technical note technical note

Analysis of Likelihood Parameter Changes in the En Route Automation Modernization

Andrew Crowell, ANG-C41

Brian S. Schnitzer, GDIT

April 2013

DOT.FAA/TC-TN13/13

Document is available to the public
through the National Technical Information
Service, Springfield, Virginia 22161

U.S. Department of Transportation

Federal Aviation Administration

William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof. The United States Government does not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the objective of this report. This document does not constitute FAA certification policy. Consult your local FAA aircraft certification office as to its use.

This report is available at the Federal Aviation Administration William J. Hughes Technical Center's Full-Text Technical Reports page: actlibrary.tc.faa.gov in Adobe Acrobat portable document format (PDF).

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Technical Report Documentation Page

1. Report No. DOT.FAA/TC-TN13/13	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Analysis of Likelihood Parameter Changes in the En Route Automation Modernization		**5. Report Date** April 2013 **6. Performing Organization Code** ANG-C41
7. Author(s) Andrew Crowell and Brian S. Schnitzer		**8. Performing Organization Report No.** DOT.FAA/TC-TN13/13
9. Performing Organization Name and Address U. S. Department of Transportation Federal Aviation Administration, William J. Hughes Technical Center Atlantic City International Airport, NJ 08405		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.**
12. Sponsoring Agency Name and Address U. S. Department of Transportation Free Flight Phase 1 Program Office, AOZ-200 Conflict Probe/URET Product Team Washington, D. C. 20590		**13. Type of Report and Period Covered** Technical Note **14. Sponsoring Agency Code** DOT

15. Supplementary Notes

The authors identified above represent the following organization: Andrew Crowell with ANG-C41 Concept Analysis Branch and Brian Schnitzer with GDIT

16. Abstract

An ongoing development and analysis task is currently being performed by the Separation Management: Modern Procedures team of the Federal Aviation Administration (FAA). In s study performed in 2011, a conclusion was made that major improvements can be realized by changing parameters of the current probe and that the likelihood threshold is a major factor. It found that the likelihood function is being underutilized at its current setting and has the potential to greatly reduce rate of the False Alerts (FAs). The study proposed an experiment to focus on the effect of changing the likelihood threshold function. An experiment with this objective has been carried out and is documented in this technical note. The current work used knowledge gained from prior experiments to construct a design of experiment that can provide recommendations for likelihood threshold settings. It was found that changing the middle parameter of the likelihood threshold function has a nearly inverse effect on False Alert performance and warning time performance. When the False Alert performance is improved, warning time performance is degraded, and vice versa. The result is that a recommendation is made to use a likelihood threshold function that provides a balance to False Alert and warning time performance, but still greatly improves upon the False Alert performance from the current live system.

17. Key Words Separation Management, Likelihood, En Route Automation Modernization, ERAM, Conflict Probe, Trajectory Prediction, conflict probability	**18. Distribution Statement** This document is available to the U.S. public through the National Technical Information Service (NTIS), Springfield, Virginia 22161. This document is also available from the Federal Aviation Administration William J. Hughes Technical Center at actlibrary.tc.faa.gov.

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 36	**22. Price**

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Executive Summary

An ongoing development and analysis task is currently being performed by the Separation Management: Modern Procedures team of the Federal Aviation Administration (FAA), which consists of a mix of FAA employees and contract staff. In 2011, a study was performed that evaluated the performance gains that could be acquired by changing parameter settings of the current version of the En Route Automation Modernization's Conflict Probe. One of the parameters of this study was the likelihood threshold function.

The conclusions of the first study were that major improvements can indeed be realized by changing parameters of the current probe and that the likelihood threshold is a major factor. It found that the likelihood function is being underutilized at its current setting and has the potential to greatly reduce rate of the False Alerts (FAs). The previous study could not make a recommendation for a setting, but proposed an experiment to focus on the effect of changing the likelihood threshold function.

An experiment with this objective has been carried out and is documented in this technical note. The current work used knowledge gained from prior experiments to construct a design of experiment that can provide recommendations for likelihood threshold settings. Only the middle parameter of the piecewise linear likelihood threshold function was altered in the experiment. It was found that changing this middle parameter has a nearly inverse effect on False Alert performance and warning time performance. When the False Alert performance is improved, warning time performance is degraded, and vice versa.

An analysis was performed that provided the distribution of likelihood values at different predicted warning times. The analysts hoped to discover a pattern that could provide additional insight into efficient settings of the likelihood value. The results of the analysis indicated that, though the likelihood function can provide a significant improvement in its current state, there may be room for improving the function even further. Several cases were discovered that indicated the likelihood function performed as designed, but the results also showed there are many opportunities for improvement. These suggested improvements are beyond the scope of this study, which focused primarily on parameter changes, and require detailed examination of the underlying algorithm.

With the likelihood algorithm in its current state, and restricted to parameter changes alone, it is recommended to continue using the 4|8|20 likelihood threshold function that had previously been recommended, as changing the threshold function from 10|20 reduces FAs by 10% and Late Alerts by 4% [Crowell, et al, 2012a]. In comparison with the baseline 10|20 Likelihood with 2.5 Lateral and 1.5 Longitudinal, using the 4|8|20 Likelihood with 1.0 Lateral and 1.25 Longitudinal reduces the False Alerts by 37% and the Late Alerts are increased by 7%. The 7% increase to Late Alerts translates to an increase of a single Late Alert and is attributable to error in the model.

Table of Contents

EXECUTIVE SUMMARY ..... **VII**

TABLE OF CONTENTS ..... **VIII**

LIST OF FIGURES ..... **IX**

LIST OF TABLES ..... **X**

1 INTRODUCTION **1**

1.1 PREVIOUS WORK 1

2 DESCRIPTION OF EXPERIMENT **2**

2.1 SCENARIOS 4

3 EXPERIMENT RESULTS **5**

4 LIKELIHOOD FUNCTION ANALYSIS **13**

4.1 FLIGHT EXAMPLES 19

4.1.1 Example 1 – Inappropriately High Likelihood *19*

4.1.2 Example 2 – Appropriately High Likelihood *20*

4.2 SUMMARY OF LIKELIHOOD FUNCTION ANALYSIS 22

5 CONCLUSION & RECOMMENDATIONS **23**

6 LIST OF ACRONYMS AND ABBREVIATIONS **24**

REFERENCES **25**

List of Figures

Figure 1. Piecewise linear functions of the likelihood experiment design. 4
Figure 2. Warning time (color) for each middle setting of the likelihood function. 7
Figure 3. False Alerts (color) for each middle setting of the likelihood function. 8
Figure 4. Prediction profiler showing the 4|8|20 likelihood used in Experiment 3. 9
Figure 5. Prediction profiler modeling the best setting for FA performance. 10
Figure 6. Prediction profiler modeling the best setting for warning time performance. 11
Figure 7. Prediction profiler modeling the setting that best balances warning time and false alert performance. 12
Figure 8. Distribution of likelihood values with respect to predicted warning time, colored by Fuzzy FA 14
Figure 9. 10|20 baseline likelihood function overlaid on a distribution of likelihood values. 15
Figure 10. 4|8|20 likelihood function overlaid on a distribution of likelihood values. 16
Figure 11. Valid Rate grouped by likelihood value and PWT. 17
Figure 12. 4|8|20 likelihood function overlaid on a distribution of uniform random values. 17
Figure 13. Distribution of differences in FAs between the likelihood values and the uniform random values. 19
Figure 14. Two aircraft shown at the time of prediction with a likelihood value of 1.0. 20
Figure 15. Two aircraft shown at the time of prediction with a likelihood value of 1.0. 21

List of Tables

Table 1. D-optimal design for a likelihood experiment modifying only the middle value of the likelihood threshold function. 3
Table 2. Results of the ZAU runs. 5
Table 3. Warning time performance of the 16 runs. 6
Table 4. Comparison between application of the 4|8|20 function to a distribution of likelihood values and to a distribution of uniform random values. 18

1 Introduction

Functional Area 18 (FA18) is an ongoing development and analysis task being performed by the Separation Management: Modern Procedures team of the Federal Aviation Administration (FAA), which consists of a mix of FAA employees and contractors. In 2011, a study was performed that evaluated the gains that could be acquired by changing parameter settings of the current version of the En Route Automation Modernization's (ERAM) Conflict Probe (CP) [Crowell, et al, 2011a]. One of the parameters of this study was the likelihood threshold function.

The conclusion of the first study (Experiment 1) was that major improvements can indeed be gained by changing parameters of the current probe. Likelihood threshold was found to be a major factor in reducing False Alerts, but it is being underutilized at its current setting of 10|20. No recommendation for setting of the likelihood threshold function could be made, since more exploration is required. This technical note documents the results of the fifth in a series of planned experiments (Experiment 5) performed specifically on the likelihood threshold function.

1.1 Previous Work

Two experiments previously performed varied the likelihood threshold function to determine some effects. Experiment 1 used the legacy ZDC scenario and varied the likelihood threshold function from the baseline 10|20 setting to two new settings of 3|8|10 and 3|8|20. For both of the new settings, the middle parameter was at a likelihood value of 0.9. The conclusion of this experiment was that allowing the likelihood threshold function to delay an alert until only three minutes prior to the predicted warning time resulted in additional Late Alerts being generated. However, both the 3|8|10 and 3|8|20 settings showed significant improvement in False Alert performance.

Due to the findings of Experiment 1, the settings used for Experiment 3 were slightly different. Once again, the baseline 10|20 setting was used, and then one other setting of 4|8|20 was also used. Once again, the middle parameter was set at a likelihood value of 0.9. This experiment also showed significant improvements in the performance of the Conflict Probe on False Alerts at the 4|8|20 setting, but without the negative impact to Late Alert performance observed in the previous experiment.

Another experiment, Experiment 2 [Crowell et al, December 2011b], was used to analyze the prototype enhancements proposed for ERAM. Though this experiment observed the interactions of the prototype enhancements with changes in the likelihood threshold function, the purpose of this experiment was to examine the effects of the prototype enhancements, and did not focus on the likelihood threshold function. The only results of Experiment 2 that are relevant to this experiment are that there was little to no interaction between the prototypes and the likelihood threshold functions.

2 Description of Experiment

This experiment, Experiment 5, uses conclusions from previous experiments to determine the settings of the Conflict Probe. Since little interaction was observed between the likelihood threshold function and any of the prototypes, the prototypes were disabled, and it is assumed that any conclusion made in this experiment will be transferable to a probe with enabled prototypes.

The plan for this experiment is to perform 16 runs each on two scenarios for a total of 32 runs. The 16 runs' settings are repeated for both scenarios and were chosen strategically to create a D-optimal design of experiment. Only the likelihood threshold settings are modified in this experiment. All other settings were chosen to be the recommended settings from previous experiments, as follows:

- 1.00 NM Lateral Conformance Bound
- 1.25 NM Longitudinal Conformance Bound

Also in experiments 1 and 3, some lessons were learned about the likelihood threshold function. It was observed that a 3 minute minimum was too low and generated additional Late Alerts, whereas a 4 minute minimum provided a significant decrease to False Alerts, but did not affect Late Alerts. One of the settings used in experiment 3 had a maximum time of 10 minutes. The decision was made that this was too low, since it prevents the CP from probing any further than 10 minutes into the future, whereas the current CP probes 20 minutes into the future. As a result of these observations, the decision was made to keep the 4 minute minimum and 20 minute maximum as constants in this experiment. Only the middle point of the likelihood threshold function will be adjusted in this experiment.

The middle value of the likelihood threshold function is adjusted on both axes. On the *likelihood* axis, it is adjusted from a value of 0.5 likelihood to 0.9 likelihood. On the *time* axis, it is adjusted from 4.1 minutes to 12 minutes. Two midpoint values are chosen on each axis, which results in the design shown in Table 1.

Table 1. D-optimal design for a likelihood experiment modifying only the middle value of the likelihood threshold function.

	Lower		Mid		Upper	
Run	Minutes	Llh Val	Minutes	Llh Val	Minutes	Llh Val
1	4	0	4.1	0.5	20	1
2	4	0	4.1	0.6	20	1
3	4	0	4.1	0.8	20	1
4	4	0	4.1	0.9	20	1
5	4	0	6	0.5	20	1
6	4	0	6	0.6	20	1
7	4	0	6	0.8	20	1
8	4	0	6	0.9	20	1
9	4	0	10	0.5	20	1
10	4	0	10	0.6	20	1
11	4	0	10	0.8	20	1
12	4	0	10	0.9	20	1
13	4	0	12	0.5	20	1
14	4	0	12	0.6	20	1
15	4	0	12	0.8	20	1
16	4	0	12	0.9	20	1

Figure 1 shows the piecewise linear functions of each of the 16 runs in the design of this experiment. A detailed explanation of this graph can be found beginning on page 31 of [Crowell et al, December 2011a]. In short, if the calculated likelihood value at a given time falls above the piecewise linear function, then it will be alerted to the controller. Conversely, if it falls below, then it will be delayed until a time at which it falls above the line. Given the different functions overlaid on the graph, it can be seen where each function differs and where some functions will alert while others will not.

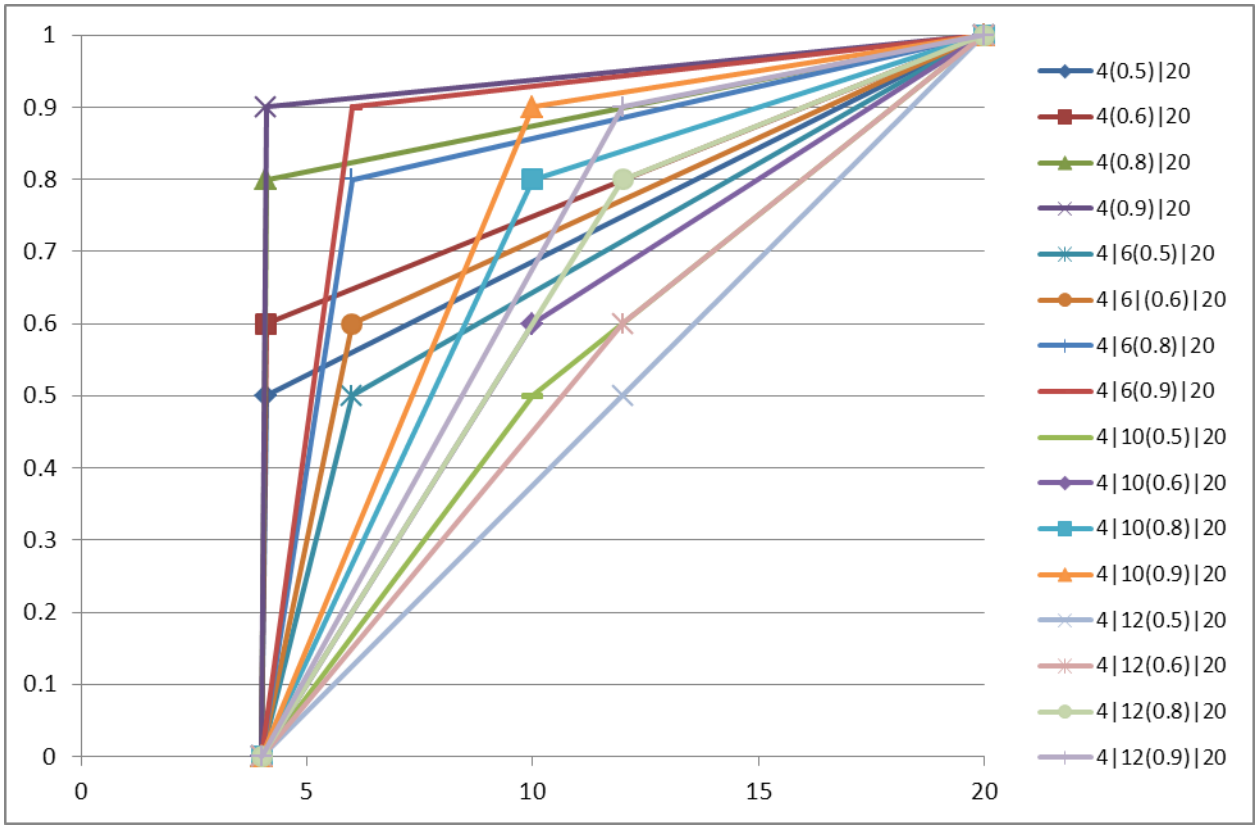

Figure 1. Piecewise linear functions of the likelihood experiment design.

2.1 Scenarios

Two scenarios are used in this experiment. The first scenario is based in the Chicago (ZAU) Air Route Traffic Control Center (ARTCC). It is a six-hour recording of traffic from February 11, 2010 at peak hours. The flights are then time shifted using a genetic algorithm technique [Paglione, 2003] in order to induce conflict events in the scenario. This scenario contains 2,234 flights and 198 conflict events.

The second scenario is based in the Washington (ZDC) ARTCC. It is also a six-hour recording of traffic data at peak hours and was recorded on April 30, 2010. Once again, the flights were time shifted in order to induce conflict events. The scenario contains 2,734 flights and 239 conflict events.

3 Experiment Results

The raw results of the ZAU runs are shown in Table 2. Several of the runs have slightly different counts of Valid Alerts (VA). However, these differences are minor and are due only to the noise inherent in the live system that is used for simulation. This noise can cause track data to differ slightly between runs. Therefore, a difference of one or two VAs between runs is within the expected noise threshold.

The Late Alert (LA) and Missed Alert (MA) counts are identical in all 16 runs. This suggests that the 4 minute minimum for the likelihood function is large enough that additional late alerts are not generated. Therefore, increasing the middle likelihood setting will have no effect on the LA count.

The different likelihood settings do have a modest effect on the False Alert (FA) count. The settings for Runs 3 and 4 produce the least amount of FAs with counts of 794 and 793, respectively. This difference of 1 FA is within the margin of error due to noise, as mentioned above. The middle likelihood settings of 4.1 and 0.8/0.9 are the most radical of all the runs. Using these values removes almost all alerts that do not reach a 0.8/0.9 likelihood before the 4 minute threshold, which accounts for the low FA count.

Runs 13 and 14 show a virtually identical performance with FA counts of 862 and 863, respectively. These are the highest FA counts of the 16 runs. The difference of 1 FA between the two runs is again within the margin of error. Run 13 is the most conservative setting, with the likelihood function a straight line from {4, 0.0} to {20, 1.0}. Runs 3 and 4, which have the least amount of FAs, contain 8% less FAs than Runs 13 and 14.

Table 2. Results of the ZAU runs.

Run	Time (minutes)	Likelihood	VA	LA	MA	FA
1	4.1	0.5	183	10	5	817
2	4.1	0.6	183	10	5	811
3	4.1	0.8	183	10	5	794
4	4.1	0.9	184	10	5	793
5	6	0.5	182	10	5	839
6	6	0.6	183	10	5	820
7	6	0.8	183	10	5	808
8	6	0.9	182	10	5	804
9	10	0.5	183	10	5	846
10	10	0.6	184	10	5	846
11	10	0.8	183	10	5	839
12	10	0.9	183	10	5	829
13	12	0.5	182	10	5	862
14	12	0.6	183	10	5	863
15	12	0.8	183	10	5	856
16	12	0.9	181	10	5	834

In addition to having an effect on False Alerts, the likelihood settings used in this experiment also have a significant effect on warning time (Table 3). As in previous experiments, the warning time metric used is the 25th percentile of warning time (see [Crowell et al, December 2011a] for an explanation of this metric). The warning time performance consistently follows the inverse trend of the FA performance. Run 4, the most radical run, shows the worst warning time performance with a value of 234 seconds. Run 13, the most conservative run, shows the best warning time performance with a value of 255 seconds, which is a 9% increase in warning time when compared to Run 4.

In Figure 2 the warning time performance is displayed in a graphical format. The time settings for the middle parameter of the likelihood function are on the abscissa, while the likelihood settings are on the ordinate. Each point represents one of the 16 pairs of settings used. The color of each point represents the warning time performance of that setting, with the color gradient shown to the right of the graph. A dark red color indicates less warning time (degradation in performance) whereas a dark blue color indicates more warning time (improvement in performance). Though there is a difference of up to 9% in warning time between the runs with the best and worst performance, none of the settings yielded a particularly poor result in the warning time metric. The lowest warning time performance, 234 seconds, is still 54 seconds greater than the minimum warning time requirement for this experiment, which is set at 3 minutes.

Table 3. Warning time performance of the 16 runs.

Run	Time (minutes)	Likelihood	Warning Time (seconds)
1	4.1	0.5	243.0
2	4.1	0.6	238.5
3	4.1	0.8	238.0
4	4.1	0.9	234.0
5	6	0.5	245.0
6	6	0.6	242.5
7	6	0.8	244.0
8	6	0.9	245.5
9	10	0.5	251.0
10	10	0.6	252.0
11	10	0.8	251.0
12	10	0.9	249.0
13	12	0.5	255.0
14	12	0.6	252.0
15	12	0.8	251.5
16	12	0.9	246.0

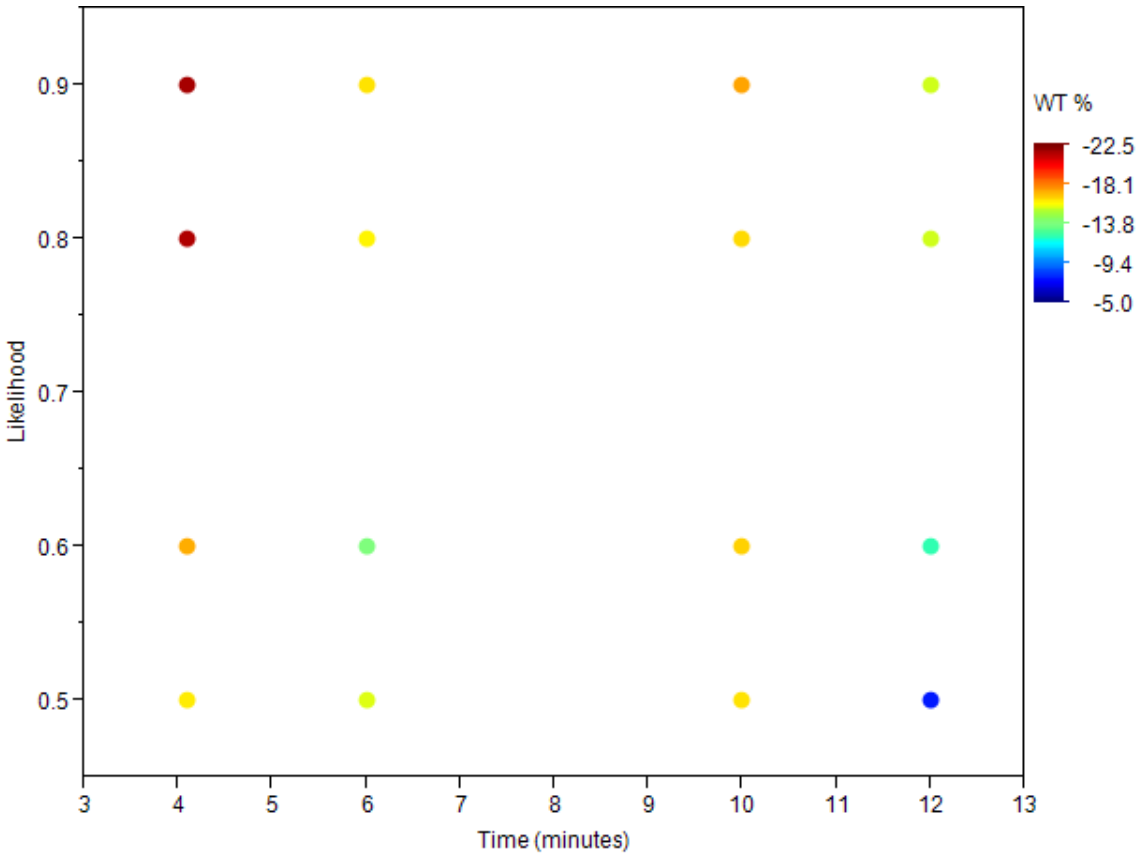

Figure 2. Warning time (color) for each middle setting of the likelihood function.

In Figure 3 the False Alert performance is displayed in a graphical format. The settings of the middle time parameter of the likelihood function are again graphed on the abscissa, while the likelihood values are on the ordinate. Each point represents one of the 16 settings used. The color of each point represents the FA performance of that setting, with the color gradient shown to the right of the graph. A dark red color indicates a greater amount of FAs (degradation in performance), whereas a dark blue color indicates less FAs (improvement in performance). The difference between the runs with the worst and best performance is 8%, as previously discussed. Note that the trend (colorization) for FA is almost exactly opposite that for warning time, which visually confirms the inverse proportionality between warning time and FA performance: as warning time performance improves (increase in warning time) FA performance degrades (increase in FA count).

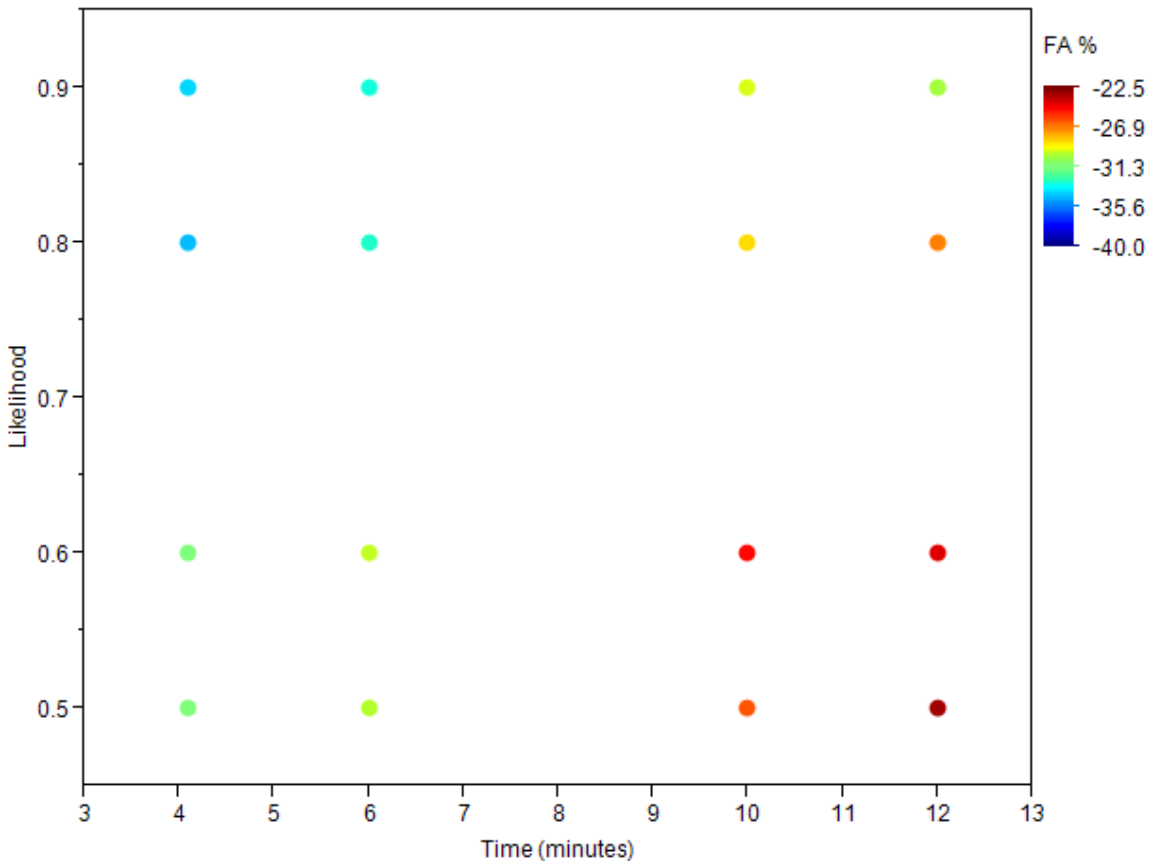

Figure 3. False Alerts (color) for each middle setting of the likelihood function.

The baseline in this experiment is used differently from previous experiments. The baseline run of ERAM was run with a likelihood setting of 10|20 with a lateral adherence bound of 2.5 nm and a longitudinal adherence bound of 1.5 nm. In this experiment, the treatment runs have different constants than the baseline, therefore the comparison made between the treatment run and the baseline does not directly show the result of changing only the likelihood value factors. Instead, the metric is used for comparing treatment runs to each other. The metrics used are percent change from the baseline as described in Eq. 1.

$$\%\,Change = \left(\frac{ExperimentalValue - BaselineValue}{BaselineValue}\right) * 100 \qquad \textbf{Eq. 1}$$

Figure 4 displays the prediction profiler: an interactive plot that depicts the results of the model when using the 4|8|20 likelihood setting from Experiment 3 - a midpoint of 8 minutes and a likelihood value of 0.9. This is the setting used in previous experiments and has been the recommended setting so far, so each setting examined in the prediction profiler will have its performance compared to the 4|8|20 setting. A positive change in warning time represents improvement, while a positive change in false alert count represents degradation. The line graphs in the prediction profiler indicate that the FA and WT performance have similar trends. That is, warning time decreases as the FA count decreases. The rate of decrease, expressed as percentage change, is approximately the same for both FA and WT, except where WT levels off between 7 to 10 minutes. Likewise, as warning time increases, FA count increases at approximately the same rate. This characteristic, an improvement in one factor in parallel with degradation in the other, makes it difficult to choose a particular setting, since the goal is to maximize one and minimize the other and no weighting priority has been defined.

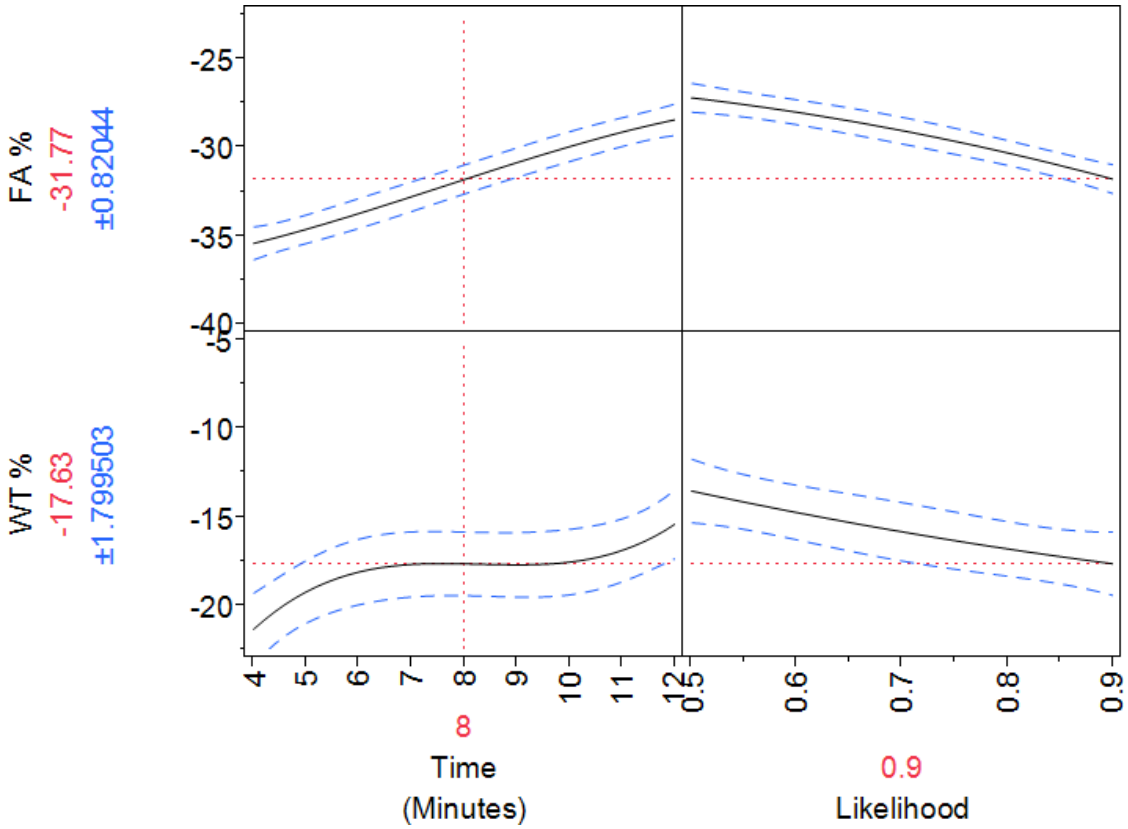

Figure 4. Prediction profiler showing the 4|8|20 likelihood used in Experiment 3.

Figure 5 depicts the prediction profiler at settings that maximize FA performance. With time set at 4 minutes and likelihood set at 0.9, false alerts are decreased from the baseline by almost 32%. Warning time is reduced from the baseline value by 17.6%. While this warning time is still above the required three minute threshold, it is up to subject matter experts to determine if the 3.6% improvement in FA performance from the 4|8|20 setting is worth the 3.7% sacrifice in warning time.

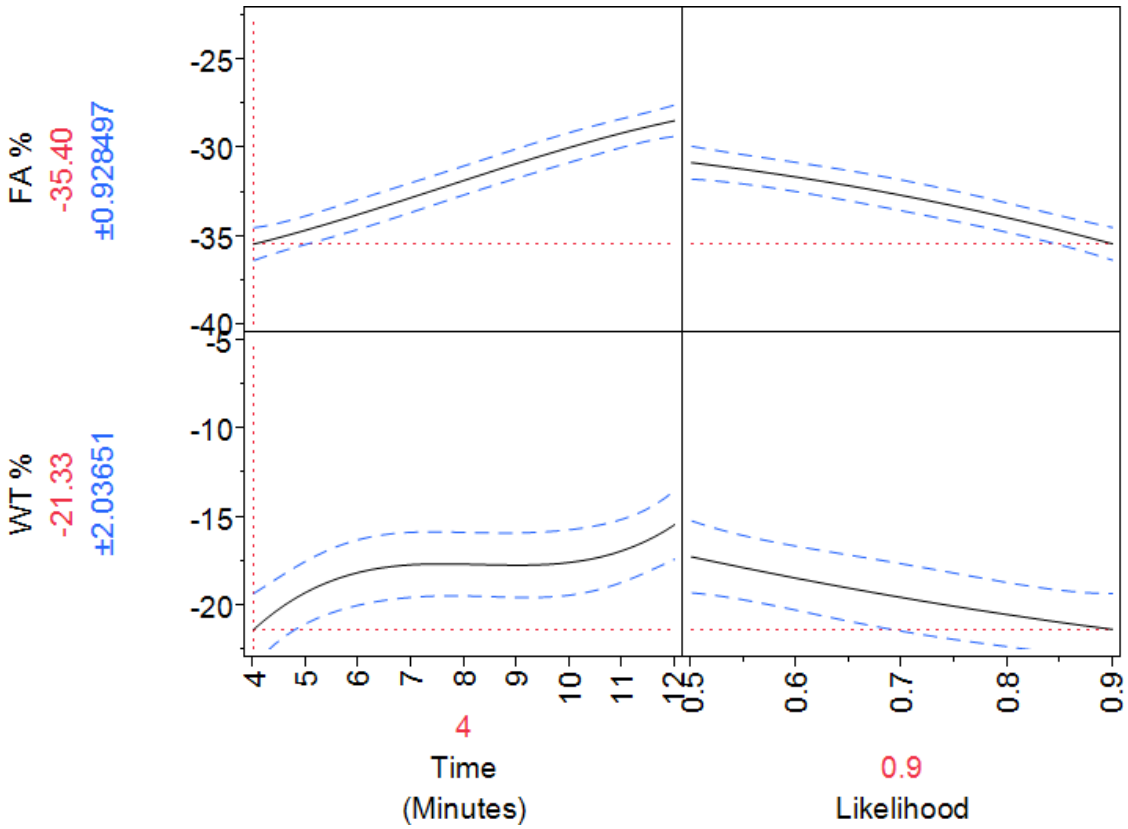

Figure 5. Prediction profiler modeling the best setting for FA performance.

Figure 6 depicts the prediction profiler at settings that maximize warning time performance. As noted previously, warning time performance follows almost the same trend as FA performance. As a result, the settings to maximize WT performance are at the exact opposite extremes of those to maximize FA performance. With these settings the FA count does show an increase of 8% with respect to the 4|8|20 whereas WT shows an increase of 6.4% with respect to 4|8|20. However, there is still a significant improvement over the baseline with a 23.8% decrease in FA. Again, it is up to subject matter experts to determine if the better warning time performance is worth the lower improvement in FA performance.

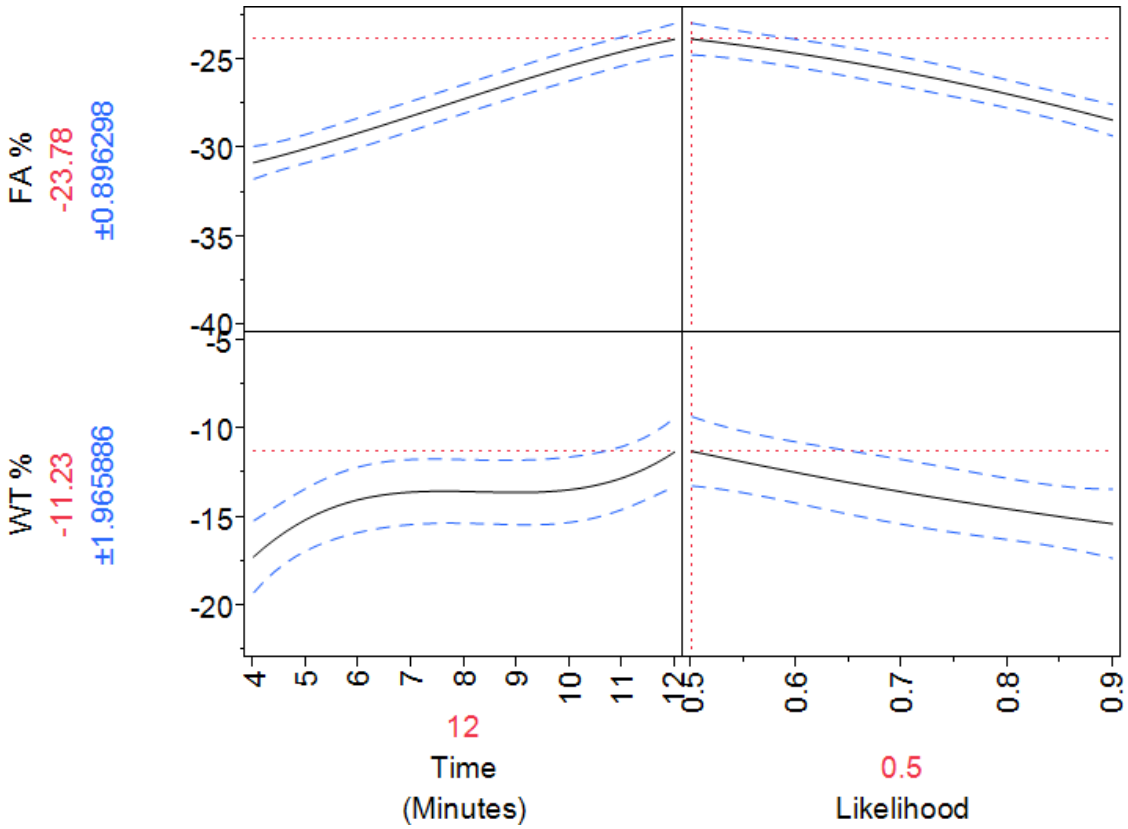

Figure 6. Prediction profiler modeling the best setting for warning time performance.

Figure 7 depicts the prediction profiler at settings that imply equal desirability between warning time performance and FA performance. A middle time setting of 7 minutes and a likelihood setting of 0.8 yields a FA decrease of 31.3% and a WT decrease of 16.8%. At these settings the FA count is increased by 0.5% with respect to 4|8|20 whereas warning time increases by 0.8%. This is not a significant improvement over the current recommendation of the 4|8|20 likelihood threshold setting.

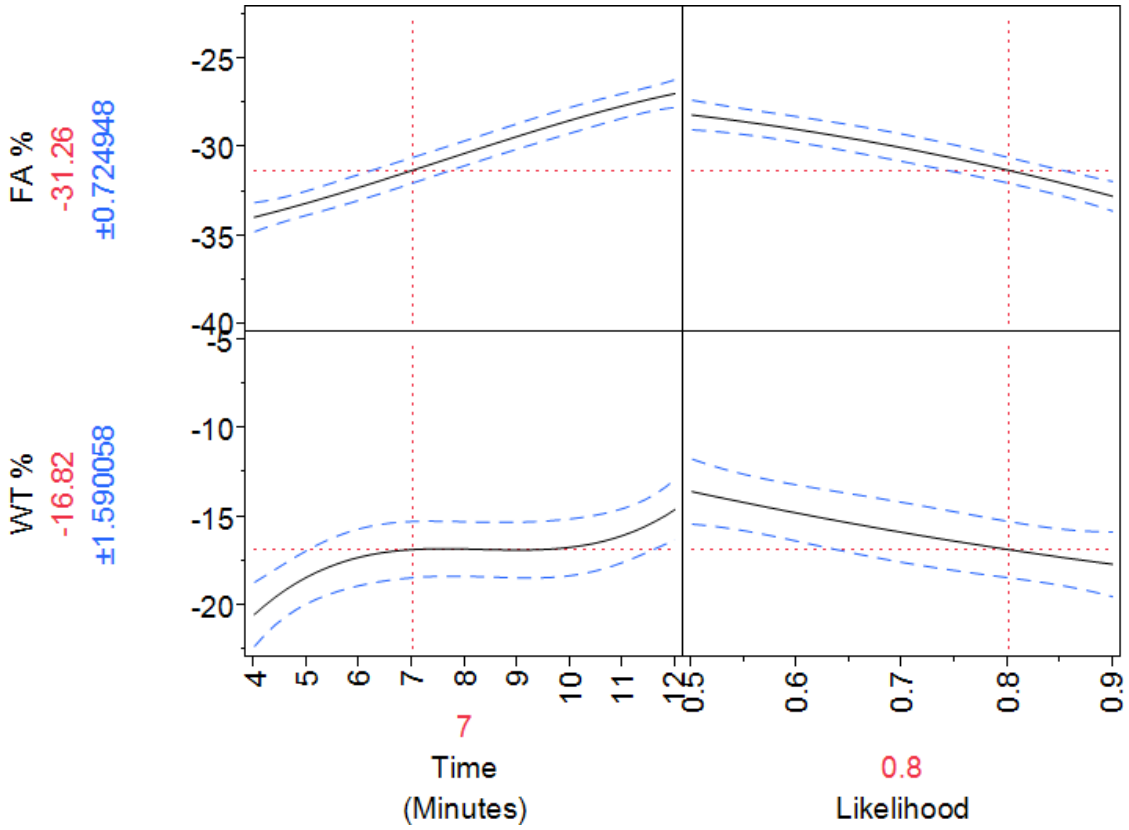

Figure 7. Prediction profiler modeling the setting that best balances warning time and false alert performance.

When the effects of middle time and likelihood settings with respect to the model are examined, it can be seen that the two factors have a solid effect on both FA and warning time. Since no preference has been given it is assumed that each performance metric has equal weight, and a setting that balances false alerts and warning time should be used. Given the nearly inversely proportional relationship between false alert and warning time performance, Concept Analysis is currently recommending to use the mid-range performance setting, a time of 8 minutes and likelihood value of 0.9, which provides similar performance to the setting indicated in the prediction profiler when balancing warning time and false alert performance. This setting, along with the previously recommended constants, provides a FA performance improvement of 31.8% with a warning time decrease of 17.6%. However, given input from subject matter experts, this recommendation may change.

4 Likelihood Function Analysis

With the previously described experiment near completion, it was decided that some analysis should be performed on the distributions of the likelihood values generated by the likelihood function. Because of the way the likelihood function is used within ERAM, the likelihood value is only recorded when an alert is notified. Therefore, if the likelihood value prevents an alert from being notified, there is no log created of this event. In order to properly analyze the distribution of likelihood, a special run was performed on each scenario that provides the maximum logs of likelihood values. Note that the term “likelihood function” refers to the algorithm in ERAM that calculates the likelihood value between 0.0 and 1.0, whereas “likelihood threshold function” refers to the piecewise-linear function in which the likelihood value is applied to determine the notification time.

The 4 minute minimum and 20 minute maximum values were used in these two runs, but the middle value used was a likelihood value of 0.01 at 19.9 minutes. This results in an extremely concave function shape, and forces likelihood to be calculated and logged in almost all situations. The likelihood value will not prevent an alert from being notified except in the case where the value is at, or very close to, 0.0.

In this analysis, a different definition is used for the term Valid. In these results, Valids consist of Valid Alerts (VA) as well as a type of Discard that is referred to as a Clearance Discard. Clearance Discards are alerts that were determined to be False Alerts, but upon further analysis it was determined that the reason for this False Alert was that the flight was given a clearance between the notification start time and the predicted conflict start time, then the alert was deleted shortly after that clearance. So, a Clearance Discard is very likely a valid problem that was resolved by the controller, causing no loss of separation to occur. In most analyses, these types of alerts are discarded from the results simply because there is no algorithmic way to validate that each alert is a valid alert. In this analysis, the conflict probe is given the benefit of the doubt and all of these discards are considered to be valid alerts.

In Figure 8, the distribution of likelihood value is presented. The abscissa is the predicted warning time (PWT) of the alert, in minutes, and each point is colored by its alert type. Red indicates FA, blue indicates VA, and green indicates Discard. The Discards and VAs should be assumed to be valid alerts. None of the Late Alerts in this scenario had likelihood values calculated, because all of them had PWT of less than four minutes.

This figure shows that there are a large number of FAs scattered throughout the range of likelihood and PWT. Many of these FAs scattered throughout will be removed or delayed by a likelihood function. The likelihood function can be drawn on this plot, in the same way as in Figure 1, and any point that falls underneath the line will at least be delayed, and possibly removed. In order to analyze the effectiveness of a likelihood function, several settings that have been used in this study and previous studies will be plotted on this scatter plot.

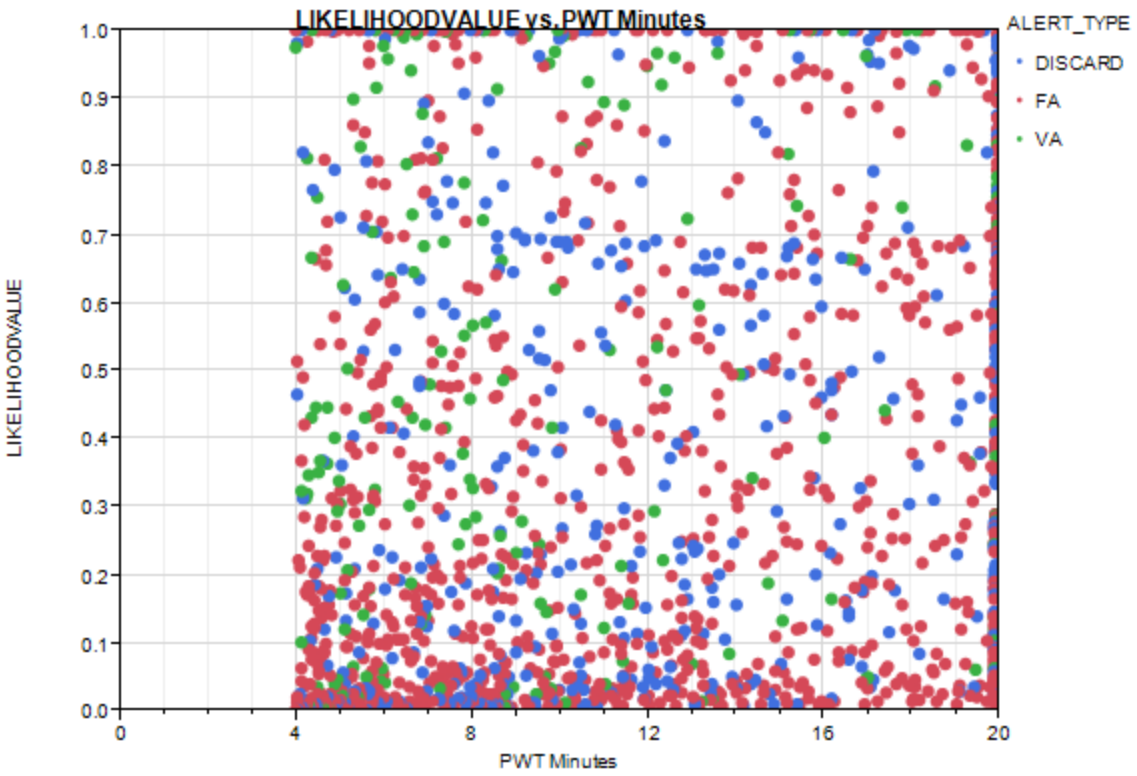

Figure 8. Distribution of likelihood values with respect to predicted warning time, colored by Fuzzy FA

Figure 9 depicts the likelihood distribution when the 10|20 likelihood function is applied. The 10|20 likelihood function is the setting being used currently in the live system. In the figure, alerts that would be removed by this likelihood function have been faded. The black line represents the likelihood function that was applied. Any points to the right of this line that have not been faded would be delayed by the function. The amount they would be delayed can be determined by tracking directly left of the point until the black line is reached. For instance, for a point that sits on the graph at 0.5 likelihood and 20 minutes warning time, tracking directly left until the black line is reached will put the warning time at 15 minutes.

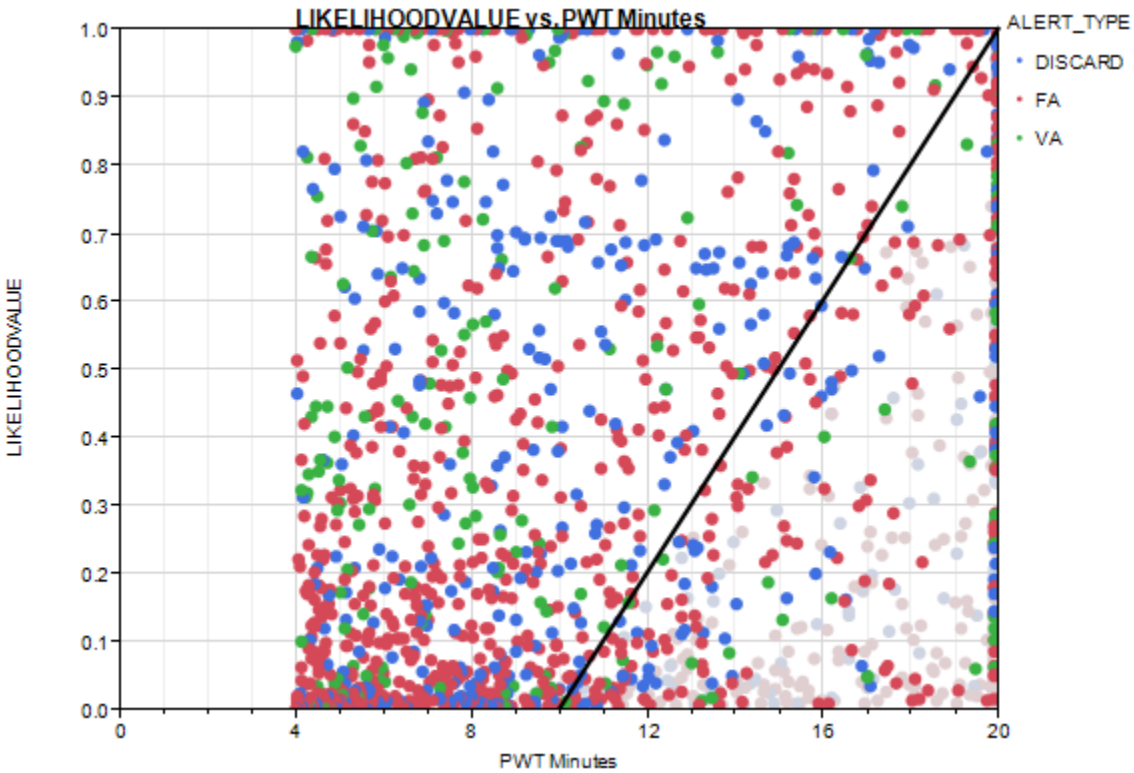

Figure 9. 10|20 baseline likelihood function overlaid on a distribution of likelihood values.

With the 10|20 likelihood, there are 306 alerts removed and 322 delayed, for a total of 628 alerts affected by the function, out of a total of 1678 alerts. 393 of the affected alerts are FAs. 213 FAs are removed by the likelihood function. The additional 180 FAs only have their notifications delayed. Of the 235 Valids, 93 are removed and the remaining 142 are delayed. There were 1007 FAs initially, so the 10|20 likelihood function reduces the FAs by 21%. Of the 671 Valids, 14% were not notified before the controller resolved them.

In contrast, Figure 10 shows the likelihood distribution when the 4|8|20 likelihood that was used in Experiment 3 is applied. This plot uses the same methods as the previous plot in that faded points represent alerts removed by the function, points to the right and below the line are delayed by the function, and the delay can be determined by tracking to the left of the point until the black line is reached. It is clear from observation of this figure that the 4|8|20 function removes many more alerts than the 10|20, which is to be expected.

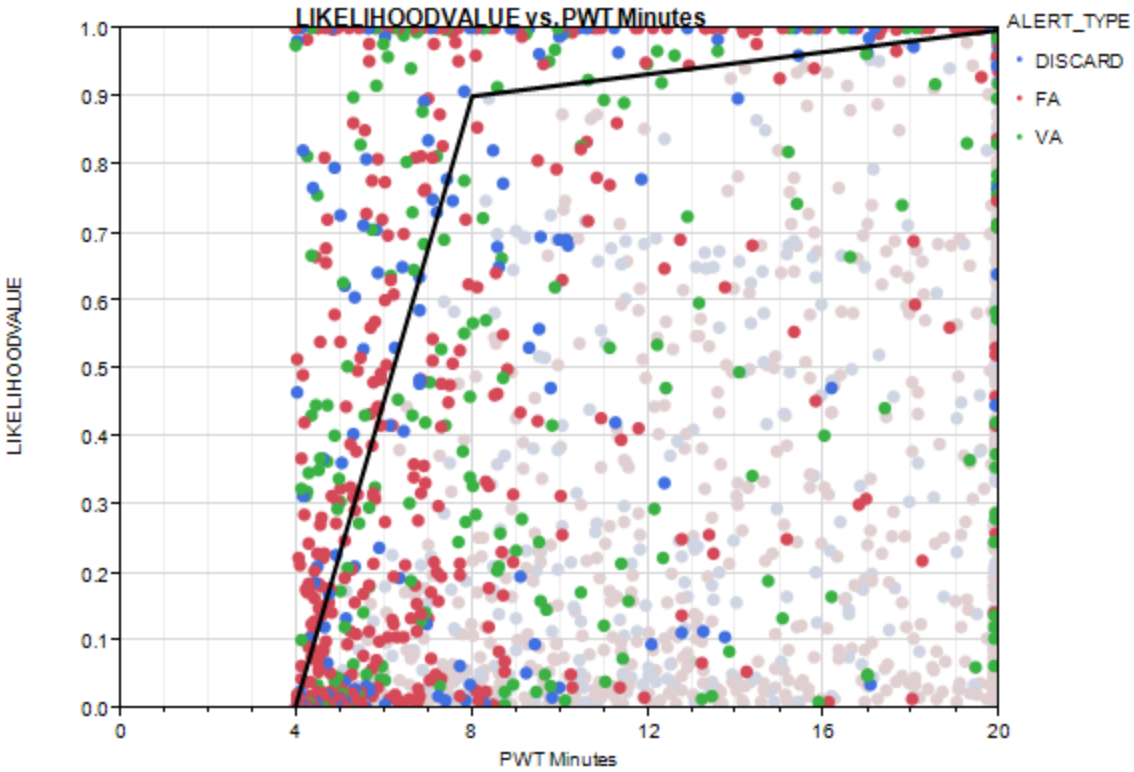

Figure 10. 4|8|20 likelihood function overlaid on a distribution of likelihood values.

The 4|8|20 likelihood function removes 906 alerts. 485 alerts are delayed, but not removed, for a total of 1391 affected alerts. 841 of the affected alerts are labeled as false. 594 of those FAs are removed by the function, whereas the remaining 247 FAs are only delayed. Of the 550 Valids affected, 312 are removed and the remaining 238 are delayed, but still notified. The 4|8|20 likelihood function reduces the FAs by 59%. However, 46% of the Valids were not notified before the controller resolved the issue.

From these two figures, it would seem that the likelihood function is an effective means of removing many FAs from the system. However, upon further inspection, it appears that it is not the likelihood function that is providing this improvement, but rather the delaying of the notification. In order to show this, the alerts are grouped into rectangular bins on the graph as shown in Figure 11. Each square in this figure contains the alerts in a 2 minute by 0.2 likelihood group. The value in the square is the number of Valids divided by the total number of alerts. The square is then colored with a gradient as shown in the legend, with a color closer to green indicating a higher valid rate and a color close to red indicating a higher false rate.

A slight trend can be observed that shows the Valid rate tends to increase as the likelihood increases. However, this trend only increases the Valid rate from 34% at 0.0-0.2 to 48% at 0.8-1.0. Though the increase seems significant, it provides no better results than applying a uniform random value in place of likelihood. To prove this is the case, this exercise was performed using a uniform random value between 0.0 and 1.0 applied to each alert. Then the 4|8|20 likelihood function was applied using the random values in place of the likelihood in the function and the results were recorded.

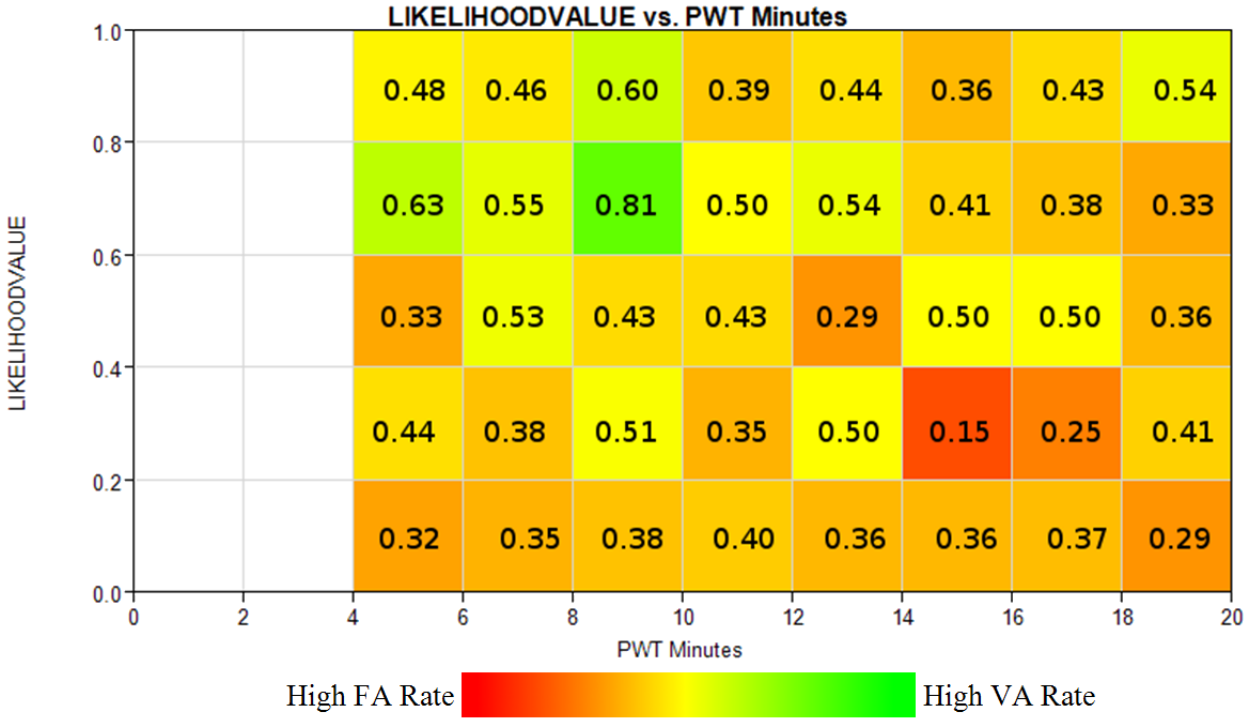

Figure 11. Valid Rate grouped by likelihood value and PWT.

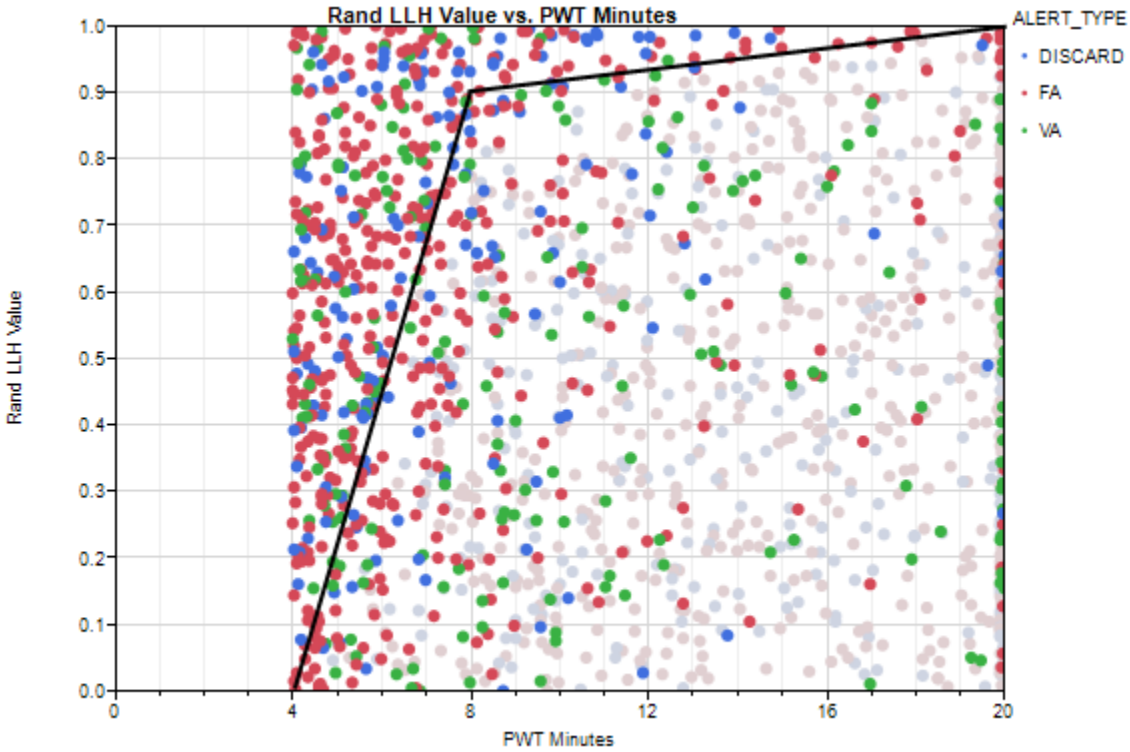

Figure 12. 4|8|20 likelihood function overlaid on a distribution of uniform random values.

Figure 12 shows the 4|8|20 likelihood function being applied to the uniform random values. Using the random values, 896 alerts are removed and 396 are delayed for a total of 1292 total affected. 773 of the affected alerts were labeled as false, and 587 of those were removed completely, whereas the other 186 were delayed. 519 Valids were affected, 309 of which were removed, and therefore not notified before the controller resolved the issue, with the remaining 210 being delayed. Using the 4|8|20 likelihood function with a random value, FAs are reduced by 58%. 46% of the Valids were not notified before the controller took action. In Table 4 these numbers are compared to those using the actual likelihood values.

Table 4. Comparison between application of the 4|8|20 function to a distribution of likelihood values and to a distribution of uniform random values.

		Likelihood	Random	Change
Alerts	**Affected**	1391	1292	-0.07
	Removed	906	896	-0.01
	Delayed	485	396	-0.18
Falses	**Affected**	841	773	-0.08
	Removed	594	587	-0.01
	Delayed	247	186	-0.25
Valids	**Affected**	550	519	-0.06
	Removed	312	309	-0.01
	Delayed	238	210	-0.12
FA	**Reduction**	59%	58%	-0.02
VA	**Reduction**	46%	46%	0.00

From Table 4 the observation can be made that the likelihood value does not provide a better performance than that of a uniform random value. Although the likelihood value provides a better performance in the delay of FAs, it also delays more Valids. To ensure that the removed FAs in the run using likelihood are not better quality removals than those removed by the random value, the Fuzzy FA evaluation is used. A better quality removal would be indicated by a higher Fuzzy FA value, indicating that it is very likely an FA.

Figure 13 shows the quality of FAs grouped by the removal status of the alert. The removal status has four different values: F_T indicates that the alert was removed by the likelihood value and not by the random value; T_F indicates the alert was removed by the random value and not by the likelihood value; SAME_F indicates the alert was removed by both values; SAME_T indicates the alert was not removed by either value.

It is more desirable for the points that are higher on the plot to be removed. Interestingly, both values result in removing mostly only FAs with greater than a 0.5 FA value. However, the additional FAs removed by the likelihood function are spread throughout 0.7 to 1.0 FA values, whereas all but three of the additional FAs removed by the random value are above 0.9 FA value. This is indication that the FAs removed by the likelihood value are not of a better quality than those removed by the random value, confirming the speculation that the current state of the likelihood function does not outperform a uniform random number generator.

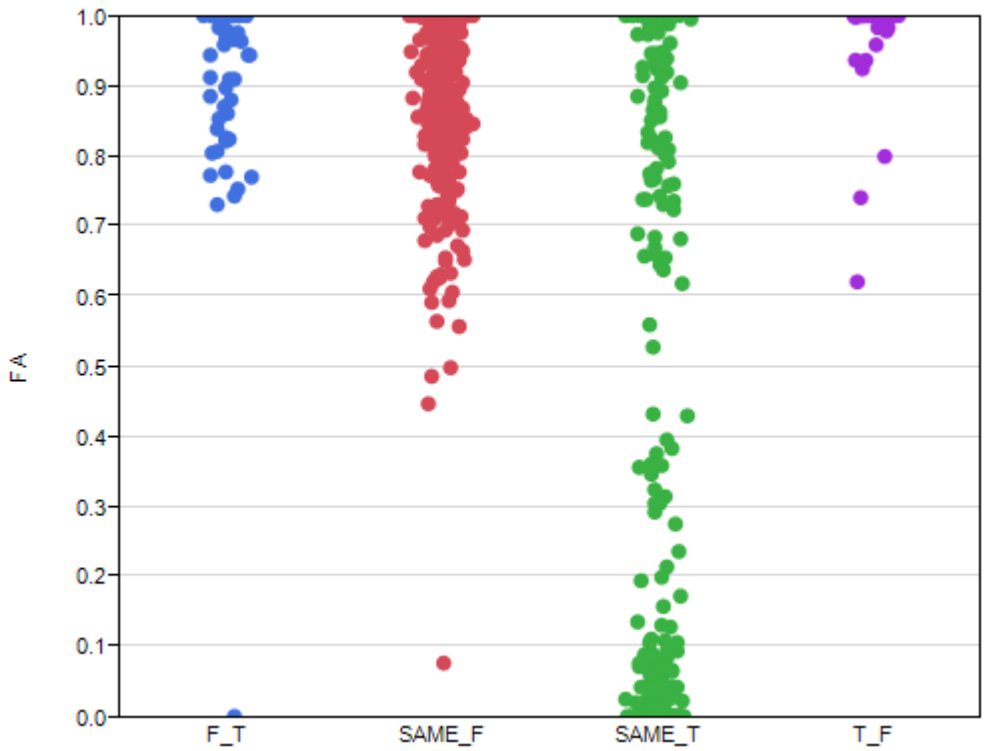

Figure 13. Distribution of differences in FAs between the likelihood values and the uniform random values.

4.1 Flight Examples

This section includes several flight examples that provide some insight into how the likelihood function is working.

4.1.1 Example 1 – Inappropriately High Likelihood

Flight Pair: BTA2074 & COA36
Routes:
BTA2074: KCHS-FLO-TUBAS-FLOPS-FOZZY-CREWE-FAK-OTT-PALEO-PEEDS-FUBRR-DQO-STEFE-SOMTO-ARD-DYLIN-MERSR-METRO-MARRT-PHLBO-KEWR
COA36: FLO-TUBAS-FLOPS-FOZZY-CREWE-FAK-OTT-PALEO-PEEDS-FUBRR-DQO-STEFE-SOMTO-ARD-DYLIN-MERSR-METRO-MARRT-PHLBO-KEWR
Predicted warning time: 1200 seconds
Likelihood value: 1.0
Predicted minimum separation: 0.0 NM
Closest point of approach: 43.6 NM, 7140 ft

This encounter is a prime example of the issues with the likelihood function in its current state. The likelihood value given by the function was 1.0 at 20 minutes, indicating that it is as certain as possible that this conflict is going to occur. However, upon examination of the encounter at this time, it is clear that this should not be the case. Figure 14 shows the two aircraft at the point in time when the prediction is made. In the figure, the aircraft positions are indicated by the black airplane icon, the actual path the aircraft follows is shown by the red or green dots, trajectory predictions are the red or green tubes, and the thick red and green lines are the current routes. In the figure, it is clear that the flights have common routes for a period of time. However, it is also

clear that neither flight is currently following the route. Even with both flights not following their routes, a 1.0 likelihood value is calculated.

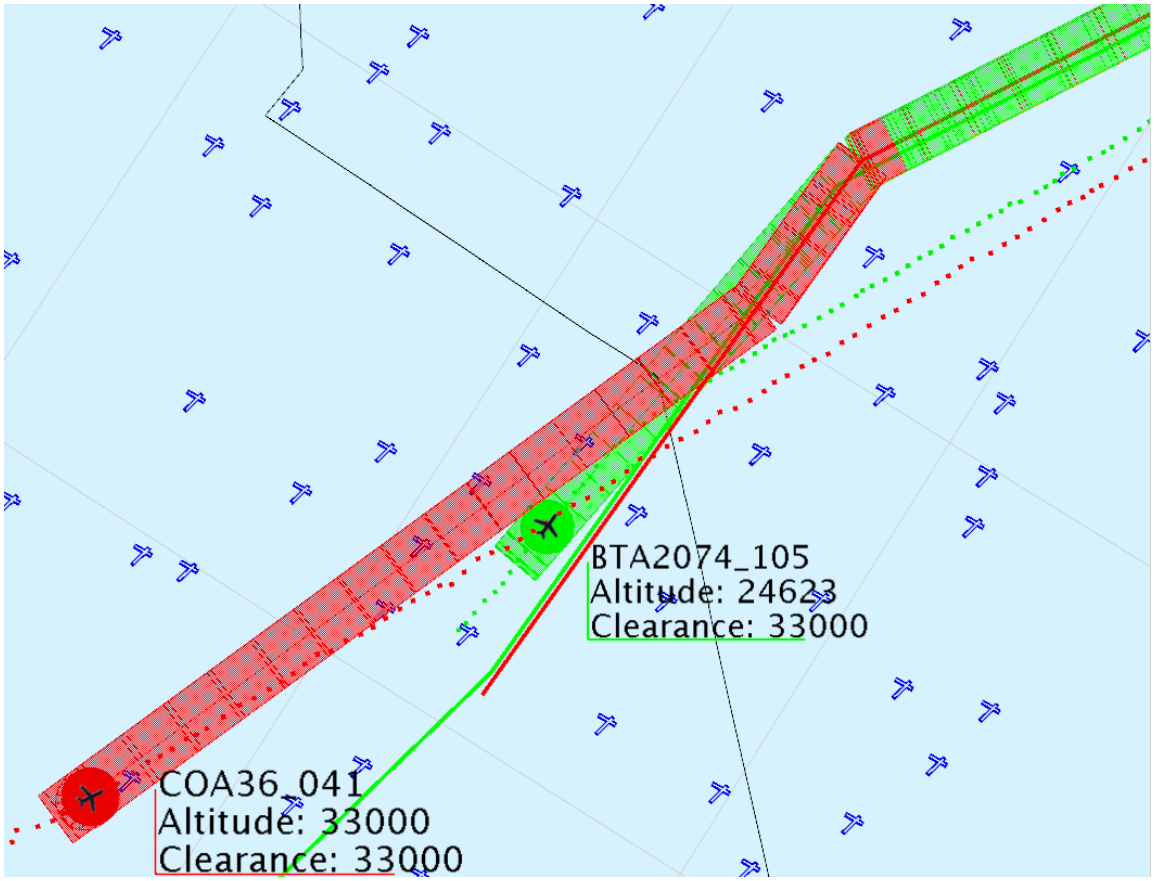

Figure 14. Two aircraft shown at the time of prediction with a likelihood value of 1.0.

At this point in time, the flights are predicted to eventually rejoin their respective routes and then continue on them. While this is a valid prediction, there is no way to be absolutely certain that this will be the case. Assuming a 1.0 likelihood value with a predicted warning time of 20 minutes is far more confident than the current situation merits. In this case, likelihood is working as it was intended to, but reevaluating the intentions of the likelihood function may be able to provide a more efficient CP.

4.1.2 Example 2 – Appropriately High Likelihood

Flight Pair: DLH466 & MXA006
Routes:

DLH466: EMJAY-ZIZZI-WARNN-SWL-SWL219012-KALDA-SAWED-ORF-EDDYS-GILMA-CLAPY-DIW-SEELO-JENKS-HOAGG-BGDOG-JORAY-OSOGY-ENVOY-YOSSI-MILSY-BOYUR-HILEY-KAINS-CIMBA-JESSS-RUBOE-KMIA

MXA006: 3502N/8021W-ORF-SAWED-KALDA-SWL219012-SWL-RADDS-SIE-BOTON-HOGGS-PANZE-KARRS-CAMRN-KJFK

Predicted warning time: 1200 seconds
Likelihood value: 1.0
Predicted minimum separation: 0.0 NM
Closest point of approach: 270.7 NM, 0 ft

This encounter shows that some 20 minute predictions can be valid. Although this particular event never develops into a conflict, from the evidence provided, it is safe to assume it would have been, if no controller action was taken.

The two flights are both following their routes and flying level at cruise altitude when the prediction is made. The automation assumes the flights will continue this way, and this is a valid assumption. The flights have a common portion of route in which the flights would be flying head on with each other. 17 minutes and 31 seconds prior to the predicted conflict start time, a clearance is given to DLH466 to climb to FL400. The alert is then deleted immediately.

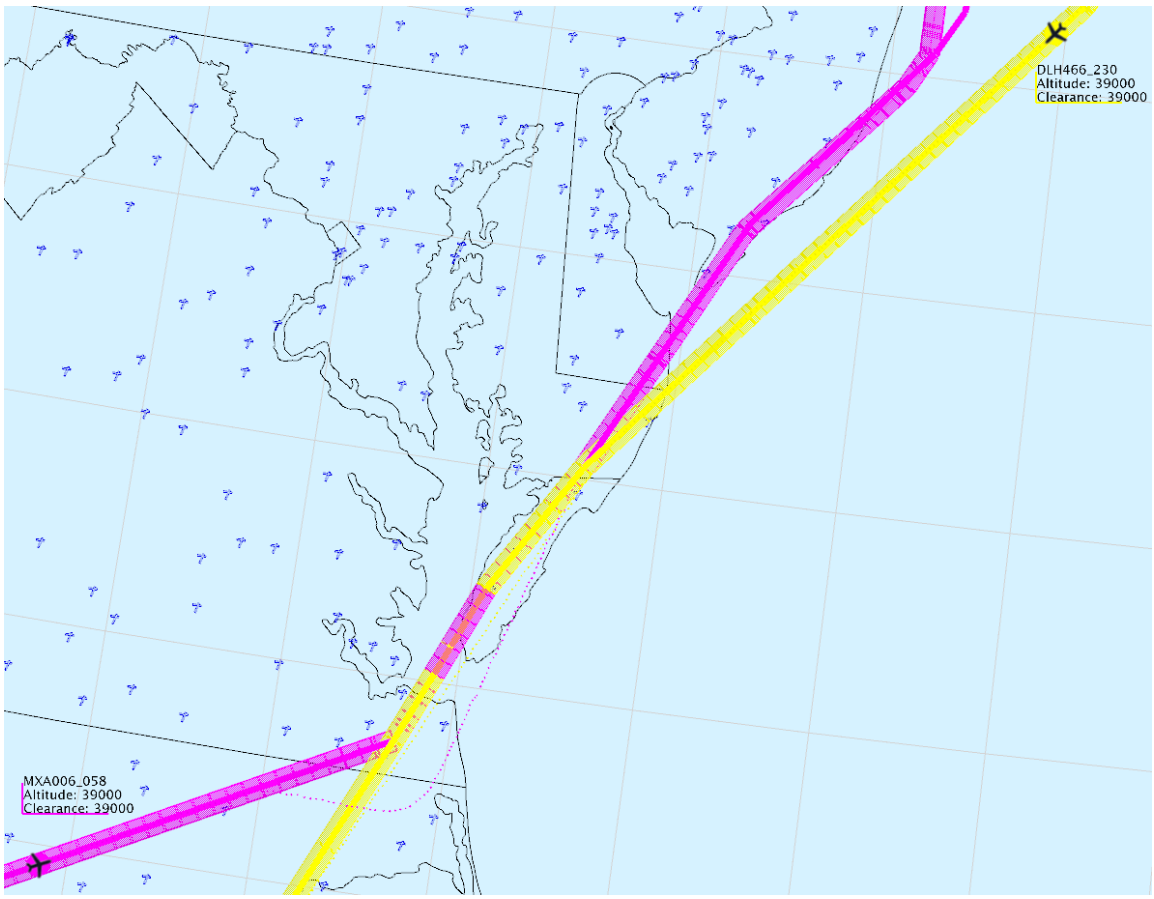

Figure 15. Two aircraft shown at the time of prediction with a likelihood value of 1.0.

After detailed analysis of this alert, it appears to be a valid prediction, and provided the controller with enough warning time to assign a new clearance altitude and avoid the conflict. The alert was then correctly deleted immediately after the controller took action. This is a perfect example of exactly how the probe should work. Although the conflict never actually occurs, the automation made assumptions that it could safely make given the current states of the two flights. If no controller action was taken, both flights would have continued at FL390 and remained on their routes, in which case, there is no doubt that a loss of separation would have occurred.

This example shows that the current likelihood function does make valid assumptions and provide a likelihood value that ultimately benefits the controller. However, from the previous example, it is apparent that there are situations in which the likelihood function may be too confident in assumptions that are made with incomplete information.

4.2 Summary of Likelihood Function Analysis

This report evaluates the effectiveness of the likelihood function in comparison with a random number generator for purposes of determining if the likelihood function can be improved from its current state. The authors are in no way suggesting a random number generator as an alternative to the current likelihood function. Furthermore, like most science related to air traffic, the likelihood function is a multivariate problem with many variables not able to be captured by statistical calculations. The statistics discussed here capture only part of the picture. They cannot, and should not, be used to suggest that the likelihood function does not perform up to standards. They can, however, indicate that the likelihood function may be able to be improved further with additional analysis.

It has already been proven that the current likelihood function provides a significant impact to the performance of the CP and, at the proper settings, can greatly reduce False Alerts. However, given the results of the analysis in this section, it is speculated that a much greater performance gain may be achievable with additional work spent on analyzing the internal workings of the likelihood function.

This report recommends a reevaluation of the likelihood function algorithms. The likelihood function has often been considered a “black box” by analysts until this point. Evaluation of the algorithms should determine if areas of the function can be improved from their current state in order to provide a greater performance increase to the CP.

5 Conclusion & Recommendations

This report documented the results of an experiment performed to determine the effects of the likelihood threshold function on the performance of the conflict probe. Sixteen runs were performed that varied the mid-point of the threshold function in order to model the effects. It was determined from this experiment that there is an inversely proportional effect of the likelihood threshold mid-point on false alert rate and warning time. A tradeoff must be made of either reducing warning time to also reduce false alert rate, or vice versa. At this time, it is recommended to set the mid-point of the threshold function to a mid-range value, therefore balancing the effect on false alerts and warning time. Upon discussion with subject matter experts, this recommendation may change.

In order to better understand the likelihood function, runs were performed that would report the calculated likelihood value for almost all alerts. This allowed the distribution of the likelihood values to be evaluated. Plotting the distribution provided insight into the calculations of the likelihood values. This insight led to the discovery that, although the likelihood provides a significant improvement to the Conflict Probe (CP), by altering the parameters in the likelihood function, there may be substantial room for improvement in the algorithm. The likelihood function has not been thoroughly analyzed or maintained for nearly a decade, so it is a recommendation of this report to analyze the current algorithms of the likelihood function in order to determine improvements that can be made in the calculation of likelihood values.

The 4|8|20 setting used in previous experiments should continue to be used, as this setting provides a significant performance improvement to FAs (10%) and LAs (4%). In comparison with the baseline 10|20 Likelihood with 2.5 Lateral and 1.5 Longitudinal, using the 4|8|20 Likelihood with 1.0 Lateral and 1.25 Longitudinal, the False Alerts are decreased by 37% and the Late Alerts are increased by 7%. The 7% increase to Late Alerts translates to an increase of a single Late Alert and is attributable to error in the model. There is expected to be a much larger performance gain if the likelihood calculation algorithms are improved. Furthermore, in order to keep the likelihood function working at peak performance, it is being recommended that a regular maintenance or evaluation schedule be implemented to keep the likelihood function current with constantly changing performance and operations in a NextGen environment.

6 List of Acronyms and Abbreviations

32BL	FA**32 B**ase**l**ine
AJE-15	FAA Domain Engineering Group
ANG-C41	FAA Concept Analysis Branch
ANSP	**A**ir **N**avigation **S**ervice **P**rovider
ARTCC	**A**ir **R**oute **T**raffic **C**ontrol **C**enter
ATC	**A**ir **T**raffic **C**ontrol
ATO-E	Air Traffic Organization En Route Program Office
BL	FA32 **B**ase**l**ine
CGS	**C**onflict **G**eometric **S**eparation
CP	**C**onflict **P**robe
DST	**D**ecision **S**upport **T**ool
ERAM	**E**n **R**oute **A**utomation **M**odernization
FA	**F**alse **A**lert
FA18	**F**unction **A**rea **18**
FA32	**F**unction **A**rea **32**
FAA	**F**ederal **A**viation **A**dministration
FAR	**F**alse **A**lert **R**ate
FTR	**F**orced **T**rajectory **R**ebuild
GAB	**G**rowth **A**dherence **B**ounds
Horz	**Hor**i**z**ontal
IBL	**I**nitial **B**aseline
IQR	**I**nter-**q**uartile **R**ange
JPDO	**J**oint **P**lanning and **D**evelopment **O**ffice
LA	**L**ate **A**lert
LAR	**L**ate **A**lert **R**ate
Lat	**Lat**eral
Llh	**L**ike**l**i**h**ood
LM	**L**ockheed **M**artin Corporation
Long	**Long**itudinal
MA	**M**issed **A**lert
MITRE	The **MITRE** Corporation
NAS	**N**ational **A**irspace **S**ystem
NC	Correct **n**o-**c**all
NextGen	**Next Gen**eration Air Transportation System
NM	**N**autical **m**iles
SME	**S**ubject **M**atter **E**xpert
TBO	**T**rajectory **B**ased **O**perations
TM	**T**rajectory **M**odeling
TRACON	**T**erminal **R**adar **A**pproach **Con**trol Center
UTC	**C**oordinated **U**niversal **T**ime
VA	**V**alid **A**lert
Vert	**Vert**ical
VHF	**V**ery **H**igh **F**requency
WT	**W**arning **T**ime

References

AuthorLast, AuthorFirst, Author Two, Author Three. January 1970. "A Document Named Something," Publication Information, Conference Proceedings, etc.

Bolczak, R., J. Celio, M. Exum, K. Viets. September 2010. "Separation Management Technology Transfer," Coordination Draft, Rev. 1, MP100138R1, The MITRE Corporation, CAASD, McLean, VA.

Crowell, Andrew, Christina Young, Ph. D., Dr. Hollis Ryan. June 2011. "Response to FA18 Interim 2 Report – Conflict Probe Analysis," FAA Memorandum, Separation Management Modern Procedures.

Crowell, Andrew, Andrew Fabian, Christina Young, Ph. D., Ben Musialek, Mike Paglione. December 2011a. "Evaluation of Parameter Adjustments to the En Route Automation Modernization's Conflict Probe," FAA Technical Note, DOT/FAA/TC-TN12/2.

Crowell, Andrew, Andrew Fabian, Christina Young, Ph. D., Ben Musialek, Mike Paglione. December 2011b. "Evaluation of Prototype Enhancements to the En Route Automation Modernization's Conflict Probe," FAA Technical Note, DOT/FAA/TC-TN12/3.

Crowell, Andrew, Andrew Fabian, Christina Young, Ph. D., Ben Musialek, Mike Paglione. December 2012. "Evaluation of Prototype Enhancements to the En Route Automation Modernization's Conflict Probe with an Updated Traffic Sample from Chicago," FAA Technical Note, DOT/FAA/TC-TN12/56.

Federal Aviation Administration. December 2004. "Next Generation Air Transportation System Integrated Plan," Joint Planning and Development Office, 1500 K Street, Suite 500, Washington, DC 20005.

Federal Aviation Administration. June 2007. "Concept of Operations for the Next Generation Air Transportation System," Version 2.0, Joint Planning and Development Office, 1500 K Street, Suite 500, Washington, DC 20005.

Lapihuska, Charlie. February 2011. "FA18 Reduced Strategic Alerts Prototype Summary Interim 2," Lockheed Martin Report, Separation Management Modern Procedures.

Lapihuska, Charlie. November 2011. "ERAM Conflict Probe Enhancement: Prototype of Reduced Nuisance Alert Options," Lockheed Martin Report, Separation Management Modern Procedures.

McKay, Edward, Ph. D. February 2011. "ERAM Trajectory Lateral Modeling Enhancements, Prototype Effort," Lockheed Martin Report, Separation Management Modern Procedures.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, December 2011.

Oaks, Robert D., Mike M. Paglione. November 2001. "Generation of Realistic Air Traffic Scenarios Based on Recorded Field Data," Air Traffic Control Association (ATCA) 46th Annual Conference Proceedings.

Paglione, Mike M., Robert D. Oaks, Karl D. Bilimoria. November 2003. "Methodology for Generating Conflict Scenarios by Time Shifting Recorded Traffic Data," Proceedings of the American Institute of Aeronautics and Astronautics (AIAA) Technology, Integration, and Operations (ATIO) Technical Forum.

Rosen, N et al. December 2008. "Analysis of En Route Problem Prediction Operational Use for Separation Management," MTR080389, The MITRE Corporation, CAASD, McLean, VA.

Ryan, Dr. Hollis F., George Chandler, Confesor Santiago, Mike M. Paglione, Shurong Liu. November 2008. "Evaluation of En Route Automation's Trajectory Generation and Strategic Alert Processing: Analysis of ERAM Performance," FAA Technical Note, DOT/FAA/TC-TN08/10.