

Research



Cite this article: Balch MS, Martin R, Ferson S. 2019 Satellite conjunction analysis and the false confidence theorem. *Proc. R. Soc. A* **475**: 20180565.
<http://dx.doi.org/10.1098/rspa.2018.0565>

Received: 21 March 2018

Accepted: 11 June 2019

Subject Areas:

statistics, space exploration,
mechanical engineering

Keywords:

belief functions, filter estimates, space
situational awareness, statistical theory and
methods

Author for correspondence:

Michael Scott Balch
e-mail: msb1@alexvalcon-llc.com

Satellite conjunction analysis and the false confidence theorem

Michael Scott Balch¹, Ryan Martin² and Scott Ferson³

¹Alexandria Validation Consulting, LLC, Alexandria, VA, USA

²Department of Statistics, North Carolina State University, Raleigh, NC, USA

³Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK

MSB, 0000-0003-4018-6471

Satellite conjunction analysis is the assessment of collision risk during a close encounter between a satellite and another object in orbit. A counterintuitive phenomenon has emerged in the conjunction analysis literature, namely, probability dilution, in which lower quality data paradoxically appear to reduce the risk of collision. We show that probability dilution is a symptom of a fundamental deficiency in probabilistic representations of statistical inference, in which there are propositions that will consistently be assigned a high degree of belief, regardless of whether or not they are true. We call this deficiency false confidence. In satellite conjunction analysis, it results in a severe and persistent underestimate of collision risk exposure. We introduce the Martin–Liu validity criterion as a benchmark by which to identify statistical methods that are free from false confidence. Such inferences will necessarily be non-probabilistic. In satellite conjunction analysis, we show that uncertainty ellipsoids satisfy the validity criterion. Performing collision avoidance manoeuvres based on ellipsoid overlap will ensure that collision risk is capped at the user-specified level. Furthermore, this investigation into satellite conjunction analysis provides a template for recognizing and resolving false confidence issues as they occur in other problems of statistical inference.

1. Introduction

A satellite conjunction is an event in which two satellites or a satellite and a piece of debris are estimated to pass near each other. The goal of satellite conjunction analysis

is to determine whether the risk of collision is high enough to necessitate a collision avoidance manoeuvre. Avoiding collision is not just important to the individual satellite operator. Everyone in the space industry has a stake in minimizing the number of in-orbit collisions. Each collision adds to the debris already in orbit, and in aggregate, this debris has the potential to make low Earth orbit un navigable [1–3]. So, while some of the issues involved in conjunction analysis may initially seem arcane, correctly resolving those issues is of existential importance to the space industry.

(a) A modern problem

Over the past 15 years, aerospace researchers have recognized a counterintuitive phenomenon in satellite conjunction analysis, known as *probability dilution* [4]. That is, as uncertainty in the satellite trajectories increases, the epistemic probability of collision eventually decreases. Since trajectory uncertainty is driven by errors in the tracking data, the seemingly absurd implication of probability dilution is that lower quality data reduce the risk of collision.

Several researchers have attempted to synthesize an *ad hoc* fix to probability dilution [4–9]. Most of them have defined an alternative collision risk metric that is increasing, or at least non-decreasing, as a function of trajectory uncertainty. Alternatively, some satellite operators have adopted a high sensitivity to small collision probabilities [10]. However, if probability dilution is indicative of a fundamental problem with collision probabilities, it is not immediately obvious that being sensitive to small collision probabilities will resolve that problem.

(b) A centuries-old argument

Satellite trajectory estimation is fundamentally a problem of statistical inference, and the confusion over probability dilution cuts to the heart of the long-standing Bayesian–frequentist debate in statistics; for a review, see [11]. Satellite orbits are inferred using radar data, optical data, GPS data, etc. which are subject to random measurement errors, i.e. noise. The resulting probability distributions used in conjunction analysis represent epistemic uncertainty, rather than aleatory variability. That is to say, it is the trajectory estimates that are subject to random variation, not the satellite trajectories themselves. In the Bayesian view, that is a distinction without a difference; it is considered natural and correct to assess the probability of an event of interest, such as a possible collision between two satellites, based on an epistemic probability distribution [12–14]. In the frequentist view, however, it is considered an anathema to try to compute the probability of a non-random event [15–17]. This prohibition is expressed in some corners of the uncertainty quantification community as the requirement that epistemic uncertainty be represented using non-probabilistic mathematics [18–20]. Despite these objections, at present, satellite navigators appear to be adopting the Bayesian position.

(c) Definitions

An *aleatory uncertainty* is an uncertainty due to random (i.e. aleatory) variation of the quantity or event being analysed. An *epistemic uncertainty* is uncertainty due to imperfect knowledge about the quantity or event being analysed. One of the many ways in which epistemic uncertainty can arise in a fixed quantity is if the data used to measure or infer it are subject to random errors.

A *belief function* is a function used to represent the amount of positive support a proposition has accrued from the evidence. A *plausibility function* merely represents the lack of evidence against a proposition. Belief and plausibility are complementary functions; that is, $\text{Pls}(A) = 1 - \text{Bel}(\neg A)$, for every proposition, A , and its complement (i.e. negation) $\neg A$. As a rule, the plausibility accorded to a proposition is always greater than or equal to the belief accorded to it. An additive belief function is a belief function that satisfies the Kolmogorov axioms and in which, as a consequence, $\text{Bel}(A) = \text{Pls}(A)$ for all propositions [21].

An *epistemic probability* is an additive belief function used to represent epistemic uncertainty. Confidence distributions,¹ fiducial distributions, and most Bayesian posterior distributions² are epistemic probability distributions. An *aleatory probability* is a probability used to represent aleatory uncertainty. It is equal to the fraction of trials in which a specified event occurs—or would occur—over an infinite number of equivalent random trials. The mathematics of aleatory and epistemic probability functions are identical; the distinction is a question of subject matter.

(d) Our contribution

In this paper, we pursue basic questions raised by probability dilution. Can satellite operators take epistemic probability at face value as a risk metric? If not, why not? What alternatives are safe for use in conjunction analysis? How can we know that those alternatives are safe? We address these questions, in part, by placing conjunction analysis in its context as a problem of statistical inference. Consequently, what we discover about probability dilution has broad implications for statisticians as well as satellite operators.

In §2, we provide a thorough exposition of probability dilution. We show that the counterintuitiveness of probability dilution is entirely due to the fact that, in conjunction analysis, probability of collision is an epistemic probability, rather than an aleatory probability. Moreover, we show that probability dilution is linked to a more severe operational defect that we call false confidence. We then evaluate the efficacy of pursuing small epistemic probabilities as a way to mitigate the problems posed by false confidence.

In §3, we show that false confidence is not limited to conjunction analysis. All epistemic probability distributions are subject to false confidence. That is, in any additive belief function used to represent epistemic uncertainty, there are false propositions that have a high probability of being assigned a high belief value. The false confidence generated by epistemic probability of collision in satellite conjunction analysis is a manifestation of this more general deficiency.

In §4, we demonstrate a framework for identifying statistical methods that are free from false confidence. We build this approach on the ‘validity criterion’ introduced in the Martin–Liu Theory of Inferential Models [26]. For a statistical inference to be valid in the Martin–Liu sense, only a true proposition may have a high probability of being assigned a high belief value. An inference satisfying this criterion will not suffer from the false confidence phenomenon described in §3. Note that the reader does not need to be familiar with Martin–Liu theory to understand the work presented in this paper. Only the validity criterion is used, and its formal definition is provided in §4.

2. Epistemic probability of collision

Epistemic probability of collision is coming into more common use as a risk metric in satellite conjunction analysis [10,27,28]. In fact, the most-cited literature on satellite conjunction analysis simply takes it for granted that epistemic probability of collision is the correct way to represent collision risk, e.g. [29–31]. However, the counterintuitiveness of probability dilution calls this practice into question, especially considering the unsettled status of epistemic probability in the statistics and uncertainty quantification communities.

This part of the paper explores issues surrounding the use of epistemic probability of collision as a risk metric for satellite conjunction analysis. Section 2a reviews the standard computation of collision probability. Section 2b describes how probability dilution arises from those mathematics. Section 2c describes a more severe operational defect in epistemic collision probabilities: false confidence. Section 2d explores the use of small thresholds for epistemic probability as a way to counteract false confidence.

¹We refer to confidence distributions in the sense of [22], as opposed to [23,24], which eschew epistemic probabilities.

²One could argue that a posterior probability derived using an aleatory prior, as in [25], is itself an aleatory probability.

(a) Computation of collision probability

As mentioned in §1b, satellite conjunction analysis is rooted in a problem of statistical inference. The inferred parameter of interest is the vector of positions and velocities of the two satellites at time of closest approach. That is,

$$\theta = (u_1, v_1, w_1, \dot{u}_1, \dot{v}_1, \dot{w}_1, u_2, v_2, w_2, \dot{u}_2, \dot{v}_2, \dot{w}_2)^T,$$

where u_1, v_1, w_1 are the position of one satellite; $\dot{u}_1, \dot{v}_1, \dot{w}_1$ are its velocity; u_2, v_2, w_2 are the position of the other satellite; $\dot{u}_2, \dot{v}_2, \dot{w}_2$ are its velocity. The true value of θ is unknown; it is estimated using tracking data, x , via an inferential algorithm called a ‘filter’ that delivers an estimate, $\hat{\theta}(x)$, along with its uncertainty, expressed as a 12×12 covariance matrix, C_θ . In this paper, uncertainty in θ is attributed to random errors in the tracking data. In reality, uncertainty also arises from errors in the dynamics model used in the filter.

Uncertainty in the trajectory estimates is usually assumed to have a multivariate normal distribution [4,29,30]. The epistemic probability density for θ is, therefore, taken to be

$$f_{\theta|x}(\theta) = \{(2\pi)^{12} \det(C_\theta)\}^{-1/2} \exp \left\{ -\frac{1}{2} (\hat{\theta}(x) - \theta)^T C_\theta^{-1} (\hat{\theta}(x) - \theta) \right\},$$

where $\det(C_\theta)$ is the determinant of C_θ . This epistemic probability distribution for θ can be rationalized in two ways. One way is to treat filter estimation as an implementation of Bayesian inference [32]. The other way is to describe this representation as a fiducial distribution for θ , in the sense of [33], or a confidence distribution for θ , in the sense of [22], on the argument that a random resampled $\hat{\theta}(X)$ would have a multivariate normal distribution with mean θ and covariance matrix C_θ , where X represents a random realization of the tracking data, drawn from the same distribution as x , the tracking data actually obtained. The opening of §4b briefly describes the conditions under which the normality assumption will and will not hold.

Assuming normality holds, the standard epistemic probability of collision calculation proceeds in three steps. First, uncertainty in θ is propagated to uncertainty in the relative offset between the two satellites at closest approach. That is,

$$\Delta u = u_2 - u_1, \quad \Delta v = v_2 - v_1 \quad \text{and} \quad \Delta w = w_2 - w_1.$$

Since the linear transform of a multivariate normal variable is itself a multivariate normal, the triple $(\Delta u, \Delta v, \Delta w)$ has a multivariate normal distribution with the following 3×3 covariance matrix:

$$C_\Delta = A C_\theta A^T \quad \text{where } A = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 \end{bmatrix}$$

or more compactly

$$C_\Delta = C_{\theta;1:3,1:3} + C_{\theta;7:9,7:9} - 2C_{\theta;1:3,7:9},$$

where $C_{\theta;m_1:m_2,n_1:n_2}$ is the sub-matrix of all entries in rows i and columns j such that $m_1 \leq i \leq m_2$ and $n_1 \leq j \leq n_2$. That is, C_Δ is the sum of the covariance matrices for the positions of the two satellites minus the covariance between the positions of the two satellites.

In the second step, uncertainty along the axis parallel to the relative velocity vector is integrated out. That is, instead of computing probability of collision as the probability of the true value of the displacement falling within the set of $(\Delta u, \Delta v, \Delta w)$ values indicative of collision, probability of collision is computed as the probability accorded to an extruded ellipse containing that three-dimensional set. The theoretical rationale behind this move is that it captures uncertainty in the timing of closest approach [30]. The primary practical effect of this step is to reduce a three-dimensional integration problem to a two-dimensional integration problem. More nuanced ways of treating uncertainty in the timing of closest approach are available [34,35], but the standard two-dimensional treatment is sufficient for our analysis.

This transformation is executed in a few sub-steps. It starts with the unit vector in the relative velocity direction, $i_{\Delta V}$, defined as

$$\Delta V = (\dot{u}_2, \dot{v}_2, \dot{w}_2) - (\dot{u}_1, \dot{v}_1, \dot{w}_1) \quad \text{and} \quad i_{\Delta V} = \Delta V / \|\Delta V\|.$$

Next, another direction in the plane perpendicular to $i_{\Delta V}$ can be defined arbitrarily. For example, it could be taken as in the same direction as the cross-product of $i_{\Delta V}$ and $(\Delta u, \Delta v, \Delta w)$, assuming that they are not aligned. Denote this arbitrary direction vector as $i_{u'}$. The final direction vector is simply the cross-product, $i_{v'} = i_{\Delta V} \times i_{u'}$. Once this system has been established, define the rotated displacement vector as

$$\begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = M \begin{bmatrix} \Delta u \\ \Delta v \\ \Delta w \end{bmatrix} \quad \text{where } M = \begin{bmatrix} i_{u'}^\top \\ i_{v'}^\top \\ i_{\Delta V}^\top \end{bmatrix}$$

is the rotation matrix. The covariance matrix for the rotated (u', v', w') vector is $C_{\Delta'} = M C_{\Delta} M^\top$. Since (u', v', w') has a multivariate normal distribution, uncertainty in the ΔV direction can be integrated out by simply dropping w' . That completes the extrusion step, reducing the problem to the two-dimensional (u', v') space, with estimate (\hat{u}', \hat{v}') and covariance matrix $C_{\Delta';1:2,1:2}$.

The third and final step in the standard calculation of collision probability is to define the probability of collision as the epistemic probability of having u' and v' such that $u'^2 + v'^2 \leq R^2$, where R is the sum of the characteristic radii of the two satellites in the conjunction event. In other words, the satellite shapes are approximated as spherical objects. Collision occurs if and only if, at closest approach, the distance between the two satellites is less than their combined size.

This probability integral can be re-expressed in a standardized probability space, by performing an eigendecomposition of the displacement covariance matrix [8], which yields

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} \hat{u}' \\ \hat{v}' \end{bmatrix} - E_S \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix},$$

where S_1^2 and S_2^2 are the eigenvalues of $C_{\Delta';1:2,1:2}$; E_S is a 2×2 matrix whose columns are the normalized eigenvectors of $C_{\Delta';1:2,1:2}$; and finally, ξ_1 and ξ_2 are independent unit normal variables, with mean zero and variance one. Note that, since E_S is the eigendecomposition of a symmetric matrix, $E_S^\top E_S$ returns the identity matrix. Defining u'' and v'' as the E_S^\top rotation of u' and v' , the magnitude of (u'', v'') is equal to the magnitude of (u', v') . That is,

$$\begin{bmatrix} u'' \\ v'' \end{bmatrix} = E_S^\top \begin{bmatrix} u' \\ v' \end{bmatrix} \implies u''^2 + v''^2 = u'^2 + v'^2, \quad \text{where } \begin{cases} u'' = \hat{u}' - S_1 \xi_1 \\ v'' = \hat{v}' - S_2 \xi_2. \end{cases}$$

Thus, the following equivalences hold for the integration condition:

$$u'^2 + v'^2 \leq R^2 \iff u''^2 + v''^2 \leq R^2 \iff (S_1 \xi_1 - \hat{u}'')^2 + (S_2 \xi_2 - \hat{v}'')^2 \leq R^2.$$

What has been accomplished is that now, instead of integrating a circle in a (u', v') space that will vary from problem to problem, we are integrating an ellipse in a standardized unit normal (ξ_1, ξ_2) space. The centroid of that ellipse is located at $(\hat{u}''/S_1, \hat{v}''/S_2)$, and its semi-major and semi-minor axes are R/S_1 and R/S_2 .

Following the logic of [36], the epistemic probability of collision is computed using a contour integral transformation, as follows:

$$\begin{aligned} \text{Bel}(\mathbb{C}) &= \iint_{(S_1 \xi_1 - \hat{u}'')^2 + (S_2 \xi_2 - \hat{v}'')^2 \leq R^2} \frac{1}{2\pi} \exp \left[-\frac{1}{2}(\xi_1^2 + \xi_2^2) \right] d\xi_1 d\xi_2 \\ &= \int_0^{2\pi} \left(\frac{1 - e^{-r_\psi^2/2}}{2\pi r_\psi^2} \right) \left(\frac{R^2}{S_1 S_2} + \frac{\hat{u}'' R}{S_1 S_2} \cos \psi + \frac{\hat{v}'' R}{S_1 S_2} \sin \psi \right) d\psi, \end{aligned}$$

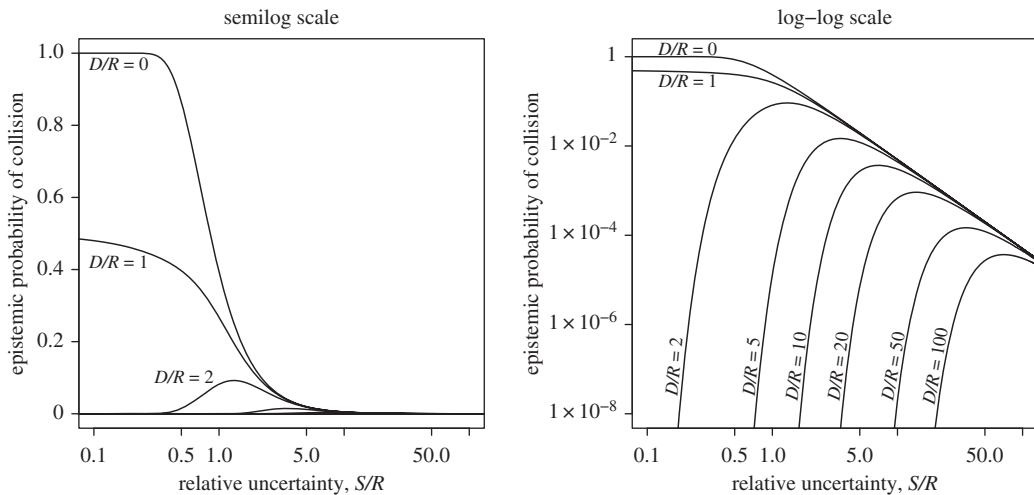


Figure 1. Epistemic probability of collision as a function of estimated relative displacement, D/R , and relative uncertainty, S/R , for the special case $S = S_1 = S_2$.

where

$$r_{\psi}^2 = \left(\frac{\hat{u}''}{S_1} + \frac{R}{S_1} \cos \psi \right)^2 + \left(\frac{\hat{v}''}{S_2} + \frac{R}{S_2} \sin \psi \right)^2.$$

This integral can be approximated to negligible error using the trapezoidal rule, provided that at least $10 \times \max(S_1/S_2, S_2/S_1)$ evenly spaced quadrature points are used [8].

Because our goal is to understand and explore how the epistemic probability of collision changes for different levels of uncertainty, we pursue the special case where $S = S_1 = S_2$. Defining the estimated displacement as $D^2 = \hat{u}''^2 + \hat{v}''^2 = \hat{u}'^2 + \hat{v}'^2$, this simplification reduces the epistemic probability of collision to a function of two ratios: estimated relative displacement, D/R , and relative uncertainty, S/R .

(b) Probability dilution

Figure 1 illustrates epistemic probability of collision as a function of S/R for several values of D/R . These curves follow a common pattern. So long as $D/R > 1$, for small uncertainties, the probability of collision is small. That much makes intuitive sense. If the satellites are estimated to miss each other, and the uncertainty in those trajectory estimates is small, then it is natural that the analyst have a high confidence that the satellites are not going to collide. Similarly, it makes sense that as uncertainty grows, the risk of collision also grows. However, something odd happens in the curves in figure 1. Epistemic probability of collision eventually hits a maximum, and past that maximum, as relative uncertainty rises, the epistemic probability of collision *decreases*.

This decrease is called probability dilution, and it has an odd implication. Since the uncertainty in the estimates of the trajectories reflects the limits of data quality, probability dilution seems to imply that lowering data quality makes satellites safer. That implication is counterintuitive in the extreme [4–8]. As a rule, lowering the data quality makes any engineering system less safe, and to claim that ignorance somehow reduces collision risk seems foolish on its face.

That being said, the mathematics of probability dilution are ironclad and relatively simple. Recall that, when expressed in unit normal space, the integration region determining the probability of collision is an ellipse whose centroid is at $(\hat{u}''/S_1, \hat{v}''/S_2)$ and whose semi-major and semi-minor axes are R/S_1 and R/S_2 . Figure 2 illustrates how the integration region changes as a function of S , when $S = S_1 = S_2$. Combined satellite size, R , and estimated distance at closest approach, D , are held fixed with $D/R = 5$. For four different values of S/R , the integration region is illustrated, along with the corresponding probability of collision.

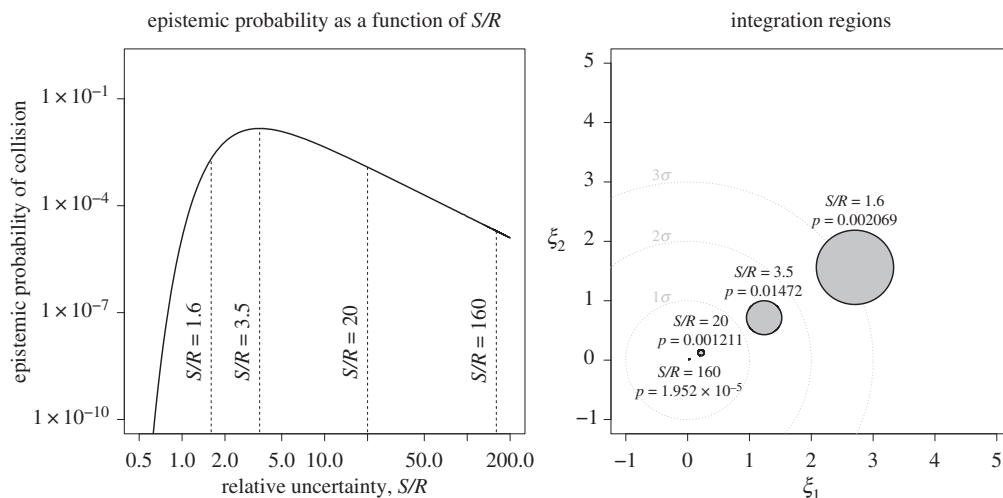


Figure 2. Epistemic probability of collision and integration region, given $D/R = 5$ and $S/R \in \{1.6, 3.5, 20, 160\}$. In the right panel, the 1σ , 2σ and 3σ contours for unit normal space are marked with dashed lines for reference. These three contours capture 39.3, 86.5 and 98.9% of the two-dimensional unit normal probability mass, respectively.

Increasing the level of uncertainty has two effects on the integration region: It shrinks and draws closer to the origin. The probability density of a unit normal space is highest at the origin, decreasing exponentially as a function of the square of distance from the origin. If $D/R > 1$, the positive effect due to getting closer to the origin, where there is more probability mass, initially outweighs the negative effect due to shrinking the integration region. Near the origin, though, the probability density curve flattens out. So, after a certain point, there is not much more probability mass to be gained by moving closer to the origin, and the negative effect of shrinking the integration region overtakes the positive effect due to shifting the location of the integration region. Past that point, probability of collision decreases as uncertainty increases. So, in total, probability dilution is due to a simple and straightforward shrinkage of the integration region relative to displacement uncertainty.

Even if the normality and shape assumptions were relaxed, the integration region would still undergo the same shrinkage phenomenon that is illustrated in figure 2. Each satellite represents a bounded set of points. The displacement values indicative of collision are therefore also a bounded set of points. If one were to relax the assumptions outlined in §2a, instead of an ellipse, one would have a different shrinking shape in a standardized three-dimensional probability space. Probability of collision would still have the same qualitative behaviour as a function of relative uncertainty. Mathematically speaking, the key factor underpinning probability dilution is that the integration region (i.e. failure domain) is bounded.

There is no way of reframing satellite conjunction analysis so that the failure domain is not bounded. Therefore, if we take the view that epistemic probability is valid, then it would seem that probability dilution is fundamental to conjunction analysis. Why, then, do aerospace researchers [4–8] find this supposed mathematical inevitability so counterintuitive?

Interestingly, were the uncertainty in the satellite trajectories aleatory, rather than epistemic, probability dilution would actually make sense. For example, suppose two satellites were known, with certainty, to be on a collision course, and the satellite operator could only impart an impulse of random magnitude in a random direction, say, via a poorly controlled manoeuvring thruster on a tumbling satellite. If the mean of this distribution were the null vector, due to a lack of control over thrust direction, then the higher the variance of the added impulse, the bigger the resulting perturbation and hence the smaller the resulting probability of collision. In this example, the mean of the resulting trajectory distribution would still have the satellites on a collision trajectory, but

if one applies a big enough impulse to one of the satellites in a random direction, that poorly controlled collision avoidance manoeuvre still has a high probability of success. In that context, higher variance in a satellite's trajectory really does reduce the risk of collision.

However, in this hypothetical example, it is a variance in the trajectory itself that makes the satellite safer, not a variance in the estimate of that trajectory. It should go without saying that, given two satellites on a sure collision trajectory, simply recomputing the trajectories with lower quality data does not make them safer. There is, therefore, an operational distinction to be made between genuine aleatory variability and epistemic uncertainty.

To summarize, the problem with probability dilution is not the mathematics. The mathematics are incontrovertible, and they even make sense in the right context. Probability dilution is only counterintuitive when orbit 'variance' reflects the limits of data quality, rather than genuine aleatory variability in the orbits themselves. So, if probability dilution is inappropriate, that inappropriateness must be rooted in a mismatch between the mathematics of probability theory and the epistemic uncertainty to which they are applied in conjunction analysis.

(c) False confidence

The final clue indicating the inappropriateness of probability dilution is the fact that, for a fixed S/R ratio, there is a maximum computable epistemic probability of collision. Whether or not the two satellites are on a collision course, no matter what the data indicate, the analyst will have a minimum confidence that the two satellites will not collide. That minimum confidence is determined purely by the data quality.

For all S/R , the maximum collision probability is obtained when $D/R = 0$; that is, when the best estimate of the two satellite paths indicates that they are on a collision course. Therefore, the curve for $D/R = 0$ in figure 1 yields the maximum computable probability of collision as a function of S/R . For example, if the uncertainty in the distance between two satellites at closest approach is ten times the combined size of the two satellites, the analyst will *always* compute at least a 99.5% confidence that the satellites are safe, even if, in reality, they are not.

A relative uncertainty of ten may sound high, but it may help the reader to keep in mind the magnitude of the trajectories being estimated, relative to the size of the objects on those trajectories. Satellite trajectories are measured in thousands of kilometres; satellites are measured in metres. Trajectory uncertainties in conjunction analysis are usually measured in hundreds of metres [37]. For conjunction analysis done more than a week in advance, those uncertainties can grow to kilometres [38]. As a consequence, S/R ratios greater than 10 are the rule, not the exception. Under those circumstances, even if two satellites are on a collision course, because of probability dilution, the epistemic probability of collision is guaranteed to be small.

The biggest problem with probability dilution is not its counterintuitiveness; instead, it is the *false confidence* that it imbues. Due to probability dilution, satellite navigators using epistemic probability of collision as their risk metric will always compute a high confidence that their satellites are safe, regardless of whether or not they really are. The only rare exceptions are conjunctions in which both satellites are extremely well-tracked and at least one of them is relatively large (for an example, see [39]). From an operational perspective, if taken at face value, epistemic probability of collision is a misleading risk metric.

(d) Pursuing small epistemic probabilities

Some satellite navigators treat relatively small epistemic probabilities of collision as though they reflect a high level of risk. So, it would be inaccurate to characterize the entire satellite operations community as taking epistemic probabilities at face value. For example, the CARA group based at NASA Goddard monitors all conjunctions with a computed probability of collision greater than 10^{-7} , and conjunctions with a computed probability of collision greater than 4.4×10^{-4} are treated as high risk. These thresholds are calibrated in an effort to achieve desired frequentist

error rates, in essence treating epistemic probability of collision as a test statistic, using historical data from past conjunctions as a reference population [10].

Unfortunately, even the idea of compensating for false confidence by setting a low threshold for epistemic probability of collision is problematic, because relative data quality modulates the relationship between epistemic probability of collision and the real aleatory probability of failing to detect an impending collision. To model this relationship, we consider the aleatory variability of predicted distance, D , between two satellites at closest approach, if the data used to produce that prediction were redrawn from the same distribution as the actual data obtained. Assuming that variability in the estimated direction of relative velocity is small enough to be neglected, the marginalization and eigendecomposition steps can proceed as in §2a. This yields

$$D^2 = \hat{U}''^2 + \hat{V}''^2 = (u'' + S_1\xi_1)^2 + (v'' + S_2\xi_2)^2,$$

where (u'', v'') is the true unknown two-dimensional relative displacement between the two satellites; \hat{U}'' and \hat{V}'' are the random resampled analogue to \hat{u}'' and \hat{v}'' from §2a; S_1 and S_2 are the principal standard deviations of the marginalized two-dimensional covariance matrix, as in §2a; and ξ_1 and ξ_2 are independent unit normal random variables.

In the case where $S_1 = S_2 = S$, this simplifies to

$$\left(\frac{D}{S}\right)^2 = \left(\frac{u''}{S} + \xi_1\right)^2 + \left(\frac{v''}{S} + \xi_2\right)^2.$$

This can be further reduced to $(D/S)^2 = \chi_{2, (D_T/S)^2}^2$, where $\chi_{j, \delta}^2$ is a non-central χ^2 random variable with j degrees of freedom and non-centrality parameter δ , and $D_T^2 = u''^2 + v''^2$ is the unknown true distance between the two satellites at closest approach.

In §2a, we showed how epistemic probability of collision can be computed as a function of D/S and S/R . Given a fixed S/R , epistemic probability of collision is a decreasing function of D/S . So, given a fixed detection threshold for epistemic probability, the aleatory probability of detecting an impending collision is equal to the probability of obtaining an observed D/S that is less than or equal to the critical D/S that yields an epistemic probability of collision equal to the threshold value. Rather than try to compute this relationship directly, it is easier to simply generate D/S values from its non-central χ^2 distribution and then compute the corresponding epistemic probabilities of collision. This yields epistemic probability threshold as a function of desired detection rate, which can be reversed to give the aleatory probability of failing to detect an impending collision as a function of a satellite operator's chosen detection threshold for epistemic probability of collision.

Figure 3 illustrates this relationship for several values of S/R in two scenarios, an impending head-on collision with $D_T = 0$ and an impending glancing collision with $D_T = R$. The two dotted vertical lines mark the lower and upper thresholds for epistemic probability of collision reported in [10]. Conjunctions with an epistemic probability of collision less than 10^{-7} are treated as lowest risk and ignored. Conjunctions with an epistemic probability of collision greater than 4.4×10^{-4} are treated as high risk events. Using these thresholds, if $S/R = 10$, an impending head-on collision has a 91.2% chance of correctly being identified as a high risk conjunction. Similarly, a glancing collision with $S/R = 10$ has a 91.1% chance of being correctly identified as high risk. So, broadly speaking, one could say that the CARA threshold has reasonably good performance for $S/R = 10$. Unfortunately, that performance degrades quickly as uncertainty increases. At $S/R = 20$, the aleatory probability of correctly identifying an impending collision as a high risk conjunction drops to 64.8%. For $S/R > 33.74$, the aleatory probability of detecting an impending collision using the 4.4×10^{-4} threshold drops to zero. For example, in a conjunction between two objects with a combined radius of 5 m, a satellite navigator using the CARA upper-risk threshold will have zero chance of detecting an impending collision unless or until their predicted displacement uncertainty drops below 170 m.

In summary, trying to compensate for false confidence by pursuing small epistemic probabilities of collision appears to be an unproductive strategy. Even small epistemic

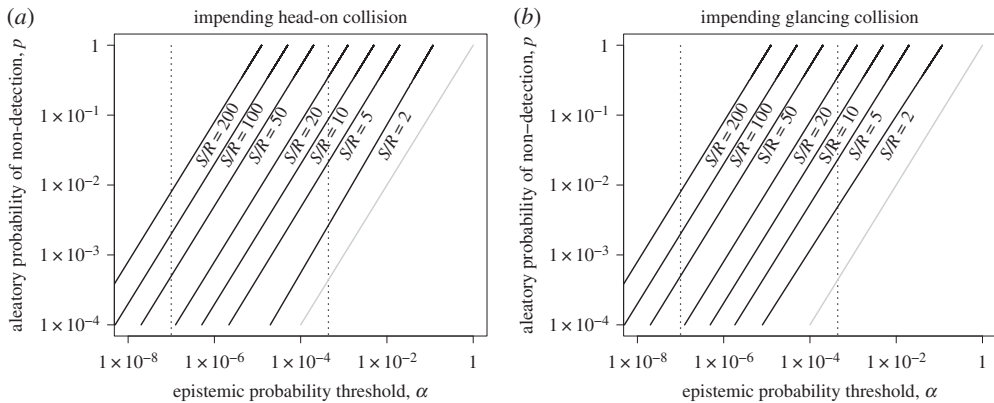


Figure 3. Aleatory probability of failing to detect an impending collision as a function of epistemic probability threshold for $S/R \in \{2, 5, 10, 20, 50, 100, 200\}$. (a) The error rate curves for a head-on collision, $D_T = 0$. (b) The error rate curves for a glancing collision, $D_T = R$. Equivalently, these plots yield the aleatory probability, p , of obtaining data such that non-collision will be assigned an epistemic probability of at least $1 - \alpha$, given the underlying reality that a collision is actually going to occur. The vertical dotted lines correspond to the CARA lower and upper thresholds reported in [10] as 10^{-7} and 4.4×10^{-4} . The grey line in each panel demarcates equality between the nominal epistemic probability threshold and the real aleatory probability of failing to detect an impending collision.

probabilities end up distorted, and they still result in false confidence under conditions that satellite navigators regularly encounter. Furthermore, the relationship between data quality and false confidence is more complicated than we have depicted in this idealized analysis. It depends not only on S/R but also on S_1/S_2 , and in reality, S_1 and S_2 are never equal. Every conjunction analysis has a unique combination of covariance matrices. So, if an analyst wants to hold the risk of failed detection at a consistent level, it is necessary to define a different epistemic probability threshold for each and every conjunction analysis. In other words, epistemic probability of collision could, theoretically, be used as a test statistic, but it cannot be taken directly as a risk metric, even if satellite navigators are sensitive to small epistemic probabilities of collision. There is no single threshold for epistemic probability that will detect impending collisions with a consistent degree of statistical reliability.

3. The false confidence theorem

What causes probability dilution and false confidence? As explored in §2b, it is not an error in the mathematics. Rather, it appears to be a mismatch between the mathematics of probability theory and the subject matter to which those mathematics are applied in satellite conjunction analysis. As mentioned in §1b, questions over the appropriateness of epistemic probability have never been settled. As shown in §2d, those questions take on an undeniably practical dimension in satellite conjunction analysis.

This part of the paper introduces the false confidence theorem. Section 3a suggests that probability dilution and false confidence are rooted in the axioms of probability theory. Section 3b posits a formal definition for false confidence. Section 3c proves that all epistemic probability distributions assign copious amounts of false confidence. Section 3d places this theorem in the context of the Bayesian–frequentist debate.

(a) Motivating insight

Given highly uncertain data, it makes intuitive sense that one would not be certain of collision. An analyst should be reluctant to assign a high degree of belief to any proposition, based on

Don't quite get this. In binary classification you would just assign

low-quality data. The problem is that, within the confines of probability theory, general non-commitment of belief is not an option. One of the axiomatic properties of probability functions is that they are additive. So, any belief not assigned to 'collision' is automatically assigned to 'not-collision', i.e. $\text{Bel}(\neg C) = 1 - \text{Bel}(C)$. While a low assignment of belief to 'collision' might seem natural when uncertainty is high, the consequently high assignment of belief to 'not-collision' seems unnatural. The implication here is profound. If additivity is the root cause of false confidence, it means that false confidence is inherent to epistemic probability distributions, due to their mathematical structure. In §3c, we show that this is exactly the case.

(b) Definition of false confidence

What even is this?

Strictly speaking, one could characterize any confidence or belief or epistemic probability assigned to a false proposition as 'false confidence'. However, the whole point of statistical inference is that one does not know, at the outset, the exact true value of the parameter being inferred. So, it is to be expected that, in the course of any given inference, some false propositions will be assigned some amount of belief. The problem described in §§2c,d is more severe: There is a fixed proposition of practical interest that is guaranteed or nearly guaranteed to be assigned a large epistemic probability, regardless of whether or not it is true.

This suggests a formal definition. To state this definition precisely, we need some additional notation. First, let $\text{Pro}_{X,\theta}$ denote the aleatory distribution of the data $X \in \mathfrak{N}^n$ depending on the fixed but unknown true value of the parameter $\theta \in \Omega_\theta \subseteq \mathfrak{N}^m$ being inferred. Consider a set-function $\text{Bel}_{\Theta;x}(\cdot)$ defined on the power set of Ω_θ , depending on data x , such that $\text{Bel}_{\Theta;x}(A)$ represents the data analyst's epistemic degree of belief in the truthfulness of the proposition $A \subseteq \Omega_\theta$ about the unknown parameter θ . Then the map $x \mapsto \text{Bel}_{\Theta;x}$ is an inferential method. We say that the inference about some proposition of interest, A , suffers from severe false confidence if, for some unacceptably high belief assignment $1 - \alpha$ where $\alpha \in (0, 1)$ and unacceptably high probability of assignment $p \in (0, 1)$, there is some putative value of $\theta \in \Omega_\theta$ such that

$$A \not\supset \theta \quad \text{and} \quad \text{Pro}_{X,\theta}(\{x : \text{Bel}_{\Theta;x}(A) \geq 1 - \alpha\}) \geq p. \quad (3.1)$$

Under this notation, figure 3 is a straightforward plot of p as a function of α , where A is the assertion that the two satellites will not collide. As relative uncertainty grows, the level of false confidence seen in satellite conjunction analysis grows arbitrarily severe. Next, we show that, when considered over the full range of measurable propositions that the analyst might assess, most epistemic probability distributions used in practice suffer from arbitrarily severe false confidence.

(c) Theorem

Most statistical inference in science and engineering involves one or more real-valued continuous parameter(s). So the epistemic probability distributions (e.g. posteriors) commonly used in practice are, for all values of the observable, continuous distributions over the parameter being inferred. That is, belief assigned by the additive belief function $\text{Bel}_{\Theta;x}$ can be represented via an epistemic probability density function, say, $f_{\Theta;x}(\theta)$, with respect to Lebesgue measure λ on Ω_θ , depending on the observation, x . This is the case for most Bayesian posteriors, confidence distributions [22–24], and fiducial distributions [40,41] used in practice. The satellite collision problem described in §2 is one such example. The false confidence theorem below states that all such epistemic probability distributions suffer from arbitrarily severe false confidence.

Theorem 3.1. Consider an additive belief function $\text{Bel}_{\Theta;x}$ characterized by an epistemic probability density function $f_{\Theta;x}$ on Ω_θ such that $\sup_\theta f_{\Theta;x}(\vartheta) < \infty$ for $\text{Pro}_{X,\theta}$ -almost all x , for all θ . Then for any $\theta \in \Omega_\theta$, any $\alpha \in (0, 1)$, and any $p \in (0, 1)$, there exists a set $A \subset \Omega_\theta$ such that (3.1) holds.

Proof. For fixed x , set $\bar{f}_x = \sup_{\vartheta} f_{\vartheta;x}(\vartheta)$. Then for any bounded set $B \subset \Omega_{\theta}$,

$$\text{Bel}_{\vartheta;x}(B) = \int_B f_{\vartheta;x}(\vartheta) d\vartheta \leq \bar{f}_x \lambda(B). \quad (3.2)$$

As a function of $X \sim \text{Pro}_{X;\theta}$, for fixed θ , \bar{f}_X has a distribution; let $\eta = \eta_{\theta,p} \in (0, \infty)$ satisfy

$$\text{Pro}_{X;\theta}(\{x : \bar{f}_x \leq \eta\}) \geq p.$$

Choose a neighbourhood $B \ni \theta$ with measure $\lambda(B) = \alpha/\eta$, so that $\lambda(B)\eta = \alpha$. Define A as the complement of B . For any belief function satisfying the Kolmogorov axioms, we have

$$\text{Bel}_{\vartheta;x}(A) = 1 - \text{Bel}_{\vartheta;x}(B). \quad (3.3)$$

By definition, $A \not\ni \theta$, and by (3.2) and (3.3), we have

$$\bar{f}_x \leq \eta \implies \text{Bel}_{\vartheta;x}(B) \leq \lambda(B)\eta \iff \text{Bel}_{\vartheta;x}(B) \leq \alpha \iff \text{Bel}_{\vartheta;x}(A) \geq 1 - \alpha.$$

By definition of η , the left-most event occurs with $\text{Pro}_{X;\theta}$ -probability at least p and, therefore, the right-most event occurs with $\text{Pro}_{X;\theta}$ -probability at least p , proving the claim. ■

Theorem 3.1 is an existence result; so, our proof proceeds by constructing the simplest possible example. This is achieved by defining a neighbourhood around the true parameter value that is so small that its complement—which, by definition, represents a false proposition—is all but guaranteed to be assigned a high belief value, simply by virtue of its size. In practice, no one would intentionally seek out such a proposition, but that is beside the point.

Every real-world risk analysis problem involves a proposition of interest that is determined by the structure of the problem itself; e.g. ‘Will these two satellites collide?’. Just as the practitioner will not seek out propositions strongly affected by false confidence, neither do practitioners have the option of avoiding such propositions when they arise. What the false confidence theorem shows is that, in most practical inference problems, there is no theoretical limit on how severely false confidence will manifest itself in an epistemic probability distribution, or more precisely, there is no such limit that holds for all measurable propositions. Such a limit can only be found for a specific proposition of interest through an interrogation of the belief assignments that will be made to it over repeated draws of the data. That is the type of analysis pursued in §2d, which reveals a severe and pernicious practical manifestation of false confidence.

Additionally, the proof of the false confidence theorem has a direct connection to the practical manifestation of false confidence seen in satellite conjunction analysis. They are both driven by the fact that the size of a measurable set puts an upper bound on the belief that can be assigned to it, which due to additivity, puts a lower bound on the belief that will necessarily be assigned to its complement. The quantitative relationship between epistemic probability and the relative size of the proposition to which it is assigned is perhaps best illustrated by figure 2. This relationship holds whether or not the proposition of interest (e.g. collision) includes the true parameter value, but in the proof of the false confidence theorem, it is necessary to show that it does indeed hold, even if $\theta \in B$. That being said, it is important not to over-interpret this commonality.

Practical manifestations of severe false confidence are not limited to problems in which the proposition of interest corresponds to an obviously small set. This is explored in recent and on-going work referenced in §6. Even in the present work, consider the fact that—while the proposition that the two satellites will collide is Lebesgue-measurable in displacement space—in the original state space, collision would correspond to an infinite hypercylinder, which has an unbounded Lebesgue measure. The false confidence theorem puts a lower bound on how widely severe false confidence will manifest itself in practice, but it does not in any way put an upper bound on the pervasiveness of this phenomenon.

At the end of the day, any risk analyst who plans to represent the results of a statistical inference using an additive belief function must determine whether their proposition of interest

(e.g. Will these two satellites collide?) is one of those affected by false confidence. If so, the next question is, 'How severe?', because there is no general limit to how bad it can get. That is the message of the false confidence theorem.

(d) Broader implications

False confidence is the inevitable result of treating epistemic uncertainty as though it were aleatory variability. Any probability distribution assigns high probability values to large sets. This is appropriate when quantifying aleatory variability, because any realization of a random variable has a high probability of falling in any given set that is large relative to its distribution. Statistical inference is different; a parameter with a fixed value is being inferred from random data. Any proposition about the value of that parameter is either true or false. To paraphrase Nancy Reid and David Cox,³ it is a bad inference that treats a false proposition as though it were true, by consistently assigning it high belief values. That is the defect we see in satellite conjunction analysis, and the false confidence theorem establishes that this defect is universal.

This finding opens a new front in the debate between Bayesian and frequentist schools of thought in statistics. Traditional disputes over epistemic probability have focused on seemingly philosophical issues, such as the ontological inappropriateness of epistemic probability distributions [15,17], the unjustified use of prior probabilities [43], and the hypothetical logical consistency of personal belief functions in highly abstract decision-making scenarios [13,44]. Despite these disagreements, the statistics community has long enjoyed a truce sustained by results like the Bernstein–von Mises theorem [45, Ch. 10], which indicate that Bayesian and frequentist inferences usually converge with moderate amounts of data.

The false confidence theorem undermines that truce, by establishing that the mathematical form in which an inference is expressed can have practical consequences. This finding echoes past criticisms of epistemic probability levelled by advocates of Dempster–Shafer theory, but those past criticisms focus on the structural inability of probability theory to accurately represent incomplete prior knowledge, e.g. [19, Ch. 3]. The false confidence theorem is much broader in its implications. It applies to all epistemic probability distributions, even those derived from inferences to which the Bernstein–von Mises theorem would also seem to apply.

Simply put, it is not always sensible, nor even harmless, to try to compute the probability of a non-random event. In satellite conjunction analysis, we have a clear real-world example in which the deleterious effects of false confidence are too large and too important to be overlooked. In other applications, there will be propositions similarly affected by false confidence. The question that one must resolve on a case-by-case basis is whether the affected propositions are of practical interest. For now, we focus on identifying an approach to satellite conjunction analysis that is structurally free from false confidence.

4. The Martin–Liu validity criterion

The false confidence theorem establishes that, to be structurally free from false confidence, it is a necessary condition that a statistical inference be represented non-additively. However, while necessary, non-additivity is not sufficient. It is trivially easy to imagine a non-additive belief function that is just slightly non-additive and therefore suffers from a false confidence problem as bad or almost as bad as that depicted in §3. What we need is a criterion, by which to constrain statistical inference, that is sufficient for eliminating false confidence or, more precisely, managing the severity with which false confidence manifests itself.

³The exact quote reads, '[E]ven if an empirical frequency-based view of probability is not used directly as a basis for inference; it is unacceptable if a procedure yielding regions of high probability in the sense of representing uncertain knowledge would, if used repeatedly, give systematically misleading conclusions'. [42, p. 295]

The Martin–Liu validity criterion states that the aleatory probability with which a false proposition will be assigned some amount of belief must be no more than one minus that level of belief [26]. Stated mathematically,

$$\text{Pro}_{X;\theta}(\{x : \text{Bel}_{\theta;x}(A) \geq 1 - \alpha\}) \leq \alpha, \quad \forall \alpha \in [0, 1], A \subset \Omega_\theta, \text{ s.t. } A \not\supset \theta. \quad (4.1)$$

Since the Martin–Liu validity criterion explicitly limits the rate at which belief is assigned to false propositions, it is almost a tautology that a statistical approach satisfying this criterion will not suffer from the severe false confidence phenomenon described in §§2–3.

This part of the paper implements the Martin–Liu validity criterion as a way to identify statistical methods that are free from false confidence. Section 4a proves the Martin–Liu validity of confidence intervals and confidence regions. Section 4b shows that $K\sigma$ uncertainty ellipsoids can often be interpreted as confidence regions. This provides a statistically reliable tool for managing collision risk in satellite conjunction analysis. Section 4c provides confidence ellipses on the two-dimensional displacement between the two satellites at closest approach.

(a) Confidence regions

Frequentist confidence intervals and confidence regions have traditionally been rationalized via coverage probability. That is, a $(1 - \alpha) \times 100\%$ confidence interval or confidence region is accepted as valid because, over an infinite number of repeated independent draws of the data, at least $(1 - \alpha) \times 100\%$ of the resulting confidence intervals or confidence regions would cover the true parameter value. However, if evaluated as tools for assigning belief to propositions, i.e. sets, confidence intervals and confidence regions also satisfy the Martin–Liu validity criterion.

Let $\Gamma_\alpha(x)$ be the $(1 - \alpha)$ confidence region for θ given the realized data x . The natural assignment of belief and plausibility made by this confidence region are

$$\text{Bel}_{\theta;x}(A) = \begin{cases} 1 - \alpha, & A \supset \Gamma_\alpha(x) \\ 0, & A \not\supset \Gamma_\alpha(x) \end{cases} \quad \text{and} \quad \text{Pls}_{\theta;x}(A) = \begin{cases} 1, & A \cap \Gamma_\alpha(x) \neq \emptyset \\ \alpha, & A \cap \Gamma_\alpha(x) = \emptyset \end{cases}.$$

That is to say, a confidence region represents the simple assertion that we are $1 - \alpha$ confident that the true value of θ is somewhere inside $\Gamma_\alpha(x)$. Any sets containing $\Gamma_\alpha(x)$ inherit that confidence; all other sets accrue no positive confidence. It is a coarse way to represent a statistical inference, perhaps more so than is necessary, but it is also safe in the sense that the Martin–Liu validity criterion holds. Indeed, by the coverage probability definition of a confidence region, the event $\{\Gamma_\alpha(x) \ni \theta\}$ has $\text{Pro}_{X;\theta}$ -probability at least $1 - \alpha$. So, for any set $A \not\supset \theta$, the event $\{\Gamma_\alpha(x) \subset A\}$ has $\text{Pro}_{X;\theta}$ -probability at most α . Therefore, for any false proposition, i.e. any set A such that $A \not\supset \theta$, the probability that said proposition will be assigned a confidence of at least $1 - \alpha$ is less than or equal to α , hence (4.1) holds.

When interpreted this way, a confidence region represents a consonant confidence structure, i.e. a possibility distribution with coverage probability properties [22]. A recent paper already proves that all consonant confidence structures satisfy the Martin–Liu validity criterion [46]. The reason that we have included the special case proof for confidence regions is that most readers are likely to be unfamiliar with possibility distributions and confidence structures. In contrast, simple confidence intervals and confidence regions are well established, well-disseminated tools of statistical inference. As proven above, they do not suffer from false confidence.⁴

(b) Uncertainty ellipsoids in satellite conjunction analysis

Uncertainty ellipsoids are already well established as a screening tool in conjunction analysis [10,47]. We propose that they can be used as a principal risk management tool. $K\sigma$ ellipsoids, also known as covariance ellipsoids, can often be treated as approximate confidence regions. The case

⁴It may be helpful to remind the reader that, as pointed out in §3, confidence distributions in the sense of [22] do suffer from false confidence, despite also being rationalized in terms of coverage probability. Consonance and coverage probability together are adequate to ensure Martin–Liu validity, but coverage probability alone is not.

for this interpretation starts by treating a satellite's true state, θ , as a fixed unknown with $\hat{\theta}(x)$ as an estimate subject to random errors. If a sufficiently large amount of data is used, the error between a satellite's filter-derived state estimate and its true state can often be treated as approximately normal, even if the observation errors are drawn from non-normal distributions [48] and even if the filter used is sampling-based [49]. That is, in conjunction analysis, one can often assume

$$\hat{\theta}(X) - \theta \sim n(0, C_{\theta}),$$

where $n(0, C_{\theta})$ is a multivariate normal distribution with zero mean and covariance matrix C_{θ} . This assumption may break down if a conjunction analysis is done too far in advance [38] or if too few data are used to estimate the satellite state or if there is an unaccounted for bias error in the filter algorithm or data. For simplicity, we persist in the normality and unbiasedness assumptions. Relaxing these assumptions is left as a topic for future work.

This representation of aleatory variability in $\hat{\theta}(X) - \theta$ can be eigen decomposed as

$$\hat{\theta}(X) - \theta = E_{C_{\theta}} \Lambda_{C_{\theta}}^{1/2} \xi,$$

where X is a (hypothetical) random realization drawn from the same distribution as the data used to estimate the orbits, $E_{C_{\theta}}$ is a matrix whose columns are the eigenvectors of C_{θ} , $\Lambda_{C_{\theta}}$ is a diagonal matrix whose entries are the corresponding eigenvalues of C_{θ} , and ξ is a unit normal vector of the same dimensionality as θ . This can easily be rearranged into a pivot, as

$$\Lambda_{C_{\theta}}^{-1/2} E_{C_{\theta}}^{\top} (\hat{\theta}(X) - \theta) = \xi.$$

A pivot is a function of the parameter being inferred and the random data used to infer it whose output is a random variable with a known fixed distribution [50]; in this case, ξ is a unit normal vector. Pivots are special because they can be used to derive confidence regions.

In particular, we are interested in $K\sigma$ uncertainty ellipsoids. A $K\sigma$ uncertainty ellipsoid on θ is generated by projecting a sphere in ξ -space centred at the origin. Let $K > 0$ be some constant. A confidence region of the form

$$\Gamma_{\alpha}(x) = \left\{ \vartheta : \xi = \Lambda_{C_{\theta}}^{-1/2} E_{C_{\theta}}^{\top} (\hat{\theta}(x) - \vartheta), \sum_{i=1}^{\dim(\theta)} \xi_i^2 \leq K^2 \right\}$$

will map to θ as a $\dim(\theta)$ -dimensional ellipsoid. The confidence, $(1 - \alpha)$, associated with this uncertainty ellipsoid can be computed as

$$1 - \alpha = \text{Pro}_{X, \vartheta}(\{x : \Gamma_{\alpha}(x) \ni \theta\}) = \text{Pro}_{\xi} \left(\left\{ \xi : \sum_{i=1}^{\dim(\theta)} \xi_i^2 \leq K^2 \right\} \right) = F_{\chi^2, \dim(\theta)}(K^2),$$

where $F_{\chi^2, j}$ is the central χ_j^2 distribution function.

Software is commercially available that is capable of computing three-dimensional $K\sigma$ uncertainty ellipsoids representing uncertainty in the predicted position of a satellite and then determining whether or not the resulting ellipsoids for two satellites intersect [47,51]. If two ellipsoids overlap in a conjunction analysis, a collision avoidance manoeuvre may be needed. If they do not, the two satellites are considered safe, to within the confidence level associated with the uncertainty ellipsoids.

There are a couple of caveats to the approach as currently implemented in the field. First, as traditionally conceptualized, ellipsoid overlap approaches fail to account for the physical size of the two satellites [52]. To correct for this, effective ellipsoid overlap should be defined as occurring when the minimum distance between the two uncertainty ellipsoids is less than the combined radius of the two satellites. So long as the position uncertainties are much larger than the satellites' combined size, the distinction is negligible. However, satellite size should not be ignored in problems with relatively small position uncertainties.

The other caveat is that, because the two satellites are treated separately, the joint confidence associated with having captured both satellite positions in their respective uncertainty ellipsoids

is usually slightly lower than the confidence individually assigned to each ellipsoid. If each ellipsoid has a $1 - \alpha$ confidence attached to it, assuming independence between the data used to infer the two satellite trajectories, the confidence attached to simultaneous coverage is

$$1 - \alpha' = \text{Pro}_{X_1, X_2; \theta_1, \theta_2}(\{x_1, x_2 : \Gamma_{\alpha,1}(x_1, A) \ni \theta_1, \Gamma_{\alpha,2}(x_2, A) \ni \theta_2\}) = (1 - \alpha)^2.$$

In reality, the errors in the data used to estimate the two satellite positions may not be entirely independent, but the specter of non-independence is easily resolved. Using Fréchet bounds [53], even under totally unknown dependence between the data,

$$1 - \alpha' \geq \max(0, 1 - 2\alpha).$$

Note that if α is small, $(1 - \alpha)^2 = 1 - 2\alpha + \alpha^2 \approx 1 - 2\alpha$. So, using $1 - 2\alpha$ to represent the joint confidence associated with the two uncertainty ellipses is both conservative and extremely close to the independence case. Thanks to the result about confidence regions presented in §4a, so long as one performs a manoeuvre whenever the two uncertainty ellipsoids intersect, the rate at which collisions occur over a large number of conjunctions—i.e. the operational aleatory probability of collision—will be capped at $\alpha' = 2\alpha$.

For example, suppose one were to use a pair of 4σ ellipsoids to represent uncertainty in the positions of the two conjuncting satellites. This represents a 99.9% confidence region around each satellite position. Moreover, one can say with 99.8% confidence that both satellite positions are contained within their respective ellipsoids. And, more importantly, if a satellite operator performs a collision avoidance manoeuvre whenever those 4σ ellipsoids overlap, the operational probability of collision will be capped at 0.227%.

All of that being said, ellipsoid overlap does not mean that two satellites are definitely going to collide. Rather, it means that collision is still plausible in light of the data. Performing a collision avoidance manoeuvre is one way of driving down the plausibility of collision, but it is not the only way. For conjunctions predicted several days in advance, prudent navigators will opt to gather more data, which will allow them to shrink the uncertainty ellipsoids before finally deciding whether or not to make a collision avoidance manoeuvre. So long as they do not wait too long, this is usually the most reasonable course of action.

(c) Uncertainty ellipses on displacement

An alternative approach based on confidence regions is to derive an uncertainty ellipse on the two-dimensional displacement introduced in §2a. Assuming that variability in the direction of the relative velocity vector is small enough to be ignored, the (random) estimate of displacement at closest approach, (\hat{u}', \hat{v}') , has a bivariate normal distribution whose mean is equal to the unknown true displacement, (u', v') . That is,

$$\begin{bmatrix} \hat{u}' \\ \hat{v}' \end{bmatrix} = \begin{bmatrix} u' \\ v' \end{bmatrix} + E_S \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix},$$

where ξ_1 and ξ_2 are independent unit normal variates and E_S , S_1 , and S_2 are the eigendecomposition of the two-dimensional displacement covariance matrix, $C_{\Delta,1;2,1;2}$, defined in §2a. A $K\sigma$ confidence ellipse can be constructed on (u', v') by projecting a circle, $\xi_1^2 + \xi_2^2 \leq K^2$, onto (u', v') . The confidence associated with this ellipse is computed from a χ^2 distribution with two degrees of freedom, as outlined in §4b. Recall that collision is associated with the condition $u'^2 + v'^2 \leq R^2$, where R is the combined 'radius' of the two satellites. So long as the confidence ellipse for (u', v') does not intersect this circle, then one can say with $(1 - \alpha)$ confidence that collision will not occur.

The advantage of this approach is that it will yield a lower false alarm rate than the ellipsoid overlap approach described in §4b. The disadvantage is that it involves two assumptions that the ellipsoid overlap approach does not, namely that variability in the direction of the estimated relative velocity vector is negligible and that the covariance between the position estimate errors

for the two satellites is known. This second assumption is often taken for granted, but it is non-trivial. Additionally, supporting software for the two-dimensional confidence ellipse on displacement is not yet widely available to satellite navigators.

5. Conclusion

The work presented in this paper has been done from a fundamentally frequentist point of view, in which θ (e.g. the satellite states) is treated as having a fixed but unknown value and the data, x , (e.g. orbital tracking data) used to infer θ are modelled as having been generated by a random process (i.e. a process subject to aleatory variability). Someone fully committed to a subjectivist view of uncertainty [13,44] might contest this framing on philosophical grounds. Nevertheless, what we have established, via the false confidence phenomenon, is that the practical distinction between the Bayesian approach to inference and the frequentist approach to inference is not so small as conventional wisdom in the statistics community currently holds. Even when the data are such that results like the Bernstein-von Mises theorem ought to apply, the mathematical form in which an inference is expressed can have large practical consequences that are easily detectable via a frequentist evaluation of the reliability with which belief assignments are made to a proposition of interest (e.g. 'Will these two satellites collide?').

Our rationale for framing conjunction analysis in frequentist terms is that, as established in the opening of §1, everyone in the space industry has an interest in limiting the number of in-orbit collisions. Thus, our goal has been to help satellite operators identify tools adequate for limiting the literal frequency with which collisions involving operational satellites occur. Framing the problem in frequentist terms enables us to do that, whereas framing the problem in Bayesian terms would not. The only circumstance in which a Bayesian analysis could directly enable satellite operators to control the frequency with which collisions occur is if it were based on an aleatory prior [25] on the satellite states, and this prior would need to reflect the underlying risk of collision due to orbital crowding. As currently practiced, no such aleatory prior is used in conjunction analysis, nor, in our estimation, is one likely to become available in the coming years. An estimate of the aggregate collision risk per unit time seems feasible, but how to parse that into priors on the satellite states is non-obvious. Such an operation may not be well-posed. So, for someone interested in limiting the literal frequency with which collisions occur, it is necessary to treat the satellite states in each conjunction as a fixed, albeit uncertain, reality and then to assess how reliably a proposed risk metric performs. That is the analysis pursued in §2d, and under that analysis, epistemic probability of collision does not appear to be a viable risk metric.

There are other engineers and applied scientists tasked with other risk analysis problems for which they, like us, will have practical reasons to take the frequentist view of uncertainty. For those practitioners, the false confidence phenomenon revealed in our work constitutes a serious practical issue. In most practical inference problems, there are uncountably many propositions to which an epistemic probability distribution will consistently accord a high belief value, regardless of whether or not those propositions are true. Any practitioner who intends to represent the results of a statistical inference using an epistemic probability distribution must at least determine whether their proposition of interest is one of those strongly affected the false confidence phenomenon. If it is, then the practitioner may, like us, wish to pursue an alternative approach.

Fortunately, the Martin–Liu validity criterion provides a formal benchmark by which to identify statistical methods that are inherently free from severe false confidence. For example, in §4a, we proved that frequentist confidence intervals and confidence regions satisfy the Martin–Liu validity criterion, thus identifying a widely disseminated and well-established statistical tool that is free from severe false confidence. Technically, this is a special case of a theorem already proved in [46], but our special case proof is offered in the hope that it will be more accessible to a wider readership. In satellite conjunction analysis, this means that $K\sigma$ uncertainty ellipsoids and uncertainty ellipses, which can often be rationalized as confidence regions, are statistically reliable in the same way that epistemic probability of collision is not.

6. Future and on-going work

Much work remains to be done exploring the range of real-world uncertainty quantification problems that are practically impacted by false confidence. Risk analysis problems with bounded failure domains are strong candidates for severe false confidence. Recent work demonstrates that false confidence can also arise as the result of nonlinear uncertainty propagation, even if a marginalization-specific reference posterior is used [54,55].

As mentioned in §4a, though statistically reliable, confidence intervals and confidence regions are also coarse. More nuanced possibilistic solutions to problems of statistical inference can be derived using consonant confidence structures or Martin–Liu inferential models. In satellite conjunction analysis, the next natural step is to evaluate the statistical performance of the alternative collision risk metrics mentioned in §1a to see which, if any, satisfy the Martin–Liu validity criterion.

Finally, a recent conference paper [56] asserts that the ellipse overlap approach will motivate too many collision avoidance manoeuvre. The authors of [56] argue for epistemic probability of collision as a risk metric on the grounds that it is reasonable to treat conjunctions with diluted collision probabilities as akin to the large number of unrecognized conjunctions that occur daily with objects that are too small to be tracked at all, for which no mitigation actions are possible. They view both situations as part of the background risk that is simply accepted as part of operating a satellite. This argument may initially seem attractive, but we see it as promoting an ill-advised tolerance to blind risk. Regardless of the burden incurred by acting on it, the approach we advocate enables satellite operators to quantify and potentially manage their collision risk exposure in the face of high relative uncertainties (i.e. high S/R ratios); epistemic probability of collision does not. We recognize the natural desire to balance the goal of preventing collisions against the goal of keeping manoeuvres at a reasonable level, and we further recognize that it may not be possible to achieve an acceptable balance between these two goals using present tracking resources. Nevertheless, committing to a risk metric that cannot reliably distinguish ambiguity from confirmed safety is not, in our view, a durable resolution to this impasse.

Data accessibility. Scripts for generating figures 1–3 in R are available at www.alexvalcon-llc.com/examples/satellite-collision-risk.

Authors' contributions. M.S.B. did the initial work on satellite conjunction analysis, formulated the false confidence theorem and led the drafting of this manuscript; R.M. proved the false confidence theorem and contributed to the writing of the manuscript; S.F. provided recommendations to improve the manuscript.

Competing interests. Alexandria Validation Consulting, LLC has developed a proprietary algorithm for computing collision risk in satellite conjunction analysis, based in part on insights derived from the research presented in this paper. To protect the financial interests of M.S.B. and his co-investors, this algorithm is currently being withheld from publication.

Funding. There are no funding sources to report.

Acknowledgements. Thanks go to the anonymous reviewers for the critical-yet-helpful feedback; to Salvatore Alfano for providing M.S.B. with an introduction to the literature on satellite conjunction analysis and probability dilution; and to William Oberkampf for his insights and periodic reminders of the systemic threat posed by in-orbit collisions. Special—though, anonymous—thanks go to the satellite navigator who initially brought satellite conjunction analysis to M.S.B.'s attention.

References

1. Kessler DJ, Johnson NL, Liou J, Matney M. 2010 The Kessler Syndrome: implications to future space operations, AAS-10-016. In *33rd Annual AAS Guidance and Control Conf.* Breckenridge, CO: American Astronautical Society.
2. Liou JC, Johnson NL. 2006 Risks in space from orbiting debris. *Science* **311**, 340–341. (doi:10.1126/science.1121337)
3. Liou JC, Johnson NL. 2008 Instability of the present LEO satellite populations. *Adv. Space Res.* **41**, 1046–1053. (doi:10.1016/j.asr.2007.04.081)
4. Alfano S. 2003 Relating position uncertainty to maximum conjunction probability, AAS-03-548. In *AAS/AIAA Astrodynamics Specialists Conf., Big Sky, MT, 3–7 August*. AIAA.

5. Frigm C. 2009 A single conjunction risk assessment metric: the F -value. *Adv. Astron. Sci.* **135**, 1175–1192.
6. Plakalović D, Hejduk M, Frigm R, Newman L. 2011 A tuned single parameter for representing conjunction risk, AAS-11-430. In *2011 AIAA/AAS Astrodynamics Specialist Conf.*, Girdwood, AK: 31 July–4 August. San Diego, CA: Univelt Inc.
7. Sun Z, Luo Y, Niu Z. 2014 Spacecraft rendezvous trajectory safety quantitative performance index eliminating probability dilution. *Sci. China Technol. Sci.* **57**, 1219–1228. (doi:10.1007/s11431-014-5523-3)
8. Balch MS. 2016 A corrector for probability dilution in satellite conjunction analysis, AIAA-2016-1445. In *18th AIAA Non-Deterministic Approaches Conf.*, San Diego, 4–8 January. AIAA.
9. Carpenter JR *et al.* 2017 Relevance of the American statistical society's warning on P -values for conjunction assessment, AAS-17-614. In *AAS/AIAA Astrodynamics Specialist Conf.*, Stevenson, WA, 20–24 August. AIAA.
10. Newman LK, Frigm RC, Duncan MG, Hejduk MD. 2014 Evolution and implementation of the NASA robotic conjunction assessment risk analysis concept of operations. In *Advanced Maui Optical and Space Surveillance Technologies Conf.*, Wailea, HI.
11. Barnett V. 1999 *Comparative statistical inference*, 3rd edn. Chichester, NY: Wiley and Sons.
12. Laplace PS. 1814 *A philosophical essay on probabilities* (Translated 1902). New York, NY: Wiley and Sons.
13. de Finetti B. 1935 *La prévision: ses lois logiques, ses sources subjectives*. Paris, France: Annales de l'Institut Henri Poincaré.
14. Robert CP. 2007 *The Bayesian choice: from decision-theoretic foundations to computational implementation*. New York, NY: Springer Science & Business Media.
15. Boole G. 1854 *An investigation of the laws of thought* (Online release 2005). Chapel Hill, NC: Project Gutenberg. See www.gutenberg.net.
16. Neyman J. 1937 Outline of a theory of statistical estimation based on the classical theory of probability. *Phil. Trans. R. Soc. Lond. A* **236**, 333–380. (doi:10.1098/rsta.1937.0005)
17. Mayo DG. 1996 *Error and the growth of experimental knowledge*. Chicago, IL: University of Chicago Press.
18. Ferson S, Ginzburg LR. 1996 Different methods are needed to propagate ignorance and variability. *Reliab. Eng. Syst. Safety* **54**, 133–144. (doi:10.1016/S0951-8320(96)00071-3)
19. Salicone S. 2007 *Measurement uncertainty: an approach via the mathematical theory of evidence*. New York, NY: Springer Science & Business Media, LLC.
20. Oberkampf WL, Roy CJ. 2010 *Verification and validation in scientific computing*. New York, NY: Cambridge University Press.
21. Shafer G. 1976 *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.
22. Balch MS. 2012 Mathematical foundations for a theory of confidence structures. *Int. J. Approx. Reason.* **53**, 1003–1019. (doi:10.1016/j.ijar.2012.05.006)
23. Schweder T, Hjort NL. 2016 *Confidence, likelihood, probability: statistical inference with confidence distributions*. New York, NY: Cambridge University Press.
24. Xie M, Singh K. 2013 Confidence distribution, the frequentist distribution estimator of a parameter: a review. *Int. Stat. Rev.* **81**, 3–39. (doi:10.1111/insr.2013.81.issue-1)
25. von Mises R. 1942 On the correct use of Bayes' formula. *Ann. Math. Stat.* **13**, 156–165. (doi:10.1214/aoms/1177731603)
26. Martin R, Liu C. 2016 *Inferential models: reasoning with uncertainty*. Boca Raton, FL: CRC Press.
27. Alfano S, Oltrogge D. 2016 Probability of collision: valuation, variability, visualization and validity, AIAA-2016-5654. In *AIAA/AAS Astrodynamics Specialist Conf.*, Long Beach, CA, 13–16 September. AIAA.
28. Frigm RC, Hejduk MD, Johnson LC, Plakalovic D. 2015 Total probability of collision as a metric for finite conjunction assessment and collision risk management. In *Proc. of the Advanced Maui Optical and Space Surveillance Technologies Conf.*, Wailea, HI.
29. Alfried KT, Akella MR, Frisbee J, Foster JL, Lee DJ, Wilkins M. 1999 Probability of collision error analysis. *Space Debris* **1**, 21–35. (doi:10.1023/A:1010056509803)
30. Patera RP. 2001 General method for calculating satellite collision probability. *J. Guid. Control Dyn.* **24**, 716–722. (doi:10.2514/2.4771)
31. Chan FK. 2008 *Spacecraft collision probability*. El Segundo, CA: Aerospace Press.
32. Meinhold RJ, Singpurwalla ND. 1983 Understanding the Kalman filter. *Am. Stat.* **37**, 123–127.
33. Hannig J. 2009 On generalized fiducial inference. *Stat. Sinica* **19**, 491–544.

34. Hall DT, Hejduk MD, Johnson LC. 2017 Time dependence of collision probabilities during satellite conjunctions, AAS-17-271. In *AAS/AIAA Space Flight Mechanics Meeting, San Antonio, CA, 5–9 February*. San Diego, CA: Univelt Inc.
35. Analytical Graphics, Inc. 2018 STK: nonlinear probability tool. See <http://help.agi.com/stk/index.htm#cat/Cat03-08.htm>.
36. Patera RP. 2005 Calculating collision probability for arbitrary space-vehicle shapes via numerical quadrature. *J. Guid. Control Dyn.* **28**, 1326–1331. (doi:10.2514/1.14526)
37. Sabol C, Sukut T, Hill K, Alfriend KT, Wright B, Li Y, Schumacher P. 2010 Linearized orbit covariance generation and propagation analysis via simple monte carlo simulations. Technical Report, Texas Engineering Experiment Station, College Station, TX.
38. Ghrist RW, Plakalovic D. 2012 Impact of non-Gaussian error volumes on conjunction assessment risk analysis, AIAA-2012-4965. In *AIAA/AAS Astrodynamics Specialist Conf., Minneapolis, MN, 13–16 August*. AIAA.
39. Newman LK. 2010 The NASA robotic conjunction assessment process: overview and operational experiences. *Acta Astronaut.* **66**, 1253–1261. (doi:10.1016/j.actaastro.2009.10.005)
40. Hannig J, Iyer H, Lai RC, Lee TC. 2016 Generalized fiducial inference: a review and new results. *J. Am. Stat. Assoc.* **111**, 1346–1361. (doi:10.1080/01621459.2016.1165102)
41. Fisher RA. 1935 The fiducial argument in statistical inference. *Ann. Eugen.* **6**, 391–398. (doi:10.1111/j.1469-1809.1935.tb02120.x)
42. Reid N, Cox DR. 2015 On some principles of statistical inference. *Int. Stat. Rev.* **83**, 293–308. (doi:10.1111/insr.v83.2)
43. Fisher RA. 1930 Inverse probability. *Proc. Cambridge Phil. Soc.* **26**, 528–535. (doi:10.1017/S0305004100016297)
44. Savage LJ. 1972 *The foundations of statistics*, 2nd edn. New York, NY: Dover Publications Inc.
45. Van der Vaart AW. 2000 *Asymptotic statistics*. New York, NY: Cambridge University Press.
46. Denoeux T, Li S. 2018 Frequency-calibrated belief functions: review and new insights. *Int. J. Approx. Reason.* **92**, 232–254. (doi:10.1016/j.ijar.2017.10.013)
47. 2018 STK: defining threat volume. See <http://help.agi.com/stk/index.htm#cat/Cat03-01.htm>.
48. Spall JC, Wall KD. 1984 Asymptotic distribution theory for the Kalman filter state estimator. *Commun. Stat. Theory Methods* **13**, 1981–2003. (doi:10.1080/03610928408828808)
49. Del Moral P, Guionnet A. 1999 Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.* **9**, 275–297. (doi:10.1214/aoap/1029962742)
50. Casella G, Berger RL. 2002 *Statistical inference*, 2nd edn. Pacific Grove, CA: Thomson Learning.
51. Alfano S, Chan FK, Greer ML. 2004 Eigenvalue quadric surface method for determining when two ellipsoids share common volume for use in spatial collision detection and avoidance. US Patent 6,694,283.
52. Alfano S, Greer ML. 2003 Determining if two solid ellipsoids intersect. *J. Guid. Control Dyn.* **26**, 106–110. (doi:10.2514/2.5020)
53. Ferson S, Nelsen RB, Hajagos J, Berleant DJ, Zhang J, Tucker WT, Ginzburg LR, Oberkampf WL. 2004 Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis, SAND-2004-3072. Technical Report, Sandia National Laboratories Albuquerque, NM.
54. Carmichael I, Williams J. 2018 An exposition of the false confidence theorem. *Stat* **7**, e201. (doi:10.1002/sta4.201)
55. Martin R. 2019 False confidence, non-additive beliefs, and valid statistical inference. (<http://arxiv.org/abs/quant-ph/1607.05051>).
56. Hejduk MD, Snow DE, Newman LK. 2019 Satellite conjunction assessment risk analysis for “dilution region” events: issues and operational approaches. In *Space Traffic Management Conf., Austin, TX, 26–27 February*.