Setting the Bar for the Replacement of the Probability of Collision Metric

Matthew D. Hejduk$^{(1)}$, Daniel E. Snow$^{(2)}$

$^{(1)}$The Aerospace Corporation
Washington DC, USA
Matthew.d.hejduk@aero.org

$^{(2)}$Omitron Inc.
Colorado Springs, CO USA
Daniel.snow@omitron.com

***Abstract* – Recent research in satellite conjunction risk assessment has levelled a number of criticisms at the probability of collision (Pc) parameter as a durable statement of satellite collision likelihood, and a number of different alternatives to this calculation have been proposed. Many of these proposals, however, stop at the outlining of the theory and do not discuss the additional philosophical and practical issues that must be confronted in evaluating such proposals for adoption. The present work seeks to outline some of these philosophical and practical considerations, and therefore the kinds of analyses and profiling that will be needed before the Pc, which is at present nearly universally adopted in the conjunction risk mitigation community, can be replaced with a new (and superior) paradigm and associated metric.**

I. INTRODUCTION

In the early days of space surveillance operations, before precision orbit updates with associated covariances were widely available, satellite conjunction risk assessment was performed using only the miss distance between the two satellites at the time of closest approach (TCA) as the risk metric; and with only this criterion available, risk assessment was difficult. A given miss distance could be safe if the two satellites' uncertainties were very small but could be quite dangerous if the uncertainties were larger. With no insight into the actual orbital uncertainties, large miss distance thresholds were required in determining which events to mitigate, prompting large and frequent mitigation actions; and the "false alarm rate" of the enterprise (the frequency of mitigation actions for situations that, after the fact, were determined not actually to have presented a significant collision risk at all) was high enough that conjunction risk assessment was performed only in special circumstances.

The situation changed during the Shuttle program with the development of the probability of collision (Pc) as a collision risk metric [1]; this metric considered the two predicted orbits' uncertainties in its calculation of the likelihood that the relative miss vector would fall within the two objects' combined size. The virtues of this new approach were obvious; and once a precision space catalogue with associated covariances became available, the Pc became the standard conjunction risk assessment metric and is the recommended parameter in all of the current Conjunction Assessment (CA) best practices handbooks [2 3 4].

As early as 2005, however, carefully researched critiques of the Pc began to appear in print, with a number of such treatments published in the last decade. This development is indeed welcome as the CA industry matures, for a healthy public examination of risk assessment approaches and improvement recommendations should serve to move the discipline to, one hopes, both safer and less operationally invasive practices. At the same time, many of the researchers and commentators on this question do not have extensive operational experience with CA, and therefore there is a danger that approaches that may appear desirable *in vacuo* will not address the actual operational situation nearly so helpfully. The purpose of the present effort, therefore, is to offer, from the point of view of CA operations experts, a set of considerations that should be addressed before a different risk assessment parameter or practice can be meaningfully offered as a replacement for the Pc. The paper begins with a short discussion of astrodynamics calculation housekeeping issues; then addresses the question of the CA null hypothesis and how this contextualizes risk; then profiles the prevalence of certain types of Pc shortcomings historically to establish the scope of the problem, as well as examines the operational burden of certain known Pc alternatives; then finally discusses the need for linking the per-event CA risk assessment metric to a lifetime risk evaluation.

II. ASTRODYNAMICS HOUSEKEEPING

The astrodynamics calculations that stand behind the calculation of CA risk assessment metrics are not considered by seasoned flight dynamics engineers to be particularly complex. However, the applied statistics aspects of the orbital safety problem are attracting more and more professional statisticians to the discipline, and many of these individuals have little to no aerospace engineering background. Practitioners of flight dynamics well understand how easy it is to make errors in dimensional projections, coordinate system and frame transformations, and Monte Carlo realizations. To aid research of this type, the NASA Conjunction Assessment Risk Assessment (CARA) program has open-sourced a large amount of standard astrodynamics and CA risk assessment calculation software on their public repository (https://github.com/nasa/CARA_Analysis_Tools), and an initiative is currently underway to obtain release authorization for a larger set of tools and test data**.** This should allow researchers to compare their initial results for the standard astrodynamics calculations to established truth values and thus ensure that the astrodynamics basics of their work are operating correctly.

III. THE CA NULL HYPOTHESIS

It has been remarked that CA risk assessment as typically practiced bears not a small resemblance to traditional hypothesis testing: a test statistic is computed; the statistic is compared to a threshold, thus defining a critical region; and based on the results of that comparison, a default posture towards risk mitigation is accepted or rejected, which will lead to certain types of action (or non-action). This default posture is a null hypothesis of sorts, even if many operational CA practitioners employ it unwittingly; and while many CA evaluations proceed by simply comparing a test statistic to a threshold—a straightforward procedure for which a default posture would not seem to matter—there are ancillary considerations that both make a null hypothesis necessary and affect the decision of whether to employ one that is risk-adverse or risk-tolerant.

In seeking guidance on null hypothesis construction, there is surprisingly little extended treatment in the applied statistics discipline; most statistics texts do little more than provide brief (and almost casual) statements that do not address whether the null should be the standard assumption set, whether it should be what one sets out to "disprove," whether it should be positively or negatively formulated, &c. Nickerson [5] performed a large survey on research practice regarding the formulation of the null hypothesis and concluded both that there is no standard procedure in the industry, and that criticism can be levelled at any particular null hypothesis selection philosophy embraced. Choosing "what one expects to see" as the null prompts the question of why one is running the test in the first place (if indeed one expected not to see evidence of the alternative hypothesis)—a criticism perhaps applicable to experimental design but less so to CA, where the purpose of the test is to adjudicate each day's fresh set of individual conjunctions. Choosing the expected outcome if no direct evidence is offered, which seems quite similar to choosing the Bayesian *a priori*, must be evidence-based and not just the result of one's subjective orientation [6]; and in the case of CA, one would need to decide whether the CA screening run to identify candidates for CA risk assessment, which identifies a small(ish) set of conjunctions as potential problems and thus worthy of risk assessment, does change the character of the prior (*i.e.*, from no possibility of collision to some possibility of collision). One clear takeaway, however, is that the practitioner has broad latitude in choosing the null and should therefore tailor its choice to the particulars of the problem to which the associated statistical test will be applied.

There are also subtleties in "accepting" or "rejecting" the null hypothesis. To be precise, the statistical test either rejects the null or fails to reject it, since a test statistic that does not violate the threshold does not establish the null but merely establishes that it should not be rejected in favour of the alternative hypothesis. If one is facing a binary decision (such as the CA decision of whether or not to manoeuvre), then this distinction may appear irrelevant: whether the null is "accepted" or "not rejected," one is still led to the same particular action, namely whatever the null hypothesis counsels. There are, however, two sets of ancillary considerations for which this distinction does matter. First, there is the question of how to proceed if the test itself fails, meaning not so much that the test statistic cannot be calculated at all but that there is little faith in the information used to generate it. In CA it is not uncommon to have orbit determination outputs that are the result of misapplied data intervals, questionable tracking, and space weather prediction shortcomings, all of which conspire to add uncertainty to the orbit determination solution yet will not be fully reflected in its associated covariance and thus will affect the value of the collision likelihood test statistic. In such cases, the test statistic and inferences drawn from it would be unreliable; so those results, regardless of their message, could not constitute good evidence to reject the null hypothesis. The null hypothesis would thus need to govern these situations. Second, there is the issue of what is called "lethal non-trackable debris," which are debris objects large enough in a collision to render a satellite inoperable (greater than 1cm typically) but too small to be tracked by current tracking networks (smaller than ~10cm). As part of a NASA CARA study effort in 2019, the NASA Orbital Debris Program Office (ODPO) used their debris models to perform a series of Monte Carlo runs to establish a set of hypothetical LEO space catalogues containing objects down to 1cm in size, and NASA CARA selected from these a "medianish" catalogue that contained a total of 300,000 objects. Given that ~25,000 of those objects constituted presently-tracked objects in LEO, ~92% of the hypothesized objects with which a collision would be fatal to a satellite are not tracked and thus not available for CA, constituting a collision risk that must simply be accepted as part of operating a satellite in space. These objects can be considered to be satellites with a random state and an infinitely large covariance: they are there, but one does not know precisely where they are at any given moment or where they will be in the future. The null hypothesis and the action counselled when information on a conjunction is inadequate should not introduce a strong dissonance with the way the enterprise is forced to respond to the lethal non-trackable debris, which are undoubtedly producing far more conjunctions every day than are those objects in the maintained space catalogue of tracked objects.

Null hypothesis selection for conjunction assessment has been varied. Carpenter and Markley [7] felt that the question of “disproving” should govern the choice: “It is a general scientific practice to associate the null hypothesis with the condition that one is seeking to disprove”; presumably seeing CA as an enterprise to establish a certain level of safety, they embraced a null hypothesis of presuming a collision, for which the burden of proof would lie with establishing that the conjunction would result in a miss. It seems, however, that one desiring a more risk-tolerant null hypothesis could use the same justification to propose a null hypothesis of presuming a miss and holding that one would need to disprove that outcome in order to require a mitigation action, especially since this is nearly ubiquitously the actual outcome of satellite close approaches. Elkantassi and Davison [8] similarly elected to embrace a null hypothesis of presuming a collision. They remark that this approach is more consistent with how hypothesis testing generally proceeds, namely that the null hypothesis is rejected when the test statistic’s likelihood falls below the (p-value) threshold. They are certainly correct that this arrangement aligns better with standard statistical analysis practice, but it is really an aesthetic consideration rather than a necessary condition for the proper exercise of the test. They further point out that their formulation is more conservative because it counsels a mitigation action in situation in which there is any ambiguity, and therefore it is an appropriate posture for a safety application. While *prima facie* attractive, such a position has historically been detrimental to safety engineering because it tends to push conservativism into every data and decision juncture in the decision process, to the degree that one loses any sense of how conservative the final result becomes, especially if this conservative posture requires so many mitigation actions so as to have a regular deleterious effect on space mission performance. One should first ascertain what false alarm rates result from such an approach, and therefore what the effects on mission execution actually are, before concluding that embedded conservatism of this type is appropriate.

In addition, as stated above, the implications for failed tests and commensurability with the presence of lethal non-trackable debris must also be assessed. When poor OD results make a durable risk assessment impossible, is the expectation that mitigation actions would nonetheless be pursued in such cases, as a null hypothesis that presumed a collision would require? In those cases, what would be the basis for designing the associated mitigation actions and certifying that they would actually make the situation safer, given that the information at hand about the conjunction is suspect? Are not such situations similar to that encountered perpetually with lethal non-trackable debris (objects believed to exist but with essentially unknown conjunction details)? This last point would also seem to apply more generally to excessively conservative null hypotheses: they would foist mitigation on situations in which high risk is not clearly established but cannot be ruled out, resulting both in more and larger mitigation actions (because less well-determined situations require larger actions in order to ensure a safer outcome). This introduces a dissonance with the lethal non-trackable debris background risk: when one has little or unreliable information about a specific conjunction, the null hypothesis requires full and large mitigation actions to address it; when one has no information, then no mitigation activity can be performed. Should O/Os be asked to make Herculean efforts, which will interrupt satellite mission performance and reduce on-orbit lifetime with large fuel expenditures, for situations in which data are questionable and risks not clearly established, when the risk associated with 87% of the collisions that could render their spacecraft inoperable are simply accepted as a reality of operating a satellite in space, with no remediation actions possible?

NASA CARA has implemented a null hypothesis that presumes a miss, and this null hypothesis is rejected only if there is good evidence of a potential collision, which at present is defined as a Pc > 1E-04. Some might argue that a 1-in-10,000 chance of a collision is not sufficient evidence to reject the null hypothesis, and in that sense this more risk-tolerant null hypothesis still ends up being conservative; but this threshold has struck an acceptable balance between mitigating risk and not burdening missions with excessive mitigation actions. Using such a null hypothesis presents a cleaner solution when orbit determination data are questionable (*i.e.*, no good evidence to establish high collision risk and drive a mitigation action) and is more strongly in continuity with the necessary acceptance of risk from lethal non-trackable debris. But it should be reiterated, as was mentioned above, that the choice of the null hypothesis is contextualized most strongly by the missed detection rate (to the degree this can be established) and the false alarm rate associated with the CA risk calculation. A balance must be struck between keeping missed detections at an acceptable level while not allowing false alarm rates to impinge unduly on space operations. If such a balance can be struck with a conservative null hypothesis, then there is no enduring reason not to

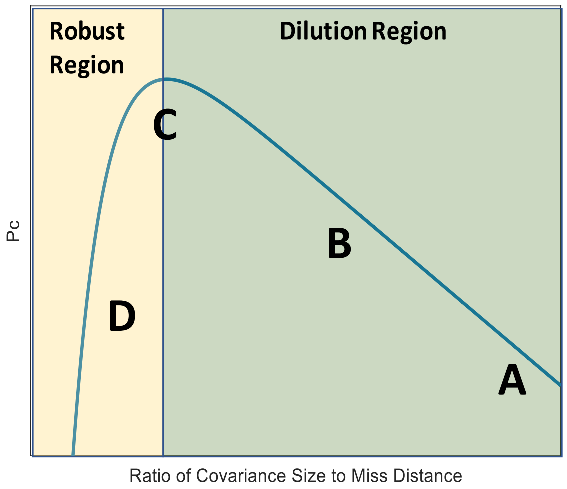

Fig. 1: Standard Dilution Region Curve

deploy it; but if it cannot, a more risk tolerant null hypothesis may be a helpful avenue to allowing a favourable balance to be reached and at the same time incorporating the ancillary aspects of the decision-making (*e.g.*, failed tests, lethal non-trackable debris) in a consistent way**.**

IV. MITIGATION RATES

A. The "Dilution Region" Phenomenon Explained

The concept of collision probability dilution entered the literature almost 20 years ago in a seminal paper by Alfano [9], in which he pointed out that there were two ways in which a collision probability could be low. The first, and by far the more desirable, is through the uncertainties on the two satellites' positions at the time of their closest approach (TCA) being very small. Small position uncertainties indicate that the estimate of the miss distance between the two satellites is a good predictor of the actual miss; and if this distance is appreciably larger than the combined sizes of the two satellites, one is strongly assured that the satellites will in fact miss each other, hence the low Pc. However, a second way in which the Pc can be low is if the uncertainties on the two satellites' positions are very large. In this case, the estimate of the miss between the two satellites is a poor predicter of the actual miss, which could take on a large number of values. The number of possible values that are smaller than the two satellites' combined size is therefore small compared to the full range of possibilities, thus resulting in a small Pc. This latter situation is, despite the low Pc value, still of operational concern because if more tracking data were available and the uncertainties smaller, it might be discovered that the situation is actually dangerous—with this danger masked by a Pc "diluted" by the large uncertainties.

The overall dilution region phenomenon is more fully illustrated by a dilution region curve shown in Fig. 1. The vertical axis is the calculated Pc value, and the horizontal axis is the ratio of the size of the combined covariance of the two satellites to the miss distance. When the combined covariance is very large, the Pc value is strongly diluted by these large uncertainties and takes on position A in this figure. As the covariance is shrunk (either by increasing the amount of tracking data or propagating less far into the future), the ratio of covariance size to miss distance shrinks, and the calculated Pc increases; this trend continues until this ratio reaches the value of $1/\sqrt{2}$ and the Pc reaches its peak value (position C). At this point, the conjunction moves from the diluted region to the "robust" region, for which the Pc now represents a durable statement of likelihood. As uncertainties are reduced in this region, the Pc value either rapidly drops off to a small value (position D) or, if the situation represents a true expected collision, remains high and in the idealized case pushes up to unity.

As implied in the above explanation, the graph can also be interpreted temporally as a statement of ideal conjunction evolution. Reading right to left (somewhat unfamiliarly, but this configuration matches Alfano's original publication), when a conjunction is first discovered typically seven to ten days in the future, Pc values are low because of the large covariances produced by such a long propagation into the future (position A). As the TCA comes closer and the objects are being updated, both fresh tracking and shorter propagations increase the Pc value (position B) until a peak is reached (position C), and then even more data and shorter propagations produce a precipitous drop in the Pc value (position D). This latter phenomenon is why it is advantageous for a satellite owner/operator (O/O) to delay the "manoeuvre commitment point" (MCP) as long as possible to try to take advantage of the expected Pc drop-off and thus avoid unnecessary conjunction risk mitigation actions. Of course, if the encounter truly is dangerous, the notable drop-off will not occur; and the O/O should perform a mitigation action.

Operationally, what should one do if at the MCP the conjunction of interest is in the dilution region? A number of different commentators believe that the dilution region phenomenon reveals a severe weakness of the Pc as a collision likelihood metric. Alfano, in the aforementioned paper [1], thought that when in the dilution region one should shrink the covariance systematically until the peak Pc value for the conjunction (position C in Fig. 1) is revealed, and one should use this as a proxy for the calculated Pc, since in principle a Pc calculated from estimates that used more tracking data could be as large as this. Balch *et al.* [10] called this dilution region behaviour an example of "false confidence" and proposed as a thought-experiment presuming the true miss distance to be zero

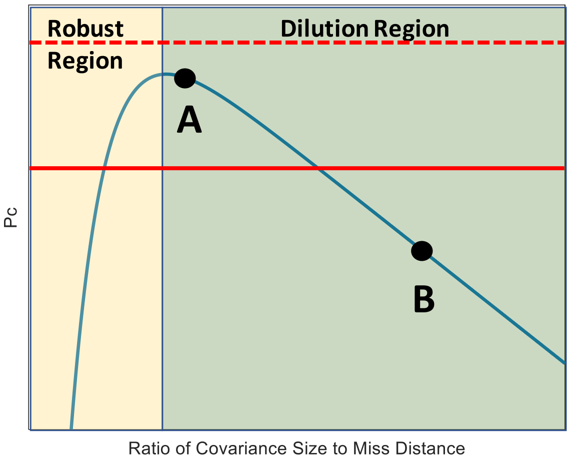

Fig. 2: Dilution Region Curve with Red Thresholds

and recalculating the Pc, noting that in many cases such a situation would not violate the usual Pc mitigation threshold of 1E-04: any methodology that does not counsel a mitigation action for situations of zero miss is problematic. As an initial proposal, he offered when in the dilution region instead using a very conservative technique of ensuring that both satellites' three-sigma covariance ellipsoids not intersect at TCA; if the predicted behaviour of the two satellites did result in an intersection, then a manoeuvre would be required in order to achieve a sufficiently large separation. Delande *et al.* [11] introduced a "collision plausibility" construct that essentially counselled a mitigation action whenever a collision was considered plausible; this has the practical effect of calling for mitigation actions whenever notably in the dilution region. Carpenter [12] and Elkantassi and Davison [8] both point out risk underreporting biases introduced by the use of the Pc and instead promote using confidence intervals on the miss distance, with the latter publication offering a full developed theory and some practical examples. All of these commentators inveigh against, in the dilution region certainly and in some cases ubiquitously, the uncritical use of the Pc as a collision likelihood metric.

B. Problem Severity and Mission Imposition

While the dilution region phenomenon is not the only issue related to the use of the Pc, it is the principal one; so in profiling how often issues with the Pc appear, it is prudent to begin with examining this dynamic. There are several different strains of Pc dilution, as illustrated in Fig. 2 and the associated commentary.

The first situation that can be encountered is represented by position A: the conjunction is formally in the dilution region, but just barely; there is hardly any difference between the peak Pc value and the calculated Pc value. If these values differ by 1/10th of an order of magnitude or less, then the conjunction is called *formally diluted*; while it meets the official definition for dilution, this status is not considered to have any operational import.

The second situation is represented by position B and a Pc mitigation threshold ("red" threshold) shown by the dotted red line. The conjunction is significantly in the dilution region, to be sure; but the peak Pc value is below the threshold for taking a mitigation action; so even with the worst possible combination of the current miss distance and any possible size of the combined covariance, the operational decision is the same whether working from the peak Pc value or the calculated value shown at B. This situation is called *significantly diluted* because the amount of dilution is appreciable, but it is not necessarily operationally problematic because it does not change the expected operational decision.

The third situation is represented by position B and a Pc mitigation threshold shown by the solid red line. In this situation, there is a difference in expected operational decision between the peak Pc, which would counsel a mitigation action; and position B, which would not. This situation is thus labelled *critically diluted* because it can visit a critical effect on the risk assessment process, namely changing the expected mitigation decision.

In the general application of risk analysis, risk is defined as the combination (usually the product) of event likelihood and event consequence [13]; the conjunction risk assessment community has been slow to embrace this construct, instead focusing nearly entirely on event likelihood, largely because O/Os understandably focus principally on the preservation of their own satellite mission, thus seeing a very high consequence in any possible collision. However, if one considers that the broader purpose of conjunction assessment is to preserve the usability of the space environment, then a second paradigm for consequence emerges, namely the amount of debris that the conjunction would produce were it to result in a collision. Techniques exist for estimating the number of trackable fragments that a satellite collision would produce [14]; and if this number exceeds a critical value, one can label it a high consequence collision and, if it were critically diluted, now label it *environmentally critically diluted*. This last category would be the most operationally concerning: the calculated Pc value is depressed, more data could well produce a Pc value well above the mitigation threshold, and a resulting collision could constitute a major debris production event.

Profiling of a large set of historical conjunction data messages (CDMs—the notification that gives the state and covariance of two conjuncting satellites at their TCA) to determine the prevalence of the different strains of dilution region presence can accomplish a number of objectives at once. First, it will show how frequently problematic dilution region states actually arise, both in an absolute sense and in comparison to situations in which the calculated Pc value alone is used to determine risk. Second, it will show the number of additional mitigation actions that the Alfano-suggested method of using the maximum Pc construct, which is the most widely used alternative to using the calculated Pc value alone, will introduce. Third, it will also serve as a reasonable proxy for the Delande *et al.* plausibility construct; while a fully-formed proposal for the regular calculation of this construct is still under development, it is believed that it will largely track the Alfano maximum Pc approach. As an aside, it is worthy of mention that the presence of lethal non-trackable debris presents special philosophical issues for collision plausibility. The presence of these debris objects make a collision plausible both before and after any trackable conjunction event; so following the plausibility construct strictly, there is no difference in the plausibility states before and after any discrete conjunction (because the lethal non-trackable debris forces at all times a "yes" to the question of whether a collision is plausible), so under a plausibility construct one can never formally justify a mitigation action for a discrete conjunction. Along with this dilution-region-based profiling, an additional calculation is included to determine which events secure a non-impinging of their three-sigma covariances, as suggested by Balch *et al.* [10] as a safer alternative to the Pc. For the purposes of this study, a proxy for the three-dimensional calculation was used, namely a determination whether the three-sigma combined covariance projected into the conjunction plane impinged on the Hard Body Radius circle (a circular area that represents the combined sizes of the two satellites. The miss-vector confidence interval approach of Elkantassi and Davison [8] requires additional interpretive apparatus to deploy and for this reason is left to a more detailed future study.

The dataset used for the profiling is the set of the CDMs received by NASA CARA from October 2018 through December 2023; this is a period of time that has included both changing space catalogue characteristics (as the number of active satellites has significantly increased) and a substantial temporal advance towards solar maximum in 2024 (which notably increases atmospheric drag). Two different perigee height bands for the primary satellites were used: 600-800 km, which represents a low-drag LEO environment; and 350-500 km, which is more heavily affected by drag and results in larger prediction uncertainties. Manoeuvre commitment points investigated run from 0.5 to 3 days prior to TCA, in half-day increments; the larger values can be used to assess mission impact for mitigation action planning, which usually must begin some notable period ahead of the point at which a mitigation action decision would actually be rendered. A trackable debris production threshold of 100 or more pieces > 5cm in size was used to determine which critically diluted conjunctions were considered environmentally critically diluted. Twelve primary objects are represented in the higher perigee height band and sixteen in the lower band; all of these primary objects were present in each of their bands for the entire data capture period. Results are normalized to show the number of each indicated event per spacecraft per year. The results of this profiling are given in Table 1.

Table 1: Event Rate Profiling Results

	Event Rates (Events/Spacecrft/Year), 600-800 km Perigee Height							
MCP (days)	Robust	Formally Diluted	Sig. Diluted	Crit. Diluted	Env. Crit. Diluted	Sum of Crit.	Pc Red at MCP	3-σ Ell. Violation
0.5	655.17	11.24	30.40	0.33	0.79	1.13	0.95	75.43
1	647.92	11.59	30.43	0.33	0.79	1.13	1.14	78.06
1.5	640.68	12.00	30.60	0.40	0.76	1.16	1.32	80.54
2	631.37	12.46	30.94	0.56	0.79	1.35	1.73	84.29
2.5	620.48	12.86	31.35	0.52	0.73	1.25	2.00	87.51
3	608.03	13.33	31.17	0.76	0.62	1.38	2.13	90.06

	Event Rates (Events/Spacecrft/Year), 350-500 km Perigee Height							
MCP (days)	Robust	Formally Diluted	Sig. Diluted	Crit. Diluted	Env. Crit. Diluted	Sum of Crit.	Pc Red at MCP	3-σ Ell. Violation
0.5	1178.49	27.08	106.60	0.75	3.13	3.88	8.06	247.98
1	1153.31	27.52	105.44	0.46	3.07	3.54	7.83	247.36
1.5	1124.61	26.86	104.81	0.43	2.88	3.31	7.73	246.77
2	1093.60	26.58	103.10	0.24	2.95	3.19	7.73	242.00
2.5	1055.69	25.65	99.73	0.25	2.88	3.13	7.25	236.30
3	1017.10	25.08	95.90	0.25	2.63	2.88	6.93	229.20

The most helpful way to describe the performance of different risk assessment paradigms is through the concepts of missed detection and false alarm rates, perhaps by constructing a true ROC (receiver operating characteristic) confusion matrix. Fortunately for space environment preservation but unfortunately for constructing such a confusion matrix, actual collisions almost never occur; so generating true missed detection rates is not possible, and true false alarm rates are not actually reconstructable either because mitigation actions obscure what the non-mitigated outcome of such situations would have been. Instead, because of the very low possibility of an actual collision, the number of red events assigned by the selected risk assessment methodology can be said to serve as a *de facto* false alarm rate; perhaps one could call it a mission imposition rate to make clear that it is not presuming to be an actual false alarm rate but does properly represent a serious impact to most satellite missions.

Given the above, it is most meaningful to look at the critically and environmentally critically diluted event rates, since these represent the main situations in which some sort of alternative risk assessment metric might be deployed and whose recommendations might be at variance with those drawn from the traditional Pc approach. The column titled "Sum of Crit." is the sum of both critically and environmentally critically diluted event rates; and it is placed conveniently next to the "Pc Red at MCP" column, which represents the number of events in which the Pc exceeds the standard "red" threshold of 1E-04 and would presumably counsel action under the current Pc-only paradigm.

For the 600-800 km perigee height band, at the lower MCPs there is approximately one event of critical dilution per payload per year, and similarly one red event by the traditional Pc rendering per payload per year. Adding an alternative risk assessment technique that tracked largely with the critical dilution region categories would in principle double the number of serious events; but given the low event density in the first place, that level of increase would not be considered a major operational perturbation. It is also a low enough density that it could probably be handled by exception processing: if the dilution region tests exposed the problem, then one could potentially stay within the Pc framework and address such situations as an adjunct to it. The ellipse overlap situation, however, is different; the number of mitigation events required by this risk assessment paradigm is substantially higher and in fact beyond the capabilities and/or desires of most O/Os who fly spacecraft in this orbit regime.

For the 350-500 km perigee height band, due to the greater difficulties in predicting orbits and therefore greater uncertainties, both the two types of critically diluted events and the traditional red Pc events manifest rates notably higher than those for the 600-800 km band. Despite the overall event increases, it is interesting that in this band the dilution region augmentation is smaller in a relative sense—only about half the event rate as that for the traditional red-level Pc approach. The absolute numbers are more problematic, but again the increase that would be asked here is probably not beyond what could be absorbed operationally, especially given that, in this regime, adjudicating risk mitigation manoeuvres by moving or slightly modifying station-keeping manoeuvres that need to be executed anyway is a common practice. Similar to what was observed for the 600-800 km band, a mitigation action rate based on a three-sigma ellipse overlap exclusion is beyond what would, or perhaps even could, be borne operationally.

The results shown here are simply examples for a pair of proposed Pc alternatives, with certain assumptions made about the implementing concept of operations. The purpose is not to embrace either of these specific proposals for, or exclude either from, operational deployment; rather, it is to show the kind of operational impact analysis that needs to be conducted before proposing any particular alternative to the Pc.

V. LIFETIME COLLISION RISK

Choosing a risk assessment approach different from the Pc is likely to decrease miss detections but increase false alarms; in the end, how might one know whether such a modification would be operationally desirable? The most meaningful criterion overall is the effect on a satellite's lifetime risk of collision, from the vantage point of either loss of mission or debris production; but such determinations are problematic because of a natural incommensurability between flux- and event-based calculation approaches. Debris-flux models are used to calculate the probability, over a given period, that a satellite will be struck by an object greater than a given size, but they give only a stochastic answer and do not include the actual operational response to CA events and the associated residual risk not fully mitigated by those operations. Results from the CA operations performed on a series of individual CA events can be amalgamated to give an overall estimate of composite risk [15]; but this is an estimate only of *perceived* risk, since the events in question have already occurred with no resultant collisions (therefore possessing an actual risk of zero), and this type of estimate changes substantially with different presumed MCPs.

A recent publication by Sweetser *et al.* [16] proposes an interesting methodology to bridge this incommensurability and give an answer of the type that the current investigation is seeking. Like Balch, the researchers here are interested in how often Pc-based risk assessment could produce missed detections. The usual CA problem decomposes the encounter into the conjunction plane, places the combined covariance on one end of the relative miss vector and the combined sizes of the two satellites on the other, and then evaluates the likelihood that the true relative miss vector might terminate within the structure representing the two objects' combined sizes. Sweetser *et al.* instead reverse the problem and, presuming a direct hit (actual miss distance of 0) and placing the combined covariance directly on top of the combined size structure, use it to determine all the possible miss vectors that the OD process could have determined, calculate the Pc that results from each, and determine the portion of these Pc values that lie above the red threshold. This proportion produces a probability of detection (Pd)—a probability that the Pc-based framework would, in the presence of a direct hit, counsel a mitigation action; a value of unity would indicate that the Pc-based methodology worked precisely as hoped (all such possible renderings of this collision would result in a Pc above the red threshold and thus manoeuvres); a low value would mean that, given the objects' sizes, the degree of orbit determination uncertainty, and the Pc red threshold, missed detections would take place rather frequently for the direct-hit case hypothesized. The reader is encouraged to read the full development of the methodology in [16] and note the surprisingly simple formula for calculating the Pd (equation 8).

Not only does this Pd give some sense of the diagnostic power of the Pc for certain conjunctions, the Pd can be used nicely in conjunction with flux-based modelling results to calculate lifetime risk. If a lifetime flux-based result for trackable-sized objects can be calculated, then the product of the complement of the Pd and the flux-based result will give the lifetime risk value with the particular CA approach that generated the Pd. For example, if on average a particular CA paradigm produced a Pd of 0.9 and a flux-based model result for a satellite's lifetime risk is 4E-04, then the lifetime risk in the presence of the CA approach is (1 – 0.9) * 4E-04, or 4E-05, an order of magnitude improvement. The Pd values resulting from different CA approaches can be compared directly to show relative improvement; and if a particular lifetime risk level is sought, each can be combined with flux-based results for the non-remediation case to determine which approaches yield acceptable lifetime risk values. Both sets of results can be set against the "mission disruption rate" from Section IV to adjudicate the needed balancing between collision risk reduction and space mission orbital safety burden.

VI. CONCLUSIONS

The conjunction risk assessment discipline is always seeking improved approaches, but in order to be usable by space operators, these approaches must be philosophically consistent with other aspects of the problem, not place an undue burden on satellite space operations, and ultimately improve the satellite collision risk posture in a demonstrable way. In discussing issues of astrodynamics validation, the choice of the CA null hypothesis, the importance of CDM profiling to determine mission disruption, and the need for evaluating effects on lifetime satellite collision risk, the authors hope to enable researchers to perform fully-formed evaluations of their proposed constructs and therefore propose improvements truly and immediately helpful to satellite conjunction assessment operations.

VII. REFERENCES

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

[2] Space Safety Coalition. "Best Practices for the Sustainability of Space Operations." Version 2.35, November 2023. https://spacesafety.org/wp-content/uploads/2023/12/SSC_Best_Practices_for_Space_Operations_Sustainability_v2.35.pdf.

[3] AIAA. "Satellite Orbital Safety Best Practices." September 2022. https://www.ascend.events/outcomes/satellite-orbital-safety-best-practices-by-iridium-oneweb-spacex-aiaa/.

[4] NASA. "NASA Spacecraft Conjunction Assessment and Collision Avoidance Best Practices Handbook. Version 2.0. https://www.nasa.gov/cara/.

[5] Nickerson, R.S. "Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy." *Psychological Methods*, Vol. 5 (2000), pp. 241-301.

[6] Sivia, D.S. and Skilling, J. *Data Analysis: A Bayesian Tutorial.* Oxford University Press, 2006.

[7] Carpenter, J.R. and Markey, F.L. "Wald Sequential Probability Ratio Test for Space Object Conjunction Assessment." *Journal of Guidance, Control, and Dynamics* Vol. 37 No. 5 (September-October 2014), pp. 1385-1396.

[8] Elkantassi, S. and Davison, A.C. "Space Oddity? A Statistical Formulation of Conjunction Assessment." *Journal of Guidance, Control, and Dynamics*, Vol. 45, No. 12 (December 2022), pp. 2258-2274.

[9] Alfano, S. "Relating Position Uncertainty to Maximum Conjunction Probability." *Journal of the Astronautical Sciences*, Vol. 53 No. 2 (April-June 2005), pp. 193-205.

[10] Balch, M.S., Martin, R., and Ferson, S. "Satellite Conjunction Analysis and the False Confidence Theorem." *Proceedings of the Royal Society A* (Mathematical, Physical, and Engineering Sciences), Vol. 475, No. 2227 (July 2019).

[11] Delande, E.D., Jones, B.A., and Jah, M. "Exploring an Alternative Approach to the Assessment of Collision Risk." Journal of Guidance, Control, and Dynamics 46 (2227; November 2022), PP. 1-16.

[12] Carpenter, J.R. "Covariance Realism is not Enough." 2019 AAS/AIAA Astrodynamics Specialist Conference (Paper # 19-605), Portland ME, August 2019.

[13] Kaplan, S. and Garrick, B. "On the Quantitative Definition of Risk." Risk Analysis, Vol. 1 No. 1 (1981), pp. 11-27.

[14] Hejduk, M.D., Laporte, F., Moury, M., Kelso, T.S., Newman, L.K., and Shepperd, R. "Consideration of Collision 'Consequence' in Satellite Conjunction Assessment and Risk Analysis." 26th International Symposium on Space Flight Dynamics, Matsuyama, Japan, June 2017.

[15] Hall, D.T. "Determining Appropriate Risk Remediation Thresholds from Empirical Conjunction Data using Survival Probability Methods." 2019 AAS/AIAA Astrodynamics Specialist Conference (Paper #19-631), Portland ME, August 2019.

[16] Sweetser, T.H., Braun, B.M., Acocella, M, and Vincent, M.A. "Quantitative Assessment of a Threshold for Risk Mitigation Actions. Journal of Space Safety Engineering 2020 https://doi.org/10.1016/j.jsse.2020.07.009.