(Preprint) IAA-AAS-DyCoSS1 XX-XXX

TRENDING IN PROBABILITY OF COLLISION MEASUREMENTS

J.J. Vallejo[*], M.D. Hejduk[†], and J.D. Stamey[‡]

A simple model is proposed to predict the behavior of Probabilities of Collision (P_c) for conjunction events. The model attempts to predict the location and magnitude of the peak P_c value for an event by assuming the progression of P_c values can be modeled to first order by a downward-opening parabola. To incorporate prior information from a large database of past conjunctions, the Bayes paradigm is utilized; and the operating characteristics of the model are established through a large simulation study. Though the model is simple, it performs well in predicting the temporal location of the peak (P_c) and thus shows promise as a decision aid in operational conjunction assessment risk analysis.

INTRODUCTION

The problem of deciding whether to maneuver a satellite which is in conjunction with another space object is often not straightforward, and a serious collision threat often involves the deliberation and cooperation of various parties.[1] Quantifying the risk for any such conjunction is typically accomplished through the use of the predicted miss distance at time of closest approach (TCA) and the calculated probability of collision P_c at that same time. These measurements are generally taken over a few days to a week before TCA. The calculated P_c value is affected by the uncertainty in the positions of the space objects, an uncertainty that generally decreases as one approaches TCA. This decrease in uncertainty typically yields a particular kind of behavior in P_c values, which we shall refer to as the "canonical behavior". We seek to incorporate the shape of this canonical behavior into our understanding of the P_c values, with the goal of making predictions about future P_c values, as well as making inferences about the location of the highest P_c value.

Although there has been considerable work in calculating the probability of collision[2,3,4,5] far less work has focused on detecting trends in repeated measurements of the P_c. Notably, Carpenter and Markley have proposed various implementations of Wald's Sequential Probability Ratio Test in deciding whether to accept the hypothesis that a new measurement on the P_c is identical in information content to the previous measurement[6,7,8] Among the advantages of this method are its simplicity and its inherent modeling of false alarms and missed detections. While a considerable advance in P_c predictive methods, this approach is not without limitations. For instance, although the WSPRT tests consecutive measurements, it has no way of directly incorporating the times at which the measurements were taken; it considers measurement time only indirectly through the accumulation of data in forming the total information matrices from which it works. In general, P_c measurements are not taken at equidistant time intervals, suggesting a potential loss of information in the WSPRT approach.

[*]a.i. solutions Inc.,10001 Derekwood Lane, Lanham, MD 20706
[†]Astrorum Consulting LLC, 10006 Willow Bend Drive, Woodway, TX 76712
[‡]Department of Statistics, Baylor University, P.O. Box 97140, Waco, TX 76798

In this paper, we propose a simple method to detect the trend in repeatedly-measured P_c values. Our approach has the advantage of directly incorporating the time between observations, which is allowed to be irregular. Additionally, we use the Bayesian paradigm in order to incorporate prior information gathered from past conjunctions. More sophisticated methods are certainly possible, but we wished to determine how much predictive power could be rendered by a simple and straightforward foundational approach.

METHODS FOR TRENDING IN PROBABILITY OF COLLISION

Calculating the Probability of Collision

The basic procedure for calculating the probability of collision between two space objects is to obtain positions and positional covariances propagated to the time of closest approach (TCA), and using this information, determine the probability of the two objects' passing within a chosen small distance (called the hard body radius or HBR) of each other. In order to reduce the computational cost of the problem, one generally employs a few assumptions. First, one assumes that the errors associated with positional uncertainty are trivariate Gaussian. This implies that the positional covariances are ellipsoidal and that the mean of the distribution of each object is taken to be the calculated position at TCA. These covariances are presumed to be uncorrelated, implying that the total positional uncertainty can be calculated simply by the sum of the two covariances (after having been rotated to be in the same coordinate system). Traditionally, one takes this combined covariance and centers it about the secondary object. Likewise, one sums the radii of circumscribing spheres about each object to create a single combined hard body sphere, which is placed at the location of the primary object. This problem is equivalent to the original problem involving two separate Gaussian densities due to the assumption that the covariances are uncorrelated.

One last assumption generally made is that of rectilinear motion near the time of conjunction, so that the dimensionality of the problem may be reduced. If the conjunction between the two satellites takes place at high velocity, then the relative motion in the neighborhood of the conjunction will be rectilinear; and a collision, should it take place, will occur in a plane normal to the relative velocity vector between the two objects. One can thus project the combined covariance and the hard body sphere into this plane and consider the situation as a two-dimensional problem: one has a circle, resulting from the projection of the hard body sphere, and a covariance ellipse, resulting from the projected combined covariance. One is then interested in the probability of the area swept out by the circle in the probability density formed by the ellipse.[9]

It is clear that the probability of collision depends heavily on the size and shape of the combined covariance ellipsoid. Alfano[10] investigated this relationship for various miss distances, hard body volumes, covariance sizes and shapes. He reported that for a given miss distance, hard body volume, and covariance shape, there is a covariance size which maximizes the probability of collision, with the probability decreasing slowly if uncertainty is increased (that is, the size of the objects' covariances are increased) and decreasing very rapidly if this uncertainty is decreased. We seek to incorporate this known behavior into a statistical model in order to better calibrate each measured probability of collision. In practice, one typically observes a decrease in the size of the covariance as the event moves closer to TCA, producing what we will refer to as a "canonical behavior". We aim to try to recover this behavior beneath all the other "noise" of the problem and ultimately identify the point of maximum probability of collision, in order to make better judgments regarding the degree of continued monitoring that the conjunction merits.

Canonical Behavior

As noted above, changes in P_c generally follow a canonical behavior with respect to a decreasing state estimate uncertainty; and the the parameter used to illustrate this phenomenon is the ratio of covariance radius to miss distance. Figure 1 depicts what we have called the "canonical behavior" of an event's P_c: an initial increasing change in order of magnitude in P_c as uncertainty decreases, followed by a subsequent drop off when the uncertainty becomes even smaller. The decrease in probability as uncertainty increases is what Alfano[10] referred to as "dilution in probability" because it was caused not by improvements in knowledge of satellite positions but by a lack of positional knowledge that renders any conclusion of high risk impossible. Note that these are generally particularly small probabilities, and consequently one is usually concerned with changes in orders of magnitude. That is, one is interested in changes in $\log_{10} P_c$ as opposed to simply changes in the P_c value. In the following development, we let y denote the $\log_{10} P_c$ value.

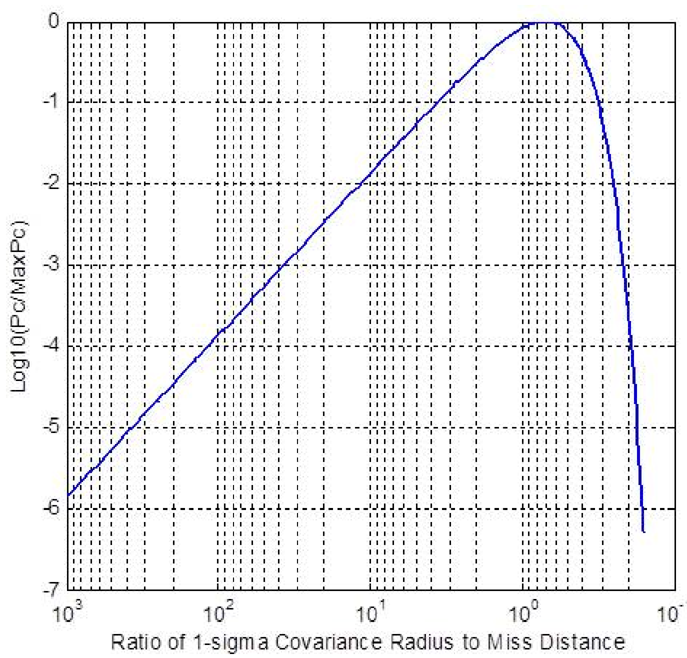

Figure 1.

Though informative, using the ratio of covariance size to miss distance as a predictor variable is difficult in practice. Although the size of the combined covariances tend to shrink over time, the rate is not the same for each event. In some cases, the value of this ratio, which appears monotonic before and after the peak point in the figure above, actually increases and decreases several different times before reaching its final value, making modeling a trend even more difficult. Furthermore, the miss distance calculated on the initial Conjunction Data Message (CDM, which is the message issued by the Joint Space Operations Center to document the states and covariances of the two objects at the time of closest approach) is subject to change on subsequent CDMs, and there is often no obvious trend in these updates. As a result, one never knows what the next ratio value will be, even if one knows at which time a CDM would be received. Thus, to use this ratio as a predictor in a statistical model, one would need to regress the ratio on some quantity one could predict, such as time. This is especially difficult because the relationship between the ratio and the $\log_{10} P_c$ value is different for each event, as is the relationship between the ratio and time. To see the difficulty in this kind of modeling, let x_{kt} be the ratio of covariance radius to miss distance for the k^{th} event at time t. We assume that both y_{it} and x_{it} is measured with errors e_{it} and ϵ_{it}, respectively. Then this model is a state-space model[11] and can be written as

$$
\begin{aligned}
y_{it} &= f(x_{it}) + e_{it} \\
x_{it} &= g(t) + \epsilon_{it} \\
e, \epsilon &\sim h(e, \epsilon|\alpha)
\end{aligned}
$$

where e and ϵ are error terms with a joint distribution $h(\cdot|\alpha)$. It is clear that when attempting to calculate y via the ratio, in practice one needs to specify not only the relationship f between the ratio x_{it} and y_{it} but also specify the relationship g between x_{it} and t.

We leave the exploration of this kind of hierarchical model for later research. In this paper, we use time as the predictor variable. We assume that the y values still follow a similar canonical behavior with respect to time as they do to with respect to the ratio. For simplicity, we attempt to model this behavior with a downward opening parabola, with the aim of correctly predicting the location and, less critically, the magnitude of the peak y value. Though model fit is important, our main goal is to correctly identify the peak y location and value, whether or not the other y values are predicted accurately.

Vertex Model

In order to compute this predicted P_c parabola, we use constrained optimization[12] to enforce a downward-opening behavior. There is also precedent for constrained inference in the Bayesian paradigm, as Gelfand[13] introduced an approach to Gibbs sampling in constrained parameter and truncated data problems. Specifically, Gelfand considers problems with ordered parameters, constrained parameters, and censored data. Considering the general equation for a parabola below, our problem is seen as one involving constrained parameters, as we know that $\beta_2 < 0$ and (as discussed subsequently) $\beta_0 < 0$.

$$y = \beta_0 + \beta_1 t + \beta_2 t^2$$

We also show that this induces a constraint on β_1. Implementing these constraints is another way in which we can "inform" the model. Utilizing these constraints along with an informative prior structure allows us to include a maximal amount of prior information, which we believe to be essential, as many of the events we consider contain only 3 or 4 data points, and we wish a durable prediction as early as possible within the event.

To allow our model to incorporate prior information from past events, we use the Bayesian paradigm.[14] Let y_{ij} be the $\log_{10} P_c$ from the j^{th} CDM from the i^{th} event. Similarly, let t_{ij} be the time (in days) until TCA for the j^{th} CDM from the i^{th} event. We assume that the observed P_c values over t follow the relationship

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, \sigma_{ij}^2)$. Furthermore, we assume this parabola will be downward opening, to attempt to model the expected canonical behavior. Utilizing the Bayesian paradigm will allow us to incorporate information about where the peak y value usually is, and how quickly the y values tend to drop off. This incorporation is accomplished by specifying informative prior distributions for the parameters, which are considered random variables in the Bayesian paradigm. Another consequence of treating parameters as random variables is that one can make predictions by taking into account not only the error in the observations but also the error in the estimates of the parameters, a Bayesian feature that is not fully possible with a frequentist approach. Finally, using the Bayesian paradigm allow us to make predictions in this four-parameter model even with only two or three observations by utilizing the prior distributions of the parameters to help identify the likely values of the parameters.

Bayesian Inference

Bayesian inference relies on the posterior distribution of the parameters. To see how the posterior distribution is calculated, let $\theta = (\beta_0, \beta_1, \beta_2, \sigma^2)' \in \Theta$ be a vector of unknown parameters. Suppose one has data $\mathbf{y}$, with joint distribution $f(\mathbf{y}|\theta)$. Let $\pi(\theta)$ be a prior distribution on θ with CDF P_θ. Treated as a function of θ for fixed $\mathbf{y}$, the joint distribution becomes the likelihood, $l(\theta|\mathbf{y})$, defined on Θ. The posterior distribution of θ, given by Bayes' theorem, is

$$\pi(\theta|\mathbf{y}) = \frac{l(\theta|\mathbf{y})\pi(\theta)}{\int l(\theta|\mathbf{y})\pi(\theta)d\theta}.$$

This is the distribution of the parameters after having seen the data vector $\mathbf{y}$. Thus, the prior beliefs about parameters and their distributions are updated after encountering the actual data. The posterior distribution often does not have a closed form and must be approximated using Markov Chain Monte Carlo (MCMC) methods.

The usual prior structure for regression coefficients in linear regression is an independent normal prior for each regression coefficient.[14] We amend this structure to incorporate the constraints we know to exist in our problem. We know that the parabola must open downwards, so that $\beta_2 < 0$. As a consequence of this constraint and the fact that all y values are less than or equal to 0 by definition (since they represent the base 10 logarithm of values between 0 and 1), we also know that $\beta_0 < 0$.

We show that the parameter β_1 must also be constrained. Because $y_{ij} \leq 0$ for all i, j, it follows that the peak y value should also be less than or equal to zero. It is easy to show that the location of the peak is $h = -\beta_1/2\beta_2$, and that the magnitude of the peak is $b = \beta_0 - \beta_1^2/4\beta_2$. In order to force the magnitude of the peak b to be less than or equal zero, we must have

$$\begin{aligned} \beta_0 - \beta_1^2/4\beta_2 &\leq 0 \\ 4\beta_2\beta_0 - \beta_1^2 &\geq 0 \\ \beta_1^2 &\leq 4\beta_2\beta_0, \end{aligned}$$

where the second line follows since $\beta_2 < 0$. This implies that $\beta_1 \in [-2\sqrt{\beta_0\beta_2}, 2\sqrt{\beta_0\beta_2}]$. Implementing these constraints in conjunction with the usual prior structure, we have

$$
\begin{aligned}
&y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \epsilon_{ij} \\
&\epsilon_{ij} \sim N(0, \sigma_i^2) \\
&\beta_0 \sim Normal(\mu_0, \sigma_0^2) I_{(-\infty,0)} \\
&\beta_1 \sim Normal(\mu_1, \sigma_1^2) I_{(-2\sqrt{\beta_0\beta_2}, 2\sqrt{\beta_0\beta_2})} \\
&\beta_2 \sim Normal(\mu_2, \sigma_2^2) I_{(-\infty,0)} \\
&\sigma^2 \sim InverseGamma(a, b),
\end{aligned}
$$

where $I()$ is the indicator function. Thus, we fit a downward opening parabola to the $\log P_c$ values over time for each event. This implies that each event has $\log P_c$ values which will rise and fall over time and that each event is allowed to have its own rate of increase/decrease. Eliciting informative priors on the regression coefficients will allow us to borrow information about what the shape of this parabola is for most events, and how much it is prone to vary. Though there are other ways to borrow information, $e.g.$ a mixed model, we find this to be a simple and straightforward way to allow the model to be flexible enough to fit all of the events. On a more technical note, attempting a mixed model in this setting is not particularly straightforward, as any random effects specified in the model would also have to be constrained. Furthermore, at least two random effects would be necessary (a random intercept and a random slope), as we desire a model which can have a different peak location and value for each event. Ultimately, we favor a more simple model that is interpretable and flexible.

We note a few additional attributes of this model here. First, although we know that the regression coefficients are necessarily correlated, we choose not to incorporate this correlation in our prior structure, principally because the prior for β_1 depends on other regression coefficients. Although estimates may be slightly more efficient by including more information, we believe that independent priors are sufficient in this case. Additionally, it is worth noting that because the regression coefficients are defined on half the real line (β_0 and β_1) and a closed interval (β_1), other prior distributions could be chosen. For instance, the Gamma distribution is defined on $(0, \infty)$, so theoretically it could be used as a prior distribution for $-\beta_0$ or $-\beta_2$. Similarly, the Beta distribution could be considered for β_1. However, our testing of these priors showed problems with their use. The sampling generally exhibited a high amount of autocorrelation and/or slow convergence, which is not the case with the truncated normal distributions.

Because P_c values can assume very small values, including the value of 0 to machine precision, using these data in an unbounded way introduces a very large dynamic range in the observed values. Operationally, there is little interest in events with a P_c below 1E-07 and essentially none with a P_c below 1E-10; so it is quite reasonable to truncate (left-censor) the dataset by resetting the values of P_c data $<$ 1E-10 to the 1E-10 value. Of course, in such a case one must accept the cognitive dissonance of the model predicting P_c values less than 1E-10. However, this is acceptable to us for a few reasons. One reason is that we are mainly concerned with finding the peak y value and knowing with some certainty when it will occur. The other reason is more practical: we are not particularly concerned with prediction for smaller values of y. Because the y values represent orders of magnitude, we are far less worried about prediction error for small values of y than we are for large values of y.

Lastly, we admit that our model cannot capture the rare occurrence that the y values initially decrease and then increase, i.e. an upward opening parabola. We do not concern ourselves with this case, as in such a case our model would fit essentially a horizontal line, indicating no discernible peak value. Though the shape of the data is not preserved, our end goal is: we seek significant statistical evidence of the size and location of the peak, and in this situation its size and location are unclear.

Inference for the Peak Value The supposed canonical behavior suggests that the order of magnitude of the P_c value increases as the uncertainty decreases and drops off after a certain point. In general, uncertainty tends to decrease with time. Thus, we expect that this relationship holds with reference to time as well. Though some events exhibit this behavior, many events only exhibit the decline in order of magnitude of the P_c value. That is, if we believe the $\log_{10} P_c$ truly increases in time initially, this increase is censored within many events–the earlier small $\log_{10} P_c$ values lie outside of the 7-day screening window or outside of the physical screening volume and were thus not reported. Similarly, because we are only able to observe a few $\log_{10} P_c$ values, we are unlikely to observe the true peak. Thus, it is difficult to measure the accuracy of any prediction of the peak we might make. Because we are not certain of being able to observe the true peak, we take the highest observed $\log_{10} P_c$ value to be the peak.

We can infer the distribution of the location of the peak by utilizing the well-known identity that the peak is located at $x_{max} = -\beta_1/2\beta_2$. We estimate this distribution by collecting the posterior samples of β_1 and β_2 from the MCMC output, and transforming them as x_{max} is defined. From the empirical distribution of x_{max}, we can compute a point estimate and a 95% credible interval for x_{max}. For the point estimate, we utilize the posterior mode of x_{max}, which is found by fitting a kernel density to the samples of x_{max} and finding the most likely value. For the credible set, since we define time as time until TCA, we are mainly concerned with the lower bound. Here, the lower bound represents, with 95% probability, the latest time at which our model predicts a peak will occur. This is operationally useful, as one is often interested if the peak will occur before 48 hours until TCA. Thus, if we can say that the peak will occur before this time with 95% probability, then the operator may be able to use this information to make a more informed decision regarding the importance of continuing to follow the event. Similarly, we can construct bounds for the magnitude of the peak. The distribution for the magnitude of the peak y_{max} is computed in the same way as for the location but instead using the transformation $y_{max} = \beta_0 - \beta_1^2/4\beta_2$.

NUMERICAL RESULTS

Data

The dataset used for tuning ($i.e.$, setting the parameters for the informative prior distributions) and testing the model is taken from the NASA Conjunction Analysis and Risk Assessment historical CDM database. One thousand events' worth of data from calendar year 2013 was used for model tuning, and the tuned model was evaluated against approximately 3000 events from 2014; so there was no overlap in terms of time-period or actual data between the two datasets. Data were taken from conjunctions against primaries in the orbital region defined by a perigee height between 500 and 750 km and an eccentricity less than 0.25. As described above, data flooring at a $\log_{10} P_c$ value of -10 was performed on the dataset; leading or trailing values of -10, except for the values directly before a value higher than this, were also eliminated. The removal of large groups of $\log_{10} P_c$ data fixed at this -10 value at the beginning or end of an event makes sense intuitively, as such a situation greatly hampers reasonable curve-fitting in the more interesting part of the response; and it also is reasonable operationally, as there is no operational interest in values so small. To qualify for use in tuning or evaluation, an event must have had at least two CDMs with a $\log_{10} P_c$ greater than -10.

Simulation Setup

We test our model on an archive of past conjunctions. To estimate the hyperparameters for our priors, we select 1000 events from the tuning dataset and run a least squares fit using the restricted parameter space. For each event, we calculate the maximum likelihood estimate (MLE) of each of the parameters. Using the 1000 regression estimates, we match the mean and variance of their distribution to a normal distribution with the same mean and variance. We do the same for the estimates of the precision and the Inverse-Gamma distribution. Our model is implemented in JAGS ("Just Another Gibbs Sampler," a widely-used MCMC utility).[15]

The prediction procedure for a given event is as follows. We attempt to make predictions for the peak y value only after the second received CDM. We are interested in estimating the maximum $\log_{10} P_c$ value y_{max} and its location t_{max}. For each of these values, we take the predicted value to be the mode of the predictive distribution of each of these quantities' distributions. We also record lower and upper bounds from 95% credible sets, estimated by calculating the empirical $2.5th$ and $97.5th$ percentiles of said distributions. These bounds can be used to calculate coverage.

In addition to coverage, we are also interested in the location of some of these bounds. For instance, as mentioned earlier, if the lower bound of the location of the peak is greater than 2, then the operator can say with 95% probability that the peak will have passed by the time a decision must be made. Additionally, one may be interested in the magnitude of the peak or the next observation. In practice, one is often not greatly concerned operationally with $\log_{10} P_c$ values that are -7 or less and may consider values between -7 and -4 to be worthy of continued monitoring. Thus, an upper bound of -8 for the peak would suggest to the operator that there is a 95% probability that the $\log_{10} P_c$ values will never exceed -8, which would add statistical evidence to support the low-risk assessment of the event. Conversely, a high upper bound may suggest that the event ought to be continued to be followed until the risk has decreased to what the operator considers to be an acceptable level.

A note on the informativeness of the priors Although we choose our priors to be informative, we do not wish them to be so informative that they overwhelm the data. In particular, the prior structure should not be such that the posterior estimates of the parameters are the same, regardless of the data. To see that our priors are not overwhelmingly informative, we present their specific distributions below.

$$
\begin{aligned}
&\beta_0 \sim Normal(-11.39, 33.33) I_{(-\infty,0)} \\
&\beta_1 \sim Normal(1.11, 12.5) I_{(-2\sqrt{\beta_0\beta_2}, 2\sqrt{\beta_0\beta_2})} \\
&\beta_2 \sim Normal(-0.025, 0.54) I_{(-\infty,0)} \\
&\sigma^2 \sim InverseGamma(0.94, 1.02)
\end{aligned}
$$

We can interpret these priors in terms of the transformed parameters of interest. Specifically, we check the induced distributions of the peak y value and its location. To do this, we draw 10,000

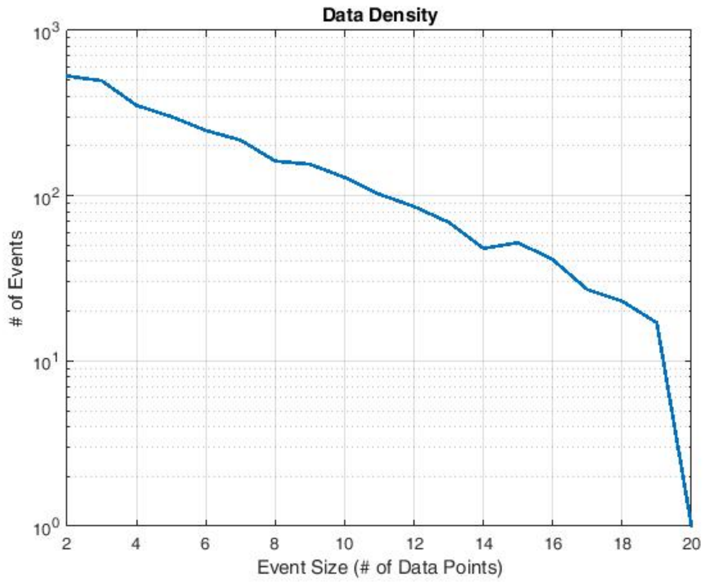

Figure 2.

Monte Carlo samples from each of the distributions independently (as we have assumed their independence in the prior structure), and check the distributions of the transformed parameters. We find that the distribution of the location x_{max} has a mean of 0.56 with a lower 2.5% quantile of -7.91 and an upper 97.5% quantile of 9.56. Similarly, we find that the distribution of the peak y_{max} has a mean of -8.93 with a lower 2.5% quantile of -20.60 and an upper 97.5% quantile of -0.47. Lastly, we find that the variance has a mean of 1.90 with a lower 2.5% quantile of 0.53 and an upper 97.5% quantile of 7.07. In practice, we find these priors to be sufficiently diffuse, so that they do not overwhelm the data in the posterior distributions.

Results

We present some operating characteristics of our model in terms of number of CDMs observed and time until TCA. It is worth noting that since we test our model on historical data, the number of events with any distinct number of CDMs is not under our control. In Figure 2, we plot the frequency of events with varying counts of CDMs. From the figure, it is clear that events with a higher number of CDMs are more rare. In fact, there are only $10^{1.4} \approx 25$ events with 16 CDMs, and fewer events with more CDMs. We must bear in mind this sparsity as we interpret the results, as it may introduce a bias.

Since our main interest is peak prediction, we first consider the bias of our estimator for y_{max}. From Figure 3, we can see that all predictions of the peak are within half an order of magnitude, regardless of how many CDMs have been received. Furthermore, most of the biases are negative, implying that the model is usually predicting a peak higher than the actual observed peak. Though this feature was not planned for, a conservative estimate of the maximum $\log_{10} P_c$ is operationally desirable, as one would rather be too pessimistic than too optimistic in terms of risk assessment. One feature worth noting is that events with more CDMs tend to have a positive bias when more CDMs are received. This is a result of the simplicity of the model and the nature of the data itself.

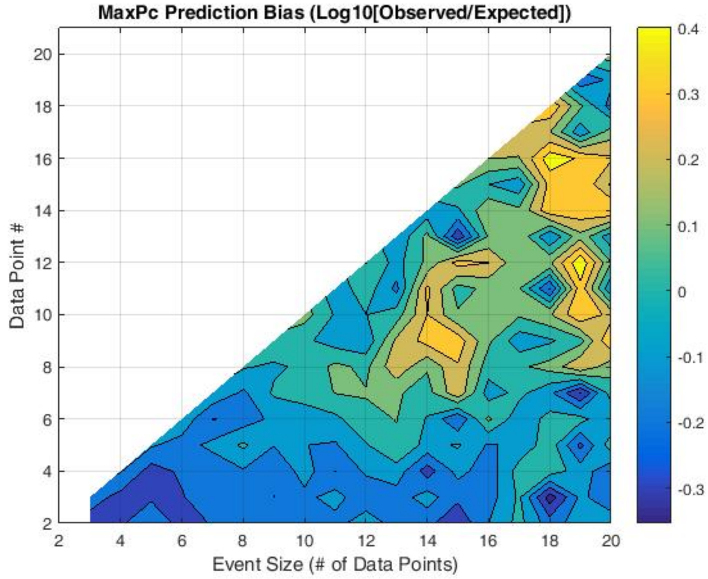

Figure 3.

As one observes more CDMs, one is more likely to observe low $\log_{10} P_c$ values or $\log_{10} P_c$ values of the floor value of -10, so that an event with many CDMs will have a few higher $\log_{10} P_c$ values near the beginning of observation time and many low $\log_{10} P_c$ values afterwards. This has the effect of "flattening out" the parabola, and pulling the vertex slightly below the peak value. Thus, for longer events, the parabola model is less likely to capture the true behavior of the $\log_{10} P_c$ values, although it still provides a reasonable estimate of y_{max}; and this flattening has only a subsidiary effect on the estimate of peak location.

In addition to the parabola model possibly being too restrictive for longer events, we also note that we have fewer events with a large number of CDMs, so that the results for such cases may also be less credible. In Figure 4, we restrict our attention to event sizes for which there were a larger number of instances in our database. Restricting the display to events with 13 CDMs or fewer guaranteed at least forty events for each event size level. As before, we note that we are generally conservatively predicting the peak. We also note that for this subset of the simulation, the prediction bias is within 0.3 orders of magnitude.

Proper prediction of the peak value, rather than the temporal peak location, was never the expected principal application of the model; but it is more natural to treat it here, as it follows naturally from the preceding discussion of peak prediction bias and is a reasonable measure of overall model fit. Figure 5 plots the 50th percentile of the $\log_{10} P_c$ residual absolute values by event size and number of CDMs received. We may interpret this figure as a central measure of how far in magnitude the predicted y values differ from the observed values. Overall model performance is reasonable, as half the predicted values fall within 0.6 of an order of magnitude. However, the operational utility of this aspect of the model ($i.e.$, ability to predict the actual $\log_{10} P_c$ value of the next CDM) will be governed by the upper-tail residual performance. Figure 6 shows these results for the 95th percentile residual set. Most of the results-space shows at least a one and one-half order-of-magnitude 95th percentile residual size, with some peaks as high as 2.5 to 3 orders of magnitude. Such results do

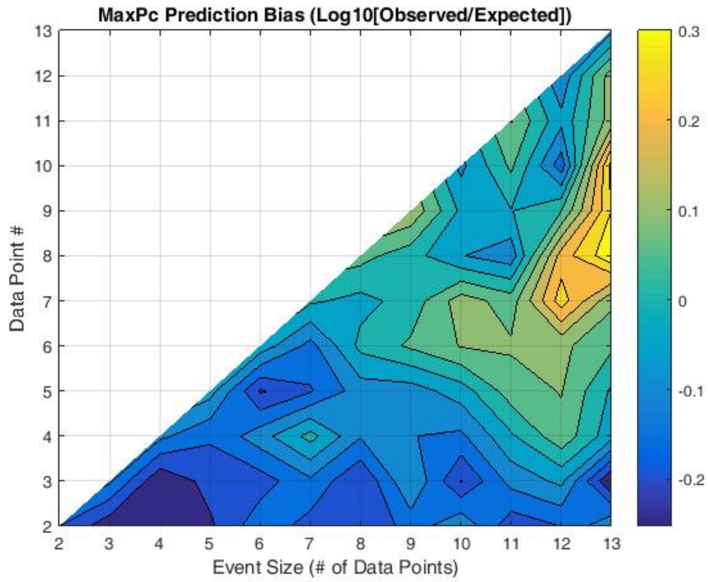

Figure 4.

not indict the model *per se*, but they are large enough probably to make this application of the model not suitable for operational use. In making Pc predictions, an order of magnitude is about as large an uncertainty as can be tolerated in order for the predicted value to be useful for decision-support.

Fortunately, the principal application of the model was not to predict the Pc peak value but rather Pc peak location; and these results are much more encouraging. The best way to evaluate the model's ability to predict the peak location is to walk through each event (CDM by CDM), obtain the model's prediction of the peak location based on the event data received so far, and perform a binary evaluation of this prediction. More specifically, as each CDM from a single event is fed to the model, the model determines whether the peak Pc has already occurred; and this prediction is compared to the actual historical outcome for that event, namely whether the peak Pc had actually occurred by that point. Figure 7 shows the results of this investigation. The color represents the percentage of the time that the model correctly predicted whether the peak Pc had already occurred. For example, for an event that will eventually receive ten CDMs, by the 6th CDM the model will correctly predict 60% of the time whether the peak Pc has already passed; and by the 8th CDM it will predict correctly 70% of the time.

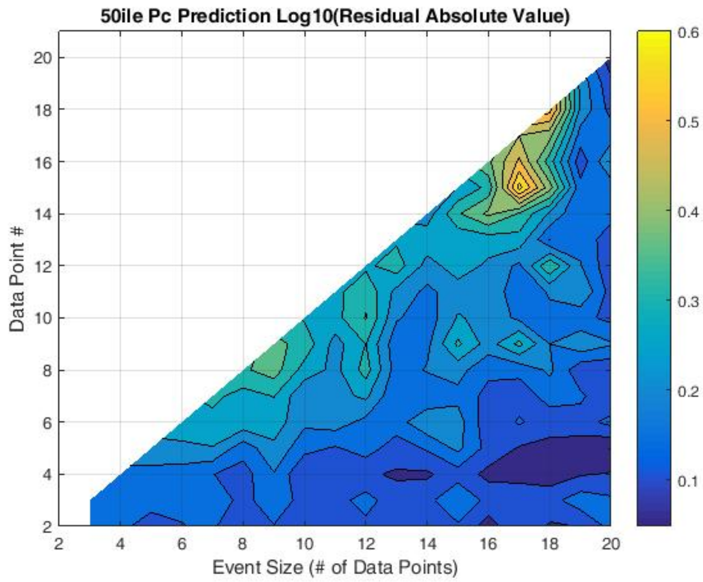

Figure 5.

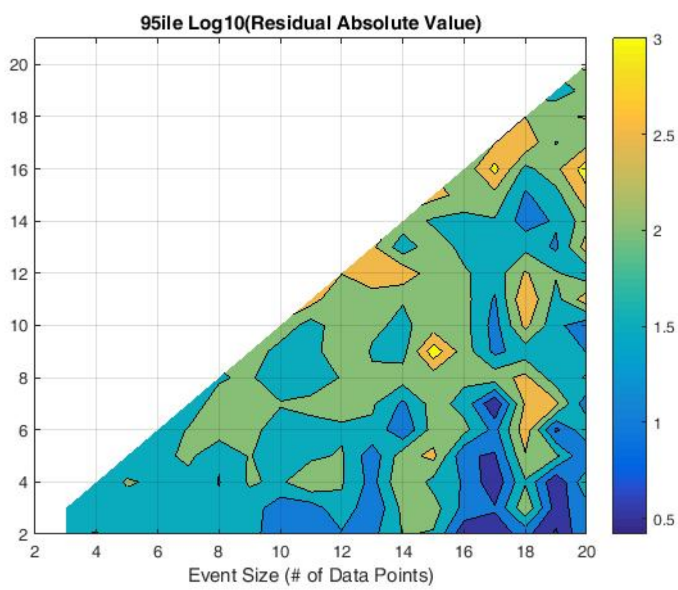

Figure 6.

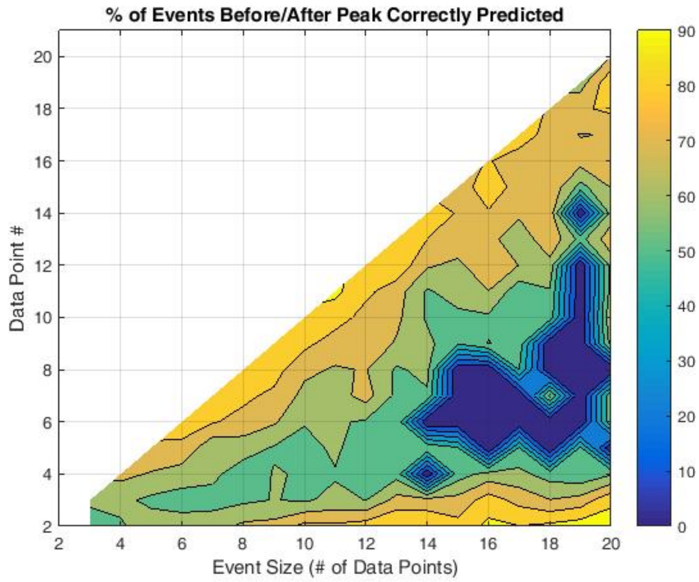

Figure 7.

It is somewhat difficult from these response statistics as framed to determine the operational utility of the model because, in the midst of a developing event, the number of CDMs that will be generated is not known and the relationship of the number of OCMs recieved to the time at which a satellite conjunction mitigation decision must be made is not established. These response statistics were thus reformulated to tabulate the model's performance in predicting peak location as a function of the number of days to each event's TCA. The same binary evaluation approach as used in Figure 7 is used here in Figure 8, but the results are shown as a bar graph, in which the blue bars show the percentage of cases in which the peak's passage is properly predicted and the yellow bars show the percent of events for which results at that particular time point are available. An event might not be "processable" at a certain point because two CDMs were not yet received by that particular time point ($e.g.$, it is not particularly unusual, for example for events six days from TCA not yet to have received any CDMs) or because the MCMC failed to converge. Typically, final conjunction mitigation decisions are made 2 to 3 days before TCA, and some systems allow the luxury of only a day's lead-time being necessary. We can see from the graph that in this time window (within 3 days of TCA) 90% or greater of the cases are processable, and of these the success rates of the model range from 75% to 85%.

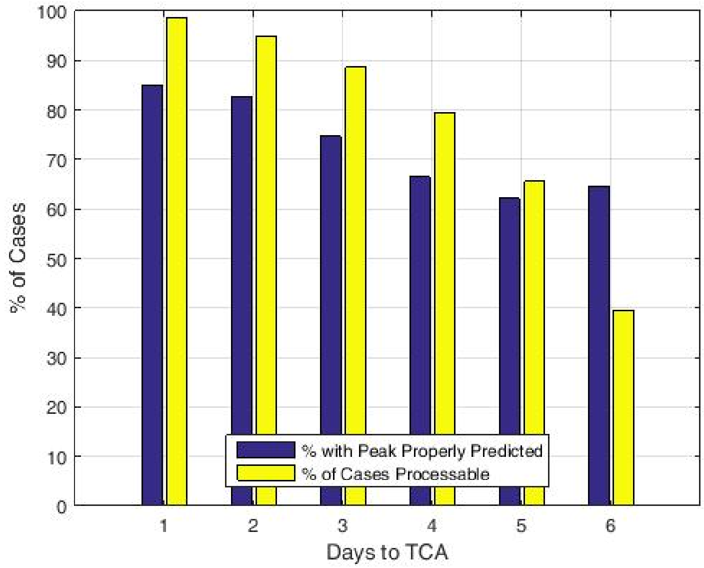

Figure 8.

These results are promising enough to suggest a more in-depth study of the approachs potential operational utility. By the two-day point, almost 80% of the cases appear to have a correct decision rendered on whether the P_c peak has passed and for which the P_c will only continue to decrease with time. In such a case, if the P_c value were very close to but not in violation of a threshold for action, one could with additional confidence close out such an event to further monitoring. While of course a result from a single test such as this cannot serve as the principal or perhaps even a first-tier risk assessment criterion, it does contribute to the cumulative case for certain types of mitigation responses. It certainly supports the positing of the Bayesian inferential method as a promising approach to this type of decision-support problem.

As part of the kind of operational suitability evaluation described above, one would wish to characterize different event history types in terms of the behavior of P_c versus time to TCA and evaluate the tools performance against those different morphologies. Some of these are straightforward and easy to evaluate, such as an initially high P_c dropping to essentially zero at four days to TCA and remaining at that level until the event expires; and others are more challenging, such as a schizophrenic P_c that bounces around significantly until about three days until TCA and then falls off only slightly through the end of the event. If the tool performs well against only the former situation and hardly exceeds a predictive capability of 50% for the latter, then it has not added any significant operational value. The algorithm should also be evaluated against operator rules of thumb that inhere in most situations and thus guide the present heuristic decisions, such as the observation that events with a P_c of X at three days to TCA are unlikely further to degrade. The tool need not necessarily outperform such heuristics in order to be useful, as it provides a technical foundation for behaviors that inhere but are presently understood only intuitively; but a truly helpful tool would exceed the capabilities of the current operational practices.

CONCLUSION

In this paper, we investigated the performance of a simple statistical model for inferring the location and magnitude of the peak $\log_{10} P_c$ value. We chose to use the Bayesian paradigm in order to incorporate information from past events. This allowed us a natural way to update the model and to incorporate the uncertainty of the model parameters and by both of these recover to some degree the "canonical" behavior believed to underlie most events. Our results show that the model is adept at estimating the peak $\log_{10} P_c$ location, which is quite likely to be operationally useful if its capabilities expand reasonably to non-trivial cases.

Our future work is focused on utilizing more information in a model which can more credibly predict the next $\log_{10} P_c$ value, and we intend to explore more powerful methods for borrowing information across events. Once two or three models have been defined and exist in prototype, a comprehensive cross-comparison of model power, with all competing directly against analyst rules of thumb and other intuitive methods, can be accomplished to determine the true operational utility of any of these models and their results taken together.

NOTATION

P_c	Probability of collision
y_{ij}	The log probability of collision for the i^{th} event at the j^{th} CDM
y_{max}	The maximum log probability of collision for the i^{th} event
t_{ij}	Time until TCA for i^{th} event at the j^{th} CDM
t_{max}	Time until TCA for the maximum log probability of collision for the i^{th} event
CDM	Conjunction data message
TCA	Time of closest approach

REFERENCES

[1] R. C. Frigm, J. A. Levi, and D. C. Mantziaras, "Assessment, planning, and execution considerations for conjunction risk assessment and mitigation operations," *Proceedings of SpaceOps 2010 Conference: Delivering on the Dream, Huntsville, Alabama*, 2010, pp. 25–30.

[2] M. R. Akella and K. T. Alfriend, "Probability of collision between space objects," *Journal of Guidance, Control, and Dynamics*, Vol. 23, No. 5, 2000, pp. 769–772.

[3] R. P. Patera, "General method for calculating satellite collision probability," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 4, 2001, pp. 716–722.

[4] K. Chan, "Improved analytical expressions for computing spacecraft collision probabilities," *Advances in the Astronautical Sciences*, Vol. 114, 2003, pp. 1197–1216.

[5] X. Xu and Y. Xiong, "A Method for Calculating Collision Probability Between Space Objects," *arXiv preprint arXiv:1311.7216*, 2013.

[6] J. R. Carpenter, F. L. Markley, and D. Gold, "Wald Sequential Probability Ratio Test for Analysis of Orbital Conjunction Data," *AIAA Guidance, Navigation, and Control (GNC) Conference*, 2013.

[7] J. R. Carpenter, F. L. Markley, and D. Gold, "Sequential Probability Ratio Test for Collision Avoidance Maneuver Decisions," *The Journal of the Astronautical Sciences*, Vol. 59, No. 1-2, 2012, pp. 267–280.

[8] J. R. Carpenter, F. Markley, K. Alfriend, C. Wright, and J. Arcido, "Sequential probability ratio test for collision avoidance maneuver decisions based on a bank of norm-inequality-constrained epoch-state filters," *AAS/AIAA Astrodynamics Specialist Conference, Girdwood, AK*, 2011.

[9] F. K. Chan, *Spacecraft collision probability*. Aerospace Press El Segundo, CA, 2008.

[10] S. Alfano, "Relating Position Uncertainty to Maximum Conjunction Probability©," 2005.

[11] N. Ravishanker and D. K. Dey, *A first course in linear model theory*. CRC Press, 2001.

[12] K. Lange, *Numerical analysis for statisticians*. Springer Science & Business Media, 2010.

[13] A. E. Gelfand, A. F. Smith, and T.-M. Lee, “Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling,” *Journal of the American Statistical Association*, Vol. 87, No. 418, 1992, pp. 523–532.

[14] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*, Vol. 2. Taylor & Francis, 2014.

[15] M. Plummer *et al.*, “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling,” *Proceedings of the 3rd international workshop on distributed statistical computing*, Vol. 124, Vienna, 2003, p. 125.