Conjunction Assessment Risk Analysis

Peak Pc Prediction in Conjunction Analysis

J.J. Vallejo, a.i. solutions
M.D. Hejduk, Astrorum Consulting
J.D. Stamey, Baylor University
The Problem

- Satellite conjunction risk typically evaluated through the probability of collision (Pc)
  - Considers both conjunction geometry and uncertainties in both state estimates
- Conjunction events initially discovered through JSpOC screenings, usually seven days before Time of Closest Approach (TCA)
  - However, JSpOC continues to track objects and issue conjunction updates
  - Changes in state estimate and reduced propagation time cause Pc to change as event develops
  - These changes a combination of potentially predictable development and unpredictable changes in state estimate / covariance
- Operationally useful datum: the peak Pc
  - If it can reasonably be inferred that the peak Pc value has passed, then risk assessment can be conducted against this peak value
  - If this value below remediation level, then event intensity can be relaxed
- Can the peak Pc location be reasonably predicted?
Conjunction Event “Canonical Progression”

- Conjunction typically first discovered 7 days before TCA
  - Covariances large, so typically Pc below maximum
- As event tracked and updated, changes to state estimate are usually relatively small, but covariance shrinks
  - Because closer to TCA, less uncertainty in projecting positions to TCA
- Theoretical maximum Pc encountered when 1-sigma covariance size to miss distance ratio is $1/\sqrt{2}$
  - After this, Pc usually decreases rapidly
- Behavior shown in graph at right
  - X-axis is covariance / miss distance
  - Y-axis is $\log_{10} (P_c/\max(P_c))$
  - Order of magnitude change in Pc considered significant, thus log-space more appropriate
- How might this behavior be modeled?
  - Underlying progression in presence of noise
Proposed Choice of Modeling Variables

• **Dependent variable is log10 value of Pc**
  – Need to address problem of very small and 0 values for Pc
  – Majority of Pc values for purposes of operations “essentially 0”: < 1E-10
    • Small values of Pc can be “floored” at 1E-10
    • Furthermore, long trains of leading or trailing 1E-10 values can also be eliminated from dataset for model tuning and evaluation; really just a function of when updates happen to occur.

• **Independent variable is time before TCA (usually in fractional days)**
  – Canonical behavior curve uses independent variable as ratio of covariance size to miss distance
  – Problematic independent variable for fitting
    • Not monotonic with time (but it does correlate at least moderately to time)
    • Need temporal independent variable in order to map to operational timelines
  – Thus, use time before TCA as independent variable for model
Bayesian Vertex Model

- Approximate theoretical progression of $\log(P_c)$ values using a downward-opening parabola
  - Equation in vertex form: $Y = a (x - h)^2 + b$
  - Can be recast as: $Y = \beta_0 + \beta_1 x + \beta_2 x^2$
  - Location of peak more important than peak value, so need not match functional form precisely
- With regression analysis of training dataset, can establish prior distributions of set of $\beta$ values
- Drawing from these priors, can use Bayes’ theorem to construct posterior distributions
  - This allows priors to be combined with unfolding data from current event
  - Can then estimate $\log(P_c)$ from mean values from parameter posteriors
Using Frequentist Methods

- If we refit the line each time we receive a new OCM using frequentist methods (i.e. least squares), we would see something like this.
Using Bayesian Methods

- If we use Bayesian methods, it is possible to incorporate prior information into the estimates.
Comparisons Between the Bayesian and Frequentist Models

- Using the Bayesian methods, we can make predictions using only two OCMs (though generally these are not particularly informative)
  - This is not possible with the frequentist model
- The frequentist model fits the points as closely as possible, whereas the Bayesian model incorporates prior information, compromising between the current and previous data
- The fits are generally similar, but the Bayesian fit is generally more conservative
  - The Bayesian model takes into account the uncertainty of the estimates, thus it is less likely to fit the data “too well”
  - As a result, the Bayesian model generally has wider error bounds, which are usually more realistic
  - The frequentist approach tends to chase the action, whereas the Bayesian approach is more realistically predictive (because it considers prior information)
• We can calculate what is known as the posterior density of the parameters given the data
  
  – \( p(\beta|y) \propto p(y|\beta) \times p(\beta) \)
  
  – Thus, we specify a prior distribution for the beta parameters \( p(\beta) \), update it with the data that we have seen \( p(y|\beta) \), and get an updated probability distribution of the beta parameters given the data \( p(\beta|y) \)

• Now, we can force the parabola to open downwards by choosing priors the allow only this shape
  
  – Consider the model \( Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \), where \( \varepsilon \) is the noise in the measurement
  
  – If we force \( \beta_0 \) and \( \beta_2 \) to be negative, this will ensure a downward opening parabola will be fit each time and ensure that the vertex be realizable (e.g., not have a y-value greater than 1, which is not possible for a Pc value)
  
  – This presents one potential hazard with the model: what if the observed data actually had the shape of an upward opening parabola? It would be fit with a horizontal line, which is not the correct shape
• The resulting constraints are
  \[
  \beta_0 < 0 \\
  (\beta_1)^2 < 4 \beta_0 \beta_2 \\
  \beta_2 < 0
  \]

• In order to specify these priors, we use truncated Normal distributions, so that
  \[
  \beta_0 \sim \text{Normal}(\mu_0, \sigma_0^2) \mathcal{I}(-\infty, 0) \\
  \beta_1 \sim \text{Normal}(\mu_0, \sigma_0^2) \mathcal{I}(-2\sqrt{\beta_0 \beta_2}, 2\sqrt{\beta_0 \beta_2}) \\
  \beta_2 \sim \text{Normal}(\mu_2, \sigma_2^2) \mathcal{I}(0, \infty)
  \]

• While other prior distributions are possible, we find that the truncated normal have the best convergence properties
  – Gamma distributions were also attempted but exhibited high levels of autocorrelation and overall slow convergence
• Assume that $\varepsilon \sim \text{Normal}(0, \sigma^2)$
• Assume a Gamma prior on the inverse of the variance $1/\sigma^2$
  – Common practice.
• Choosing the parameters of these prior distributions
  – Use restricted maximum likelihood to estimate downward opening parabolas
    on a set of test data (we examined over 1000 events)
  – Collect all of the betas from the fits
  – Find parameters of a truncated normal distribution that is close to the observed
    distribution of each parameter by matching quantiles
• Conjunction data archive assembled for 2013-14 for well-populated orbit regime
  – Perigee height between 500 and 750 km and eccentricity < 0.25
  – Thousands of events per year
• Use part of 2013 data to “train” model—set prior distribution coefficients
• Use 2014 data as validation dataset
• Segregate performance results
  – First, by total number of data points (CDMs) in the event
    • Data-poor events may perform worse than data-rich ones
  – Second, by data point number
    • How does model perform after point 3 versus point 6 or 10?
• Probably want at least 50 events surveyed to feel confident about model performance conclusions
• This achieved only for event sizes smaller than 14 data points
• Should focus on performance results for these shorter events—sampling more plentiful
Bayesian Vertex Model: Mean Peak Estimation Error

- mean(Y – Yhat) for all the events of each size
- Value becomes unstable beginning at event sizes of about 13 observations
- Stable region shows mean values ranging from around 0 +- half an order of magnitude
- Model is biased but biased in a favorable direction
  - Overpredicting leads to conservative safety-of-flight decisions—better than the reverse
Bayesian Vertex Model: 50th and 95th percentile Peak Absolute Residual Errors

- Focus on more stable region (event sizes of 13 or fewer points)
- At the 50th percentile all of that area is less than 0.5 of an order of magnitude
  - An acceptable result
- At the 95th percentile, that area varies between 0.5 and 3 orders of magnitude
  - Probably not an acceptable result
- Model probably not useful for peak prediction
- However, could still be useful for predicting whether peak has occurred
Bayesian Vertex Model:
Peak Prediction Performance

• Operational question: has the event reached its peak Pc value?
• Plot at right shows, for all events of a certain size after a certain data point, the percent correct peak predictions
  – % of the time the model indicates the peak has already passed, and in fact it has
• In region of interest (< 14 data points), performance always better than 50% once half the event points received
  – Performance moves to 80-100% as number of points reaches total event size
• However, difficult to use result, since # of points not known in advance
  – Examine predictive force at “times to TCA” of operational interest
Bayesian Vertex Model: Peak Prediction Performance (cont’d)

- Examine situation at typical maneuver planning and commit times
  - 4, 3, 2, and 1 days before TCA
- Blue bars show percentage of correct before/after peak predictions at these time points
- Yellow bars show number of events for which prediction was possible
  - At least two points needed
  - MCMC fails to converge occasionally
- Not stunning performance, but could be an operational tool of some utility
Conclusion/Future Work

- A simple statistical model shows operational promise in determining whether the peak Pc value has occurred
- Additional areas requiring exploration
  - Event Pc histories need categorization
    • May be that algorithm performs well only in “obvious” cases; may not be helpful more ambiguous situations where greater operational need
  - Different overall functional forms may yield better results
    • For instance, the log probabilities of collision are effectively bounded between -10 and 0, suggesting a different distribution (Beta) may be more appropriate
  - Other modeling paradigms
    • Other ways of borrowing information, e.g. mixed models
  - Longitudinal data analysis, because the observations are repeated measurements on different events