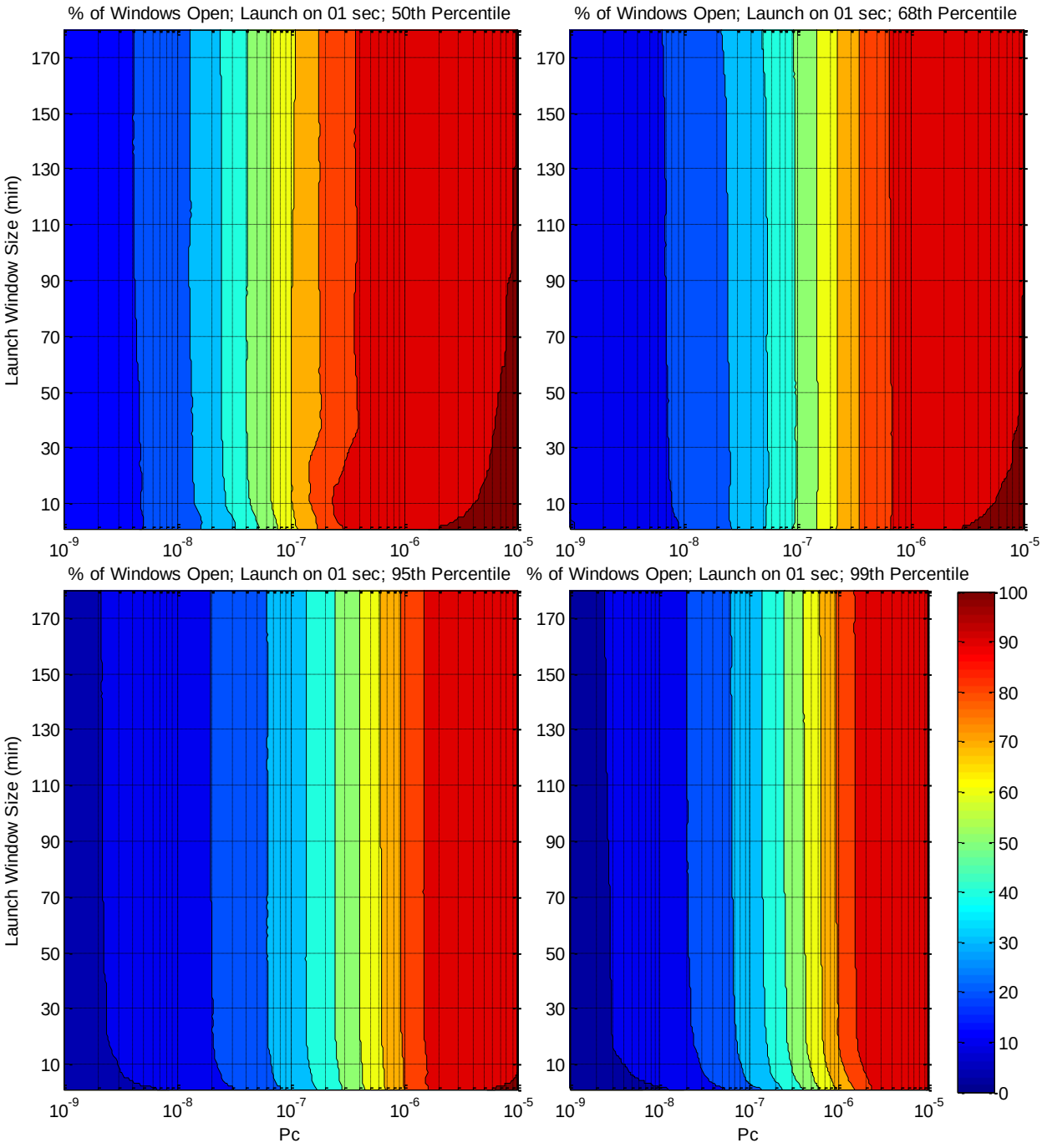


Figure 6-12: Window Open/Closed Percentage vs Launch Window Size vs Pc; SP Screenings, 1-sec Interval ("Launch on Second"), Cumulative Pc

Section 7: Miss Distance Proxy for Pc

It was remarked earlier in this report that the only durable single calculation to assess the risk of any given conjunction is the probability of collision, or Pc. This calculation incorporates all of the relevant factors in that in uses the nominal miss distance, the size of the covariances of the primary and secondary satellites' state estimates, and the estimated combined sizes of the spacecraft; and it returns an actual probability as its result—a calculation appropriate to risk assessment. However, there are times when this calculation is not possible, such as when covariance information is not available and cannot reasonably be synthetically constructed. In such cases, it is a common practice simply to conduct conjunction screening and perform risk assessment using only the miss distance between the two objects at TCA. Indeed, with this approach there is rather little "risk assessment" that can be performed: either the conjunction has a miss distance less than the screening threshold, or it does not; but the overall simplicity of the approach—simply examining the minimum miss between the two objects—is quite appealing, and the fact that covariance information is not required promotes it all the more. Because of the interest in this alternative technique, not just for those cases where key information is not available for the Pc calculation but as a simplified procedure overall, it is useful to provide some analysis to indicate what can be done with a miss-distance-based paradigm and what cannot.

The first technical point that must be raised is that the miss distance cannot serve as a substitutionary datum for the Pc: the correlation between the miss distance and the Pc is actually rather weak, or at best one-sided. While it is generally true that a smallish miss distance is required for a largish Pc, the logical entailment is only in one direction: a smallish miss distance may be a necessary condition (most of the time) for a largish Pc, but it is not a sufficient condition. The presence of a large covariance, and therefore uncertainty, about one or both of the satellites means that the miss distance (which is based merely on an estimate of the mean position for both objects) is not strongly determined; and if the computed miss distance is small, then the random selection of position points within an error ellipse about both objects would tend to generate a set of much larger miss distances, lowering the risk. If one uses a miss-distance-based screening, many—probably most—of the conjunctions will have a much smaller risk profile if evaluated with a Pc-based approach; so a large number of "false alarms" will be generated from such screenings. Furthermore, it is difficult to know what sort of risk level to ascribe to a conjunction with a given miss distance. By changing the associated covariance incrementally from a very small size to a very large one and calculating the Pc at each increment, one can obtain the maximum possible Pc for this miss distance[11]; but most of the Pc data associated with this miss distance will be much smaller than that, and some values will be essentially negligible. So when miss distance alone is used as a risk assessment criterion to determine launch window cut-outs, it is not clear what risk level is actually being invoked.

However, while there may be no direct link between risk as evaluated by Pc and the miss distance, "probability profiling" of the miss distance against the Pc can be accomplished and is in fact rather straightforward; and the results of such profiling can be used to make more definitive statements about risk from the results of a miss-based evaluation. Such a profiling is conducted by collecting a large number of events with a Pc greater than or equal to a given threshold. The miss distance values from all of these events are then collected and their distribution characterized, either by a CDF or by percentile points. This approach can provide the context needed for miss-distance results to be mapped directly to a risk level , as one can now make a statement of the form "screening

at a miss distance of x will identify 95% (or some other percentile point) of the events that would present a Pc risk of y." Because the higher percentile points (such as 99th, 99.9th, 99.99th percentile, &c.) are not well defined for the finite samples available for such profiling, if one wishes to push into the higher end of the distribution, the best approach is to fit a candidate distribution to the empirical distribution and use this candidate distribution, which has an analytical representation, to estimate the behavior of the empirical distribution at very high percentile points. This is also necessary for the present case because the experimental screenings were run with a 100 km x 100 km x 100 km miss-distance box and thus cannot report empirical data for miss distances greater than this.

The above rubric was followed with the experimental dataset of screening results. These results were arranged into groups (with replacement) of individual events in which the Pc was greater than 1E-05, 5E-06, 1E-06, *u.s.w.,* up to 5E-09. For each, a cumulative CDF was constructed; and additionally a two-parameter lognormal and two-parameter gamma distribution were fitted to each CDF, using maximum likelihood estimation for the distribution parameters. The two-parameter lognormal distribution has a PDF of the form

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}x}\exp\left(\frac{-[\ln(x)-\mu]^2}{2\sigma^2}\right), \tag{7-1}$$

and parameter estimation is straightforward if one performs a logarithmic change of variable and estimates the mean and variance in logarithmic space, using standard methods. The two-parameter gamma distribution is of the form

$$f(x;\beta,m) = \frac{1}{\beta^m\Gamma(m)}x^{m-1}\exp\left(\frac{-x}{\beta}\right), \tag{7-2}$$

in which β is the scale parameter and m is the shape parameter; the maximum likelihood estimation equations for these parameters are given in standard tests on distribution parameter estimation.[12] Figures 7-1 and 7-2 show the three curves (CDF plot of actual data, lognormal fit of data, and gamma fit of data) for each of the Pc thresholds enumerated above. As can be seen, for nearly all cases the gamma fit is excellent, especially for the GP cases—good enough that a formal goodness-of-fit test was judged not to be necessary, given the approximate nature of the enterprise. The quality of this fit is not entirely surprising, as Chan and others have remarked that the miss-distance distribution for a single event should follow a non-central chi-squared distribution,[13] which is part of the gamma distribution family; so the two-parameter gamma distribution suggested itself as a candidate from the beginning. Tables 7-1 and 7-2, which appear after the figures, give the actual miss distance values at the indicated percentile points for the Pc levels investigated.

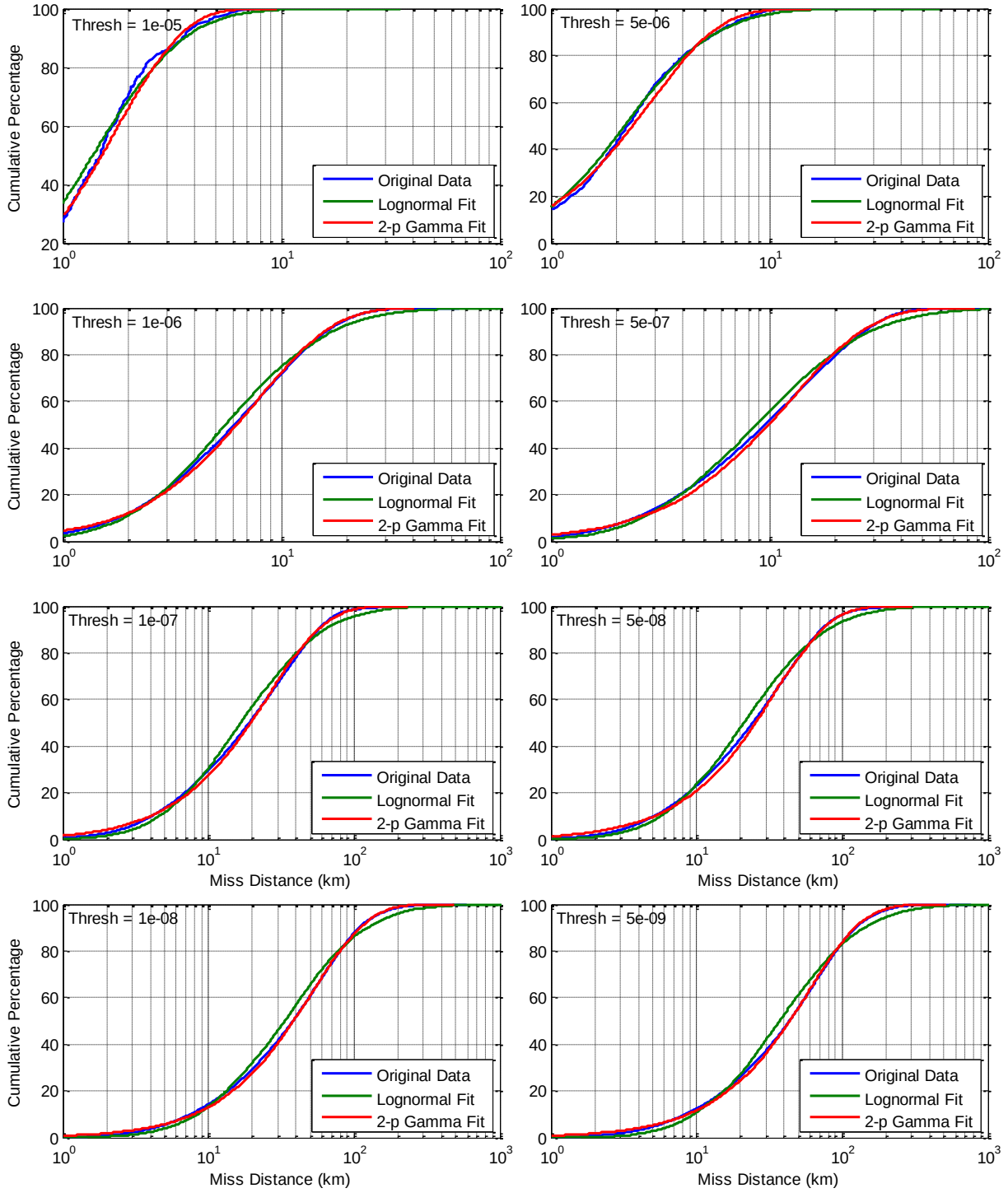


Figure 7-1: GP Miss Distance Distributions and Fits to Lognormal and Gamma Distributions for a Range of GP Pc Thresholds

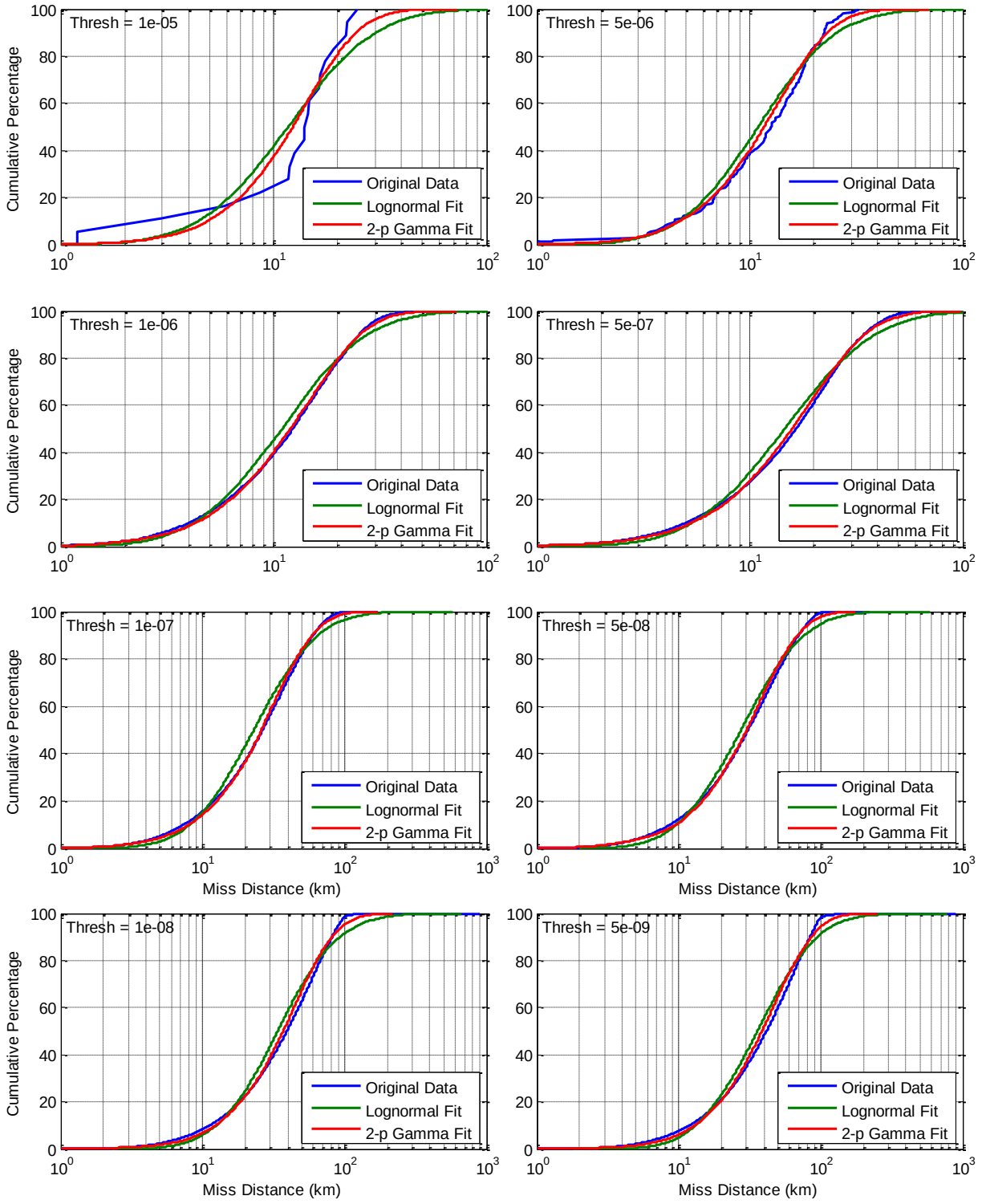


Figure 7-2: SP Miss Distance Distributions and Fits to Lognormal and Gamma Distributions for a Range of SP Pc Thresholds

Table 7-1: GP Miss Distance Screening Equivalents (km) for given Pc Values at Various Confidence Levels

	GP Pc							
Percentile	1.0E-05	5.0E-06	1.0E-06	5.0E-07	1.0E-07	5.0E-08	1.0E-08	5.0E-09
50	1.5	2.4	6.4	9.9	19.8	25.2	38.3	43.5
68	2.1	3.3	9.2	14.8	30.5	38.8	59.0	67.1
95	4.0	6.7	19.7	32.9	72.2	92.2	140.1	159.8
99	5.6	9.3	28.1	47.9	107.1	136.9	208.0	237.6
99.9	7.7	12.9	39.8	68.6	156.3	199.9	303.7	347.1
99.99	9.7	16.4	51.2	89.1	205.1	262.4	398.5	455.6
99.999	11.7	19.9	62.5	109.3	253.5	324.5	492.8	563.5
99.9999	13.6	23.3	73.7	129.5	301.7	386.3	586.7	671.0

Table 7-2: SP Miss Distance Screening Equivalents (km) for given Pc Values at Various Confidence Levels

	SP Pc							
Percentile	1.0E-05	5.0E-06	1.0E-06	5.0E-07	1.0E-07	5.0E-08	1.0E-08	5.0E-09
50	12.3	11.8	11.9	15.8	25.7	29.9	36.9	38.8
68	16.2	15.5	16.1	21.7	35.9	41.7	50.9	53.2
95	29.4	27.7	30.8	42.2	72.7	84.1	100.0	103.9
99	39.3	37.0	42.1	58.3	101.8	117.6	138.4	143.5
99.9	52.7	49.3	57.5	80.1	141.7	163.5	190.7	197.4
99.99	65.4	61.2	72.3	101.2	180.5	208.1	241.4	249.6
99.999	77.9	72.7	86.8	121.9	218.6	251.9	291.2	300.7
99.9999	90.1	84.0	101.1	142.3	256.3	295.2	340.3	351.2

Tables such as these can present a real temptation to the CA risk analyst: one can perform a simple screening based on miss distance, free of the data-release hassle of covariance-enabled state estimate data and knowing that it would be essentially as effective as the more cumbersome Pc calculation approach. Additionally, there are occasional situations in which the Pc approach, due to the lack of the proper data, is simply not possible; and in such cases, tables such as these can be helpful. However, the false-alarm rate, as pointed out earlier, is quite high; and such a quantity of false alarms will have a significant effect on the ability to keep launch windows open. To illustrate the difficulty, two quad-chart presentations similar to those used in Section 6, giving window open percentage as a function of window size and in this case miss distance, were generated and appear as Figures 7-3 and 7-4.

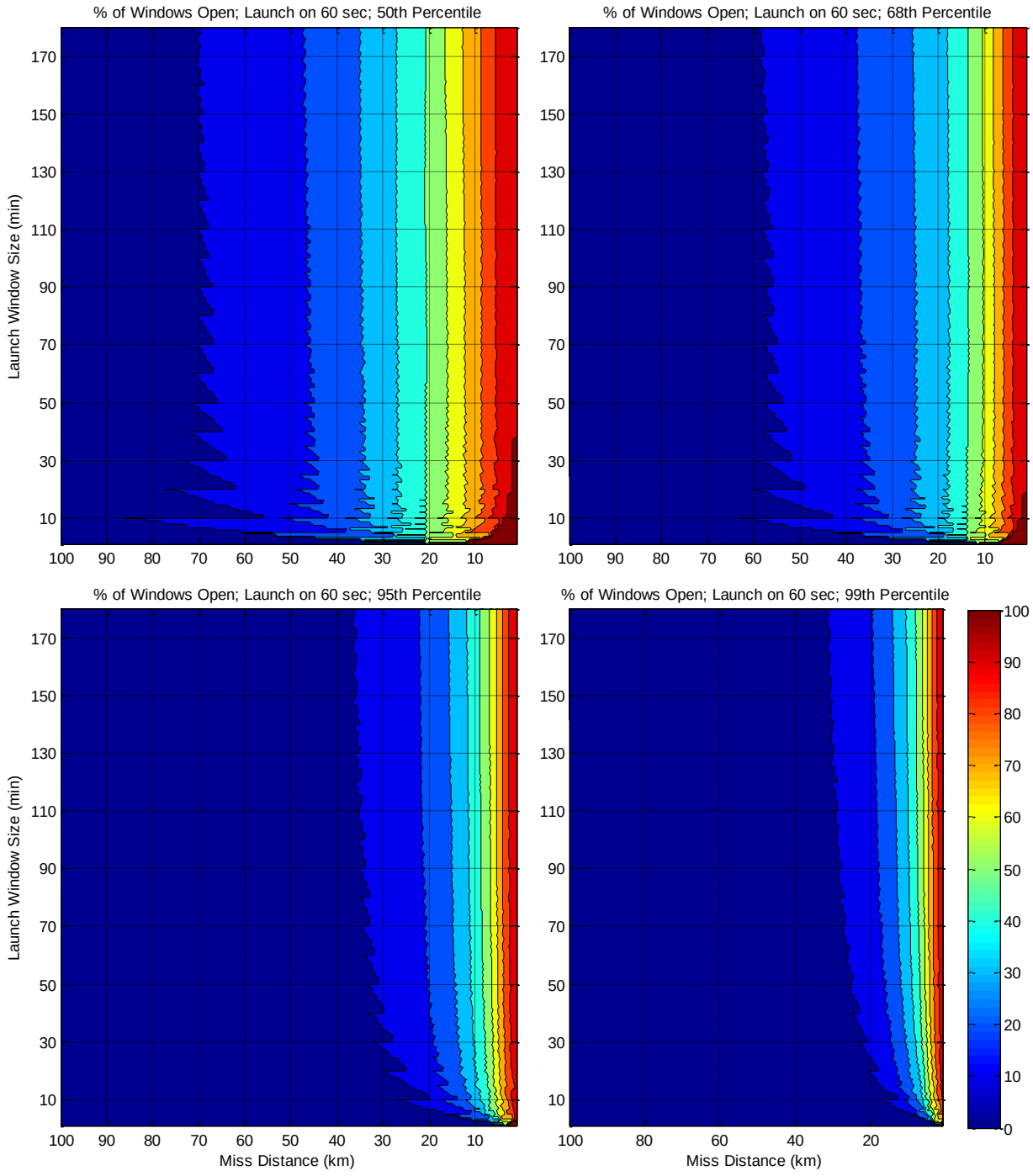


Figure 7-3: Window Open/Closed Percentage vs Launch Window Size vs Miss Distance; GP Screenings, 60-sec Interval ("Launch on Minute")

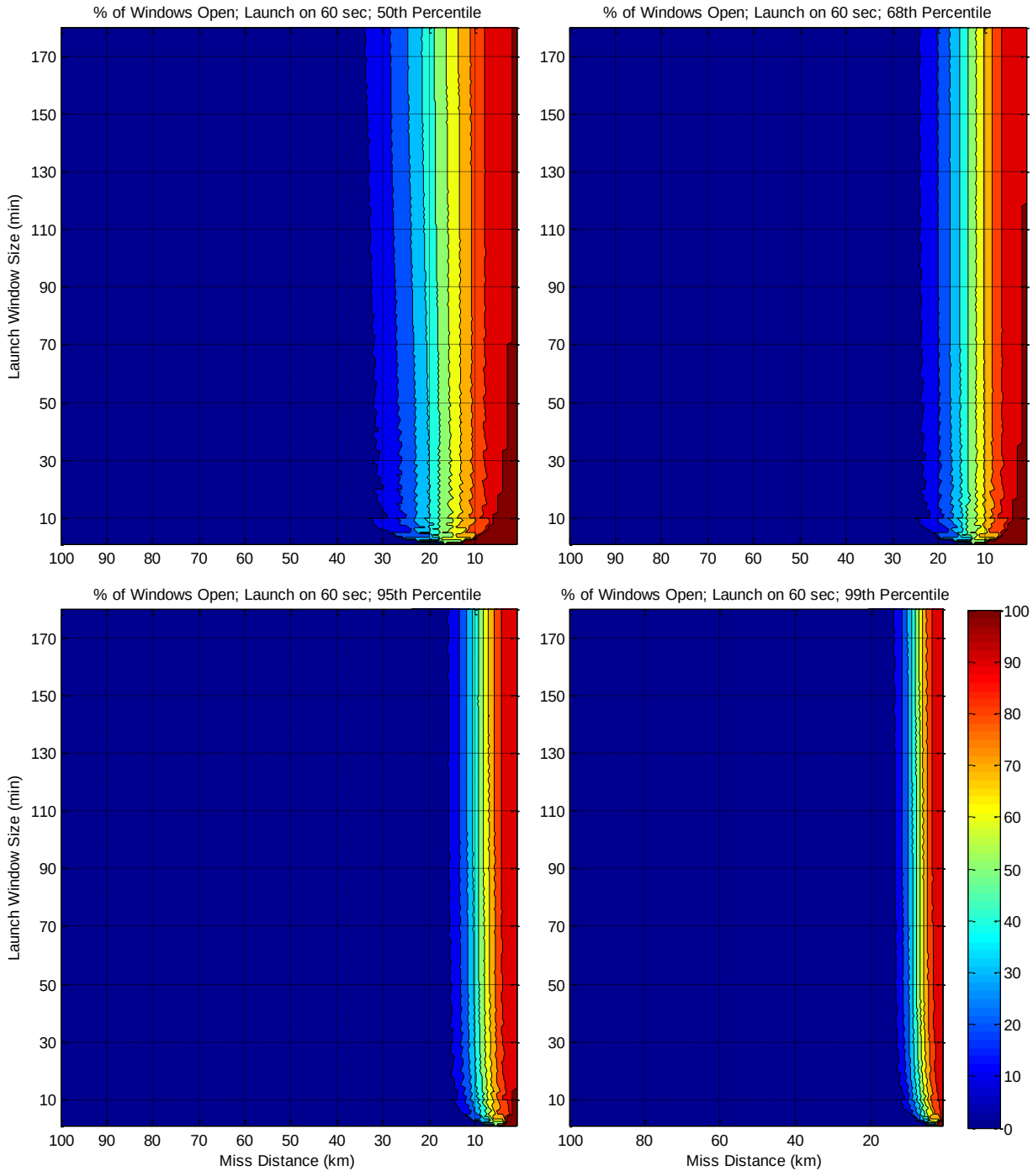


Figure 7-4: Window Open/Closed Percentage vs Launch Window Size vs Miss Distance; SP Screenings, 60-sec Interval ("Launch on Minute")

These charts make clear very quickly that screening at anything close to a large miss distance will have very serious effects on launch window availability. Looking back at Table 7-1, one observes that for a Pc equivalent screening of 1E-06, a miss distance screening of 19.7 km is needed to meet a 95% confidence value and a miss distance screening of 28.1 km to meet a 99% confidence level. In taking these numbers to Figure 7-3, for the 19.7 km figure, the launch window would be 50% open at the 50th percentile and only about 25% open at the 95th percentile. If one wishes the 99% confidence level and moves to the 28.1km figure, one now obtains windows 25% open at the 50th percentile and only 10% open at the 95th percentile—values that are clearly unacceptable. The situation in SP is slightly more sanguine at the 50th percentile but about equally bad at the 95% percentile. Moving to either a more demanding equivalent Pc threshold or a higher confidence value brings the situation to a point where the launch windows are almost entirely closed.

In summary, it is possible through this type of profiling to develop miss-distance screening threshold equivalents to Pc-based screening thresholds; but the operational use of these thresholds will come at a heavy price, namely a very large number of screening false alarms that quickly closes unacceptably large portions of launch windows. In occasional circumstances it may be necessary to perform screenings with this approach, but clearly such situations need to be rare. The technique may have more utility if only a limited number of satellites are to be screened—such as a relatively short list of defended assets. This particular approach to the LCOLA problem is explored in the next section.

Section 8: Defended Assets Screening

In Sections 2 and 3 of this report it was established that, even though the predicted launch trajectories often manifested large errors, their covariances properly reflected these errors, allowing a durable calculation of associated probabilities of collision. However, while the legitimacy of a Pc calculated under these circumstances is generally recognized, there is more question about the actionability of such information.[14] There are two basic ways a Pc can be small: the covariances can be relatively small and with little overlap, meaning that the state estimates are determined with little uncertainty and thus one can say that the two spacecraft trajectories do not in fact pass all that near to each other; or the covariances can be very large and overlap each other substantially, meaning that the state estimates are so poorly determined that the calculated miss distance is not reliable and thus there is actually a low probability of the two objects' colliding. The first scenario presents a low-risk situation because the conjunction information is well determined; the second scenario presents a low-risk situation because there is not enough reliable information to consider it high risk. Some have worried that, if the Pc threshold is set at a relatively small (i.e., restrictive) value, the risk assessment process will become overrun with events of the latter type, in which launch windows are substantially encumbered by conjunction events that in truth are not truly actionable.

If one is truly concerned about this possibility, one fairly activist reorientation of the risk assessment enterprise in response is substantially to limit the scope of the screenings, namely to focus them only on defended assets. Such a group of assets could be, in probably its most expansionist sense, all active satellites; and in its most restrictive sense only manned-flight vehicles actually occupied by astronauts. Once an approved list of defended assets has been established, this limited screening approach is easy enough to implement: one simply screens launch trajectories against only those satellites. It is no surprise that this approach can significantly improve the degree to which launch windows are open, and the goal of this section is to document the degree of improvement that such a reorientation would bring.

While it is not the purpose of this discussion to try to adjudicate the philosophical issue of whether this kind of reorientation is desirable, it is only responsible to state that such an activity should not be pursued lightly. First, any launch-related collision results in loss of mission for the launching spacecraft; and while it is understandably worse if an active on-orbit asset is lost also, it is still a quite serious situation to lose a launching asset; and this latter eventuality is what screening against the entire catalogue seeks to prevent. Second, regardless of how one feels about the loss of a the launching asset to a collision, such a collision will still be a significant debris-producing event, perhaps even much worse than the on-orbit collisions chronicled to date because it may consist of both a hyperkinetic impact and subsequently an exploding rocket motor, creating that much more fragmentation debris.

Figures 8-1 through 8-8 detail the expected launch window availability situation for the reduced screening approach described above. Open-source information was used to create a list of about 500 "active satellites" and a subset of these that are United States government or military satellites (about 150). A third subset, manned-flight-related objects, is so small (only a few objects) that screening against them would have essentially no encumbering effect on any launch windows, so it did not merit a separate set of graphs. As it is, the results for the two more demanding cases show a very large amount of launch window availability. For GP-based screenings against all active satellites, except for the

very shortest of window lengths all windows are at least 70% open at the 95th percentile, for both Pc values up to 1E-09 and miss distances up to 100km, but the latter does require context, as will be explained presently. For US government/military satellites, the results are even more sanguine—below the 95th percentile the graphing routine fails to produce contour graphs at all because the entire graph-space is at the 100% open value. For SP-based screenings, the Pc results are very similar to those for GP; but the miss-distance results look quite different. This is not due to any underlying difference in performance but rather to systematic differences in the data. The Aerospace GP screening routine filters out any results with Pc values less than 1E-12. Typically this makes no difference to the analysis, as there is little interest in events with risk values that small; but some portion of those events do have miss distances less than the 100km threshold reported in these graphs. For the SP screenings, there was no *a priori* exclusion; so all of the low-Pc events that nonetheless do have miss distances smaller than 100km are retained and shown here. Thus, it must be said that the SP results reflect a more realistic case than those for GP; but in either case, the results for miss distance values of 25km or less, which is the largest value that has been proposed to date for a miss-based screening against unmanned assets, show window-open percentages of 70% or greater.

If one wishes to "solve the problem" of encumbered launch windows, this approach of screening against only defended assets appears to be quite effective. Even miss-distance-based screenings, notorious for their heavy-handed window-closing potential, can be used up to the 25km miss distance range while leaving windows mostly open. Of course, one should reflect separately on whether it is desirable to neglect the debris-based aspect of collision avoidance, both to improve the survivability of the launching vehicle and to avoid any additional debris-producing events. Having said that, there are occasional situations that might benefit from this approach. Small-value payloads with extremely tight launch windows might pursue a defended-asset screening as a fratricide-minimization approach, recognizing that the launched payload itself is not particularly valuable and the debris-creation risk will need to be borne anyway given the brevity of the launch window. One would, however, expect such situations to be extremely rare.

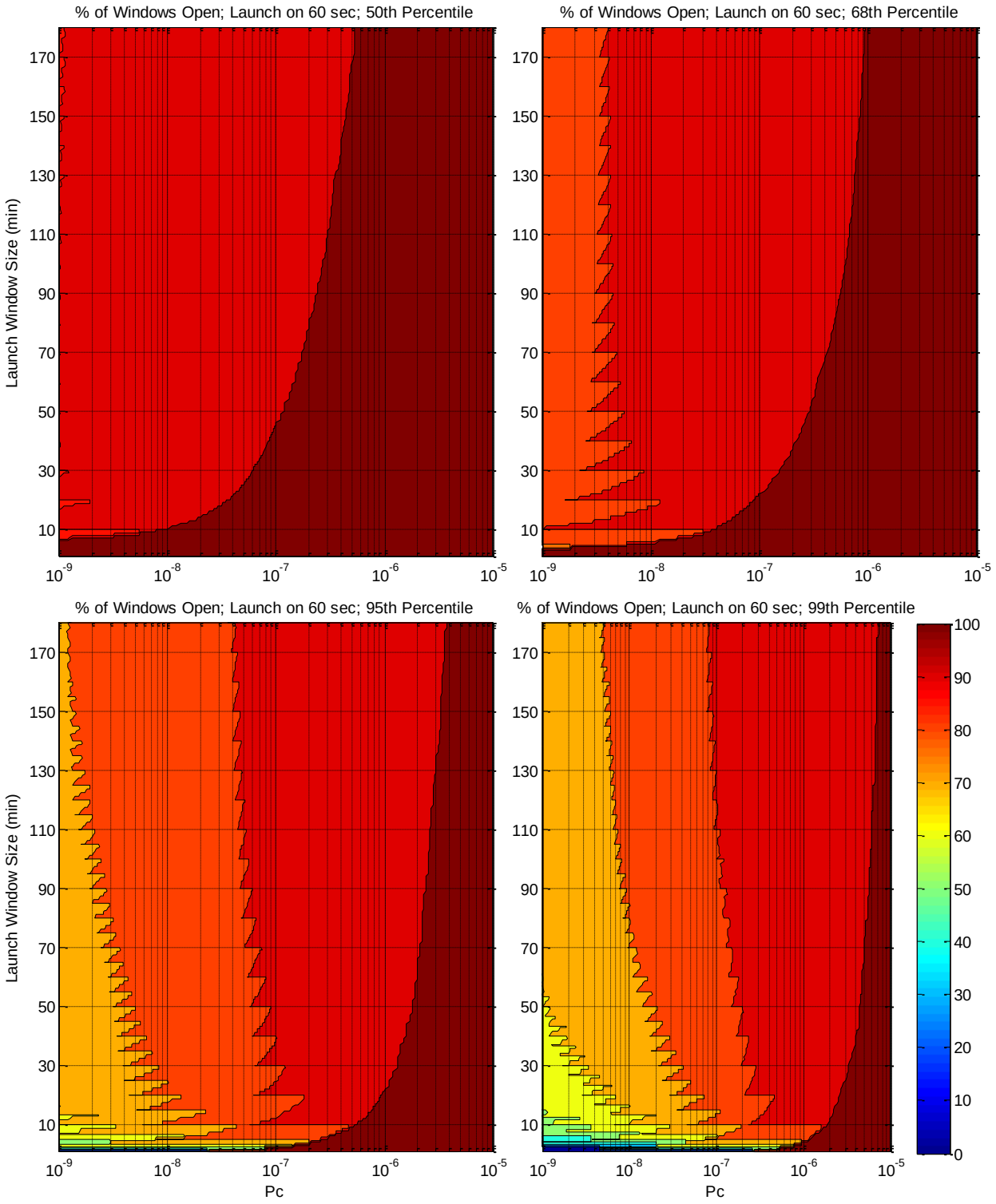


Figure 8-1: Window Open/Closed Percentage vs Launch Window Size vs Pc; GP Screenings, Launch-on-Minute, all Active Satellites

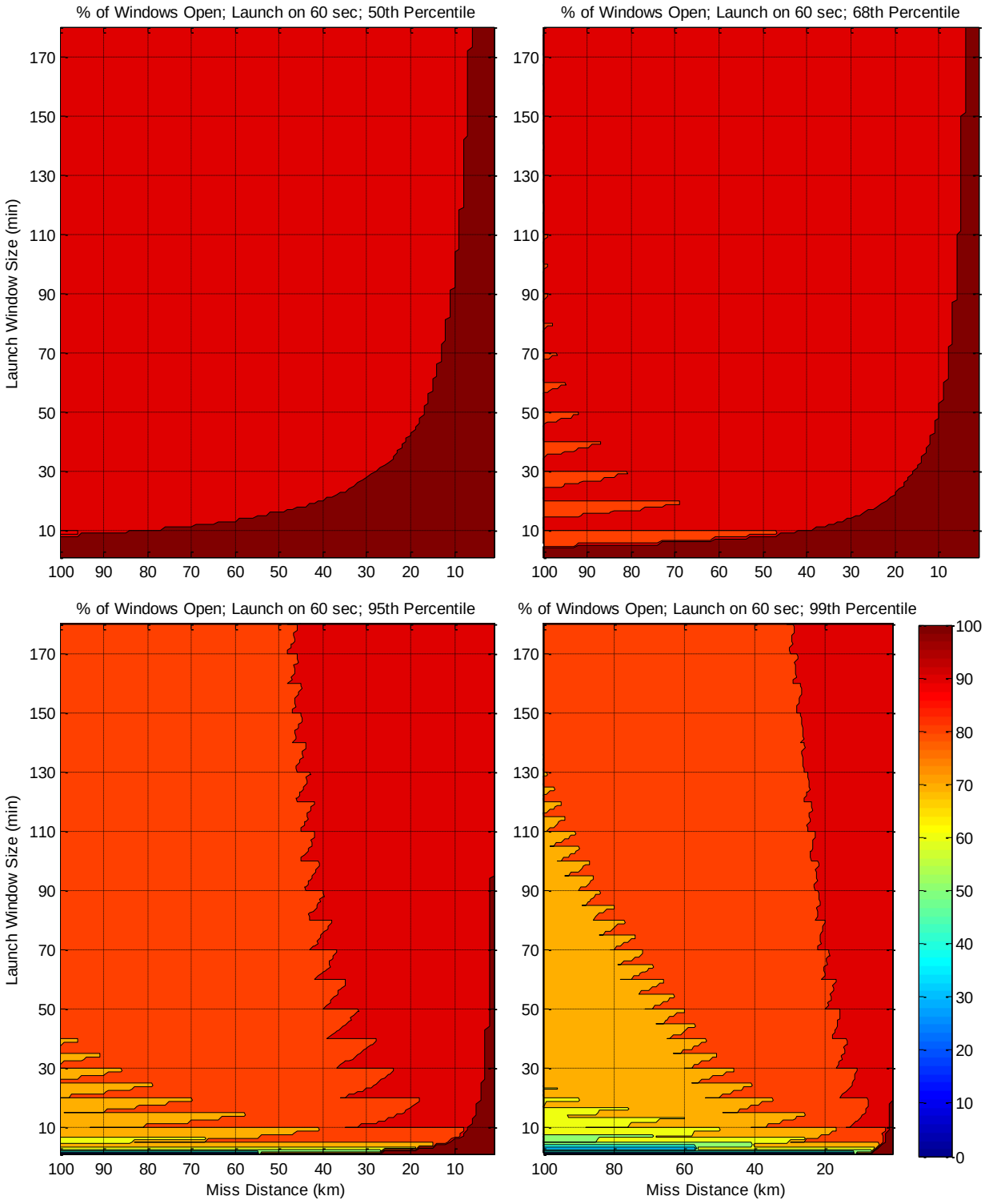


Figure 8-2: Window Open/Closed Percentage vs Launch Window Size vs Miss Distance; GP Screenings, Launch-on-Minute, all Active Satellites

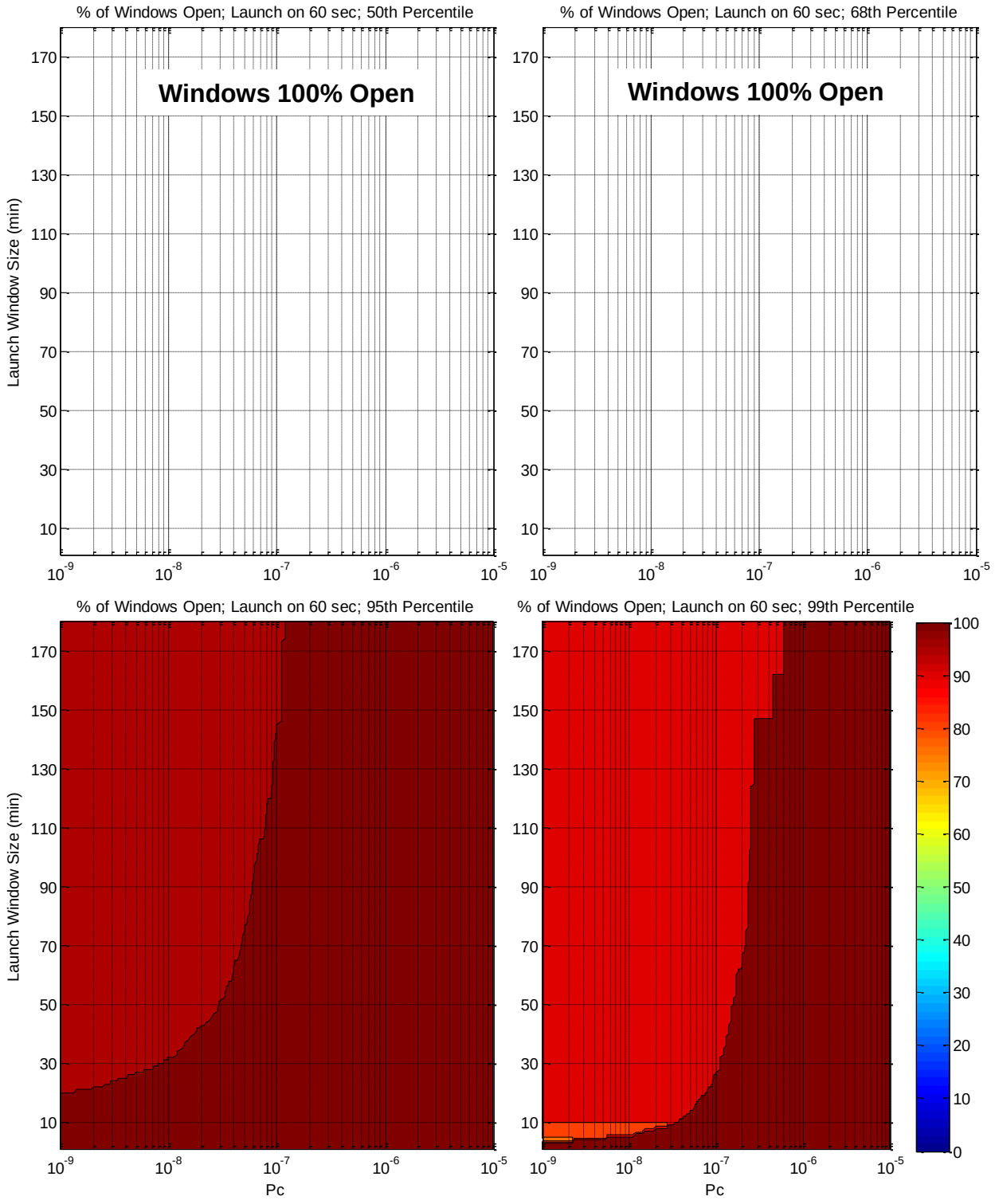


Figure 8-3: Window Open/Closed Percentage vs Launch Window Size vs Pc; GP Screenings, Launch-on-Minute, US Military/Government Satellites only

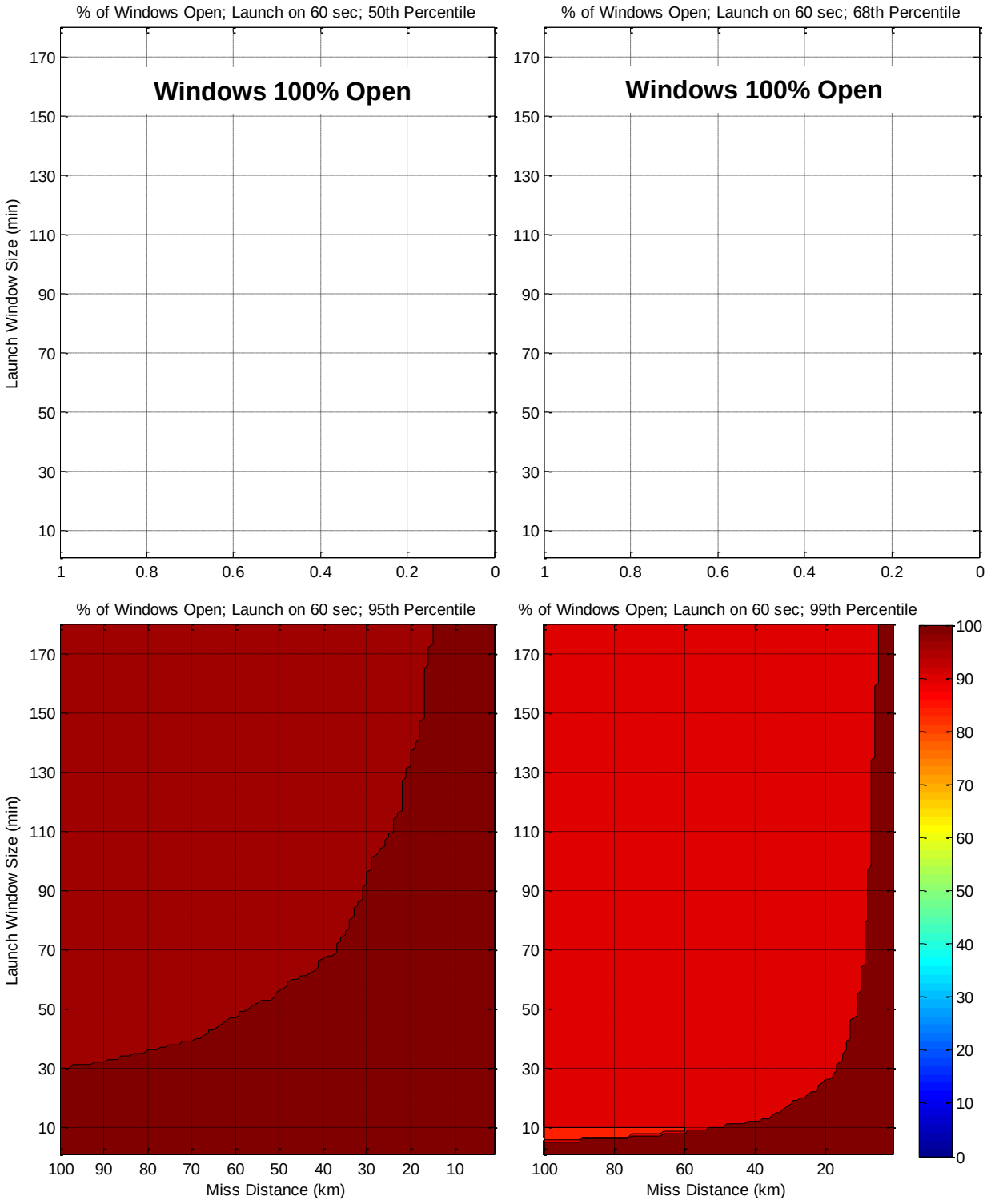


Figure 8-4: Window Open/Closed Percentage vs Launch Window Size vs Miss Distance; GP Screenings, Launch-on-Minute, US Military/Government Satellites only

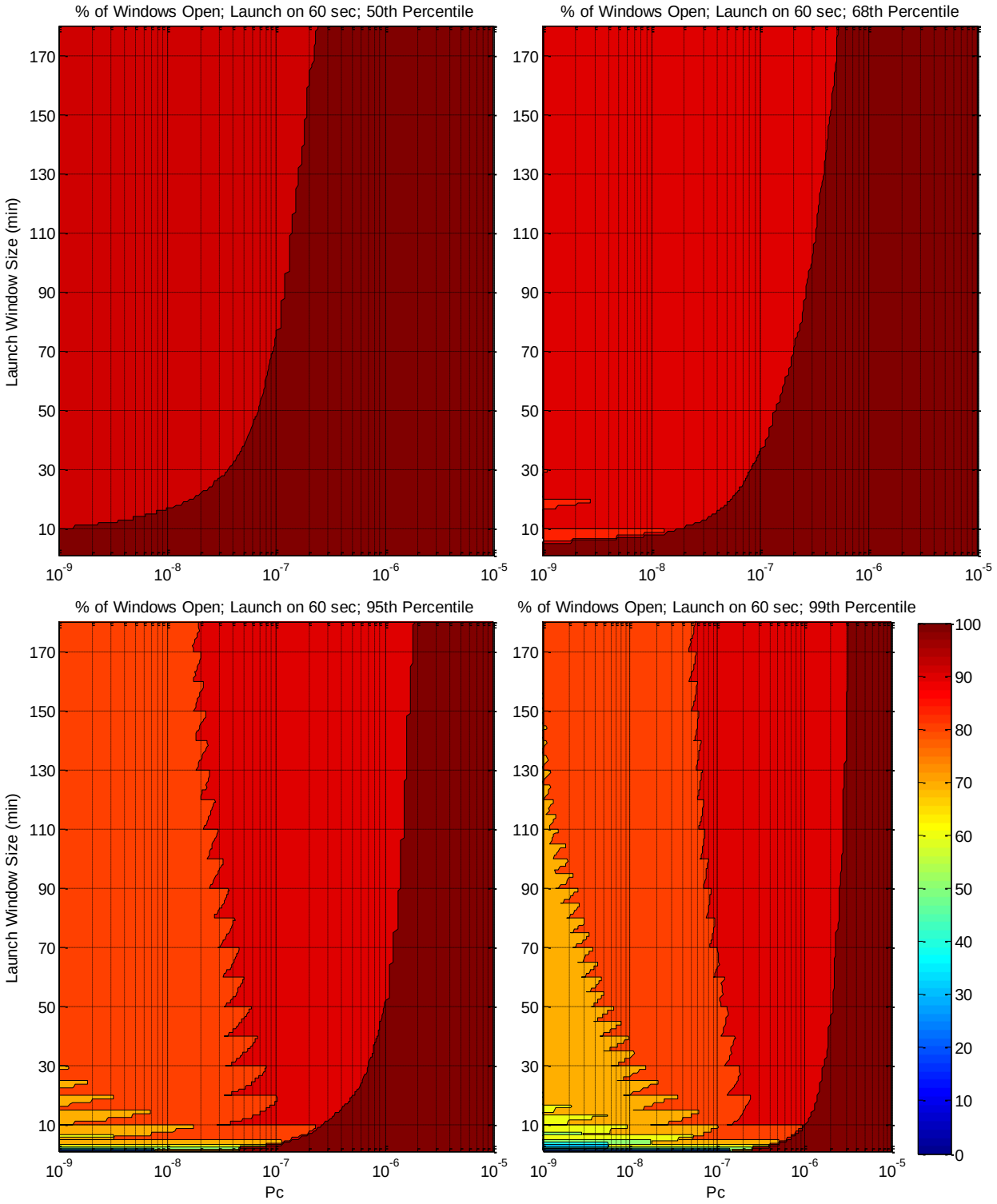


Figure 8-5: Window Open/Closed Percentage vs Launch Window Size vs Pc; SP Screenings, Launch-on-Minute, all Active Satellites

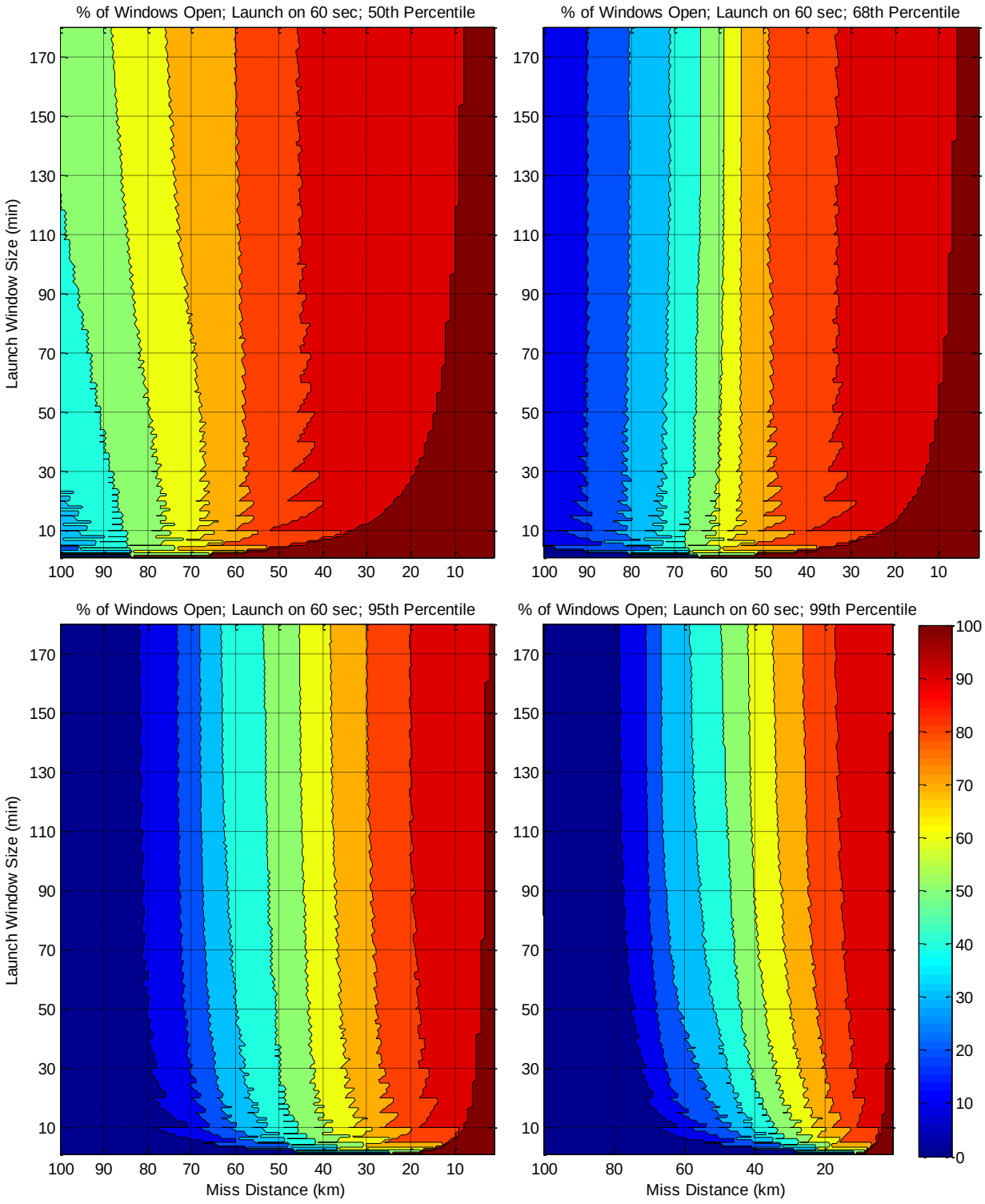


Figure 8-6: Window Open/Closed Percentage vs Launch Window Size vs Miss Distance; SP Screenings, Launch-on-Minute, all Active Satellites

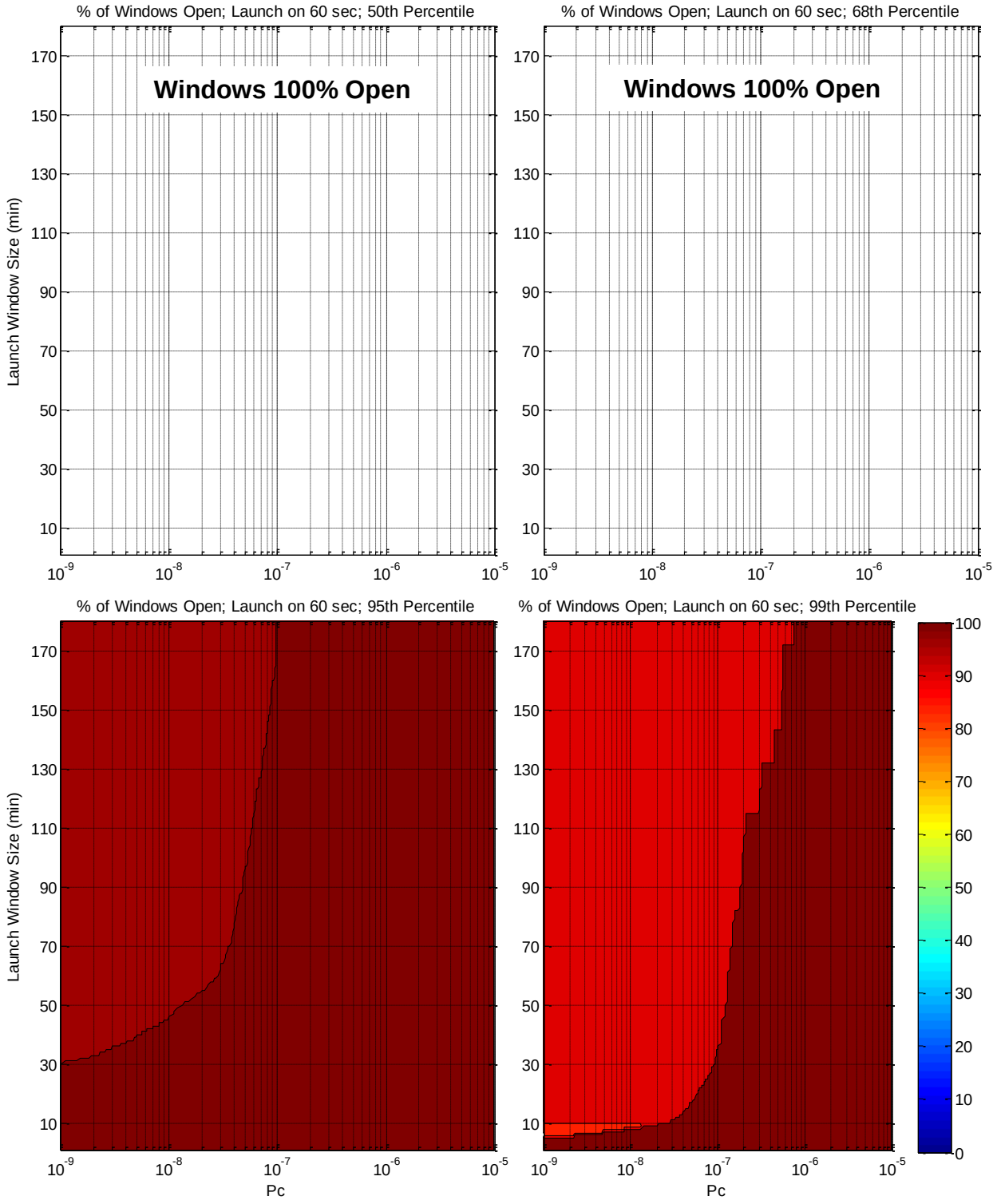


Figure 8-7: Window Open/Closed Percentage vs Launch Window Size vs Pc; SP Screenings, Launch-on-Minute, US Government/Military Satellites only

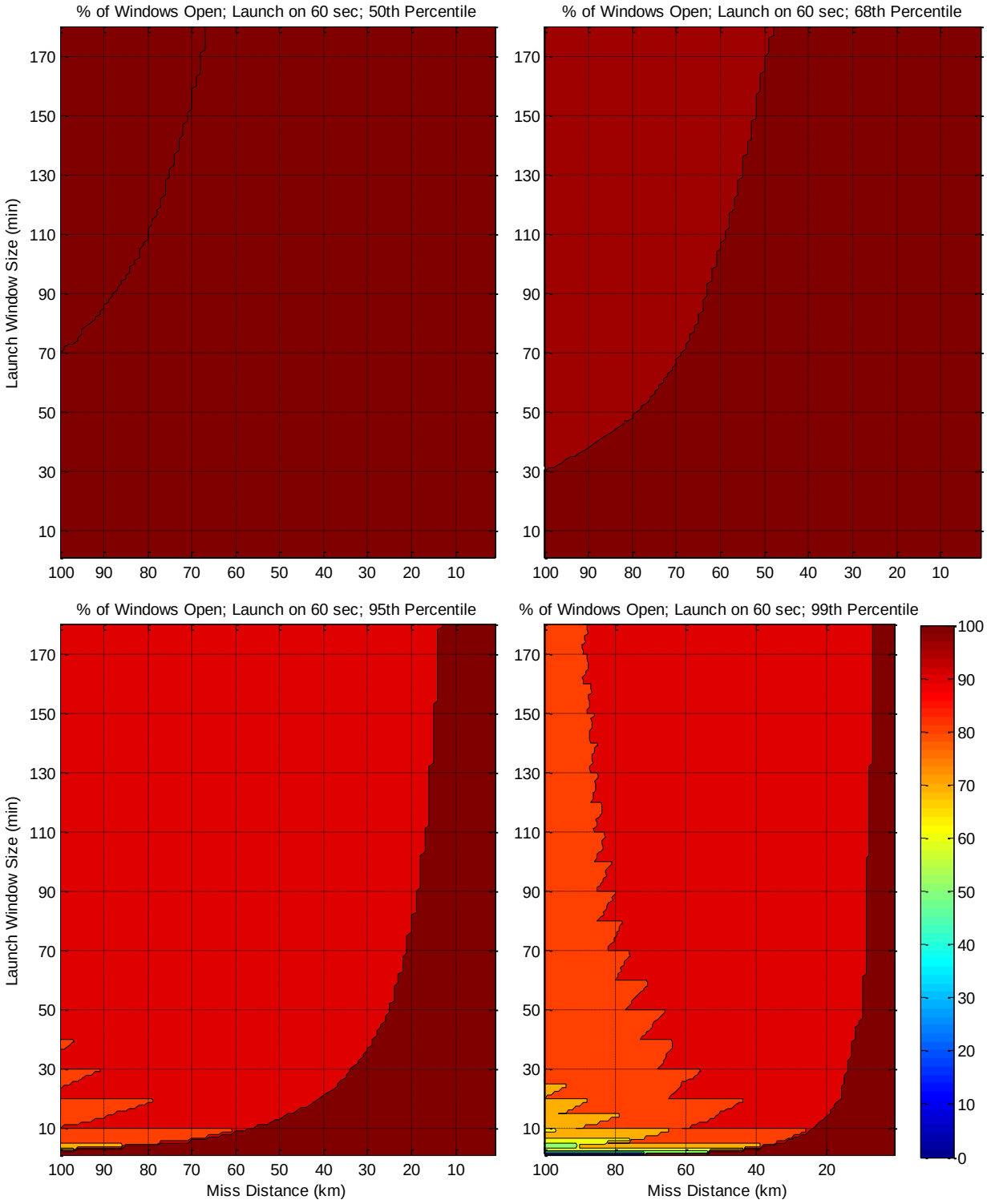


Figure 8-8: Window Open/Closed Percentage vs Launch Window Size vs Miss Distance; SP Screenings, Launch-on-Minute, US Government/Military Satellites only

Section 9: Lifetime LCOLA Risk Abatement

A question that is perpetually present to the collision avoidance and risk assessment enterprise is whether all of the effort, attention, and cost of these activities produce any tangible benefit. There have been a few verified collisions, to be sure; and some have even resulted in loss of mission for one of the colliding objects. But the number of collisions over 50 years of *conquête spatiale* is actually very small, the "sky" is relatively large, where one sets the Pc screening threshold is essentially an arbitrary decision, and remediation of even high-risk potential collisions can be waived by the proper authorities; and all of these facts pertain to on-orbit collisions, for which the total time on orbit is vastly larger than the relatively small amount of ascent time addressed by the LCOLA activities. All of this has led some observers to question the value of the entire process. What is needed is a way to evaluate the degree to which LCOLA practices improve the overall collision risk posture.

One approach is to calculate the lifetime risk posed to the launch enterprise by performing no LCOLA activities at all over a large number of launches and to compare this to the risk assumed by executing LCOLA screenings at a variety of different Pc values. For screenings at any given value, one can then directly weigh the improvement in risk posture against the encumbrances that those screenings would introduce. The following discussion outlines the calculation methodology for this approach and then presents the results in a number of different distillations.

To calculate the cumulative risk, in the absence of any LCOLA activity, of a launch-related collision, one begins with the assumption that any single launch screening from the large screening experiment that has provided all of the data for the last several sections of this report can be used to represent a typical screening result for any given launch. One draws 300 such events at random (with the number 300 representing the maximum launch density of 12 launches per year over a 25-year period) from the screening database (with replacement) and calculates the set of cumulative Pc values determined by sequentially stepping through the sample of 300: Pc of trajectory 1, Pc of trajectories 1 and 2 taken together, Pc of trajectories 1, 2, and 3 taken together, etc. This will yield a curve of the running cumulative probability of a launch-related collision as a function of the number of launch events. This experiment is then conducted as a series of 100,000 Monte Carlo trials of this calculation, and the Pc median (and variability about that median, if desired) at each sequential point (1 through 300) is tabulated. The result is a curve giving the cumulative Pc as a function of number of launches. The number-of-launches datum can be turned into a time-frame if one will postulate the average number of launches expected to be sustained each year.

This same approach can also be followed for the situation in which LCOLA screenings at a certain Pc value are performed. Each of the trajectories in each group of 300 draws is examined, and any trajectory that has a maximum Pc greater than the screening threshold is discarded and replaced by another from the pool of trajectories; this continues until an entire set of 300 trajectories that passes the screening threshold is assembled. The idea is to emulate the activity that would actually be performed at the launch site: if a particular launch opportunity failed the screening threshold, it would be cut out and a different opportunity selected. Of course, this approach presumes that there always would be enough compliant trajectories in any given launch window that an acceptable one could be identified; for short windows and very demanding screening levels, this assumption becomes questionable.

The results of these calculations are displayed in Figure 9-1, in which the top graph represents the GP results and the bottom the SP; the x-axis is the cumulative number of launch events. If, for example, one presumes twelve launches per year for a twenty-year period, the result is 240 cumulative launch events. The y-axis is the LCOLA screening level imposed. The very top of the graph corresponds to a screening level of 1E-03; and because no events arose in the experiment that even approached this level of risk, this line can be considered to represent the position of running no LCOLA screenings at all. The color indicates the level of cumulative risk, with dark red signifying a high risk and blue a low risk. The contours are evenly spaced in a base-10 logarithmic framework (contour 1 is for exponents between -4 and -4.5, contour 2 between -4.5 and -5, etc.); when this plot is converted to linear units, the spacing appears uneven (1E-04, 3.2E-05, 1E-05) but is easier to interpret than fractional exponents. As “practice” in reading this chart, one can follow in the upper (GP) graph the horizontal line corresponding to a LCOLA screening level of 1E-06 on the y-axis. For 100 cumulative launch events, the color band is a light red, corresponding to a cumulative risk level between 1E-05 and 3.2E-05. At 250 cumulative events, the color band is a darker but not the darkest red, corresponding to a cumulative risk level between 3.2E-05 and 1E-04. The SP graph is similar in appearance and behavior, except it is depressed about half an order of magnitude in terms of cumulative risk—the risks for a given Pc screening level and number of cumulative launch events are about half an order of magnitude smaller than the corresponding GP risk levels. This behavior is consistent with other GP-SP relationships encountered in other areas of this investigation.

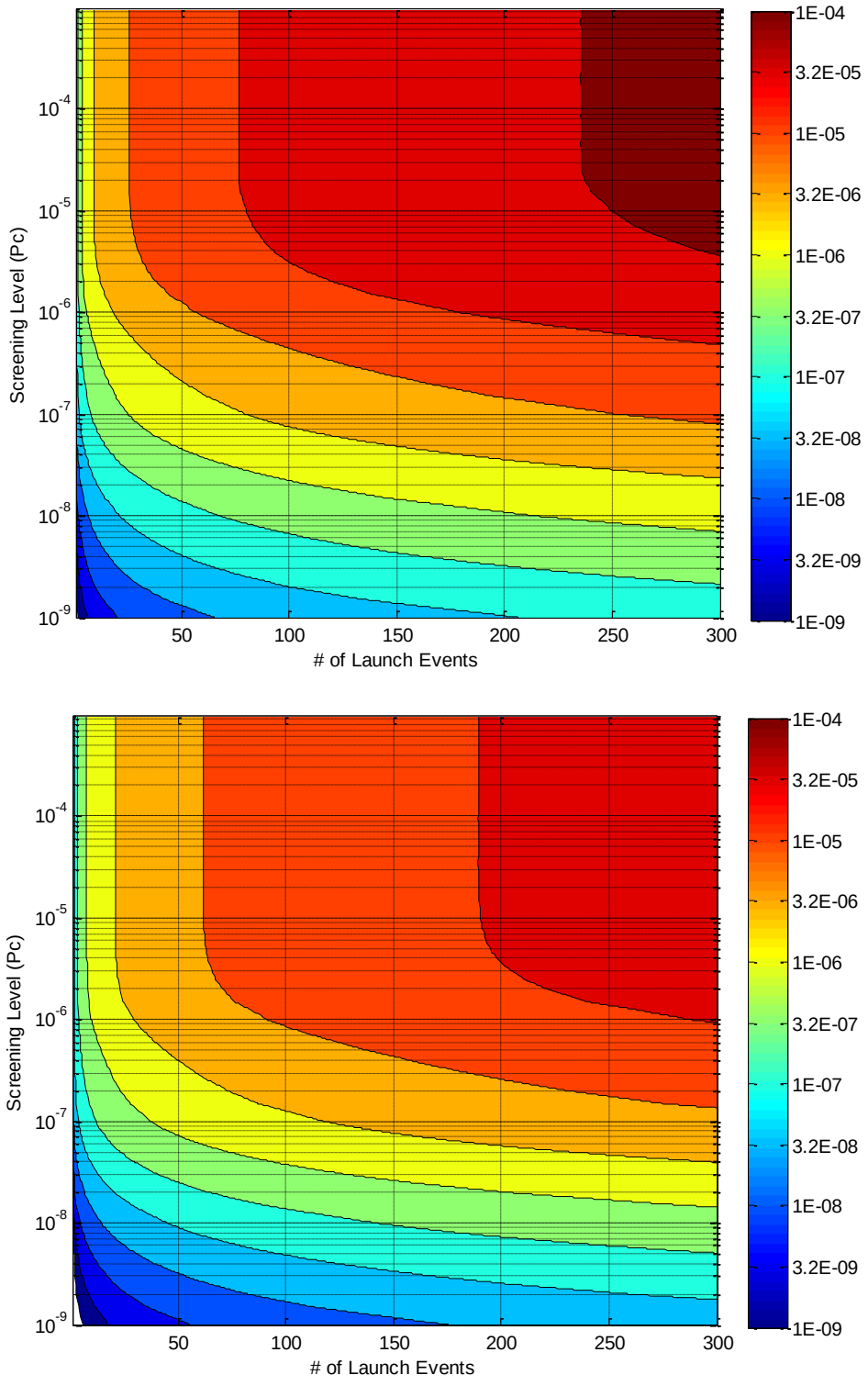


Figure 9-1a and b: GP (top) and SP (bottom) Cumulative Risk as a function of LCOLA Screening Level and Total Number of Launch Events

It will be noticed that, in the y-value range from approximately 1E-05 to 1E-03 on the GP chart and 5E-06 to 1E-03 on the SP chart of Figure 9-1, the contours are nearly vertical. Given that the level of 1E-03 is functionally equivalent to performing no LCOLA activities, this fact indicates that no risk posture improvement is wrought at all until one begins screening at the level of 1E-05 in GP and 5E-06 in SP. LCOLA policies that mandate screening levels more permissive than these are essentially the equivalent of doing nothing at all, and one should thus consider whether there is any compelling purpose to keep such policies in place.

The difference in cumulative risk between performing no LCOLA and screening at a given Pc level can be obtained through subtraction by eye from the charts in Figure 9-1, but for convenience this subtraction has been conducted and the results presented in Figure 9-2. The x- and y-axes are the same as for Figure 9-1, but the colors now represent the difference in risk posture, shown as the base-10 log of the ratio of the screened risk level to the level obtained by performing no screening at all. A level of -1 would indicate an order of magnitude difference between the LCOLA screening cumulative risk and the "big-sky" risk; a level of -2 would indicate two orders of magnitude difference, &c.

It is to be expected that, just as in Figure 9-1, the GP and SP parts of Figure 9-2 have the same look and feel, with a similar half-order-of-magnitude shift "down" for the SP results; and the interpretation of both plots is straightforward. The values plotted on this chart have been rounded to the nearest tenth of an order of magnitude in order to eliminate a small amount of chatter that would make the graph more difficult to read; so the dark red section represents risk reductions less than one-tenth of an order of magnitude—essentially a value of zero. For a value of 240 total launch events (twelve events per year for twenty years), in the GP case no improvement at all is observed until one reaches a screening level of about 8E-06, where the dark red band ends; and one stays between 0.1 and 0.5 of an order of magnitude until crossing into the dark yellow band, which occurs at a level of about 6E-07; for SP, to achieve the half-order-of-magnitude improvement requires screening at about 2.5E-07. A single order-of magnitude improvement requires levels of 1E-07 for GP and 6E-08 for SP.

While this information is useful, it is still somewhat abstract in the sense that these screening levels are not linked directly to launch operations encumbrance. The charts in Figure 9-3 attempt to do this by linking levels of risk abatement directly to launch window closure levels. This is accomplished by determining the window closure rates associated with LCOLA screenings at given Pc screening levels and linking to these the cumulative risk abatement results for those same screening levels. The result eliminates the "middleman" datum of Pc screening level and produces a direct relationship between total cumulative risk level and launch window closure level. Of course, certain assumptions were necessary here in order to reduce the dimensionality of the problem: regarding window closures, a launch-on-minute scenario with a 30-minute launch window was presumed, and 95th-percentile window closure data from Section 6 (GP and SP; Figures 6-1 and 6-7) were used; and for cumulative risk, a cumulative launch number of 240 (twelve launches per year for ten years) was employed.

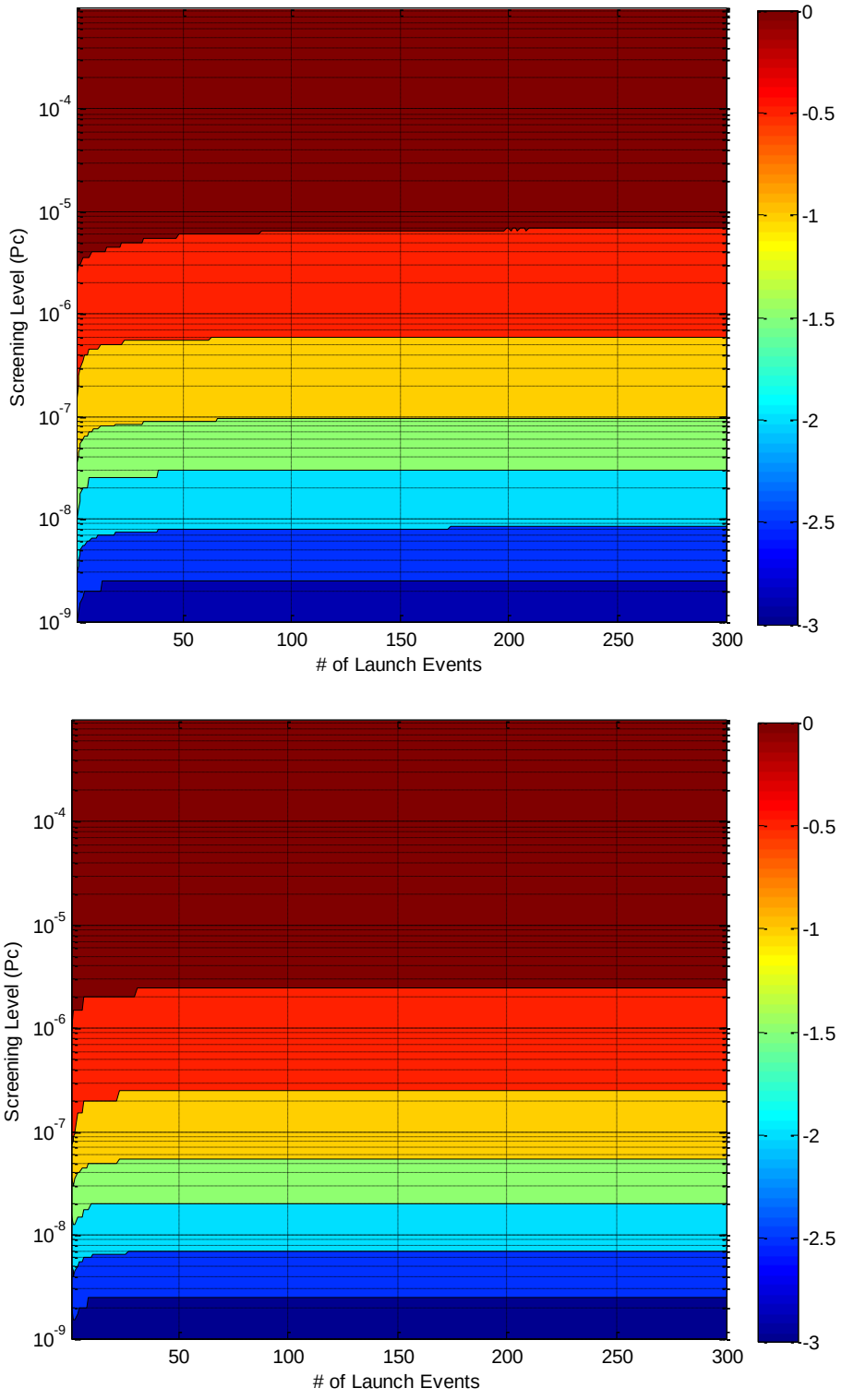


Figure 9-3a and b: GP (top) and SP (bottom) Cumulative (log10) Risk Reduction as a function of LCOLA Screening Level and Total Number of Launch Events

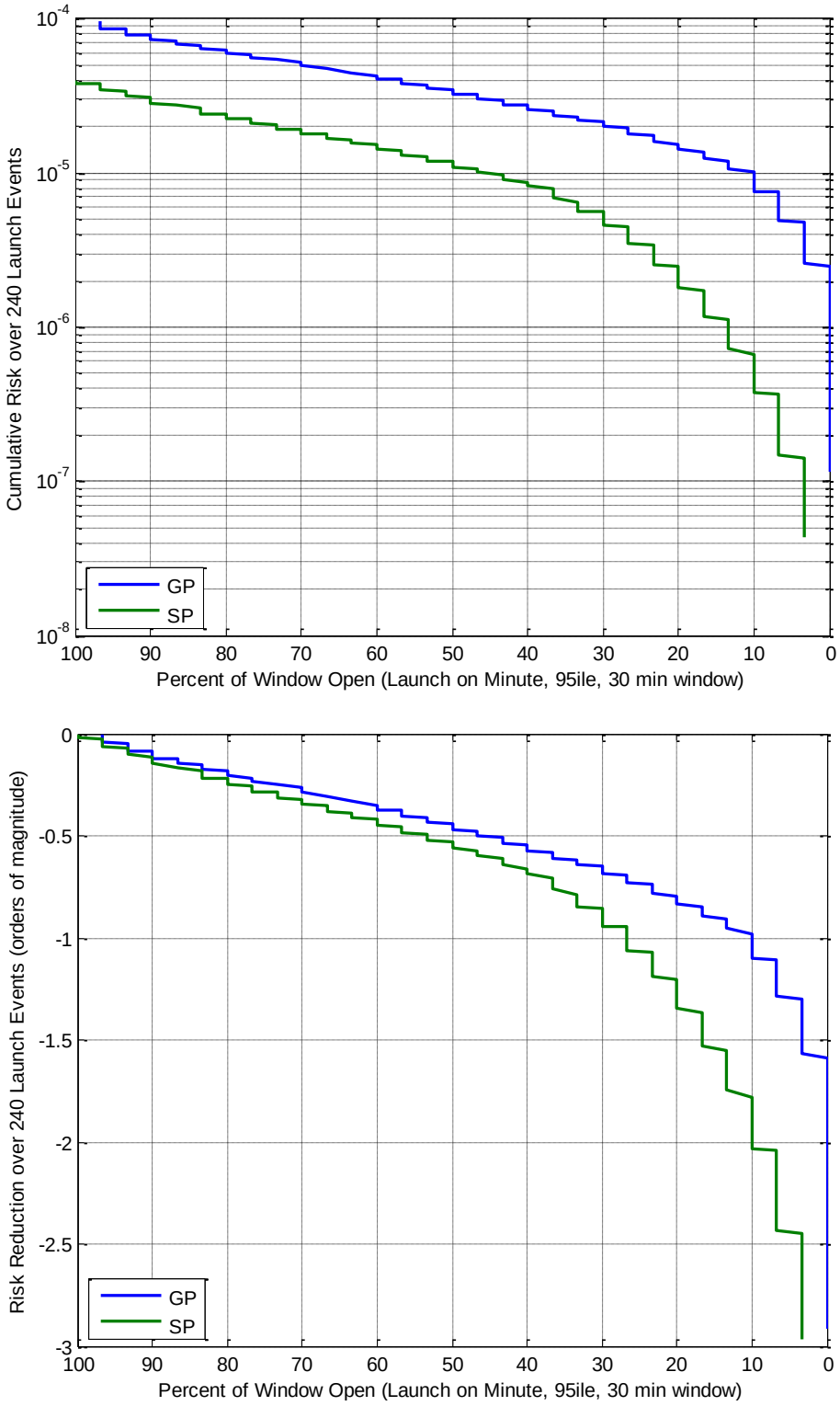


Figure 9-2a and b: Cumulative Risk versus Percent of Windows Open (top) and Total Risk Reduction versus Percent of Windows Open (bottom)

The upper graph in Figure 9-3 gives the cumulative risk as a function of window closure; the x-axis is the percent of the launch window that is open, so the situation becomes less favorable the more to the right one moves; and the y-axis is the cumulative risk over the 240 launch events. The lower graph plots the same information but as a risk decrement from the risk posture encountered by performing no LCOLA at all, so here one can see the improvement in risk posture (in orders of magnitude, just like the charts in Figure 9-2) as a function of window closure levels. Because a 30-minute launch window is used, the range of values for window closure levels is in discrete units of one thirtieth; so the graph appears somewhat jagged.

In examining this latter chart, one sees that one must accept a 50% window closure level to improve the overall risk posture by half an order of magnitude (about the same for GP and SP) and to levels of 75% and 90% closure to improve the situation by a full order of magnitude. A two order-of-magnitude remediation requires 90% window closure in GP and is not possible at all for SP. These values would be even less sanguine for shorter windows or more granular launch cadences (such as launch-on-30-seconds). Non-trivial improvement to the LCOLA risk posture is possible, but this improvement brings with it a considerable level of encumbrance to launch operations.

Section 10: Summary and Conclusions

A number of disparate LCOLA topics have been addressed in this study. The following is an attempt to collect and present the major findings in summary form but in somewhat more detail than the executive summary. Specific recommendations are also embedded in the discussion as appropriate.

Accuracy of Predicted Launch Trajectories (Section 2). In comparison to state estimates of on-orbit assets, predicted launch trajectories are less accurate. The Delta II trajectories, which are somewhat more consistent in response, produce position error CDFs that clump around 20-30 km Vmag error; and nearly all stay under 50 km. The Atlas V trajectories, which are somewhat less consistent, show accuracy groupings around the 10 km, 30 km, and 70-100 km error regions. Overall, this level of performance can be considered about an order of magnitude less accurate than TLE state estimates and perhaps about 1.5 orders of magnitude less accurate than eGP-produced TLE state estimates, which is the direction in which the GP catalogue is moving. Given these results, there is no real advantage from a state estimate accuracy point of view to using the SP catalogue over the GP catalogue for LCOLA screenings.

Realism of Predicted Launch Trajectory Covariances (Section 3). The estimated covariances that accompany the predicted launch trajectories are also very large, and this has led to speculation that perhaps they are excessively conservative. Classical covariance realism tests, however, verified that normalized residuals calculated with the aid of these covariances conform to the expected distributions at quite satisfactory *p*-values; so one can conclude that they do provide a realistic estimate of the trajectory errors. This finding is important, as it means that the Pc values calculated with these covariances are reliable and thus can serve as a legitimate input to LCOLA operations.

Relationship of GP to SP LCOLA Screening Results (Section 4). In comparing GP to SP screening results for the same conjunction events, 60% of the cases produced Pc values within one order of magnitude of each other (*e.g.,* 1E-07 vs 1E-06). Since an order of magnitude change in Pc is the rule-of-thumb for a significant change, the difference in calculated Pc between GP and SP for a majority of the cases is below the threshold of significance, meaning that the results are essentially the same. Furthermore, for nearly all of the remaining 40% of the cases, in which the Pc difference is greater than an order of magnitude, the GP Pc is greater (larger number) than the SP Pc. This indicates that GP Pc calculations can be used as a conservative proxy for SP screenings; there may be more window cut-outs than necessary, but there is only a very small chance of missing or under-assessing a risky event. This result is a further validation of the conclusion that GP-based screenings can appropriately be used for LCOLA.

Maximum vs Cumulative Pc Calculations (Section 5). Current LCOLA processing uses the maximum Pc from all of the discrete conjunctions in a screening to represent the overall risk level for that screening; a cumulative calculation would convolve all of the individual Pc values for every conjunction in the screening into an overall cumulative Pc value. This latter calculation, which more thoroughly represents the overall combined risk of launching along a particular trajectory at a particular moment, tends to ride above the maximum Pc calculation by a factor of 1.5 to 2. Furthermore, in about 5% to 10% of the cases (for the range of Pc values that are likely to be used for screening thresholds), the cumulative Pc exceeds the Pc screening threshold when the maximum Pc does not. These indices make

a reasonable argument for favoring the cumulative over the maximum Pc as the LCOLA screening criterion.

One approach to avoid having to change existing LCOLA screening software to accommodate a cumulative Pc calculation is to identify a maximum Pc that is somewhat more demanding (smaller number) than the "canonical" Pc against which one wishes to screen and screen against that value instead. Experimenting with this approach reveals that it is nearly impossible to identify such a modified Pc value that will function in a (mostly) equivalent manner to the larger-number cumulative Pc and not generate an unacceptable number of type 1 and type 2 errors. If the additional virtues of the cumulative Pc approach are desired, it will be necessary to modify the software to use that calculation paradigm. Fortunately, the calculation itself is not difficult; and such a modification would be easy and straightforward.

Window Closure Effects (Section 6). Launch operations typically desires launch windows of a minimum duration of twenty minutes and ideally much longer (hours), with such windows being at least 80% open, with true nervousness beginning at the 50% level. If the more conservative criterion (*i.e.,* 80% open) is to be applied, then the Pc screening levels to be used, from Table 6-3, are 2E-06 in GP (max Pc, 95th percentile) and 8E-07 in SP. However, it was pointed out that the openness of launch windows is more or less evenly distributed throughout the window; so if one were to subdivide a window of reasonable length into, say, fifths, each fifth should have about the same level of availability as the overall window. Therefore, if the true concern is the ability to find a reasonable launch opportunity in perhaps the beginning, middle, and final portion of the launch window, one can chose a Pc screening threshold that produces a much lower level of window openness than the 80% canonical value; one might even go as low as 50%. In choosing a middle ground between the 80% and 50% values and generating nice round numbers, one might promote 1E-06 for GP and 4E-07 for SP.

Miss Distance Proxy (Section 7). While there is no direct correlation between close-approach miss distance and Pc, it is possible to establish a linkage in a percentile sense—to establish the screening miss distance that must be used in order to capture a certain percentile of the events that have a Pc of a stated value; the statement formulated would be of the form "Screening at a miss distance of x would capture y percent of the events that produce a Pc of at least z"; or, to use a real example, "Screening at a miss distance of 28.1 km would identify 99% of the events with a Pc of 1E-06 or greater." Table 7-1 gives actual values for this miss distance proxy at a variety of equivalent Pc and percentile values. While undoubtedly simpler, this approach is heavy-handed in terms of launch window cut-outs: screening at miss distances greater than 10km (necessary for equivalency to most of the Pc values that would be desired) essentially makes launch operations impossible, and even the 10km value is probably too large. This approach is therefore not recommended except in the most extreme contingencies.

Screening of Defended Assets Only (Section 8). One approach to preserving openness of launch windows is to limit the LCOLA screening activities to only certain defended assets, such as all active satellites. Such an approach does dramatically increase launch window availability percentages when viewed as a function of Pc; as a function of miss distance, the situation is certainly more sanguine but still reasonably restricted when miss distances of 50km or so are desired. As such, this stratagem is probably not as useful as imagined. If one restricts the defended assets to only US active government/military satellites, the launch

windows are open virtually all of the time when viewed as a function of Pc and nearly all of the time even for miss distances up to 100km.

Overall Value of LCOLA Operations (Section 9). One way to gauge the overall value of LCOLA operations is to compare the effects of screening at a certain Pc level to the alternative of doing nothing at all to mitigate potentially dangerous conjunctions. A Monte Carlo simulation can determine the cumulative risk of a group of launch events (e.g., x events per year over y years), both with no LCOLA activities performed at all and with LCOLA screenings conducted at a given Pc threshold. Screening at a Pc threshold equal to or greater than 1E-05 (GP) or 5E-06 (SP) is functionally equivalent to doing nothing from a risk management standpoint. In order to achieve half an order-of-magnitude improvement over the "do nothing" risk posture, one must screen at a level of about 6E-07 in GP and 2.5E-07 in SP; to achieve a full order-of-magnitude improvement requires 1E-07 and 6E-08, respectively. These data can also be plotted against window closure values; to obtain the half-order-of-magnitude improvement results in window closure levels of 50% for both GP and SP; achieving the full order-of-magnitude improvement results in closure levels of 75% and 90%, respectively.

Section 12: Acronym List

Acronym	Definition
AFI	Air Force Instruction
ANOVA	Analysis of variance
ASAT	Anti-satellite
ASW	Astrodynamics workstation
CA	Conjunction analysis
CDF	Cumulative distribution function
COMBO	Computation of Miss between Orbits
DoF	Degree(s) of freedom
ECI	Earth-centered inertial
EDF	Empirical distribution function
EFG	Earth-Fixed Greenwich
eGP	Extrapolated general perturbations
ELV	Expendable Launch Vehicle
GEO	Geosynchronous orbit
GOF	Goodness-of-fit
GP	General Perturbations
GPR	Goddard Policy Requirement
GPS	Global Positioning System
GSFC	Goddard Space Flight Center
ISS	International Space Station
JPL	Jet Propulsion Laboratory
JSpOC	Joint Space Operations Center
KSC	Kennedy Space Center
LCOLA	Launch collision avoidance
LDA	Launch Decision Authority
LEO	Low-Earth orbit
LSP	Launch service provider
MA COLA	Mission assurance collision avoidance
MAJCOM/CC	Major Command Commander
MCC	Mission Control Center
NAF/CC	Numbered Air Force Commander
NASA	National Aeronautics and Space Administration
OD	Orbit determination
Pc	Probability of collision
RSO	Resident space object
SGP4	Simplified General Perturbations Theory #4
SP	Special Perturbations
TCA	Time of closest approach
TDRSS	Tracking and Data Relay Satellite System
TLE	Two-line element set
USAF	United States Air Force
USSTRATCOM	United States Strategic Command
UVW	Radial, in-track, cross-track
VCM	Vector covariance message

Section 13: References

[1] Hametz, M.E. "A Geometric Analysis to Protect Manned Assets from Newly-Launched Objects: COLA Gap Analysis." AAS/AIAA Space Flight Mechanics Conference (Kauai, HI), February 2013.

[2] Cappellucci, D.A. "Special Perturbations to General Perturbations Extrapolation Differential Corrections in Satellite Catalog Maintenance." AAS/AIAA Astrodynamics Specialists Conference (Lake Tahoe, CA), Aug. 2005.

[3] Hejduk, M.D., Ericson, N.L., and Casali, S.J. "Beyond Covariance: A New Accuracy Assessment Approach for the 1SPCS Precision Satellite Catalogue." 2006 MIT / Lincoln Laboratory Space Control Conference, Bedford, MA. May 2006.

[4] Hejduk, M.D. "Space Catalogue Accuracy Modeling Simplifications." 2008 AAS Astrodynamics Specialists Conference, Honolulu, HI, August 2008.

[5] Ghrist, R.W. and Plakalovic, D. "Impact of Non-Gaussian Error Volumes on Conjunction Assessment Risk Analysis). AAS/AIAA Astrodynamics Specialists Conference (Minneapolis, MA), Aug. 2012.

[6] D'Agostino, R.B. and Stephens, M.A. *Goodness-of-Fit Techniques*. New York: Marcel Dekker, Inc., 1986.

[7] Gist, R.G. and Oltrogge, D.L. "Collision Vision: Covariance Modeling and Intersection Detection for Spacecraft Situational Awareness." 1999 AAS/AIAA Space Flight Mechanics Conference, Girdwood, Alaska, August 1999. See also Oltrogge, D.L. and Gist, R.G. "Collision Vision: Situational Awareness For Safe And Reliable Space Operations" Paper IAA-99-IAA.6.6.07, 50th International Astronautical Congress, Amsterdam, The Netherlands, 4-8 Oct 1999.

[8] Foster, J.L. and Estes, H.S. "A Parameteric Analysis of Orbital Debris Collision Probability and Maneuver Rate for Space Vehicles." NASA/JSC-25898 (August 1992).

[9] Cerven, W.T. "Covariance Error Assessment, Correction, and Impact on Probability of Collision." AAS/AIAA Space Flight Mechanics Meeting (San Diego, CA), February 2011.

[10] KSC Report ELVL-2008-0040593, 5 NOV 2008.

[11] Alfriend, K.T. et al. "Probability of Collision Error Analysis." *Space Debris*, Vol. 1, #1 (1999), pp. 21-35.

[12] Cohen, A.C. and Whitten, B.J. *Parameter Estimation in Reliability and Life Span Models*. New York: Marcel Dekker, Inc. 1988.

[13] Chan, F.K. *Spacecraft Collision Probability*. El Segundo, CA: The Aerospace Press, 2008.

[14] Alfano, S. "Collision Avoidance Maneuver Planning Tool." AAS/AIAA Astrodynamics Specialists Conference (Lake Tahoe, CA), Aug. 2005.