note technical note tech

Evaluation of Prototype Enhancements to the En Route Automation Modernization's Conflict Probe with an Updated Traffic Sample from Chicago

Andrew Crowell, FAA ANG-C41

Andrew Fabian, FAA ANG-C41

Christina Young, Ph.D., FAA ANG-C41

Ben Musialek, General Dynamics Information Technology

Mike Paglione, FAA ANG-C41

December 2012

DOT.FAA/TC-TN12/56

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

Errata

Report No. DOT/FAA/TC-TN12/56

EVALUATION OF PROTOTYPE ENHANCEMENTS TO THE EN ROUTE AUTOMATION MODERNIZATION'S CONFLICT PROBE WITH AN UPDATED TRAFFIC SAMPLE FROM CHICAGO

December 2012

TECHNICAL NOTE

REPLACED PAGES vii, 36, 37

Released February 4, 2013

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT.FAA/TC-TN12/56		

4. Title and Subtitle Evaluation of Prototype Enhancements to the En Route Automation Modernization's Conflict Probe with an Updated Traffic Sample from Chicago	**5. Report Date** December 2012 **6. Performing Organization Code** ANG-C41
7. Author(s) Andrew Crowell, Andrew Fabian, Christina Young, Ph.D., Ben Musialek, and Mike Paglione	**8. Performing Organization Report No.** DOT.FAA/TC-TN12/56
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 organizations: Andrew Crowell, Andrew Fabian, and Mike Paglione with FAA ANG-C41, Christina Young with CSSI Inc, and Ben Musialek with General Dynamics Information Technology.

16. Abstract

The Federal Aviation Administration (FAA) is currently implementing a number of improvements to the National Airspace System (NAS) in the United States under a multi-agency initiative called the Next Generation Air Transportation System (NextGen) Program. The Separation Management and Modern Procedures Project is a NextGen initiative and its objective is to implement the En Route Automation Modernization (ERAM) strategic conflict probe on the radar controller display utilizing ERAM's Trajectory Modeling (TM) and Conflict Probe (CP) sub-systems. The FAA's Air Traffic Organization's En Route Program Office (ATO-E) has employed the FAA's Concept Analysis Branch (ANG-C41) to conduct a series of independent evaluations on prototype enhancements to the TM and CP sub-systems and has contracted the prime contractor of ERAM, Lockheed Martin, under FAA Task Orders 45 and 51 to develop these prototypes within the ERAM architecture. This paper describes the third in a series of integrated experiments to study these enhancements. The experiment consists of simulated runs using the ERAM system with various combinations of prototypes enabled and with various parameter settings. The trajectory modeling and conflict probe performance of these treatment runs are compared against a baseline run which represents the current state of the live ERAM CP. Recorded data from real flights in the Chicago Center (ZAU) was processed to create a realistic air traffic scenario sample. This technical note provides a detailed description of the analyses performed as well as the results of these analyses.

17. Key Words	18. Distribution Statement
Separation Management, Conflict Probe, En Route Automation Modernization, ERAM, Trajectory Modeling	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)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	51	

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

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Executive Summary

The Federal Aviation Administration (FAA) is currently implementing a number of improvements to the National Airspace System (NAS) in the United States under a multi-agency initiative called the Next Generation Air Transportation System (NextGen) Program. The NextGen operational concept envisions a future air traffic environment managed by aircraft trajectory with advances in ground automation like the conflict probe. The Separation Management and Modern Procedures Project is one of these NextGen initiatives and its objective is to implement the En Route Automation Modernization (ERAM) strategic conflict probe on the radar controller display. The strategic conflict probe utilizes ERAM's Trajectory Modeling (TM) and Conflict Probe (CP) sub-systems to notify air traffic controllers when aircraft will violate separation standards as much as 20 minutes in the future. The FAA's Air Traffic Organization's En Route Program Office (ATO-E) contracted the prime contractor of ERAM, Lockheed Martin, under FAA Task Orders 45 and 51 to develop these prototypes within the ERAM architecture so the FAA may evaluate their efficacy. ATO-E has employed the FAA's Concept Analysis Branch (ANG-C41) to conduct a series of independent evaluations on performance enhancements to the TM and CP sub-systems.

This paper describes the third in a series of integrated experiments to study these enhancements. The experiment consists of simulated runs using the ERAM system with different combinations of prototypes enabled and with various parameter settings. The TM and CP performance of these treatment runs are compared to that of the baseline run, which represents the current state of the live ERAM system. Each of these runs is based on the same scenario, which is generated by time-shifting real traffic data recordings to induce conflicts. The traffic data is from a 2010 recording of the Chicago Center during peak hours. This is the second scenario in a series to be analyzed with this approach. All conclusions and recommendations in this report are based solely on this scenario and should not be considered final. Final recommendations will be made once similar experiments have been run on multiple scenarios.

The addition of Function Area 32 trajectory modeling enhancements is the only treatment factor that exhibited an improvement in trajectory accuracy that was statistically and practically significant. The effects of changing the lateral adherence, longitudinal adherence, and likelihood settings are similar to those found in previous studies. Lateral adherence setting has a major impact on the performance in terms of both false alerts (FA) and late alerts (LA). Longitudinal adherence setting has a large impact on LA performance and much less of an impact on FA performance, while likelihood shows a major impact on the FA performance and little impact on the LA performance.

The Forced Trajectory Rebuild (FTR) prototype shows some improvement to the FA performance at certain settings, but always exhibits unacceptable LA performance. The direct comparisons and the model analysis are consistent with these results and there are no settings found at which FTR could provide an acceptable LA performance. The Growth Adherence Bound (GAB) prototype also shows an improvement to FA performance and at certain settings does provide acceptable results for LA performance. The Conflict Geometry Separation (CGS) prototype requires modifications in order to provide borderline significant improvement to FA performance in a small number of specialized encounters. With the correct combination of settings, the False Alerts can be reduced by up to 32% without significantly impacting Late Alerts or Warning Time. Overall, the results of this third experiment support conclusions from previous experiments: that parameter changes in the current system and trajectory prototypes provide significant improvement, while GAB and CGS conflict probe prototypes provide modest improvement.

Based on these results, it is being recommended to proceed with development and implementation activities of the trajectory improvements, parameter adjustments, and GAB and CGS prototypes.

Table of Contents

List of Figures

List of Tables

1 Introduction

The Federal Aviation Administration (FAA) currently has many projects underway for improving the National Airspace System (NAS) that fall into the realm of the Next Generation Air Transportation System (NextGen). Separation Management: Modern Procedures is concerned with the performance and usability of the strategic Trajectory Predictor (TP) and Conflict Probe (CP) of the En Route Automation Modernization (ERAM). The current goal is to improve the performance of the strategic CP by reducing the nuisance alerts to acceptable levels, without adversely affecting its performance on correct alerts. This technical note details a study performed by the Concept Analysis Branch of the FAA in support of this goal.

1.1 Background to Study

In 2011 the FAA's Concept Analysis Branch (ANG-C41) published two reports of integrated experiments that were performed on a single day of recorded, time-shifted air traffic data from a single Air Route Traffic Control Center (ARTCC, Center). The recording date was March 17, 2005, and included traffic in the Washington, D.C. (ZDC) ARTCC.

The documents reported that the lateral conformance bound being used in the current live system (2.5 nm) is inefficient and much larger than it needs to be which results in a generation of nuisance alerts. A recommendation was made to greatly reduce this bound, possibly to as low as 1.0 nm, and even lower once ADS-B is more prevalent in the NAS.

Longitudinal conformance bound was found to be much closer to a preferred value. The current value is set to 1.5 nm, and a recommendation was made to set it to 1.25 nm.

The likelihood function was determined to be used inefficiently in the current system with a mapping of 10|20 (0.0 likelihood alerted at 10 minutes, 1.0 likelihood alerted at 20 minutes). It was recommended that these values change, but no value could be recommended at the time. Instead a future study was planned, which will be performed and published in 2012.

Finally, the three prototype algorithms, Forced Trajectory Rebuild (FTR), Growth Adherence Bounds (GAB), and Conflict Geometry Separation (CGS), were studied. Only GAB was recommended for addition into the probe. FTR and CGS both showed improvement in certain circumstances but overall hindered the CP more than they helped it. Additional study was recommended for FTR and CGS.

Overall, additional analysis was recommended for all factors of the probe, since only a single day of traffic in a single center was used for the previous study. Before any of the recommended settings can be put into the live ERAM system, it must be proven that these settings perform well under all circumstances. This requires the use of additional dates and additional centers of air traffic data.

This study, named FA18 Experiment 3 (Experiment 3) applies the knowledge gained from Experiments 1 and 2, and expands the study to a day of traffic on February 11, 2010 in the Chicago Center (ZAU). This experiment, along with Experiment 4, scheduled for later in 2012, will help to determine how these changes to the CP affect traffic across the NAS.

1.1.1 Prototype Enhancements

This study analyzes the same three prototype algorithms studied in Experiment 2 [Crowell et al, December 2011b]. The algorithms evaluated are Forced Trajectory Rebuild, Growth Adherence Bounds, and Conflict Geometric Separation. Each is briefly defined in the following sections and a complete description of the algorithms can be found in [Lapihuska, November 2011].

1.1.1.1 Forced Trajectory Rebuild

The Forced Trajectory Rebuild (FTR) algorithm will trigger a trajectory update upon demand. Currently, only the subject aircraft's trajectory is guaranteed to be regenerated at the time of probing, whereas the object aircraft's trajectory may have been generated several minutes prior. Over time, the actual track of an aircraft can vary from the trajectory-predicted position. This can lead to errors in the predicted position down-route to the time when minimum separation is determined, in turn affecting the accuracy of the probe.

The FTR algorithm will trigger an update to the object (second aircraft of a flight pair) trajectory and result in more accurate probe results. Also, since the rebuilt trajectory includes other factors included in track history, the down-route accuracy of the trajectory will be further enhanced.

The FTR algorithm provides these additional trajectory build capabilities:

1.) Ensures that the probe is operating on trajectories that have the latest track information.
2.) Provides the probe with the option of delaying notification while maintaining control on how long that delay will be. The probe is no longer limited to dependence on re-adherence rebuilds and has the ability to schedule a re-probe at a predetermined future time.

1.1.1.2 Growth Adherence Bounds

The approach for the prototype Growth Adherence Bounds (GAB) algorithm is to perform a filter on the standard conformance (adherence) bounds that will modify the conformance bounds as the probe traverses temporally through the route. The algorithm will apply smaller conflict detection adherence bounds to near-in time segments in lieu of the standard conformance bounds currently used. The bounds are gradually increased as the probe proceeds further down the predicted route path until the graduated bounds reach the same size as the standard bounds.

The reasoning supporting the algorithm is the fact that flights typically only deviate gradually from the predicted path, so that near-in time segments can have smaller conformance bounds than the bounds used further down the route.

The GAB is most effective when the "age" of the trajectory (time between trajectory build start time and current time) is small – that is, when the trajectory has recently been updated. The FTR function helps ensure that condition by the timely rebuilding of trajectories. GAB is also more effective when the flights are diverging over time and if the period of conflict is limited to the next few minutes. A significant contribution of GAB is to shorten the duration of notification after minimum separation has passed and a separation of 6.2 nm has been achieved. The threat of conflict no longer exists even though the flight separation is still within conformance bound distances. Any reduction in notification time reduces controller distraction, so early elimination of conflicts that have already passed critical separation times improves nuisance rate performance.

The GAB design is conceptually based on an earlier MITRE effort [Rosen, 2008] [Bolczak, 2010] designated as "tactical check" and proposed as a NextGen Separation Management enhancement. However, there are significant differences between the two. The prototype GAB applies the growth in both lateral and longitudinal directions while the MITRE approach applied lateral only. The MITRE approach ensured, within the algorithm design, regular trajectory updates – the prototype does not. The MITRE approach applies some asymmetric lateral adaptations based on relationship of the track to the filed route. The prototype applies a symmetric growth factor with respect to the current trajectory.

Using a single lateral conformance bound for all times results in increased nuisance alerts. The GAB algorithm may be especially useful for reducing nuisance alerts in predicting a near-term conflict.

1.1.1.3 Conflict Geometric Separation

For a few specific potential conflict cases, additional processing, called Conflict Geometry Separation (CGS), will be executed. CGS examines the conflict geometry to determine whether or not a conflict should be discounted.

CGS processing will depend on the category of the specific conflict geometry. The three geometry categories are in-trail, parallel, and crossing. An in-trail conflict is a conflict that can occur on a shared segment of a route that is common between two flights and where the two aircraft are flying generally in the same direction. However, in-trail conflicts can also occur between aircraft that do not have common route segments. A parallel conflict occurs when the corresponding trajectory paths and route paths are greater than 6nm from each other and the closure angle for any of the segment corresponding pairs does not exceed 15 degrees. All remaining conflicts that do not meet the definition of either an in-trail or a parallel are defined as a crossing conflict. After categorizing the geometries of the conflicts, specific criteria are examined to determine subsequent action. The CGS algorithm is applied selectively based on the encounter geometries of the conflict flight pair.

In Experiment 2, the algorithm parameters were set to delay the alert until it had a predicted time-to-conflict of just over three minutes. However, this assumed that the closure rate would not increase over those three minutes. In many cases the closure rate did increase, reducing the time-to-conflict to less than three minutes, and in turn causing CGS to generate a Late Alert according to the definition of Late Alerts in [Crowell et al, December 2011a]. In this study, the time-to-conflict was increased to four minutes to try to avoid this situation.

1.1.2 Conflict Probe Parameters

The last three settings manipulated in the treatment runs of this experiment, as well as in Experiments 1 and 2, are parameters of the ERAM Conflict Probe. These parameters can be varied independently and affect the probe in different ways. The three parameters changed were lateral conformance bound, longitudinal conformance bound, and likelihood.

1.1.2.1 Conformance Bounds

The conformance bounds serve two purposes in ERAM. They determine when a trajectory is built for re-adherence purposes, and they determine when a conflict prediction is made based on a trajectory. The lateral conformance bound is added to the left and right side of the trajectory or flight, whereas the longitudinal conformance bound is added to front and back.

For re-adherence purposes, the conformance bounds create a box that is twice the width of the lateral and twice the length of the longitudinal. It is then placed centered on the predicted position of the flight as shown in Figure 1. If the actual position of the flight is outside of this box, then a new trajectory will be rebuilt to re-adhere to the position of the flight.

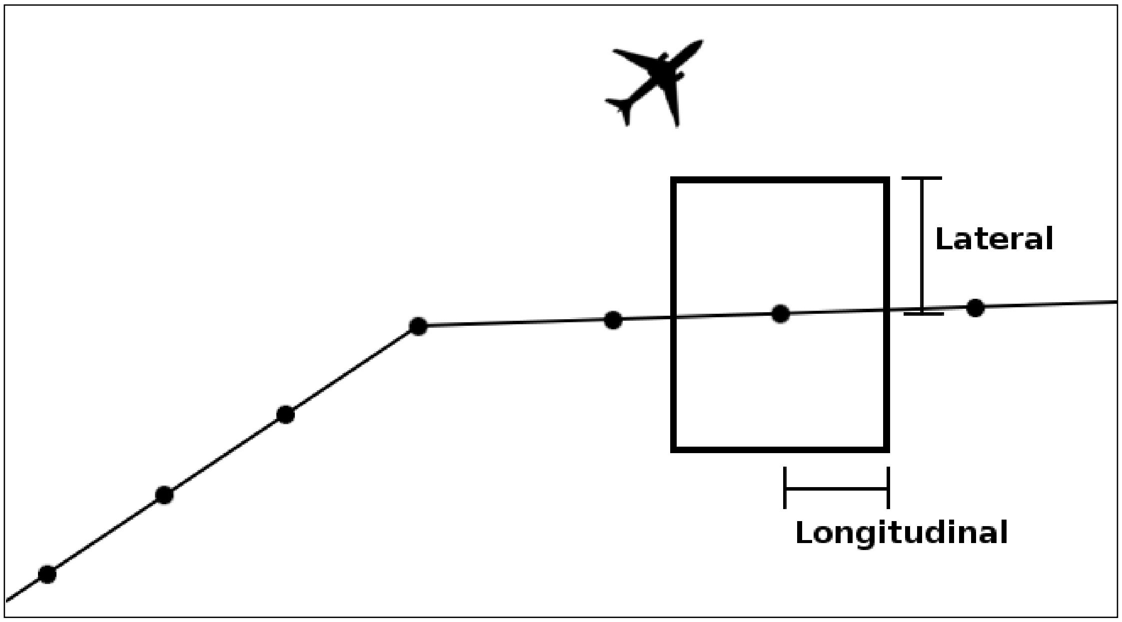

Figure 1. Conformance Bounds Used for Re-Adherence of Trajectory to the Flight Path

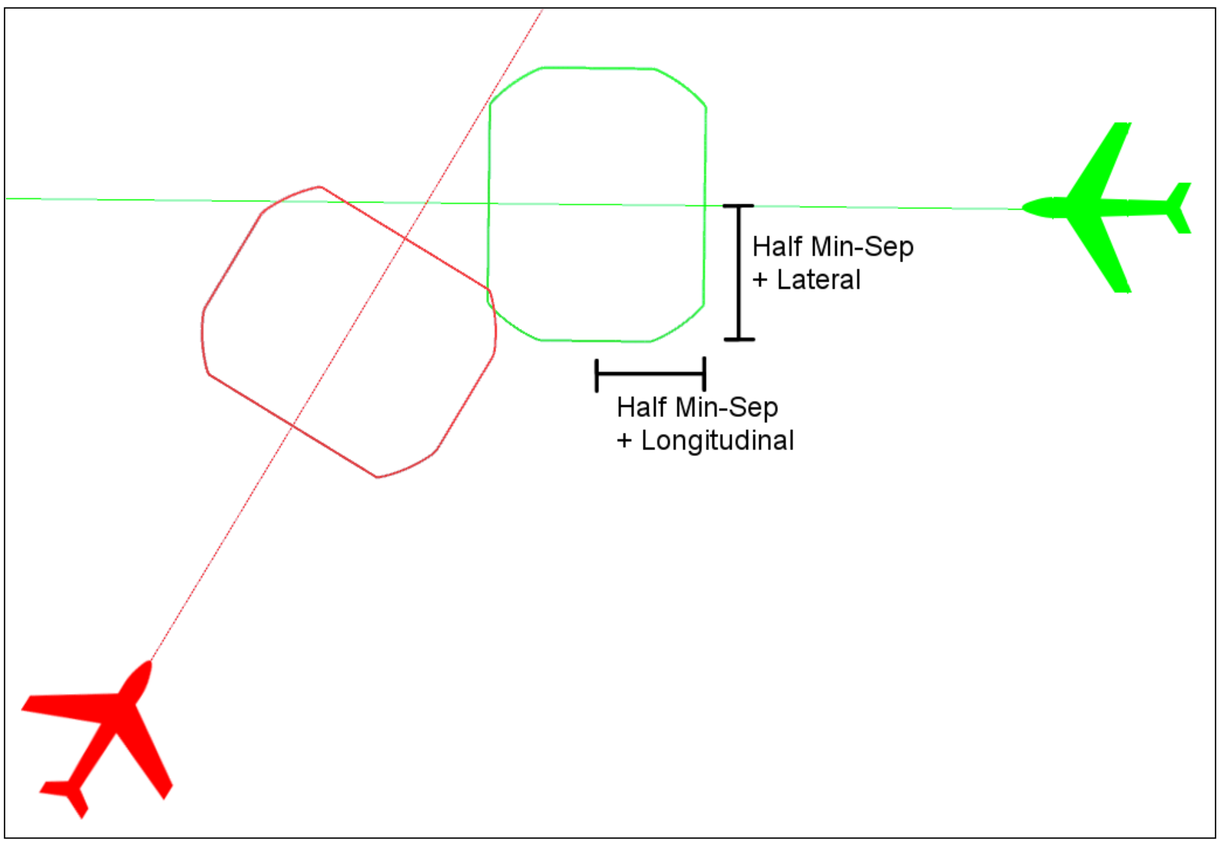

Figure 2. Depiction of Conformance Bounds Used for the Conflict Probe

The use of conformance bounds for the Conflict Probe is much more complex. Several levels of filters are used initially. Then, an octagonal shape is formed using the geometry and conformance bounds of each of the aircraft. This document will not go into the details of this algorithm, but the same results can be visualized by adding the conformance bounds to the required separation of each aircraft to create a box around the predicted position of each aircraft and then cutting off the corners of the boxes with a circular filter. The resulting shape is shown in Figure 2. If these two boxes intersect each other, then a conflict prediction is made.

1.1.2.2 Likelihood

Likelihood is a value determined by the conflict probe that represents how likely the predicted conflict is to occur. Each conflict prediction is given a value between 0.0 and 1.0, with 0.0 being very unlikely and 1.0 being very likely. The likelihood setting altered in this experiment is the likelihood threshold at which the conflict prediction will be notified to the air traffic controller.

The threshold setting is a piecewise linear function of likelihood with respect to warning time as shown in Figure 3. If the likelihood value calculated by the probe is above the line created by this function, then a notification is generated. In this experiment the likelihood setting is represented by two or three numbers in the format *a*|*b*|*c* or *a*|*c*, where *a* is the minimum time in minutes at which likelihood is considered. At or below this time, a likelihood value of 0.0 will still cause a notification. *b* is the time in minutes at which a 0.9 likelihood value is the minimum required likelihood to cause a notification. Finally, *c* is the time in minutes at which a likelihood value of 1.0 is required to cause a notification. A notification will not be generated when warning time is greater than *c*. If no *b* is included, then the function is linear from *a* to *c*, and the value of *b* can be determined via linear interpolation.

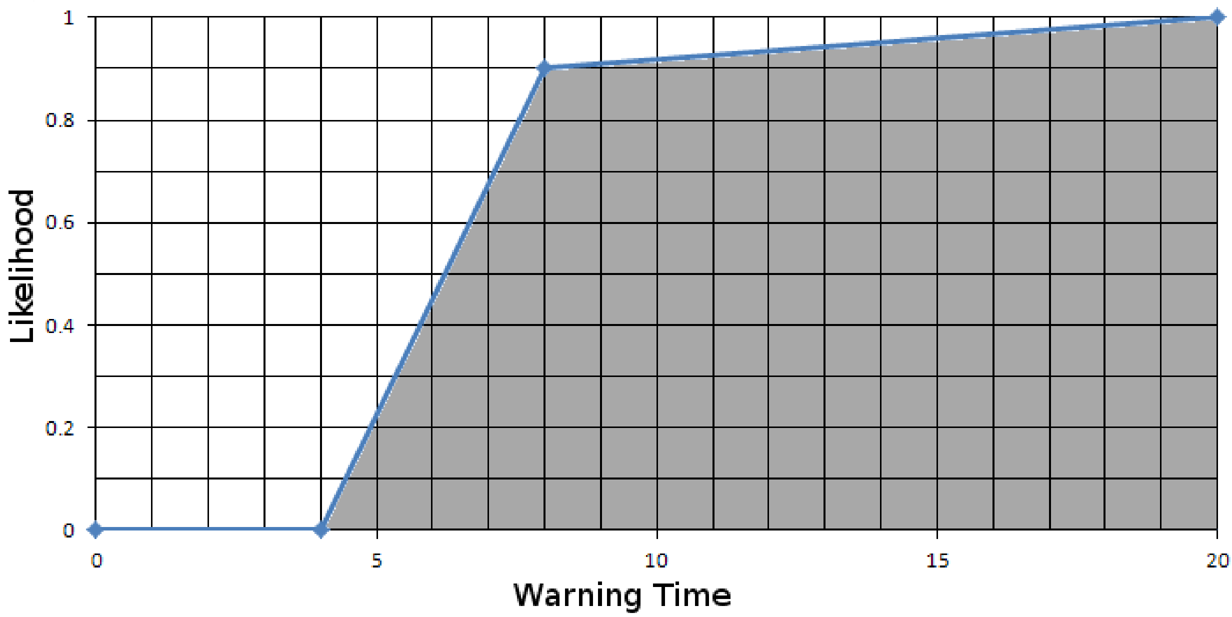

Figure 3. Example of Piecewise Linear Likelihood Function

1.1.3 Previous Work

This study, designated Experiment 3, is the first of two experiments to be performed in 2012 with the purpose of analyzing the impact of three prototype algorithms in ERAM on CP performance. Both studies are a follow up to Experiments 1 and 2 performed in 2011. Experiment 1 was performed to determine if there was a set of parameter adjustments that could be made to the conflict probe in order to improve performance. Experiment 2 analyzed each of the three prototype enhancements to determine if any of them provide a significant improvement to performance.

1.2 Scope of Study

This document reports on the results of an experiment limited to one six-hour traffic sample collected on February 11, 2010 from the Chicago Air Route Traffic Control Center (ZAU). To induce conflicts between aircraft and for evaluation purposes only, the data sample was time-shifted using the methodology documented in [Paglione, 2003].

This experiment follows a similar experiment from 2011 that used a sample of traffic from Washington Center (ZDC) recorded in May 2005. This experiment is intended to expand the findings of the previous experiment. Another experiment will follow later in 2012 that uses a more recent ZDC traffic recording. Since this experiment is an extension of a previous experiment, the findings in this study (Experiment 3) will be compared to the findings from the previous study (Experiment 2). However, any results presented in this document are still considered preliminary results. Final results will not be presented until the next experiment has been performed, due to be completed in late 2012 (deemed Experiment 4).

All of the analyses in this document were performed on a time-shifted scenario. Currently, the metrics available for analyzing performance require a time-shifted scenario to be used in order to generate actual loss of separation that would not occur under normal circumstances. This time-shifting can create some events that the conflict probe will never encounter in a live system. As a result, the reader should be careful not to take any numbers presented in this document out of context. All numbers presented in this document should be used only for comparison to other numbers included in this document, unless otherwise noted. The False Alert, Late Alert, and Missed Alert rates, as well as the warning time values presented in this document do not reflect the actual values of the live ERAM system and should not be considered as such. Because of this, most of the values presented in this document are in the form of percentage change from the baseline results. Though some raw numbers may be presented, they should be considered only in the context of this document.

1.3 Document Organization

This technical note is organized in the following sections: Section 1.1.1 provides a high-level description of the three prototype enhancements being analyzed in this study. Section 2 defines the experiment performed and describes the development of the model along with the final statistical qualities of the model. Section 3 describes the analyses that were performed to evaluate the Trajectory Modeling (TM) performance (Section 3.1) and the Conflict Probe (CP) performance (Section 3.2). Finally, Section 4 wraps up the conclusions of the performance analyses and makes recommendations based on the findings.

2 Description of Experiment

One of the most powerful inferential statistical approaches is the design, implementation, and synthesis of experiments. Experiments are performed by most researchers and scientists in practically all disciplines. An input stimulus is entered into a process with a set of controllable factors. The uncontrollable factors are not easily manipulated, but through experimental design techniques such as blocking and randomization can be removed from the experiment. The output response variables are the dependent variables of the experiment. They are often determined by application of a metric or measured by a sensor device.

Table 1. Processing Steps for the Experimental Analysis

Step	Description	Section
1 – Problem Definition	Define the problem statement	2.1
2 – Design of Experiment	Design the experiment – The factors, levels of the factors, response variables to be run, and the model to be used for analysis are defined.	2.2
3 – Execute Experiment	Execute the experiment and prepare output data – The system is configured for the experimental runs defined by the design, runs executed, and resulting output data is processed for input into model	3
4 – Implement Model	Implement statistical model defined by the experiment.	3.2.2
5 – Model Results	Examine the results of the model and discuss factor effects	3.2.2 & 3.2.3
6 – Synthesize Impact	Synthesize overall results from the model and publish conclusions.	4

There are many purposes for performing an experiment. For this study, the objective of designing and executing an experiment is to establish (1) which pre-determined factors and interactions of these factors show a statistically significant effect on the ERAM system's performance, and (2) the relative sizes of the determined significant effects. From designing the experiment to concluding on its results, a series of processing steps should be performed as identified in Table 1. The first two steps presented in Table 1 are described in this section, which documents the plan for the experimental analysis. The last four steps are described in Section 3 and Section 4, which present the results by documenting the actual execution and analysis of the experiment.

The integrated experiment used in this study slightly modifies that used in Experiment 2, based on the lessons learned from the previous experiment. The purpose of this experiment is to determine if any of the manipulated factors provide a statistically significant improvement to the performance of the Conflict Probe. In order to evaluate this, it is also necessary to determine how each of the factors interacts with one another. The main difference between this experiment and the previous is the use of two likelihood settings: 10|20 and 4|8|20, as opposed to the three used in the previous experiment.

The factors for the prototype enhancements are binary and indicate whether that particular prototype enhancement is on or off. Given the four binary factors and the two continuous factors, the total number of runs required for a full factorial design (assuming three samplings of the continuous functions) would be 144. Since each run must be performed using the live ERAM system in a simulation environment, it is necessary to reduce this number considerably. The experiment was designed using the JMP® software tool and is described in the following sections.

2.1 Definition of the Problem Statement

It must be determined if any of the three prototypes or three parameter changes can provide a significant improvement to Conflict Probe (CP) performance. CP performance is measured in False Alert, Late Alert, and warning time performance, all of which can vary separately. Low False Alerts, low Late Alerts, and high warning time are the desired qualities of CP performance. The three prototypes covered in this study are intended to improve False Alert performance. A significant improvement to CP performance will be recognized if a prototype significantly improves False Alert performance, and does not significantly degrade Late Alert performance. It is also desirable to avoid degrading warning time performance, but this is not a requirement in order for a CP performance improvement to be recognized. For this study, the problem statement is expressed as follows:

> *Through a set of purposeful runs of ERAM, input with the ZAU time-shifted test traffic scenario, the experiment shall determine the statistically significant impact that the Forced Trajectory Rebuild, Growth Adherence Bounds, or Conflict Geometric Separation prototype algorithms, or lateral adherence, longitudinal adherence, and likelihood alterations have in terms of trajectory and conflict prediction accuracy performance.*

A significant change, whether it is improvement or degradation, is defined as a change in the respective metric (False Alerts, Late Alerts, or warning time) that is greater than the confidence intervals of the statistical model. These confidence intervals are discussed in the next section.

2.2 Design of Experiment

In order to reduce the number of runs required to perform the analysis, a d-optimal design was used rather than a full factorial [NIST/SEMATECH, 2011]. A d-optimal design can be thought of as selecting the corner points on a six-dimensional hypercube created from the six factors, allowing the model to interpolate in between these corner points. For the two continuous factors, center points are also selected in those dimensions allowing a quadratic interpolation to be performed instead of just a linear interpolation. For the single ternary factor, one of the settings can be considered a center point of a two-section piecewise linear function.

2.2.1.1 Factors

The factors used in the experiment included settings of ERAM that can be changed in the current version as well as prototype upgrades. The prototype upgrades would require code enhancements to the current version of ERAM.

The lateral and longitudinal bounds of the conformance box were varied independently from each other. This variance did not include the prototype changes in FA18 Interim 2 [Crowell et al, June 2011] [Lapihuska, 2011] that decoupled the TM bounds from the CP bounds. Instead all changes to the conformance bounds affected both the TM and CP bounds. These bounds are continuous factors, modeled using a quadratic equation. Ranges of the two continuous factors are listed in Table 2.

Table 2. Continuous Factors of the Integrated Experiment

Factor	Min	Max
Lateral Bound	0.5 nm	2.5 nm
Longitudinal Bound	1.0 nm	1.5 nm

In some initial experiments, likelihood appeared to be a significant factor. In order to further understand the effects of likelihood, it was varied among three discrete values. Effects of likelihood cannot easily be modeled as a continuous function because the likelihood parameter is a function in itself. The functions used for likelihood contain either two or three parameters. When two parameters are used, the first one is the maximum time in minutes at which a likelihood value of 0.0 will generate an alert. The second parameter is the minimum at which a likelihood value of 1.0 is required in order to generate an alert. This creates a linear function similar to that shown in Figure 4. The white area above the line is where the likelihood must fall in order for an alert to be generated. When three parameters are used, it becomes a piecewise linear function, with the first parameter being the maximum time at which a likelihood value of 0.0 will generate an alert. The last parameter is the minimum time in minutes at which a likelihood value of 1.0 is required in order to generate an alert, and the center parameter is the time in minutes at which a chosen value is required in order to generate an alert. In the case of the 3-parameter settings used in this experiment, this chosen value is set to 0.9. This results in a function like the one shown in Figure 5. In this experiment, two different values for the likelihood function were used. The settings used for likelihood are shown in Table 3.

Table 3. Nominal Factors of the Integrated Experiment

Factor	Settings	
Likelihood Function	10|20	4|8|20

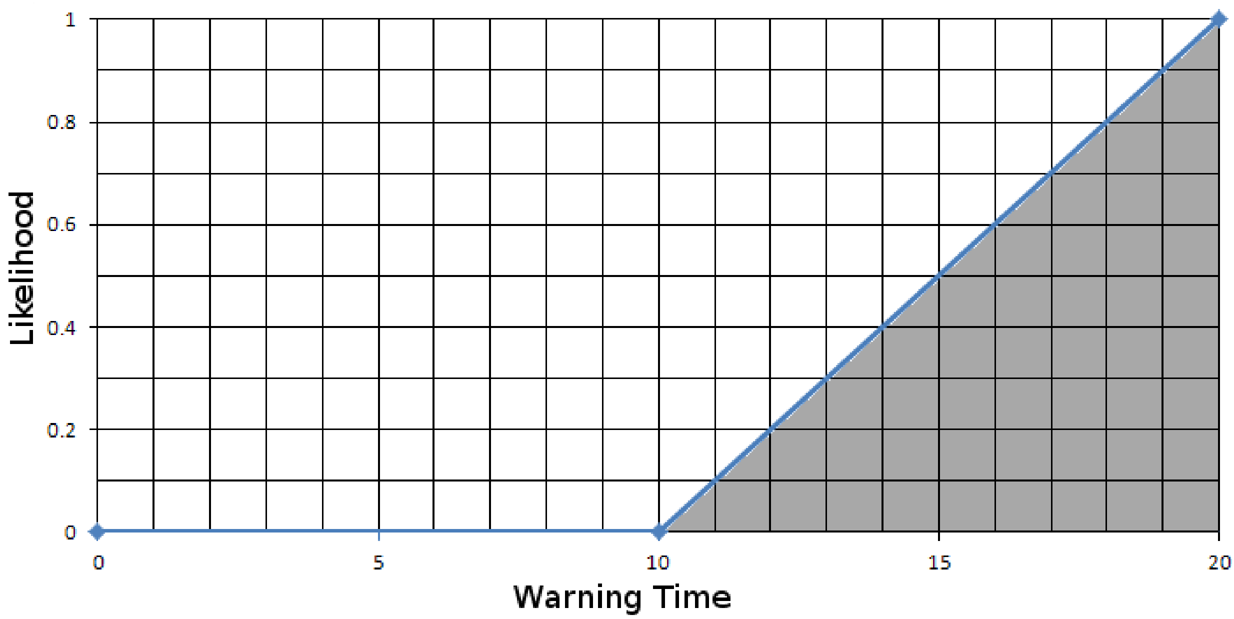

Figure 4. Likelihood Function for 10|20 Setting

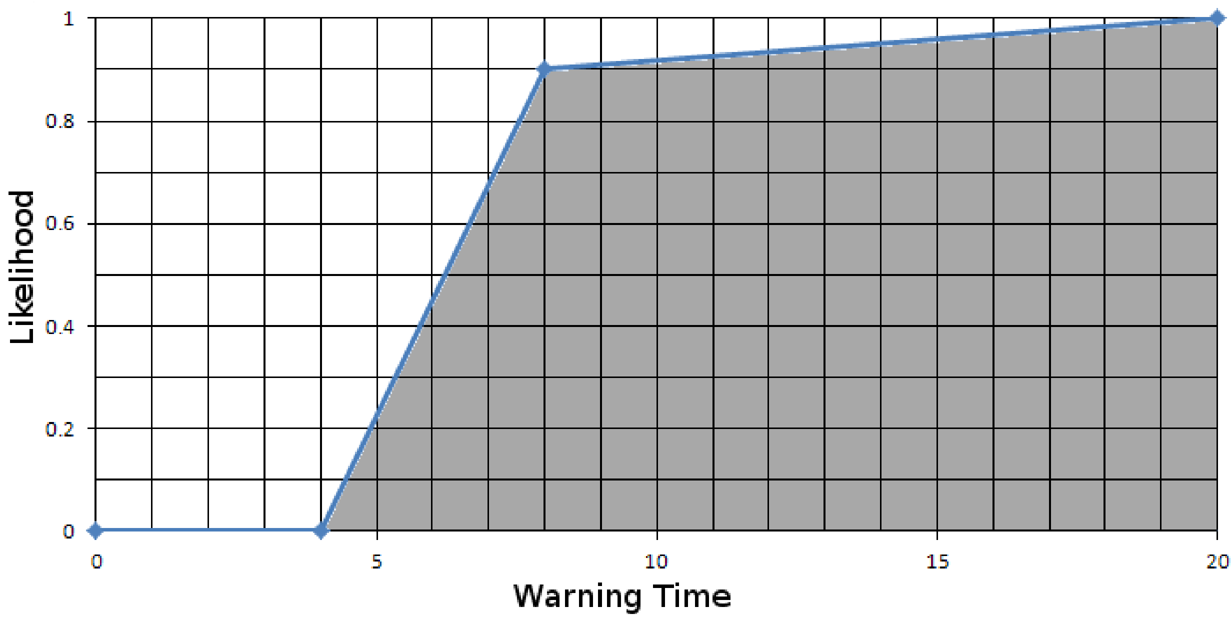

Figure 5. Likelihood Function for 4|8|20 Setting

The prototype enhancements are binary factors, either running or not running. These enhancements were described in Section 1.1.1. They include FTR (Forced Trajectory Rebuild), GAB (Growth Adherence Bounds), and CGS (Conflict Geometric Separation).

The settings described above resulted in the 40 runs shown in Table 4. Also shown in this table are the settings used currently in the deployed version of ERAM. This run is referred to as the baseline (BL) run.

Table 4. Runs for the Integrated Experiment

Run	FTR	GAB	CGS	Lat	Lon	Likelihood
1	Off	Off	Off	0.5	1.5	10|20
2	Off	Off	Off	1.5	1.5	4|8|20
3	Off	Off	Off	2.5	1.0	10|20
4	Off	Off	Off	2.5	1.3	4|8|20
5	Off	Off	On	0.5	1.0	10|20
6	Off	Off	On	0.5	1.5	4|8|20
7	Off	Off	On	1.5	1.0	4|8|20
8	Off	Off	On	2.5	1.5	10|20
9	Off	On	Off	0.5	1.0	4|8|20
10	Off	On	Off	0.5	1.5	4|8|20
11	Off	On	Off	1.5	1.3	10|20
12	Off	On	Off	2.5	1.5	10|20
13	Off	On	On	0.5	1.3	4|8|20
14	Off	On	On	0.5	1.5	10|20
15	Off	On	On	2.5	1.0	10|20
16	Off	On	On	2.5	1.5	4|8|20
17	On	Off	Off	0.5	1.0	4|8|20
18	On	Off	Off	0.5	1.5	10|20
19	On	Off	Off	2.5	1.0	10|20
20	On	Off	Off	2.5	1.5	4|8|20
21	On	Off	On	0.5	1.0	10|20
22	On	Off	On	0.5	1.5	4|8|20
23	On	Off	On	1.5	1.5	10|20
24	On	Off	On	2.5	1.0	4|8|20
25	On	On	Off	0.5	1.0	10|20
26	On	On	Off	0.5	1.5	4|8|20
27	On	On	Off	2.5	1.0	4|8|20
28	On	On	Off	2.5	1.5	10|20
29	On	On	On	0.5	1.0	4|8|20
30	On	On	On	0.5	1.5	10|20
31	On	On	On	2.5	1.3	10|20
32	On	On	On	2.5	1.5	4|8|20
33	On	On	On	1.0	1.3	4|8|20
34	On	Off	On	1.0	1.0	10|20
35	On	On	Off	1.0	1.3	4|8|20
36	Off	On	On	1.0	1.5	10|20
37	Off	On	Off	1.5	1.5	4|8|20
38	Off	On	Off	2.5	1.0	10|20
39	Off	Off	On	1.5	1.5	4|8|20
40	Off	Off	On	2.5	1.0	10|20
BL	Off	Off	Off	2.5	1.5	10|20

2.2.1.2 Model

The initial model allowed both of the continuous factors to have at most a quadratic effect. It was assumed that all factors could interact only in pairs (two-way interactions only). The constant or overall mean effect is represented in the model as μ, and $\varepsilon_{n(fghijk)}$ represents the assumption of independently normally distributed random error with a zero mean. All factors are assumed to be additive. The model is defined as in Eq. 1.

Response:

$$
\begin{aligned}
R^{0}_{fghijk} = \mu &+ Lat_i^2 + Lat_i + Long_j^2 + Long_j + Lat_i Long_j \\
&+ Like_k + Lat_i Like_k + Long_j Like_k + FTR_f Lat_i + FTR_f Long_j \\
&+ FTR_f Like_k + GAB_g Lat_i + GAB_g Long_j + GAB_g Like_k \\
&+ CGS_h Lat_i + CGS_h Long_j + CGS_h Like_k + FTR_f GAB_g \\
&+ FTR_f CGS_h + GAB_g CGS_h + FTR_f + GAB_g + CGS_h + \varepsilon_{n(fghijk)}
\end{aligned}
\qquad \textbf{Eq. 1}
$$

Where:
FTR_f = forced trajectory rebuild prototype, f = on, off
GAB_g = growth adherence bounds prototype, g = on, off
CGS_h = conflict geometric separation prototype, h = on, off
Lat_i = lateral conformance bounds in nautical miles, i = 0.5, 1.0, 1.5, 2.5
$Long_j$ = longitudinal conformance bounds in nautical miles, j = 1.0, 1.25, 1.5
$Like_k$ = likelihood, k = “10/20”, “4/8/20”
$\varepsilon_{n(fghijk)}$ = random error, n = 1, 2, … for all f, g, h, i, j, k

This model fits the data very well for the FA % response (Figure 6) and fairly well for the WT % response (Figure 7). The LA % response does not fit as well (Figure 8). However, there is no room for improvement in LA % given the current data set. The horizontal blue line in these figures represents the mean value of the samples and the red curves indicate the 95% confidence interval. The significance is quickly established in a leverage plot by determining if the confidence interval intersects the mean. No intersection indicates insignificance. All responses demonstrate statistical significance.

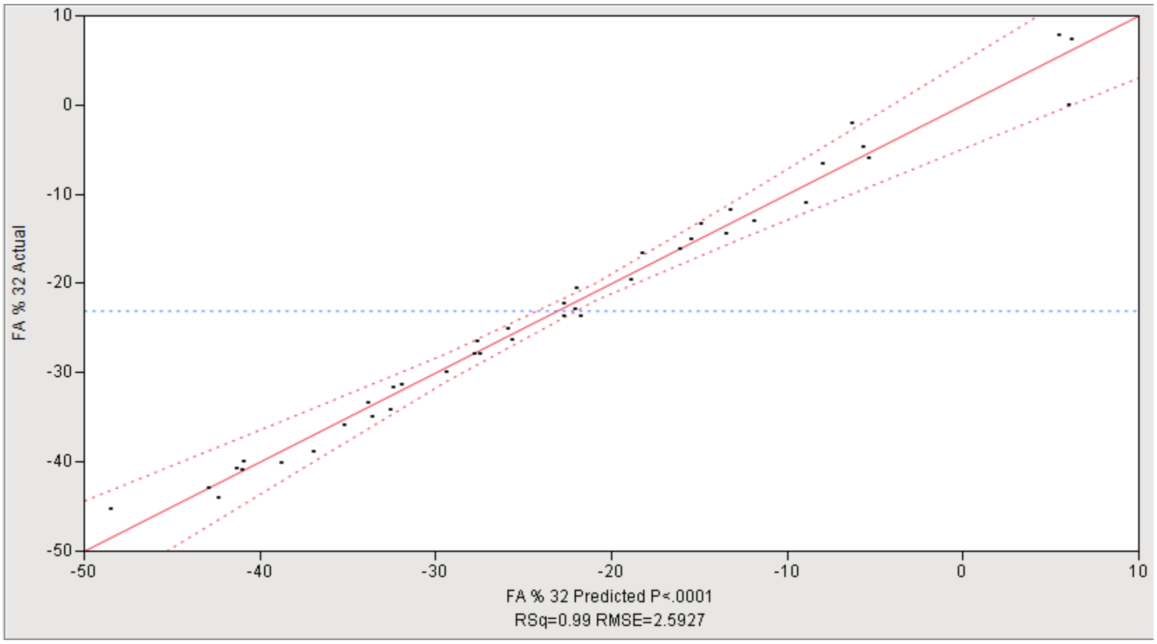

Figure 6. FA % Response Model Fit to Data (R^2 = 0.99, RMSE = 2.5927)

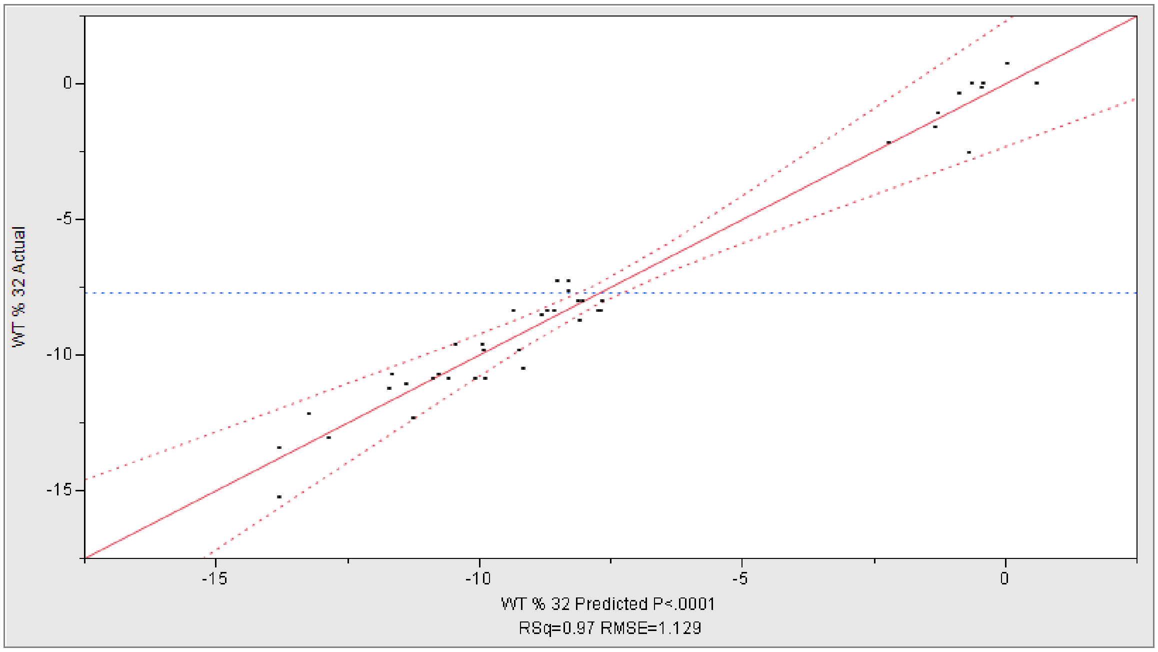

Figure 7. WT % Response Model Fit to Data (R^2 = 0.97, RMSE = 1.129)

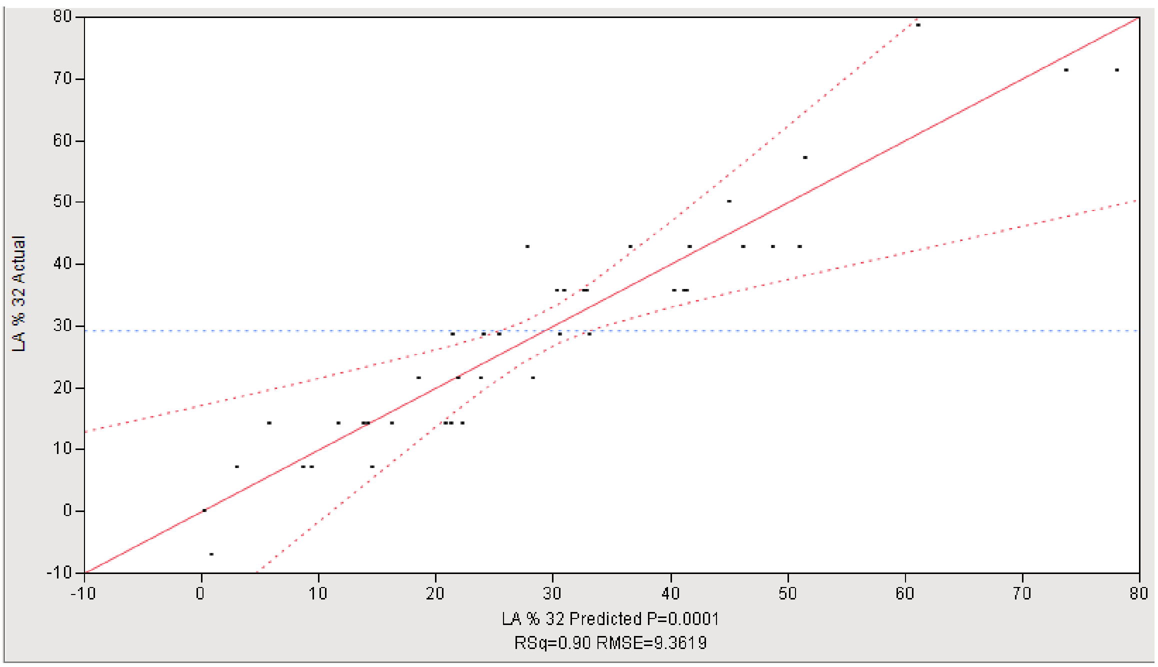

Figure 8. LA % Response Model Fit to Data (R^2 = 0.90, RMSE = 9.3619)

The model also relies on the assumption that the random error $\varepsilon_{n(ijk)}$ is normally distributed. The residual errors should therefore be tested for normality. Figure 9 shows histograms and normal quantile plots for the responses. The normal probability plots illustrate that for each response, the model errors fall within the confidence interval along the diagonal line of the plot, indicating that each residual is at least approximately normally distributed. This provides evidence that the residual errors in FA and WT % are normal and the model is indeed appropriate. The sensitivity of the data in LA % makes it harder to determine normality, but the data is not skewed enough to warrant rejecting LA % residuals as normal. Thus, the model passes the test for having normally distributed residual errors.

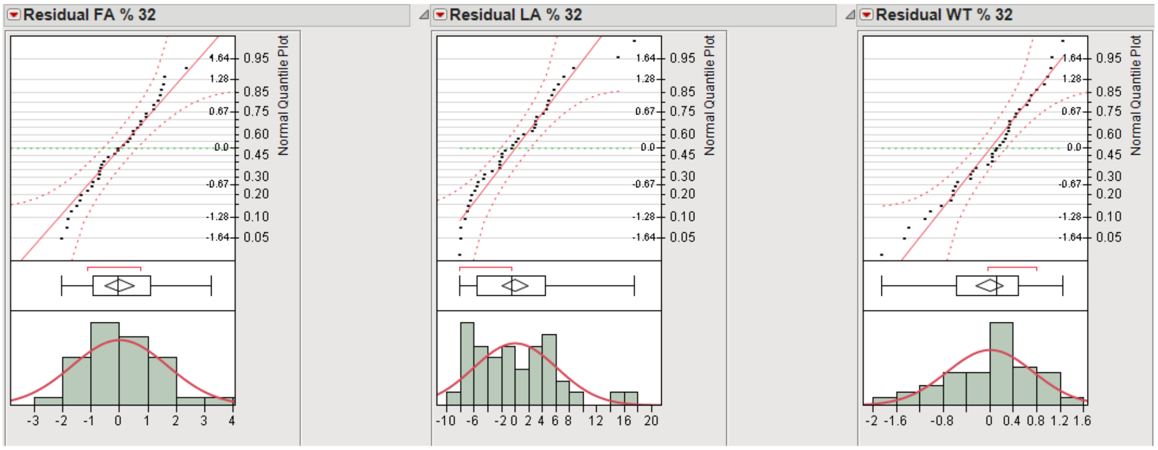

Figure 9. Residual Error Plots for FA, LA, and WT

2.3 Trajectory Modeling Enhancements

This section evaluates algorithmic enhancements to the aircraft trajectory modeling that have been implemented in the En Route Automation Modernization (ERAM) prototyping effort. The algorithmic details of the planned prototype effort can be found in the Lockheed Martin report [McKay, 2011] delivered as part of the Separation Management Task Order 51 activity for Functional Area 32 (FA32) –trajectory modeling improvement.

In earlier ERAM releases, trajectory modeling only begins at the track position on a lateral re-adherence of the trajectory or upon change of track control. Otherwise, the initial point is taken to be the track projection onto the previous trajectory. On some trajectory rebuilds, the trajectory's initial point could be up to 2.5nm from the latest track reported position. This prototype effort changes the aircraft trajectory algorithm so that it always starts at the latest track position and investigates different lateral rejoin enhancements. It provides for risk reduction, the firming up of algorithmic changes, an associated accuracy benefit, and accelerated software development and implementation strategy.

These changes to the trajectory modeling were included in all of the treatment runs for this experiment. Therefore, a new run at the baseline settings that also includes the FA32 trajectory modeling enhancements is needed. This chapter provides a comparison of this new Baseline (BL) to the Initial Baseline (IBL) run in order to provide a perspective for the treatment run analyses.

2.3.1 Analysis Description

The IBL is detailed in [Crowell and Young, 2012] and can be used to demonstrate the performance differences of the BL. The FA32 update introduces an enhanced method of trajectory modeling, and a new scenario was generated using this method. The purpose of this section is to analyze the effect of the trajectory modeling enhancements by comparing the BL to the IBL in terms of trajectory accuracy. Specifically, the accuracy of the trajectories generated in each scenario is calculated using the simulated track data as a baseline.

FA32 scenario data was provided by the Lockheed Martin Corporation and run through ANG-C41's software tools using the same process as in the IBL scenario. The conformance bounds and likelihood parameter were set to match the values in the IBL run, which were 2.5 nm lateral, 1.5 nm longitudinal and 10/20 likelihood. Therefore, the only difference between the two runs is the FA32 trajectory modeling enhancement.

2.3.2 Trajectory Accuracy Results

The trajectory analysis results are presented in this section. Average errors are calculated for the same flight in the IBL and BL runs; these values are collected and compared over many flights to illustrate any underlying differences between the runs. A matched pair analysis is performed for each of the four metrics using the average error per flight between the IBL and BL with the trajectory modeling enhancements. The analysis includes a paired *t*-test, which examines the distribution of differences in error between the two scenarios and tests if the mean of the differences is statistically different from zero. The results from the paired *t*-tests are provided in Table 5. In Figure 10, each of the four graphs plots the difference in average trajectory error per flight from the IBL to the BL against the mean of the two errors. Appendix C of Crowell et al. [2011a] provides a detailed description of the matched pair analysis and graphical output. Positive values indicate that trajectories in the IBL have more error on average than in the BL run.

Table 5. Statistical Results for Trajectory Error

Error	Mean Diff	Std Error	*p*-value
Horizontal (nm)	0.1381	0.0104	< 0.0001
Abs. Vertical (ft)	5.0931	2.3315	0.0290
Abs. Cross Track (nm)	0.1843	0.0078	< 0.0001
Abs. Along Track (nm)	0.0062	0.0150	0.6770

From the results in Table 5, the BL has less average error per flight for all four types of error. However, the observed difference in vertical and along track error is too small to be practically significant. To determine whether the test results are statistically significant, *p*-values are presented in the table. Each *p*-value is the probability of observing a discrepancy in means at least as large as that observed if there was no underlying difference in the means. A *p*-value less than 0.05 is typically considered to indicate statistical significance of a result. In this analysis the mean difference is statistically significant for horizontal, absolute vertical, and absolute cross track errors. The greatest effect is exhibited in cross track error. This improvement is also reflected in horizontal error, which is an aggregate metric.

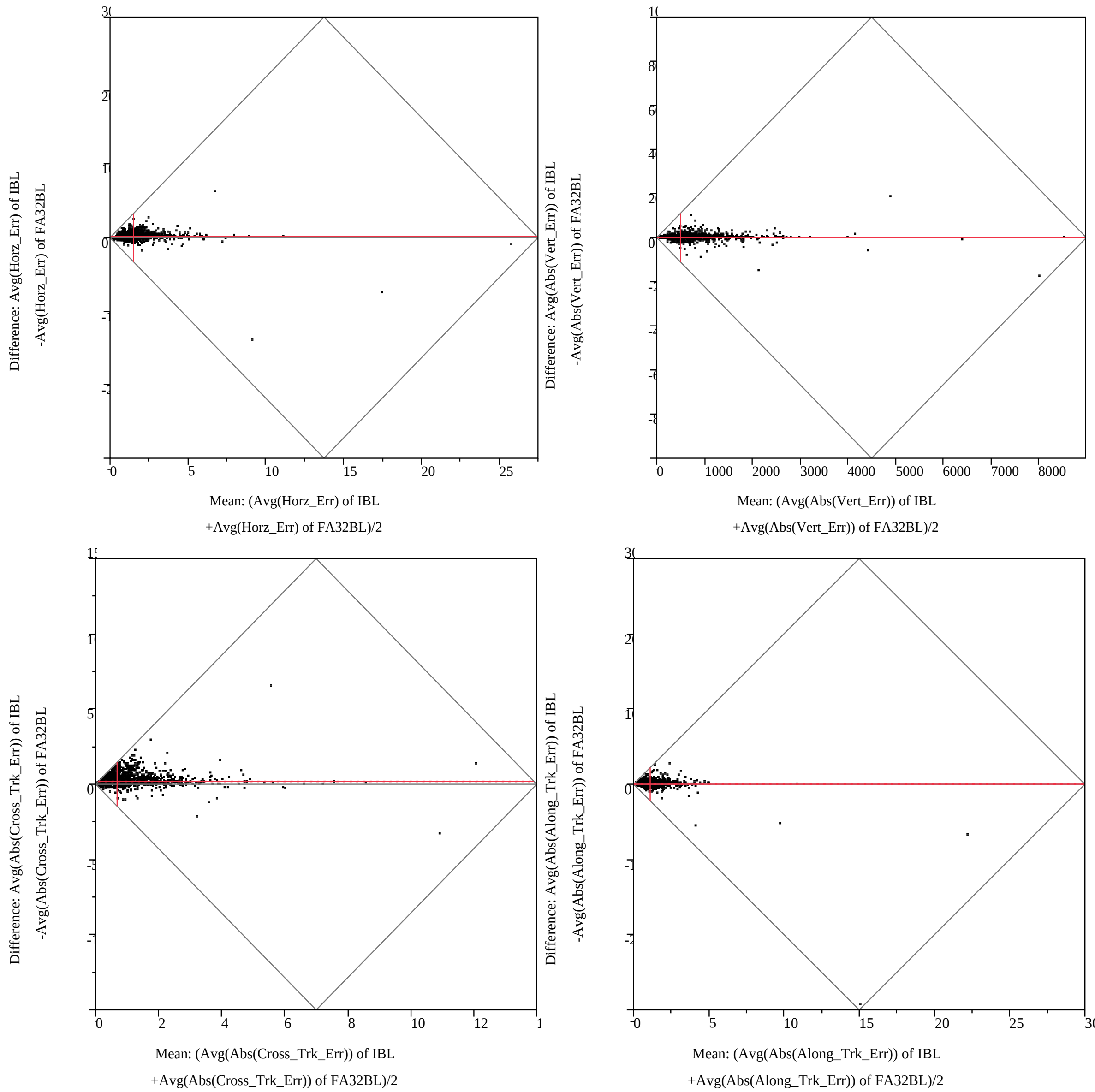

Figure 10. Matched Pair Analyses for Trajectory Error Metrics

Figure 10 illustrates the same results of the matched pair analysis graphically. It presents the measurements for each metric using a special plot from the commercial statistical software package called JMP®. It plots each paired difference, with the vertical axis being the difference of the flight's average error in the IBL run minus the respective mean in the treatment run. The horizontal axis is the average of these two measurements. The resulting plot normalizes the error differences indicating net trends between runs. For example, if the errors are predominantly above the red line at zero in the middle of the plot, this indicates the error is larger in the baseline run more often. If the errors are below this line, they indicate the error is larger in the treatment run (in this case, the BL run). Figure 10 illustrates a slight trend of error differences being above the zero line, supporting the same conclusion as drawn from the results of Table 5. The effect of increased accuracy is strongest for cross track error, indicating that the trajectory accuracy benefit of the FA32 enhancements is mainly observed in this dimension.

3 Performance Evaluation

The performance evaluation analyses used in this study are similar to those used in Experiment 2 [Crowell et al, December 2011b]. The metrics used are those described in the documentation of Experiment 1 [Crowell et al, December 2011a]. An integrated experiment was designed, similar to that used in Experiment 2 but containing different settings for the likelihood and longitudinal parameters, as described in Section 2.2.1.1. Unlike Experiments 1 and 2, this experiment was designed to allow the analysts to determine the effects of prototype enhancements and parameter changes in a single experiment. The analyses on the trajectory modeling (TM) and conflict probe (CP) performance are described in detail.

3.1 Trajectory Modeling Analysis

To observe the effect size of each individual algorithm enhancement, the matched pair analysis technique was applied to the average absolute error per flight. The average error value was compared for the same flight in a scenario with FTR on and with FTR off. Identical settings were used for the other parameters and all of the other prototypes were turned off. Table 6 presents the results of the analyses.

Table 6. Matched Pair Test Results- FTR Effect Isolated

Lat	Long	Mean Diff (Cross Track)	*p*-value (Cross Track)	Mean Diff (Along Track)	*p*-value (Along Track)
0.5	1.5	0.0035	< 0.0001	0.0470	< 0.0001
2.5	1	0.0212	< 0.0001	0.0402	< 0.0001

This process was repeated for scenario pairs that effectively isolate GAB the same way, comparing the average error value for each flight in a scenario with GAB on to the average value from the same flight in a scenario with GAB off. Results from three such pairs of scenarios are presented in Table 7. Finally, Table 8 contains results from scenarios that isolate CGS.

Table 7. Matched Pair Test Results- GAB Effect Isolated

Lat	Long	Mean Diff (Cross Track)	*p*-value (Cross Track)	Mean Diff (Along Track)	*p*-value (Along Track)
1.5	1.5	0.0004	0.7334	-0.0037	0.4344
2.5	1	-0.0003	0.8071	0.0044	0.0306
2.5	1.5	-0.0018	0.2027	-0.0010	0.6339

Table 8. Matched Pair Test Results- CGS Effect Isolated

Lat	Long	Mean Diff (Cross Track)	*p*-value (Cross Track)	Mean Diff (Along Track)	*p*-value (Along Track)
1.5	1.5	0.0028	0.0029	0.0009	0.6605
2.5	1	0.0003	0.8380	0.0042	0.0409
2.5	1.5	0.0003	0.8139	-0.0015	0.4530

Positive mean difference values in the above tables indicate an improvement in trajectory accuracy (decrease in average absolute error by flight) when the algorithm enhancement is enabled. The results in Table 6 for FTR indicate that for average absolute along track and cross track error, the mean differences between scenarios have associated *p*-values of less than 0.0001, which means the effect of the FTR enhancement in this sample data is statistically significant. However, at magnitudes below 0.1nm, these effects are not practically significant because they are on the scale of radar noise. The results in Table 7 for GAB indicate inconsistent results in terms of trajectory error, as do the results in Table 8 for the CGS enhancement. Further work is needed to provide proof of increased trajectory accuracy.

3.2 Conflict Probe Analysis

Several analyses were performed on this experiment to determine the effects that each factor has on the performance of the Conflict Probe (CP). Since this experiment focuses on both the prototype enhancements and the parameter changes, the null hypothesis is different than the previous two experiments:

> *A significant Conflict Probe performance improvement is not observed through parameter changes of the likelihood, or lateral and longitudinal conformance bounds, nor through the prototype enhancements of Growth Adherence Bounds, Forced Trajectory Rebuild, or Conflict Geometry Separation.*

A significant performance improvement is defined as a reduction in False Alert Rate greater than the confidence interval, no increase in Late Alert Rate greater than the confidence interval, and a 25th percentile of warning time above the three minute threshold. All of these requirements must be true in order for it to be considered a significant improvement.

The study documented here will attempt to reject this null hypothesis, therefore showing that these enhancements or parameter changes may indeed provide a significant improvement to the ERAM system.

In all analyses performed in this study there are two baseline runs used. The first is referred to as the Initial Baseline (IBL) and is the run utilizing the current probe being used in the live ERAM system. This baseline run does not include any prototype performance enhancements and is run at the settings currently being used in the live system. The second baseline run is referred to as the Baseline (BL). This run uses the settings of the current live ERAM system, but includes trajectory lateral modeling enhancements defined under Function Area 32 ap1 [McKay, 2011]. The result of the analysis on the FA32 ap1 was a recommendation for addition of those enhancements to the trajectory modeler. As a result, all following experiments assume those lateral modeling enhancements will be included in the system when these additional conflict probe enhancements are implemented.

Table 9 shows the alert type counts for the 40 treatment runs and the two baselines. This table is mainly here for documentation of the experiment, but a few observations can be made from it. First, Runs 19 and 3 are the only treatment runs that show increased FA count over the baseline. Run 29 shows the biggest improvement to False Alert count, but also has a significant increase to Late Alerts. Run 29 has the most extreme settings with all enhancements on, 4|8|20 likelihood, and the lowest possible conformance bound settings. Only Run 2 shows no increase to Late Alerts or Missed Alerts. Run 2 also shows a significant improvement to FA count, though the improvement is not as large as some of the other runs. This run differs only in modified parameter settings and has no prototype enhancements enabled.

Table 9. Alert Type Counts

Run	FTR	GAB	CGS	Lat	Lon	Llh	VA	FA	LA	MA
1	Off	Off	Off	0.5	1.5	10|20	183	974	9	6
2	Off	Off	Off	1.5	1.5	4|8|20	185	980	8	5
3	Off	Off	Off	2.5	1	10|20	180	1255	12	5
4	Off	Off	Off	2.5	1.25	4|8|20	180	1114	10	6
5	Off	Off	On	0.5	1	10|20	181	860	11	5
6	Off	Off	On	0.5	1.5	4|8|20	183	779	10	5
7	Off	Off	On	1.5	1	4|8|20	178	799	14	6
8	Off	Off	On	2.5	1.5	10|20	181	1031	13	4
9	Off	On	Off	0.5	1	4|8|20	181	691	13	5
10	Off	On	Off	0.5	1.5	4|8|20	181	760	10	6
11	Off	On	Off	1.5	1.25	10|20	177	1000	14	6
12	Off	On	Off	2.5	1.5	10|20	183	1145	10	5
13	Off	On	On	0.5	1.25	4|8|20	182	715	10	6
14	Off	On	On	0.5	1.5	10|20	182	908	10	6
15	Off	On	On	2.5	1	10|20	179	940	12	6
16	Off	On	On	2.5	1.5	4|8|20	180	803	12	5
17	On	Off	Off	0.5	1	4|8|20	181	654	12	7
18	On	Off	Off	0.5	1.5	10|20	178	992	11	9
19	On	Off	Off	2.5	1	10|20	178	1261	11	8
20	On	Off	Off	2.5	1.5	4|8|20	181	1091	9	9
21	On	Off	On	0.5	1	10|20	177	861	12	8
22	On	Off	On	0.5	1.5	4|8|20	179	749	11	8
23	On	Off	On	1.5	1.5	10|20	175	1013	18	6
24	On	Off	On	2.5	1	4|8|20	173	818	15	9
25	On	On	Off	0.5	1	10|20	180	842	11	7
26	On	On	Off	0.5	1.5	4|8|20	182	702	9	7
27	On	On	Off	2.5	1	4|8|20	178	770	14	6
28	On	On	Off	2.5	1.5	10|20	179	1040	12	7
29	On	On	On	0.5	1	4|8|20	179	640	13	6
30	On	On	On	0.5	1.5	10|20	175	893	12	9
31	On	On	On	2.5	1.25	10|20	178	929	12	8
32	On	On	On	2.5	1.5	4|8|20	175	699	14	8
33	On	On	On	1	1.25	4|8|20	180	667	13	6
34	On	Off	On	1	1	10|20	172	893	16	9
35	On	On	Off	1	1.25	4|8|20	181	692	12	5
36	Off	On	On	1	1.5	10|20	181	902	11	5
37	Off	On	Off	1.5	1.5	4|8|20	182	875	11	4
38	Off	On	Off	2.5	1	10|20	182	1100	10	6
39	Off	Off	On	1.5	1.5	4|8|20	179	843	14	4
40	Off	Off	On	2.5	1	10|20	179	1017	14	5
BL	Off	Off	Off	2.5	1.5	10|20	183	1317	10	4
IBL	Off	Off	Off	2.5	1.5	10|20	186	1360	8	4

Table 10. Adjusted Late Alert Value Compared to LA+MA Counts

Run	FTR	GAB	CGS	Lat	Lon	Likelihood	LA+MA	Adj LA
1	Off	Off	Off	0.5	1.5	10|20	15	11.0752
2	Off	Off	Off	1.5	1.5	4|8|20	13	9.2781
3	Off	Off	Off	2.5	1.0	10|20	17	10.2695
4	Off	Off	Off	2.5	1.3	4|8|20	16	10.9979
5	Off	Off	On	0.5	1.0	10|20	16	11.3934
6	Off	Off	On	0.5	1.5	4|8|20	15	11.1325
7	Off	Off	On	1.5	1.0	4|8|20	20	11.7178
8	Off	Off	On	2.5	1.5	10|20	17	10.1995
9	Off	On	Off	0.5	1.0	4|8|20	18	12.1353
10	Off	On	Off	0.5	1.5	4|8|20	16	11.4478
11	Off	On	Off	1.5	1.3	10|20	20	12.4343
12	Off	On	Off	2.5	1.5	10|20	15	10.3678
13	Off	On	On	0.5	1.3	4|8|20	16	12.2573
14	Off	On	On	0.5	1.5	10|20	16	11.835
15	Off	On	On	2.5	1.0	10|20	18	11.7227
16	Off	On	On	2.5	1.5	4|8|20	17	11.4188
17	On	Off	Off	0.5	1.0	4|8|20	19	12.6787
18	On	Off	Off	0.5	1.5	10|20	20	15.1183
19	On	Off	Off	2.5	1.0	10|20	19	13.5272
20	On	Off	Off	2.5	1.5	4|8|20	18	14.4026
21	On	Off	On	0.5	1.0	10|20	20	14.4719
22	On	Off	On	0.5	1.5	4|8|20	19	13.9318
23	On	Off	On	1.5	1.5	10|20	24	14.7155
24	On	Off	On	2.5	1.0	4|8|20	24	15.3088
25	On	On	Off	0.5	1.0	10|20	18	13.448
26	On	On	Off	0.5	1.5	4|8|20	16	12.498
27	On	On	Off	2.5	1.0	4|8|20	20	13.3264
28	On	On	Off	2.5	1.5	10|20	19	13.1444
29	On	On	On	0.5	1.0	4|8|20	19	13.477
30	On	On	On	0.5	1.5	10|20	21	15.3853
31	On	On	On	2.5	1.3	10|20	20	14.0269
32	On	On	On	2.5	1.5	4|8|20	22	15.6699
33	On	On	On	1.0	1.3	4|8|20	19	13.8532
34	On	Off	On	1.0	1.0	10|20	25	16.7893
35	On	On	Off	1.0	1.3	4|8|20	17	11.9922
36	Off	On	On	1.0	1.5	10|20	16	11.4846
37	Off	On	Off	1.5	1.5	4|8|20	15	9.3602
38	Off	On	Off	2.5	1.0	10|20	16	10.6265
39	Off	Off	On	1.5	1.5	4|8|20	18	11.418
40	Off	Off	On	2.5	1.0	10|20	19	11.3361
BL	Off	Off	Off	2.5	1.5	10|20	14	9.1641
IBL	Off	Off	Off	2.5	1.5	10|20	12	7.8676

As always, it is important to consider the warning time of the alerts. There are two metrics used to evaluate the warning time. The first is Adjusted LA, shown in Table 10. This metric adjusts the LA+MA count by the amount of warning time provided by the conflict probe. This will reduce the value for a LA that has a high warning time to some value between 1.0 and 0.0, whereas a MA with no warning time will count as a value of 1.0. There is a slight increase in the Adjusted LA of Run 2 over the baseline indicating that, although there are actually less Late Alerts, the

LAs and MAs within Run 2 contain less warning time. This is not an unexpected result because of the smaller conformance bounds and the reduced likelihood setting.

Table 11 shows the warning time metrics for each of the treatment runs and the baselines. Three warning time metrics are used. The median is included to give an idea of how the run performed overall. The 25^{th} percentile is the metric of most interest which illustrates how the lower end of the alerts performed with regard to warning time. This metric is used because it represents how close the lower end of warning time distribution is to being called Late Alerts. Increasing this value is much more desirable than increasing the median, which is often far above the warning time requirement of 180 seconds. The inter-quartile range (IQR) illustrates the range between the 75^{th} percentile and the 25^{th} percentile. This value can help explain some of the differences between runs. A larger IQR indicates that a conflict probe increases the warning time of those alerts that already have a lot of warning time, or decreases the warning time of those that have little. The IQR can be observed along with the 25^{th} percentile value to get an idea of the shape of the curve of warning times. From here on, the 25^{th} percentile of warning time will be used as the main warning time metric.

This table shows that overall the runs performed very well in regards to warning time. The lowest 25^{th} percentile is 233 seconds, which is still 53 seconds above the required 180 seconds, and only 42 seconds lower than the baseline.

Table 11. Median, Inter-Quartile Range, and 25th Percentile of Conflict Warning Time

Run	FTR	GAB	CGS	Lat	Lon	Llh	Med	IQR	25th %
1	Off	Off	Off	0.5	1.50	10|20	403.0	371.0	269.0
2	Off	Off	Off	1.5	1.50	4|8|20	336.0	208.0	254.0
3	Off	Off	Off	2.5	1.00	10|20	383.0	331.0	275.0
4	Off	Off	Off	2.5	1.25	4|8|20	329.0	241.0	253.0
5	Off	Off	On	0.5	1.00	10|20	362.0	351.0	252.0
6	Off	Off	On	0.5	1.50	4|8|20	316.0	201.0	241.0
7	Off	Off	On	1.5	1.00	4|8|20	317.0	198.5	244.5
8	Off	Off	On	2.5	1.50	10|20	363.0	354.5	253.0
9	Off	On	Off	0.5	1.00	4|8|20	328.0	189.5	252.0
10	Off	On	Off	0.5	1.50	4|8|20	332.0	214.0	246.0
11	Off	On	Off	1.5	1.25	10|20	400.0	368.0	268.0
12	Off	On	Off	2.5	1.50	10|20	410.5	372.0	275.0
13	Off	On	On	0.5	1.25	4|8|20	315.0	192.0	244.0
14	Off	On	On	0.5	1.50	10|20	363.0	358.0	252.0
15	Off	On	On	2.5	1.00	10|20	362.5	350.0	252.0
16	Off	On	On	2.5	1.50	4|8|20	316.5	229.0	245.0
17	On	Off	Off	0.5	1.00	4|8|20	328.0	186.0	252.0
18	On	Off	Off	0.5	1.50	10|20	397.0	357.0	272.0
19	On	Off	Off	2.5	1.00	10|20	395.0	378.5	274.5
20	On	Off	Off	2.5	1.50	4|8|20	332.0	249.0	251.0
21	On	Off	On	0.5	1.00	10|20	358.5	343.0	248.0
22	On	Off	On	0.5	1.50	4|8|20	316.0	195.5	245.5
23	On	Off	On	1.5	1.50	10|20	354.0	357.0	245.0
24	On	Off	On	2.5	1.00	4|8|20	313.5	236.5	238.0
25	On	On	Off	0.5	1.00	10|20	398.5	366.0	274.0
26	On	On	Off	0.5	1.50	4|8|20	334.0	198.0	255.0
27	On	On	Off	2.5	1.00	4|8|20	330.5	237.0	245.0
28	On	On	Off	2.5	1.50	10|20	400.0	425.5	270.5
29	On	On	On	0.5	1.00	4|8|20	305.0	204.0	233.0
30	On	On	On	0.5	1.50	10|20	345.0	346.0	245.0
31	On	On	On	2.5	1.25	10|20	363.0	373.5	248.5
32	On	On	On	2.5	1.50	4|8|20	320.0	242.0	239.0
33	On	On	On	1.0	1.25	4|8|20	316.0	195.5	241.5
34	On	Off	On	1.0	1.00	10|20	360.5	352.0	248.5
35	On	On	Off	1.0	1.25	4|8|20	330.0	234.0	248.0
36	Off	On	On	1.0	1.50	10|20	364.0	351.0	255.0
37	Off	On	Off	1.5	1.50	4|8|20	334.0	221.0	251.5
38	Off	On	Off	2.5	1.00	10|20	384.0	340.0	277.0
39	Off	Off	On	1.5	1.50	4|8|20	316.5	204.5	245.5
40	Off	Off	On	2.5	1.00	10|20	363.0	353.0	253.0
BL	Off	Off	Off	2.5	1.50	10|20	415.0	379.0	276.0
IBL	Off	Off	Off	2.5	1.50	10|20	415.0	365.0	283.0

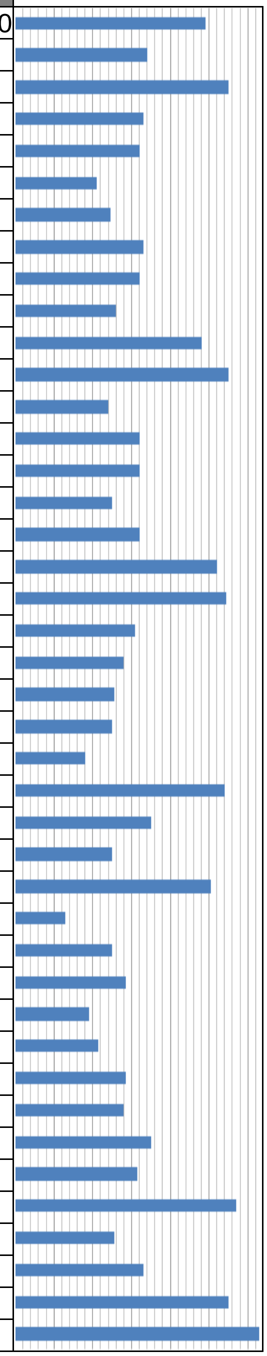

Figure 11 shows the Hit Rate [Crowell et al, December 2011a] versus the FA Rate for each of the 40 treatment runs and two baselines. Each point on the plot represents one of the 40 treatment runs or the two baselines and is labeled as such. The color of each point represents the percentage difference of the 25th percentile of warning time from the FA32 Baseline. The legend to the right of the plot shows that the color moves from blue to red as the warning time decreases. Since the goal is to not decrease warning time, a blue color is desirable. The goal is also to decrease FA Rate while retaining Hit Rate, so the most desirable location is the top-left corner of the plot.

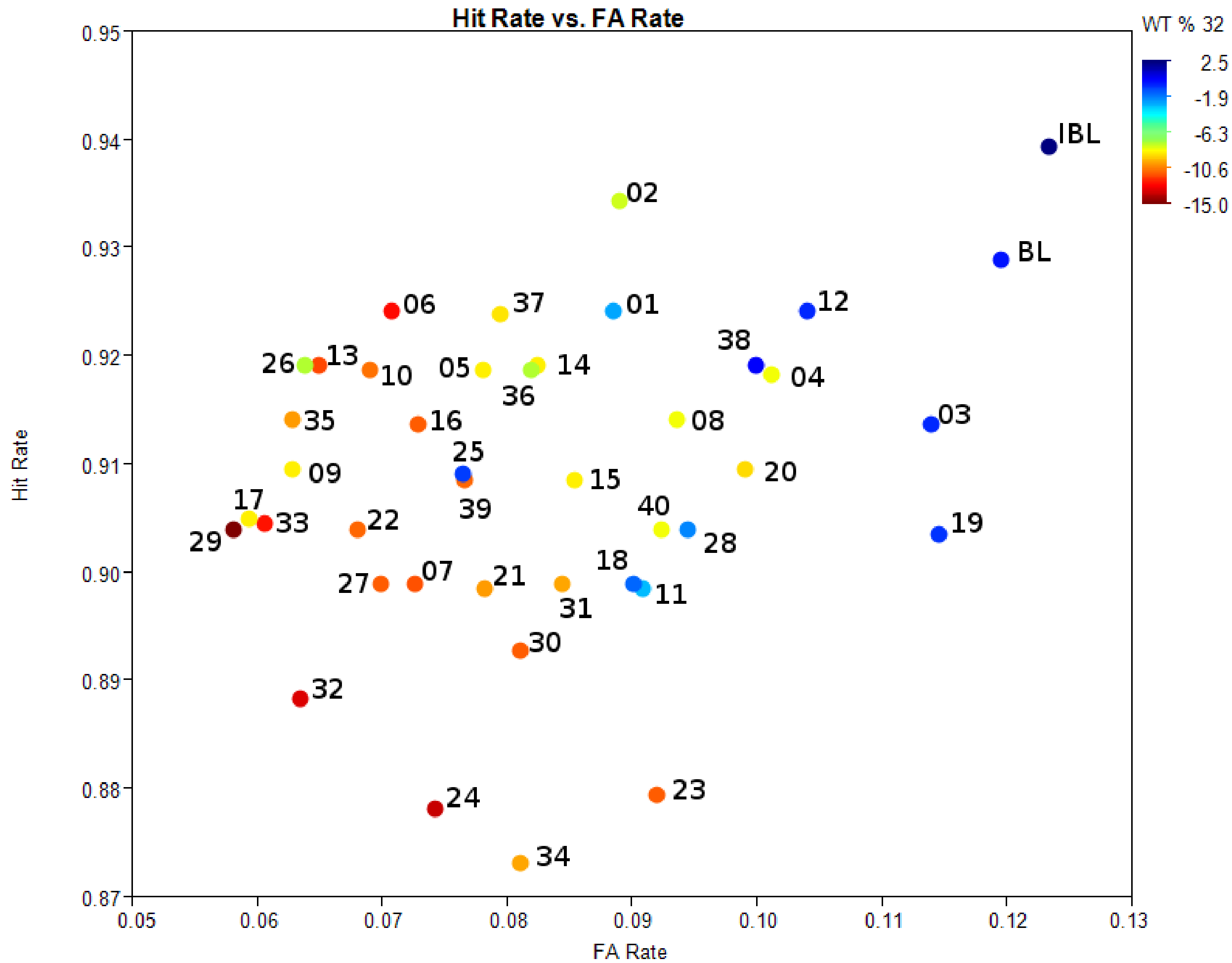

Figure 11. Hit Rate vs. False Alert Rate (Colored by Warning Time)

Unfortunately, the top-left of this plot is white space, indicating that no run stands out as the best performing run. Instead there are a few runs that perform well in some dimensions and not so well in others, or perhaps average in others. Run 2 stands out as one that performs above average for Hit Rate, and average for FA Rate and Warning Time. Run 26 stands out as one that performs above average for FA Rate and average for Hit Rate and Warning Time.

This plot is very different from the similar plot in Experiment 2. The results from Experiment 2 show a lot of runs at the top-right, a lot at the bottom-left, and many blue runs and red runs. This Experiment contains many more runs showing average performance, with most of the runs grouped in the middle of the plot and with many shades of yellow, green, and orange. There are even some dark blues mixed in with other colors, whereas in Experiment 2, there was a well-defined separation between the reds and the blues.

It is important to note that a color further from blue does not necessarily indicate a bad result. This only indicates that the warning time is less than the baseline's. As discussed earlier in this section, all runs performed very well in regards to warning time. Contrast the maximum percentage decrease of 15% in this experiment with the 46% maximum in Experiment 2. All of these runs would be shades of blue and green in Experiment 2.

3.2.1 Direct Comparisons

Several runs in the integrated experiment were strategically chosen to provide the capability of performing direct comparisons between the two runs with only a single prototype differing between them. Two sets of runs were generated for each prototype for the purpose of direct comparison.

3.2.1.1 Forced Trajectory Rebuild

There are eight sets of directly comparable runs for the FTR prototype. Of these eight runs, five of the runs showed some reduction in FA count. All but one of those five runs also showed an increase in LA count, and all runs showed an increase in MA count. Three runs showed an increase in FA count, though one was an insignificant increase and can be considered unchanged.

Table 12 depicts the set of directly comparable runs for FTR. For FA, LA, and MA, the counts are provided for when the prototype is off and when it is on. The "Diff" value provided for each metric is the following formula:

$$D = \frac{x_{on} - x_{off}}{x_{off}}$$

where x_{on} is the value of the metric (FA, LA, MA count) when the prototype is on, x_{off} is the value of the metric when the prototype is off, and D is the resulting "Diff" metric. For each of the metrics used, it is desirable to decrease the value with the prototype, so a negative value is most desirable.

The findings from the direct comparisons indicate that having FTR on with the other prototypes turned off seems to have a negative effect on the FA performance of the probe. The findings also indicate that FTR becomes less effective as the Lateral and Longitudinal conformance bounds decrease, which is to be expected due to the nature of the algorithm. The algorithm generates new trajectories when the object trajectory is older than some threshold, but smaller conformance bounds also cause new trajectories to be generated more often. Therefore, it is less likely that FTR will initiate a trajectory build when smaller conformance bounds are in effect, since the trajectories will already be generated more often.

Table 12. Results from Directly Comparable Runs for FTR

					FA			LA			MA		
GAB	CGS	Lat	Lon	Llh	Off	On	Diff	Off	On	Diff	Off	On	Diff
On	On	2.5	1.5	4|8|20	803	699	-0.130	12	14	0.17	5	8	0.60
On	Off	2.5	1.5	10|20	1145	1040	-0.092	10	12	0.20	5	7	0.40
On	Off	0.5	1.5	4|8|20	760	702	-0.076	10	9	-0.10	6	7	0.17
Off	On	0.5	1.5	4|8|20	779	749	-0.039	10	11	0.10	5	8	0.60
On	On	0.5	1.5	10|20	908	893	-0.017	10	12	0.20	6	9	0.50
Off	On	0.5	1	10|20	860	861	0.001	11	12	0.09	5	8	0.60
Off	Off	2.5	1	10|20	1255	1261	0.005	12	11	-0.08	5	8	0.60
Off	Off	0.5	1.5	10|20	974	992	0.018	9	11	0.22	6	9	0.50

3.2.1.2 Growth Adherence Bounds

Four sets of runs are available for direct comparison of the GAB prototype. All of the runs exhibit an improvement to FA performance. Two of the runs exhibit an improvement to LA count, whereas the two others show degradations. One run shows an improvement to MA performance, two show degradations, and one has no change.

Table 13 illustrates the set of directly comparable runs for GAB. For FA, LA, and MA, the counts are provided for when the prototype is off and when it is on. The "Diff" value provided for each metric is the following formula:

$$D = \frac{x_{on} - x_{off}}{x_{off}}$$

where x_{on} is the value of the metric (FA, LA, MA count) when the prototype is on, x_{off} is the value of the metric when the prototype is off, and *D* is the resulting "Diff" metric. For each of the metrics used, it is desirable to decrease the value with the prototype, so a negative value is most desirable.

The main observation that can be made from these comparisons is that GAB has a tendency to improve the FA performance, even if only by a small amount.

Table 13. Results from Directly Comparable Runs for GAB

					FA			LA			MA		
FTR	CGS	Lat	Lon	Llh	Off	On	Diff	Off	On	Diff	Off	On	Diff
Off	Off	2.5	1.0	10|20	1255	1100	-0.124	12	10	-0.17	5	6	-0.20
Off	Off	1.5	1.5	4|8|20	980	875	-0.107	8	11	0.28	5	4	-0.20
Off	On	2.5	1.0	10|20	1017	940	-0.076	14	12	-0.14	5	6	0.20
Off	Off	2.5	1.5	10|20	1317	1145	-0.131	10	10	0.00	4	5	0.25

3.2.1.3 Conflict Geometry Separation

There are five sets of runs for direct comparison of the CGS prototype. All of the runs improve the performance of the probe for FA count. All runs also degrade the performance of the probe for LA count significantly. Two runs improve the performance on MA count, but only by a single alert each.

Table 14 shows the set of directly comparable runs for CGS. For FA, LA, and MA, the counts are provided for when the prototype is off and when it is on. The "Diff" value provided for each metric is the following formula:

$$D = \frac{x_{on} - x_{off}}{x_{off}}$$

where x_{on} is the value of the metric (FA, LA, MA count) when the prototype is on, x_{off} is the value of the metric when the prototype is off, and *D* is the resulting "Diff" metric. For each of the metrics used, it is desirable to decrease the value with the prototype, so a negative value is most desirable.

The CGS prototype was modified from Experiment 2 such that in this experiment it will only delay the alert to a four minute threshold of warning time instead of the three minute threshold used in Experiment 2. This modification was expected to improve the Late Alert and Missed Alert performance, and though it has improved it over the previous experiment, it is still at unacceptable levels.

Table 14. Results from Directly Comparable Runs for CGS

					FA			LA			MA		
FTR	CGS	Lat	Lon	Llh	Off	On	Diff	Off	On	Diff	Off	On	Diff
Off	Off	2.5	1.0	10|20	1255	1017	-0.190	12	14	0.17	5	5	0.00
Off	Off	1.5	1.5	4|8|20	1100	940	-0.145	10	12	0.20	6	6	0.00
Off	On	2.5	1.0	10|20	980	843	-0.140	8	14	0.75	5	4	-0.20
Off	Off	2.5	1.5	10|20	1317	1031	-0.217	10	13	0.30	4	4	0.00
On	On	1	1.3	4|8|20	692	667	-0.036	12	13	0.08	5	6	0.20

After these findings were discussed, several of the encounters that differed between the directly comparable runs were analyzed to determine the cause of the increase in Late Alerts. It was hypothesized that the parallel algorithm used in CGS may be causing the additional Late Alerts. This parallel algorithm was removed from the prototype and five additional runs were made. Four of these runs have the CGS prototype turned on without the parallel algorithm. The last run was an additional directly comparable run.

Table 15 shows the results of the comparisons with the parallel algorithm turned off. The columns are similar to those in Table 14. The additional columns named "NP" are the counts for the run with CGS on with the parallel algorithm turned off. Also, the "Diff" columns use a similar equation as above, except the x_{on} variable in the equation is the value from the "NP" column.

The results with the parallel algorithm turned off show some improvement in the LA and MA performance over the original CGS prototype. All runs show no increase in MA count. One run shows no increase in LA count, one run increases by a single LA, one run increased by two, and one run increased by three. However, the improvement to FA performance has been decreased for all of these runs. The largest improvement is about a 2% decrease, and there is even one run in which the FA count increases.

Table 15. Results from Directly Comparable Runs for CGS with Parallel Algorithm Off

					FA				LA				MA			
FTR	GAB	Lat	Lon	Llh	Off	On	NP	Diff	Off	On	NP	Diff	Off	On	NP	Diff
On	On	1.0	1.25	4|8|20	692	667	700	0.012	12	13	15	0.25	5	6	5	0
Off	On	1.0	1.25	4|8|20	773	---	761	-0.016	10	---	11	0.10	5	---	5	0
Off	Off	1.5	1.50	4|8|20	980	843	960	-0.020	8	14	10	0.25	5	4	5	0
Off	Off	2.5	1.00	10|20	1255	1017	1229	-0.021	12	14	12	0.00	5	5	5	0

While the FA numbers do not seem impressive here, it needs to be noted that this prototype is a special-use algorithm that only comes into play in certain situations. It was determined that the CGS algorithm, with parallel off, only affected 160 encounters. When there are over 700 False Alerts in the run, affecting 160 of those may not show a large impact on the grand scale. Of those 160 encounter pairs affected by the CGS prototype, there were originally 101 False Alerts generated with no CGS prototype. There were only 86 False Alerts generated with the CGS prototype on. That's a much more impressive decrease of 14.8% in False Alerts. Though it affects only a small part of the grand scale of the scenario, that is due to the specialized purpose of this particular prototype, which is intended to decrease the alerts for in-trail encounters.

A 14.8% reduction to the affected region is a successful result, but it still leaves a large number of False Alerts that should also be removed. The prototype certainly shows much more promise now that the impact to LA and MA has been reduced, but it still requires modifications to improve its effect even further.

3.2.2 Model Analysis

The prediction profiler[1] is used to examine the results of the integrated model. Figure 12 depicts the results of the model when set to the baseline settings. Since the metrics used here are the percentage difference from the baseline, when the model is set to the baseline settings, all metrics are expected to be close to zero. FA and WT are very close to zero. LA is slightly off, with a result of -4.15%, but it is within the noise threshold of about 10% due to the small sample size for Late Alerts.

Since this experiment is a multi-dimensional problem (18 dimensions total) it cannot be fully described with a single two-dimensional figure. Figures 12 through 23 attempt to describe the shape of this 18-dimensional hypercube.

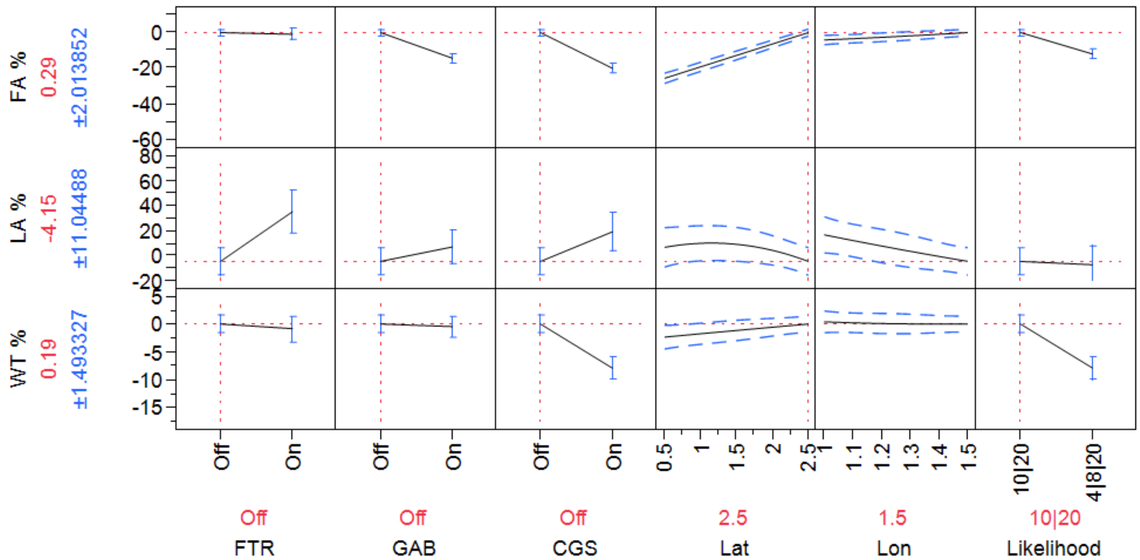

Figure 12. Model Results with Baseline Settings

Figure 13 shows the model results when all parameters are left at the baseline settings and only the FTR prototype is enabled. The major impact is in the Late Alerts which are increased by 35%. That is a statistically significant impact, and represents an increase of five Late Alerts. Interestingly, the False Alerts are not affected. It is clear that the FTR prototype alone does not provide a benefit, but it is possible that it contains interactions with other factors.

[1] The prediction profiler is an analysis tool provided by the JMP® software and is described in detail in the Experiment 1 document [Crowell et al, December 2011a].

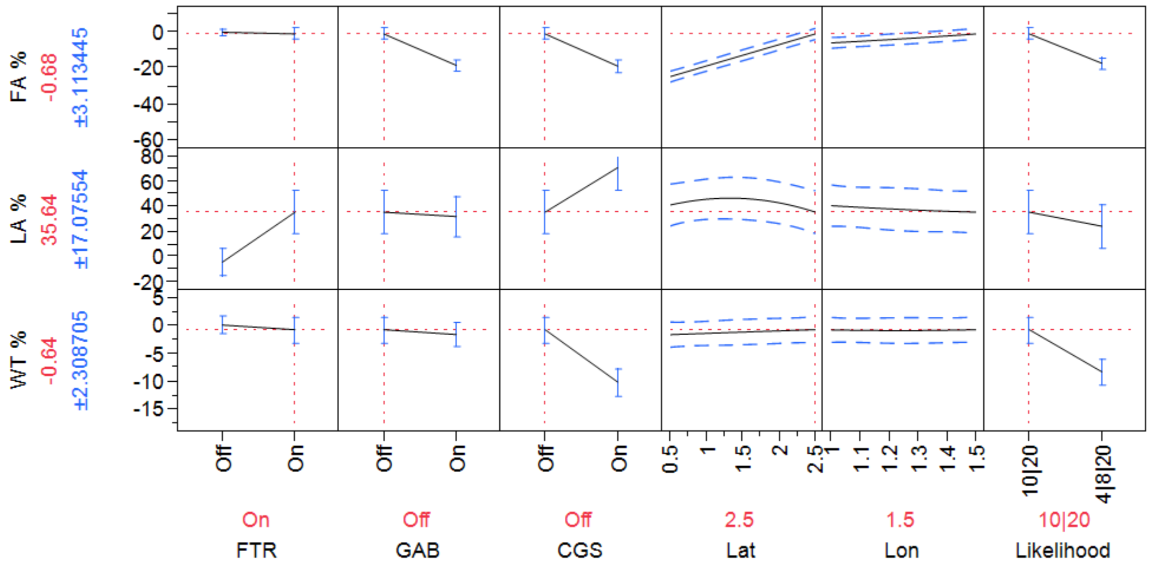

Figure 13. Model Results with FTR Prototype Enabled

In Figure 14 the baseline settings are shown once again, but this time with only the GAB prototype enabled. This shows much more promise than the FTR prototype, reducing the FAs by 14%. It also increases the LAs by 7%, but 7% is representative of less than a single alert and is not statistically significant.

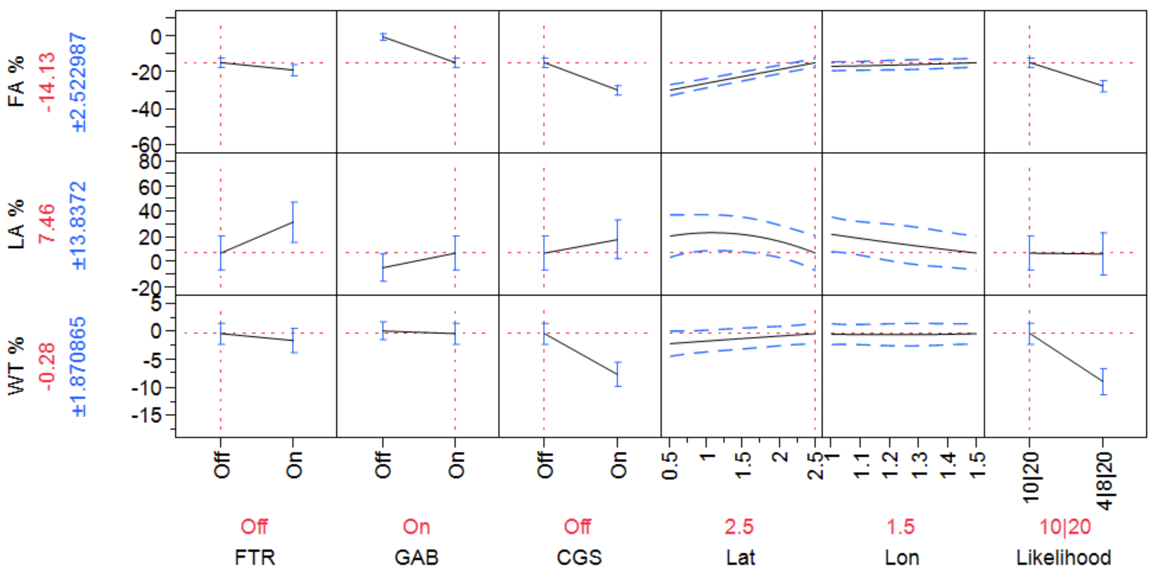

Figure 14. Model Results with GAB Prototype Enabled

The last prototype, CGS, is shown with the baseline settings in Figure 15. The CGS prototype used in the model still contains the parallel algorithm, so it is not surprising that it increases the LA by 19%, or about 3 alerts. It also reduces the False Alerts by 19%, which is also expected of the original CGS prototype from the analyses detailed in previous sections.

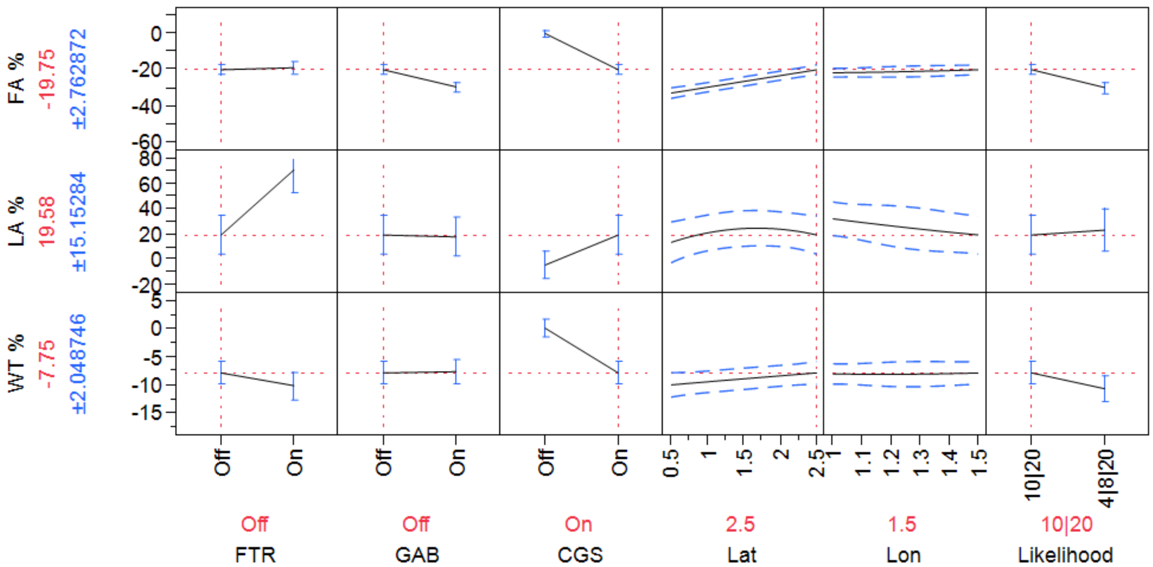

Figure 15. Model Results with CGS Prototype Enabled

Next, Figure 16 illustrates the case in which all of the prototypes are enabled and with the parameters set to the baseline settings. False Alerts are reduced by 31%, but the Late Alerts are increased by a very large 54% which represents over seven LAs. However, looking at the far left column of the graph, the effects of the FTR prototype can be seen on each of the metrics. The FA row indicates that the FTR prototype has no effect, whereas the LA row indicates that the prototype greatly increases the LAs.

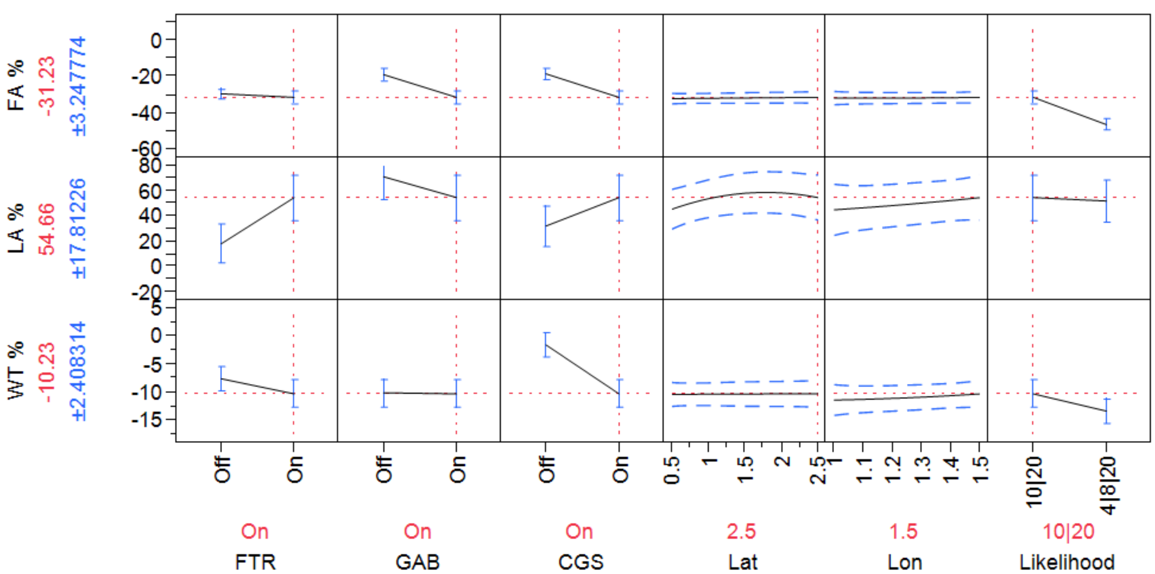

Figure 16. Model Results with All Prototypes Enabled

In Figure 17, the FTR prototype has now been disabled. The FA performance is only reduced slightly to a 29% reduction, whereas the LA performance is greatly improved over the previous settings, with an 18% increase to LAs. Though this is still a significant increase in LAs, it is improved over the previous run depicted in Figure 16. At this point it is apparent that the FTR prototype is not beneficial at the baseline parameter settings.

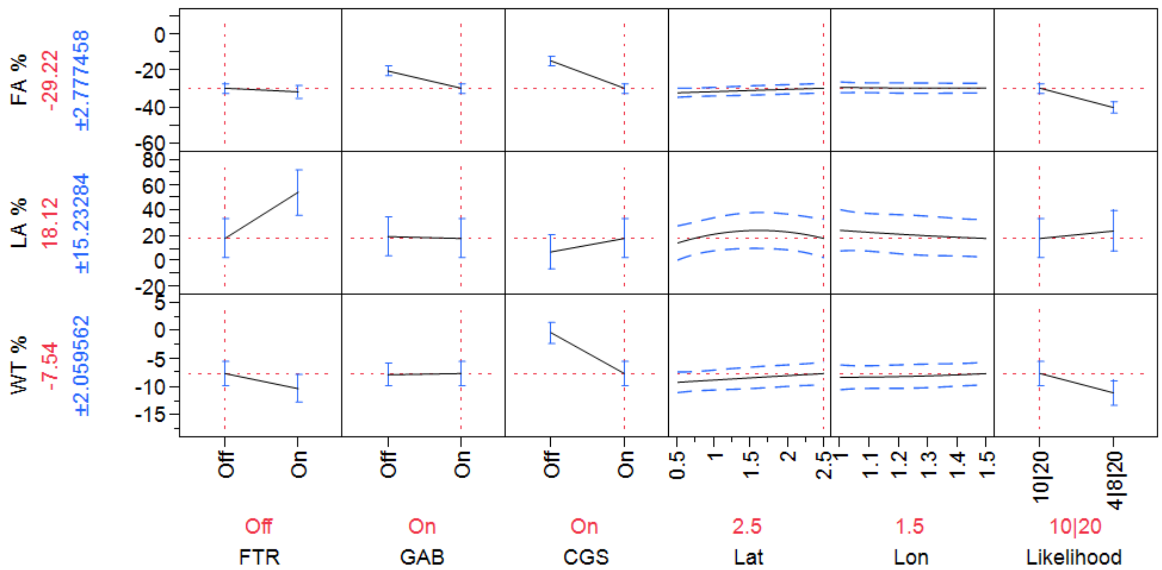

Figure 17. Model Results with GAB and CGS Prototypes Enabled

Next, the settings are optimized for each metric. Since there are three competing response metrics, it is not possible to optimize the entire model. Each response must be optimized individually, then a compromise must be made to determine desirable settings overall. In Figure 18, the model is optimized for False Alert performance. The FAs are reduced by a very significant 55%, but the LAs are also increased by 25%. The 25% increase is statistically significant and represents an increase of more than three Late Alerts.

The FTR and GAB prototypes are enabled, with a likelihood setting of 4|8|20, lateral adherence of 0.5 nm, and longitudinal of 1.0 nm. The effects of the FTR prototype at these settings are much more beneficial, showing a reduction in False Alerts and Late Alerts. So, it is now known that FTR does interact with the other parameter settings.

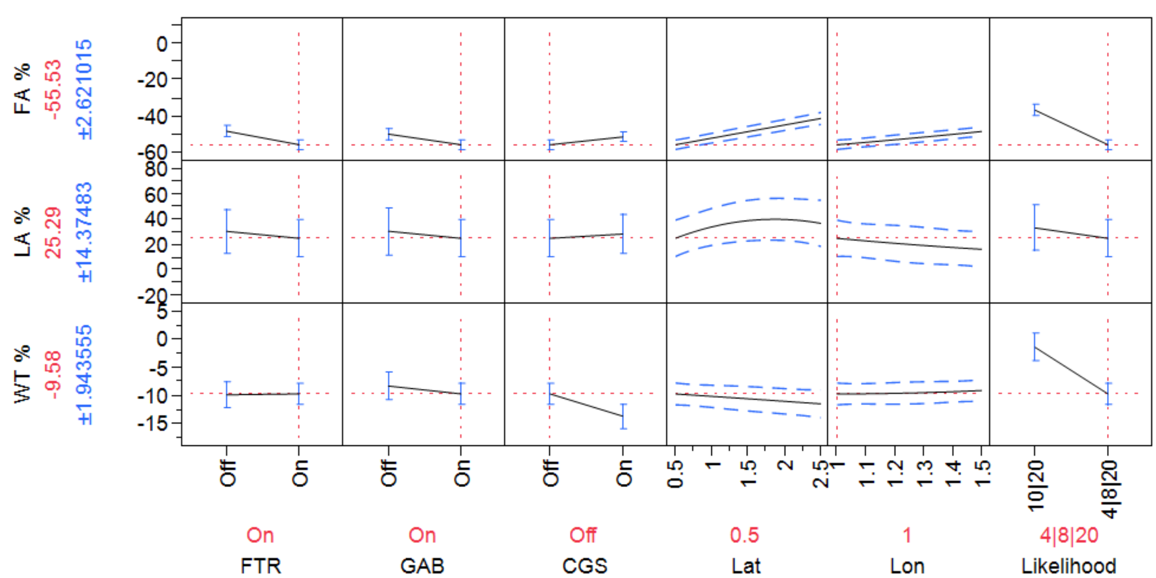

Figure 18. Optimal False Alert Performance

The settings are optimized for Late Alert performance in Figure 19. The major difference between these settings and those optimized for FA performance are the prototype enhancements. In this figure all prototypes are disabled. The longitudinal parameter is also changed from 1.0 to 1.5 nm. In this figure and in the previous figure, the longitudinal setting is seen as having an inverse effect on FA and LA performance, which makes finding the right setting difficult.

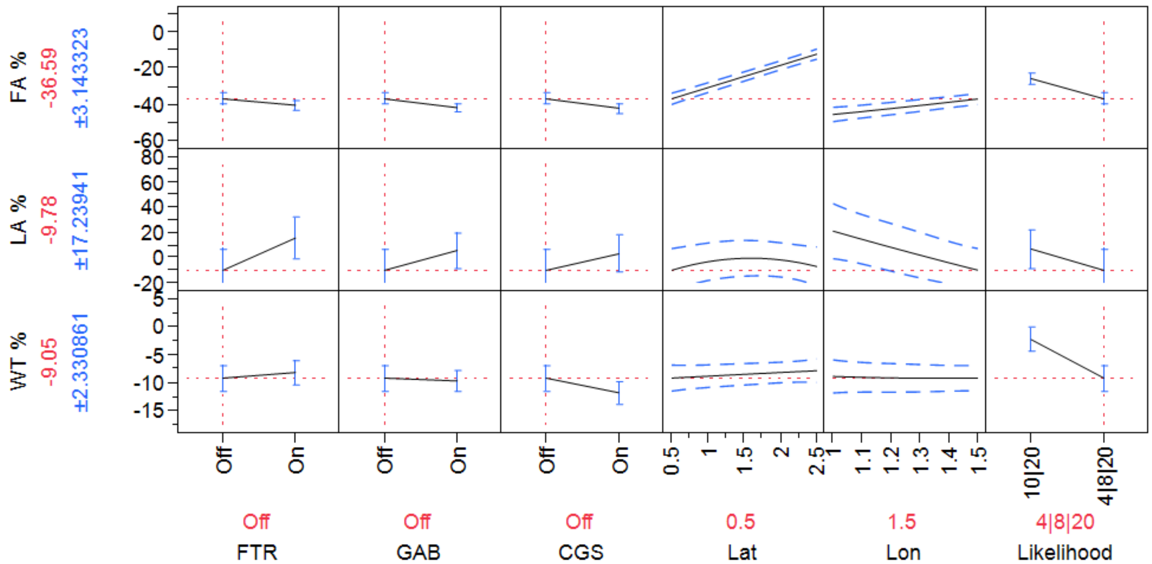

Figure 19. Optimal Late Alert Performance

An interesting finding in this figure is that when the settings are optimized to reduce Late Alerts, there is also a significant reduction in False Alerts. The Late Alerts are reduced by 9%, which represents a single Late Alert and is not statistically significant. However, the False Alerts are reduced by 36%, which is significant. Warning time is also reduced by 9%, which is statistically significant, but still provides a warning time of about 251 seconds.

Finally, Figure 20 shows the settings that optimize warning time performance. As discussed earlier in this document, there were no treatment runs that performed poorly in regards to warning time, so the model is also not expected to show any poor warning time performance. This figure is included to be consistent with previous experiments in which warning time was a much more important factor. In this experiment, because all of the treatment runs performed well in warning time, this metric will not be given much focus.

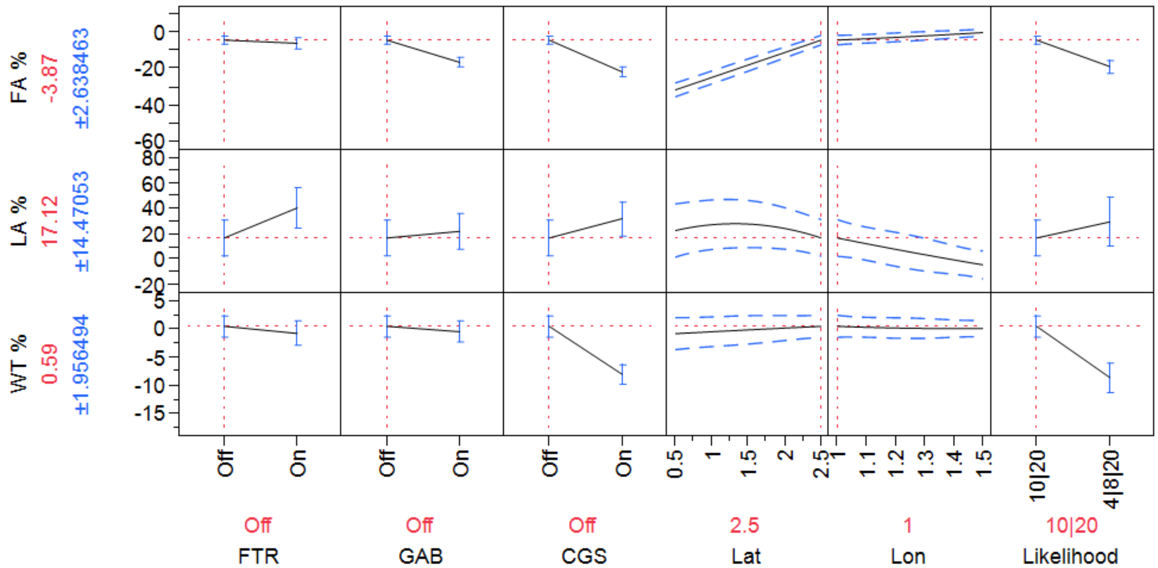

Figure 20. Optimal Warning Time Performance

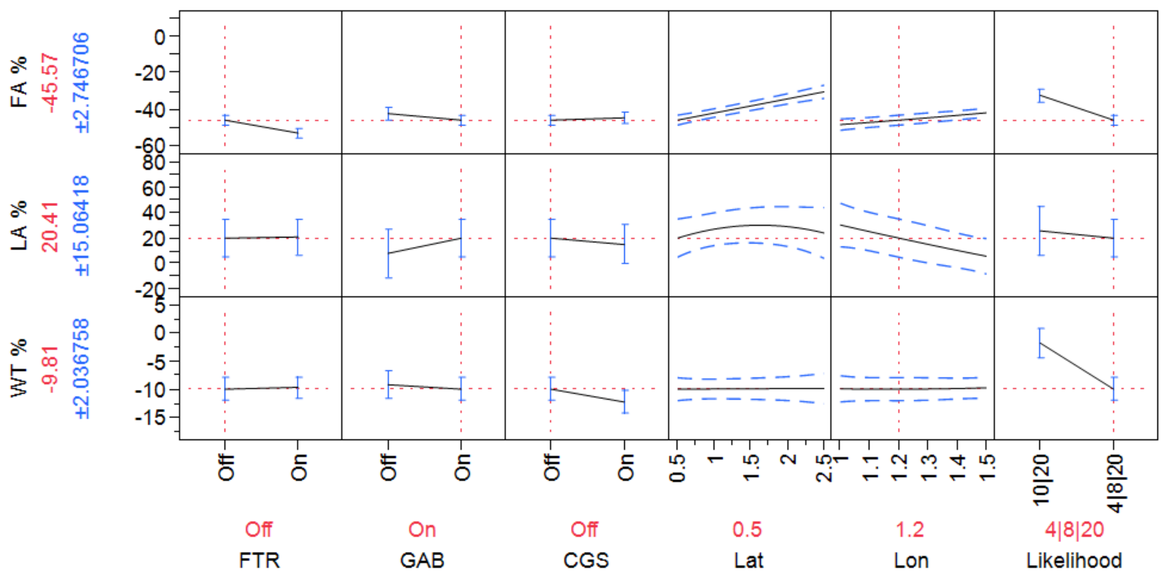

Figure 21. Values Closest to Experiment 2 Recommended Settings

The recommended settings from Experiment 2 [Crowell et al, December 2011b] were the GAB prototype on, lateral of 0.5 nm, longitudinal of 1.2 nm, and 3|8|10 likelihood. This experiment does not contain the 3|8|10 likelihood setting, so instead Figure 21 sets the parameters to the values most similar to the recommendations from Experiment 2 and uses 4|8|20 likelihood. In Experiment 2, these settings provided a 54% reduction to FAs, 2% increase to LAs, and 18% reduction in warning time. These settings in Experiment 3 provide less promising results, with a 46% reduction in FA count but a 20% increase in LA count.

Previously described in Figure 19 were the settings optimized for Late Alerts. The goal of this experiment is to reduce the False Alerts without having a negative impact on the Late Alerts. The settings for optimized Late Alert performance did accomplish these goals, but there is some flexibility in the Late Alert performance that may allow for a greater FA performance to be gained. In Figure 19, the Lateral adherence bound was set to 0.5 nm. From discussion with ATC SMEs, this is known to be too low of a setting given the current precision of radar. In Figure 22, the lateral adherence bound is raised to 1.0 nm. This, unfortunately, increases False Alerts and Late Alerts from the previous settings. False Alerts are now reduced 30% over the baseline, and Late Alerts are reduced about 3% from the baseline.

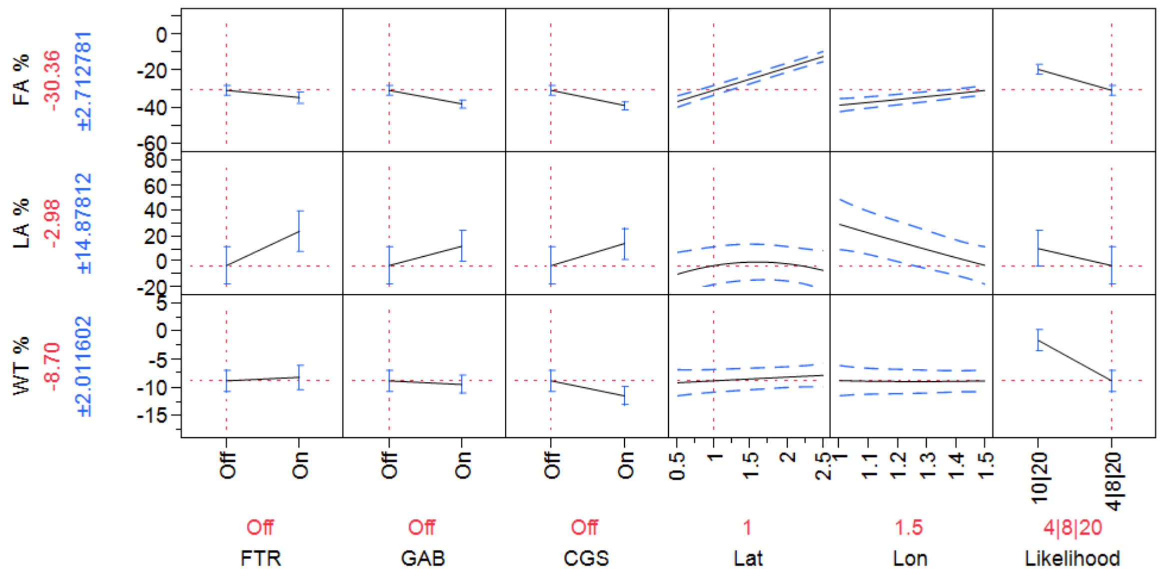

Figure 22. Lateral adherence bound raised to 1.0 nm to cope with radar precision levels

The current Late Alert performance provides a little more room for improving the False Alerts. Reducing the longitudinal adherence bound to 1.35 nm (Figure 23) brings the FA performance to a 33% reduction over the baseline. Late Alerts now increase by 6% over the baseline, but this number represents less than one LA and is not statistically significant. The longitudinal adherence bound is now closer to the previously recommended setting of 1.2 nm.

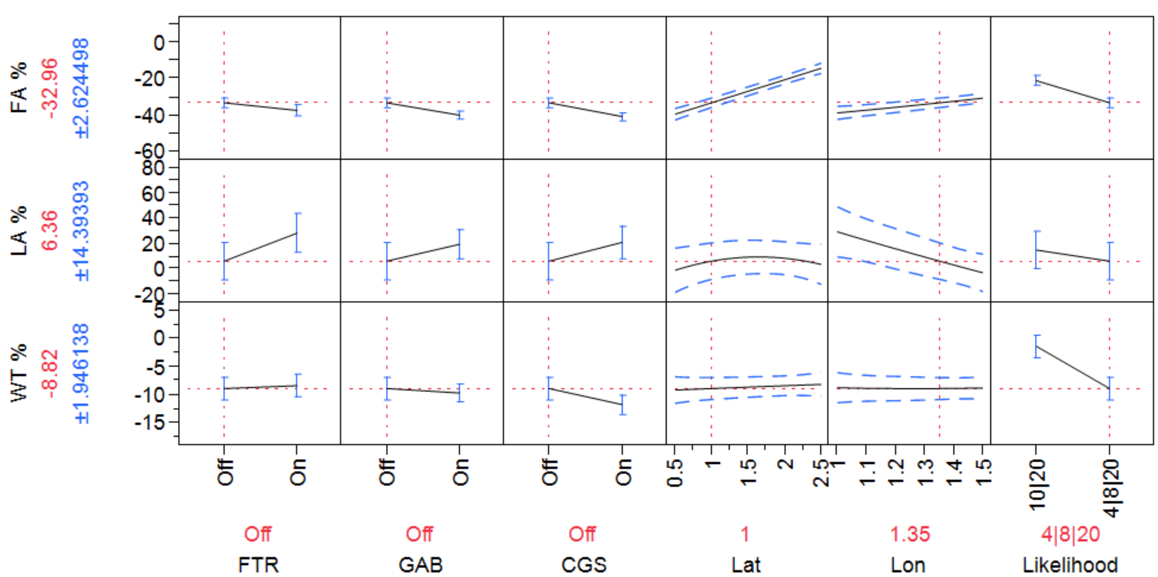

Figure 23. Setting for Good Performance in False Alerts and Late Alerts

3.2.3 Factor Effects

The analyses performed in the previous sections have helped to determine the effects of the three prototype enhancements and the three parameter settings. Though some work still needs to be done, there are some conclusions that can be made for this scenario in regards to the factors.

The Forced Trajectory Rebuild (FTR) prototype exhibited some improvement to the False Alert performance at certain settings, but always exhibited unacceptable Late Alert performance. The direct comparisons and the model analysis were consistent in these results, and there were no settings found at which FTR could help to provide an acceptable LA performance. The prototype can provide up to a 9% reduction in False Alerts, but can also cause up to a 4% increase in FAs. It may increase Late Alerts by up to 51%. It may also provide a reduction to Late Alerts by as much as 5%, but this phenomenon was only observed when the LA performance was already at unacceptable levels, and remained so with FTR on. At the baseline parameter settings, FTR generates a 2% increase to FAs and 38% increase to LAs.

The Growth Adherence Bound (GAB) prototype also showed an improvement to False Alert performance. At certain settings it did provide borderline acceptable results for LA performance. It can provide up to a 19% reduction in FAs and up to a 3% increase. For Late Alerts, it can provide up to a 13% reduction, and up to a 16% increase. At the baseline parameter settings, GAB generates a 12% reduction in FAs and a 9% increase in LAs.

The Conflict Geometry Separation (CGS) prototype was determined to once again require modifications in order to provide acceptable performance. For the version that contained the parallel algorithm and was analyzed in the model analysis, it showed some improvement to False Alerts, but this improvement corresponded with very large degradation in Late Alerts. It can provide up to a 19% reduction in FAs and up to a 2% increase. It can also provide up to a 14% reduction in Late Alerts, but also up to a 31% increase in LAs. At baseline settings, CGS generates its maximum of 19% reduction in FAs, but a 21% increase to LAs.

The version of CGS without the parallel algorithm was not represented in the model, and could only be analyzed with the direct comparisons. However, the direct comparison analysis indicated that CGS performed much better for LAs when the parallel algorithm was turned off, with a maximum of only a 25% increase to LAs as opposed to a 75% with parallel turned on. There was no increase observed to MAs as opposed to a 20% increase with parallel turned on. Though without parallel it showed less of an impact on False Alerts, this is because CGS is a specialized algorithm that only impacts a small number of encounters, as it is intended to.

As seen in both previous experiments, the lateral adherence setting has a major impact on both the FA and LA performance. Lowering the settings below the 2.5 nm baseline can have a major impact on performance. It can reduce the False Alerts by up to 30% at a setting of 0.5 nm and 23% at a setting of 1.0 nm. It can also reduce Late Alerts by up to 30% at a setting of 0.5 nm and 17% at a setting of 1.0 nm. With all other parameters left at the baseline settings, a lateral setting of 0.5 nm generates a 24% reduction in FAs and 8% increase in LAs. A lateral setting of 1.0 nm generates an 18% reduction in FAs and a 12% increase in LAs.

The longitudinal adherence setting has much less of an impact on FA performance, but a large impact on LA performance. This result is consistent with findings of previous work. It can produce up to an 11% reduction in FAs at a setting of 1.0 nm and a 5% reduction at a setting of 1.25 nm. In previous experiments, the effect of the longitudinal adherence bound on LA performance has a concave shape, with the bottom of the curve centered at around 1.25 nm. This experiment revealed a much more linear effect, with the most LAs appearing at a setting of 1.0 nm. At a longitudinal setting of 1.0 nm the LAs may be increased by as much as 34%, and at a setting of 1.25 nm by as much as 15%. At the baseline settings a lateral setting of 1.0 nm generates no change in FAs and an increase in LAs of 18%. A longitudinal setting of 1.25 nm also generates no change in FAs and generates an increase in LAs of 8%.

Likelihood once again shows a major impact on the FA performance of the probe and little impact to the LA performance. Changing threshold from 10|20 to 4|8|20 can reduce the FAs by up to 22% and never increases the FAs no matter what the other settings are. Changing the likelihood setting can reduce LAs by up to 26% but may also increase the LAs by up to 21%. With all other baseline settings, the 4|8|20 likelihood setting generates a 10% reduction in FAs and a 4% reduction to LAs.

4 Recommendations and Future Work

After Experiment 2, recommendations were made based on a single scenario of data. Now, recommendations can be made based on information gathered from multiple experiments using two different scenarios. Experiment 2 was performed on a scenario that contained traffic data from the Washington Center (ZDC) from 2005. The experiment documented in this report (Experiment 3) was performed on traffic data from the Chicago Center (ZAU) from 2010. In a future report due later in the year, a fourth experiment (Experiment 4) will document analysis performed on a traffic sample from ZDC in 2010.

For Experiment 3, the trajectory error analysis indicated that the FA32 trajectory modeling enhancements improved the trajectory accuracy in the dimension perpendicular to the direction of flight (cross track error). However, there was no consistent evidence of improved trajectory accuracy from FTR, GAB, or CGS. The FA32 enhancement is the only treatment factor that exhibited an improvement in trajectory accuracy that was statistically and practically significant.

In both Experiments 2 and 3, Forced Trajectory Rebuild (FTR) did not show a significant performance improvement to the Conflict Probe (CP). It does show up to a 9% improvement in False Alert performance, but at the sacrifice of a 51% degradation to Late Alert performance. It is being recommended at this time that FTR is not pursued any further without substantial review and changes to the algorithm. Experiment 4, planned for the end of calendar year 2012, will not include the FTR prototype.

Growth Adherence Bounds (GAB) showed a significant improvement to False Alert performance. It can provide up to a 19% reduction in FAs, and can also provide up to a 13% reduction in LAs. Although GAB can provide significant improvement, it can also increase the FAs by up to 3% and the LAs by up to 13%, depending on the settings of the other parameters. Care should be taken in using this prototype with other parameters set appropriately. Experiment 4 is required in order to make a final recommendation for the GAB prototype.

Conflict Geometry Separation (CGS) had to be altered from the version used in the previous experiment. In Experiment 2, the prototype delayed the alert down to a predicted warning time of three minutes, which turned out to be too much of a delay. In this experiment that was changed to four minutes. However, it was discovered that the prototype still generated an unacceptable level of additional Late Alerts, up to a 31% increase, with up to a 19% reduction in FAs. Upon further analysis, it was hypothesized that the additional Late Alerts were primarily due to the algorithm designed to delay alerts when two flights are flying close to parallel. This portion of the algorithm was disabled and several of the runs were recreated. The CGS prototype, without the parallel algorithm, exhibited a borderline significant improvement to the small number of specialized encounters it is designed to affect, increasing LAs by a total of 25%. There were 101 FAs that fit the criteria to be affected by the CGS algorithm, and 15% of those were removed by CGS. Given the marginal results, it is recommended to pursue the algorithm modifications made by MITRE, since these are expected to improve the performance even further.

As in Experiment 1, it is being recommended that all three parameters are changed from their current settings. The lateral conformance bound is currently set to 2.5 nm, which was determined to be much larger than required. A much more efficient setting would be about 1.0 nm, which can reduce the FAs by 23% and the LAs by 17%.

Due to the noise inherent in the current ground-radar tracking system, the longitudinal conformance bound should not be lowered much beyond its current setting of 1.5 nm. Based on the two experiments performed so far, a setting of 1.25 nm is being recommended. This setting should provide a 5% reduction in FAs with insignificant effect on LAs.

Finally, the current likelihood setting of 10|20 was once again determined to be an inefficient use of an otherwise powerful algorithm. This experiment was not designed to provide a recommended setting for likelihood. The only recommendation being made currently is to change the likelihood parameter to a setting that provides a more efficient use of the algorithm. An experiment planned for later in the calendar year 2012 will provide a recommended setting for the likelihood function.

In summary, the lateral, longitudinal, and likelihood parameters all provide significant improvements to the Conflict Probe, but the likelihood parameter already is the subject of a parallel analysis that is soon to be completed. The FTR prototype was found to be ineffective at this time and will not be pursued further. The GAB and CGS algorithms show improvement and are being recommended for implementation in the operational system, although CGS requires a small amount of additional work to improve it even further. The correct combination of all of these settings can provide a 32% reduction in False Alerts with no significant impact on Late Alerts or Warning Time. Overall, Experiment 3 was a successful study. Most of the findings of the previous experiment were confirmed. This experiment also helped to fill in some of the gaps of the previous experiment, but there is still some future work required on several of the factors studied.

5 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**ase**l**ine
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

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.

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.