technical note

Evaluation of Prototype Enhancements to the En Route Automation Modernization's Conflict Probe

Andrew Crowell, FAA ANG-C41

Andrew Fabian, FAA ANG-C41

Christina Young, Ph.D., CSSI Inc.

Ben Musialek, General Dynamics Information Technology

Mike Paglione, FAA ANG-C41

December 2011

DOT.FAA/TC-TN12/3

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

U.S. Department of Transportation

Federal Aviation Administration

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

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

NOTICE

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

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

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Technical Report Documentation Page

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

4. Title and Subtitle	5. Report Date
Evaluation of Prototype Enhancements to the En Route Automation Modernization's Conflict Probe	December 2011
	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-TN12/3
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 one of these NextGen initiatives. 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 the performance enhancements to the En Route Automation Modernization (ERAM) Trajectory Modeling (TM) and Conflict Probe (CP) sub-systems. These ERAM enhancements are required by the Separation Management and Modern Procedures Project. ATO-E contracted the prime contractor of ERAM, Lockheed Martin under FAA Task Orders 45 and 51, to develop these prototypes within the actual ERAM architecture so the FAA could evaluate their efficacy. This paper describes the second in a series of integrated experiments to study these enhancements. This experiment was performed using thirty treatment runs of the ERAM system that each had different combinations of prototypes enabled and parameter settings. These thirty runs are compared to the baseline run, which represents the current state of the live ERAM CP. This technical note provides a detailed description of the analyses performed and 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	84	

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 Separation Management and Modern Procedures Project is one of these NextGen initiatives. 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 the performance enhancements to the En Route Automation Modernization (ERAM) Trajectory Modeling (TM) and Conflict Probe (CP) sub-systems. These ERAM enhancements are required by the Separation Management and Modern Procedures Project. ATO-E contracted the prime contractor of ERAM, Lockheed Martin under FAA Task Orders 45 and 51, to develop these prototypes within the actual ERAM architecture so the FAA could evaluate their efficacy.

An experiment analysis was performed using thirty treatment runs of the ERAM system that each had different combinations of prototypes enabled and parameter settings. These thirty runs are compared to the baseline run, which represents the current state of the operational 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 2005 recording of the Washington, DC Center during peak hours. Since it is a single scenario source, all conclusions and recommendations are based solely on this scenario and should not be considered final recommendations. Final recommendations will be made once similar experiments have been run on multiple scenarios.

Three prototype software algorithms are evaluated in this study. The Growth Adherence Bound prototype alters the current standard adherence bounds to ones that are reduced at the near-time predicted position and get larger as the prediction advances in time. The Conflict Geometric Separation prototype evaluates the geometry of a predicted conflict and delays notification if specified criteria are met such as being in-trail. The Forced Trajectory Build prototype will trigger a trajectory update for aircraft predicted in conflict. Besides the prototypes, several existing CP parameters were altered in the same experiment. The lateral and longitudinal conformance bounds were evaluated at operational settings to several times lower. Different time and probability settings were evaluated with the existing likelihood algorithm. It delays notification of only those conflicts that it predicts to have a high probability of occurring. The combination of all five of these factors contributed to the thirty treatment runs performed.

Based on the scenario used in this study, the Growth Adherence Bound prototype seems the most promising, since it can decrease the False Alerts (FA) without significantly increasing Late Alerts (LA). Turning on this prototype with all other settings unchanged can provide a 9% FA performance increase, with no significant performance degradation to Late Alerts or warning time. However, a much larger FA performance increase can be gained by changing the likelihood setting, slightly reducing the longitudinal conformance bound to around 1.25 nm and greatly reducing the lateral conformance bound to 1.0 nm or lower. These parameters can be set in such a way that the LA performance is not significantly degraded, but warning time performance is adversely affected. Though the warning time is reduced by these parameter changes, it is still significantly greater than the 180 second warning time requirement that was used in this study, which is reflected by the lack of an increase in Late Alerts.

It was discovered that the Conflict Geometric Separation (CGS) prototype was run at inappropriate settings that would greatly increase the Late Alerts. CGS was found to have the

capability of reducing False Alerts by as much as 18%, but with the settings used, only at the expense of an increase to Late Alerts. From some exploration of this prototype it was found that changing these settings to appropriate values will likely allow the CGS prototype to be a valuable addition to the Conflict Probe, but in order to verify this, another study must be performed on the CGS prototype. There were 15 treatment runs with the CGS prototype turned on. Each of these should be re-run with the appropriate settings so the experiment can be repeated. Based solely on the experiment performed for this document a recommendation cannot be made for CGS.

The FTR prototype was found to provide a minor improvement to FA performance, but at some settings sacrifices LA performance for the FA performance gain. There is some interaction of the FTR prototype with the likelihood setting. At this time it is not recommended to use the FTR prototype, but it cannot yet be ruled out as a viable addition to the probe. Once the experiments have been run on multiple scenarios, a final recommendation can be made.

Table of Contents

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

TABLE OF CONTENTS ..... **IX**

LIST OF FIGURES ..... **X**

LIST OF TABLES ..... **XII**

1 INTRODUCTION **1**

1.1 BACKGROUND TO STUDY 1
1.1.1 Prototype Enhancements *4*
1.1.1.1 Forced Trajectory Rebuild 4
1.1.1.2 Growth Adherence Bounds 4
1.1.1.3 Conflict Geometric Separation 5
1.1.2 Conflict Probe Parameters *5*
1.1.2.1 Conformance Bounds 5
1.1.2.2 Likelihood 7
1.1.3 Previous Work *8*
1.2 SCOPE OF STUDY 8
1.3 DOCUMENT ORGANIZATION 10

2 DESCRIPTION OF EXPERIMENT **11**

2.1 DEFINITION OF THE PROBLEM STATEMENT 12
2.2 DESIGN OF EXPERIMENT 12
2.2.1.1 Factors 12
2.2.1.2 Model 15

3 PERFORMANCE EVALUATION **17**

3.1 TRAJECTORY MODELING ANALYSIS 17
3.2 CONFLICT PROBE ANALYSIS 18
3.2.1 Fuzzy Analysis Results *24*
3.2.2 Direct Comparisons *27*
3.2.2.1 Forced Trajectory Rebuild 29
3.2.2.2 Growth Adherence Bounds 29
3.2.2.3 Conflict Geometric Separation 30
3.2.3 Model Analysis *31*
3.2.3.1 Implementation of Statistical Model 32
3.2.3.2 Examine Model Results 48
3.2.3.3 Fuzzy Model Analysis 57
3.2.4 Factor Effects *63*
3.3 EXAMPLE FLIGHTS 63
3.3.1 Forced Trajectory Rebuild Example *64*
3.3.2 Growth Adherence Bounds Example *65*
3.3.3 Conflict Geometric Separation Example *66*

4 RECOMMENDATIONS AND FUTURE WORK **69**

LIST OF ACRONYMS AND ABBREVIATIONS **70**

REFERENCES **71**

APPENDIX A **72**

List of Figures

Figure 1. Illustrative Example of a Conflict 2
Figure 2. Conformance bounds used for re-adherence of the trajectory to the flight path. 6
Figure 3. A simplified depiction of the conformance bounds being used for the conflict probe. ... 7
Figure 4. Piecewise linear likelihood function. 8
Figure 5. Linear likelihood function for 10/20 setting. 13
Figure 6. Piecewise linear likelihood function for 3/8/20 setting. 14
Figure 7. Prediction Profiler for Average Absolute Cross Track Error 17
Figure 8. Hit Rate vs. False Alert Rate colored by Warning Time. The top-left corner of the graph and a blue color are the most desirable 24
Figure 9. Fuzzy Hit Rate vs. Fuzzy False Alert Rate 27
Figure 10. Traditional FA % and LA % metrics and WT % metric leverage plots. 32
Figure 11. Traditional FA % and LA % metrics and WT % metric leverage plots with outlier removed. 33
Figure 12. Traditional FA % and LA % metrics and WT % metric in the refined model. 37
Figure 13. Residual Error Distributions for the Response Variables from the Refined Model. 38
Figure 14. Fuzzy FA % and LA % metrics and WT % metric leverage plots. 40
Figure 15. Residual Error Distributions for the Fuzzy Response Variables from the Model. 43
Figure 16. Fuzzy FA % and LA % metrics and WT % metric refined leverage plots. 45
Figure 17. Residual Error Distributions for the Fuzzy Response Variables from the Refined Model 47
Figure 18. Prediction profiler of the model set to the settings of the baseline run. 48
Figure 19. Prediction profiler set to the baseline settings for parameters but with all three prototype enhancements turned on. 49
Figure 20. Prediction profiler set to the optimal settings to reduce Late Alerts according to Experiments 1 and 2. 49
Figure 21. Prediction profiler set to the optimal settings to reduce False Alerts according to Experiment 1. 50
Figure 22. Prediction profiler set to the optimal settings to reduce False Alerts according to Experiment 2. 50
Figure 23. Prediction profiler set to one possible combination of the optimal settings. 51
Figure 24. Prediction profiler with settings to greatly improve FA performance and only slightly adversely affect LA performance. 52
Figure 25. Interaction profile for FA % metric. 53
Figure 26. Interaction profile for LA % metric. 54
Figure 27. Interaction profile for WT % metric. 55
Figure 28. Baseline settings with only GAB prototype turned on. 56
Figure 29. Settings that provide a good performance to FA with GAB turned on. 57
Figure 30. Fuzzy metric values of the model at baseline settings. 57
Figure 31. Fuzzy values with all prototypes turned on at the baseline settings. 58
Figure 32. Fuzzy values with optimized performance for only Late Alerts. 59
Figure 33. Fuzzy values with optimized performance for False Alerts only. 59
Figure 34. Fuzzy values with mostly optimal settings for FA performance, but FTR turned off to improve LA performance. 60
Figure 35. Fuzzy values with good FA performance and FTR off and longitudinal set to 1.25 nm to improve LA performance. 60
Figure 36. Fuzzy values with only the lateral adherence bound adjusted from the previous settings to achieve the best Late Alert performance. 61

Figure 37. Settings of the probe that most reduce the False Alerts without degrading the Late Alerts, according to Fuzzy metrics. 62
Figure 38. Traditional metrics with settings of the probe that most reduce False Alerts without degrading Late Alerts according to Fuzzy Metrics. 62
Figure 39. FTR not used in conflict prediction. 64
Figure 40. FTR use increases prediction accuracy. 65
Figure 41. Baseline conflict detection (2.5 by 1.5 conformance bounds) of second aircraft pair. 66
Figure 42. Conflict detection without CGS. 67
Figure 43. Same scenario with CGS, where even though the boxes overlap, an alert does not (and will not) generate due to finding this to be a false alert while it is still waiting to present the prediction. 68

List of Tables

Table 1. Processing Steps for the Experimental Analysis 11
Table 2. Continuous factors of the Integrated Experiment. 12
Table 3. Nominal factors of the Integrated Experiment. 13
Table 4. Thirty runs for the integrated experiment and the baseline run. 15
Table 5. Matched Pair Test Results. 18
Table 6. Alert type counts of the thirty treatment runs and the two baselines. 20
Table 7. Adjusted Late Alert value of the treatment runs compared to the LA+MA counts. 21
Table 8. Overall Alert Rates and False Alert Rates by separation category. 22
Table 9. Median, inter-quartile range, and 25th percentile of conflict warning time. 23
Table 10. Fuzzy Alert values for each of the treatment runs and the baselines. 25
Table 11. Overall Fuzzy Alert Rates and Fuzzy False Alert Rates by separation category. 26
Table 12. FA, LA, and WT of Experiments 1 and 2 treatment runs and FA32 Baseline 28
Table 13. Performance comparisons for runs with all common settings except for FTR. 29
Table 14. Performance comparisons for runs with all common settings except for GAB. 30
Table 15. Performance comparisons for runs with all common settings except for CGS. 31
Table 16. The results of re-running Run E2-25 with notification delay set to four minutes. 31
Table 17. Performance comparisons for the CGS run that was re-run with four minutes notification delay and the original run with three minute delay. 31
Table 18. Initial Model Effect Test for the FA % Response Variable. 34
Table 19. Initial Model Effect Test for the LA % Response Variable. 35
Table 20. Initial Model Effect Test for the WT % Response Variable. 36
Table 21. Summary of Refined Model Effect Tests for Response Variables. 39
Table 22. Initial Model Effect Test for the Fuzzy FA % Response Variable. 41
Table 23. Initial Model Effect Test for the Fuzzy LA % Response Variable. 42
Table 24. Summary of Fuzzy Model Effect Tests for Fuzzy Response Variables. 44
Table 25. Summary of Refined Fuzzy Model Effect Tests for Fuzzy Response Variables. 46
Table 26. Comparison results of each of the 30 treatment runs to the FA32 Baseline run. 72

1 Introduction

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 one of these NextGen initiatives. 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 the performance enhancements to the En Route Automation Modernization (ERAM) Trajectory Modeling (TM) and Conflict Probe (CP) sub-systems. These ERAM enhancements are required by the Separation Management and Modern Procedures Project. ATO-E contracted Lockheed Martin under FAA Task Orders 45 and 51, to develop these prototypes within the actual ERAM architecture so the FAA could evaluate their efficacy.

This technical note describes a study performed on thirty experimental runs of ERAM. Each of the runs was performed with a combination of parameter adjustments and prototype enhancements strategically chosen for an integrated experiment. This experiment, designated Function Area 18 (FA18) Experiment 2, expands upon Experiment 1 detailed in [Crowell, 2011, B] that included only parameter adjustments with no prototype enhancements.

1.1 Background to Study

The FAA created the NAS to provide a safe and efficient airspace environment for the air transportation system in the United States. The NAS is composed of a network of air navigation facilities, air traffic control facilities, and airports, along with the technologies and the rules and regulations to operate the system. As the air transportation system in the United States has grown, the NAS has evolved by incorporating new procedures and new technologies. The projected increases in demand could lead to a greater stress and perhaps to decreased quality of service for NAS users. In response to this the United States Congress created the multi-agency Joint Planning and Development Office (JPDO) in 2003. The mission of the JPDO is to design and deploy an air transportation system meeting the nation's anticipated aviation and air traffic needs in 2025. Since its creation the JPDO has published an integrated plan [Federal Aviation Administration, 2004] and documented a concept of operations [Federal Aviation Administration, 2007] that establish a vision for the Next Generation Air Transportation System (NextGen). An integral part of this vision is Trajectory Based Operations (TBO), which represents a paradigm shift from clearance-based air traffic control to trajectory-based air traffic control. With TBO it is envisioned aircraft will fly negotiated trajectories and the air traffic control functions will move to trajectory management.

The FAA is the single Air Navigation Service Provider (ANSP) that operates the key components of the NAS, such as Air Route Traffic Control Centers (ARTCCs) and Terminal Radar Approach Control (TRACON) centers. A key function provided by an ANSP is to ensure the safe separation of aircraft within the air transportation system. Aircraft-to-aircraft separation is managed by air traffic controllers who make strategic and tactical decisions using radar displays to visualize aircraft positions and flight paths. Although more recent automated decision support tools (DSTs) have been made available to controllers that predict potential loss of separation (i.e., conflicts) and aid in their evaluation, the effectiveness of the DSTs is limited by airspace complexity, controller workload, and the use of voice communications.

NextGen envisions trajectory-based separation management that will provide precise management of the current and future positions of all controlled aircraft in the air transportation

system. This will require enhanced DSTs that not only predict future conflicts, but also provide conflict resolution that is communicated directly to the aircraft from the ANSP through digital Data Comm*. This planned separation management capability will be able to handle the anticipated increase in traffic demand and aircraft diversity with minimal impact to user-desired performance profiles and to the environment, while retaining the existing strict safety standards.

A specific separation management problem is an aircraft-to-aircraft conflict, which can be defined as a situation where two or more aircraft violate minimum separation criteria. The separation criteria depends on the aircraft's navigational equipment and the airspace (e.g., En Route, Terminal, or Oceanic) in which they are flying. A Conflict Probe (CP) is a DST that predicts when conflict situations may occur by continuously comparing projected aircraft positions over a user-specified look-ahead time.

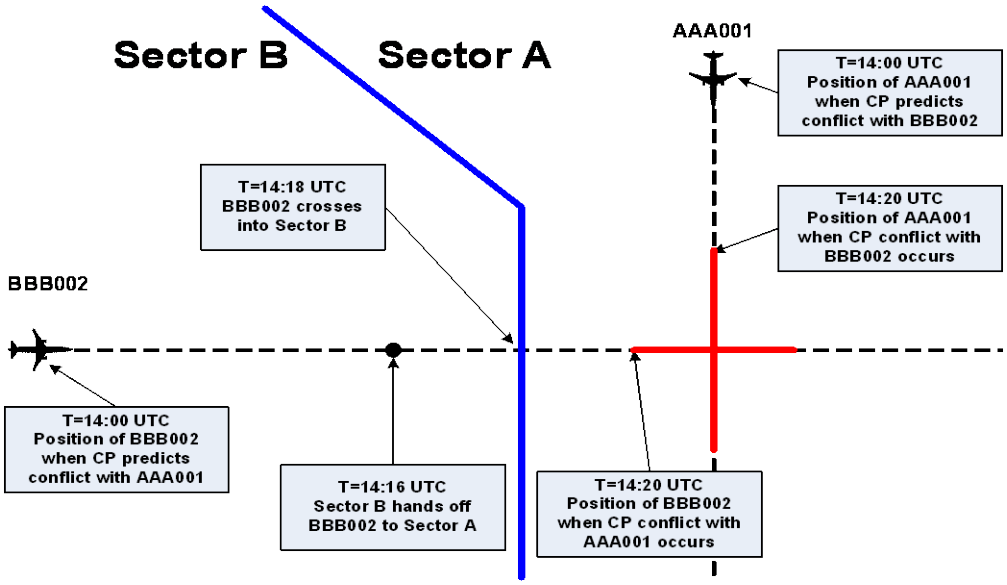

Figure 1. Illustrative Example of a Conflict

Figure 1 presents a horizontal view of an example conflict situation in which an aircraft, identified as AAA001, is flying southbound under the control of the Sector A controller. Another aircraft, identified as BBB002, is flying eastbound at the same altitude under the control of the Sector B controller. In this figure the airplane icons represent the positions of these aircraft at 14:00 UTC†, when a CP might predict a conflict in Sector A where their paths cross. The two aircraft are predicted to violate their separation criteria at 14:20 UTC, indicating a look-ahead time of 20 minutes. The situation is further complicated by the proximity of the aircraft to the sector boundary, in that the anticipated hand-off between the air traffic controllers would occur at about 14:16 UTC. This figure shows a typical separation management problem in which the flight path of the affected aircraft must be changed by heading, altitude, or speed changes.

It must be emphasized, that this figure is presented for illustrative purposes only in order to provide a touchstone for describing how conflicts are resolved. The separation of aircraft is complex because each situation presents unique circumstances, where numerous variables such as other air traffic, special use airspace that must be avoided, proximity of one of the aircraft to its

* Data Comm refers to the program and technology required to digitally communicate air traffic and weather information between the individual aircraft and the responsible ground air traffic control automation.

† UTC refers to Coordinated Universal Time, which within this study is equivalent to Greenwich Mean Time

destination, weather, and controller workload influence the action selected by the air traffic controllers. For this reason a number of assumptions are made to simplify this example. First, it is assumed both AAA001 and BBB002 are mid-flight and at their cruise altitude and AAA001 is flying at a slower rate of speed than BBB002. Secondly, it is assumed there are no thunderstorms or other abnormal weather conditions nor any other aircraft in the area.

Lastly, it is assumed the CP presents the potential conflict to the Sector A controller. Given these assumptions, this is a simple conflict with few complicating factors that can be used in this Study Plan to illustrate approaches to conflict resolution. With this in mind, here is a comparison of the current and envisioned future ways that this conflict might be resolved:

- **Current Resolution of Conflicts.** In today's NAS there are a number of alternative ways to solve this illustrative conflict; none of which would be considered the "right" way to resolve the conflict. For example, since the Sector A controller is notified of the potential conflict by the CP he/she would typically call the Sector B controller and request an early hand-off of aircraft BBB002. Once the hand-off has occurred and BBB002 has established voice communications with the Sector A controller, the Sector A controller would notify the pilot of the heading change and the conflict would be avoided. Alternatively, the Sector A controller could contact the Sector B air traffic controller and have him/her initiate a heading change to BBB002 to resolve the conflict in a timelier manner. On the other hand, if the Sector A controller's workload was heavy, he/she might request AAA001 to change their altitude, thus avoiding the time consuming coordination with Sector B. With each of these alternatives there is much time lost due to coordination between sector controllers and voice communications with the affected pilots.
- **Future Resolution of Conflicts.** Since NextGen is currently in the process of defining the alternatives that will exist in the future NAS, it cannot be stated with certainty how this example conflict will be handled in the future. But it is anticipated the function of the CP will be enhanced in several fundamental ways. First, it will not only exist at the associate controllers display like today but will be integrated within the radar controller's display as well. The data block will contain alert data and additional pullout menus will contain time ordered listings of these notifications. Next, these enhanced alerts will automatically take into account user preferences and aircraft capabilities to generate a rank-ordered set of resolutions that both resolve the conflict and meet metering constraints. The enhanced CP will then notify the Sector A controller of the predicted conflict and present a menu of recommended resolutions. The Sector A controller will select the best resolution to the DST. If the best resolution involved maneuvering AAA001, automation would send the clearance directly to AAA001 via Data Comm. However, if the best resolution involved maneuvering BBB002, the Sector B controller would be notified of the conflict and recommended resolution, and if the Sector B controller concurs, automation would send the clearance directly to BBB002 via Data Comm.

The current CP used in the NAS is very effective at detecting potential aircraft-to-aircraft conflicts that are of real concern and providing controllers with enough time to properly separate the aircraft. However, the current CP of ERAM is not effective at filtering out those events that will not become conflicts without any controller interaction. These additional alerts that are not of interest to the controller are called nuisance alerts, because they detract from the events that are of real concern. Before this CP can be used effectively by the radar controller or even by conflict resolution automation, the nuisance alert rate must be reduced to a level that is acceptable by Air Traffic Control (ATC) subject matter experts (SME). Likewise, in the process of reducing the nuisance alert rate, the rate of detecting the events that are of a real concern should not be adversely affected.

Thus, the motivation of this work is in support of the NextGen project titled, *Separation Management and Modern Procedures*. This particular project is charged with enhancing and then implementing a strategic conflict probe into the radar controller's display, required for the future NAS envisioned by NextGen. The prototype performance enhancements are required to present accurate and timely conflict predictions to the radar controller and support other NextGen advanced functions as they are implemented later.

1.1.1 Prototype Enhancements

For this study, three new algorithms, developed by Lockheed Martin (LM) engineers, to improve the performance of the conflict probe, will be evaluated to assess their impact predicting conflicts and reducing nuisance and false alerts. The new algorithms to be 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, 2011, B].

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 working from 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 through the route. The algorithm will treat near-in time segments with a smaller conflict detection conformance bounds than 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 are the same size as the standard bounds.

The reasoning for 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 timely rebuilding of trajectories. GAB is most effective

when the flights are separating 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, Conflict Geometry Separation (CGS), will be executed that examines the conflict geometry to determine if 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 going 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.

1.1.2 Conflict Probe Parameters

The last three settings varied in the treatment runs of this experiment, as well as in Experiment 1, are parameters of the ERAM Conflict Probe. These parameters can each 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 2. 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.

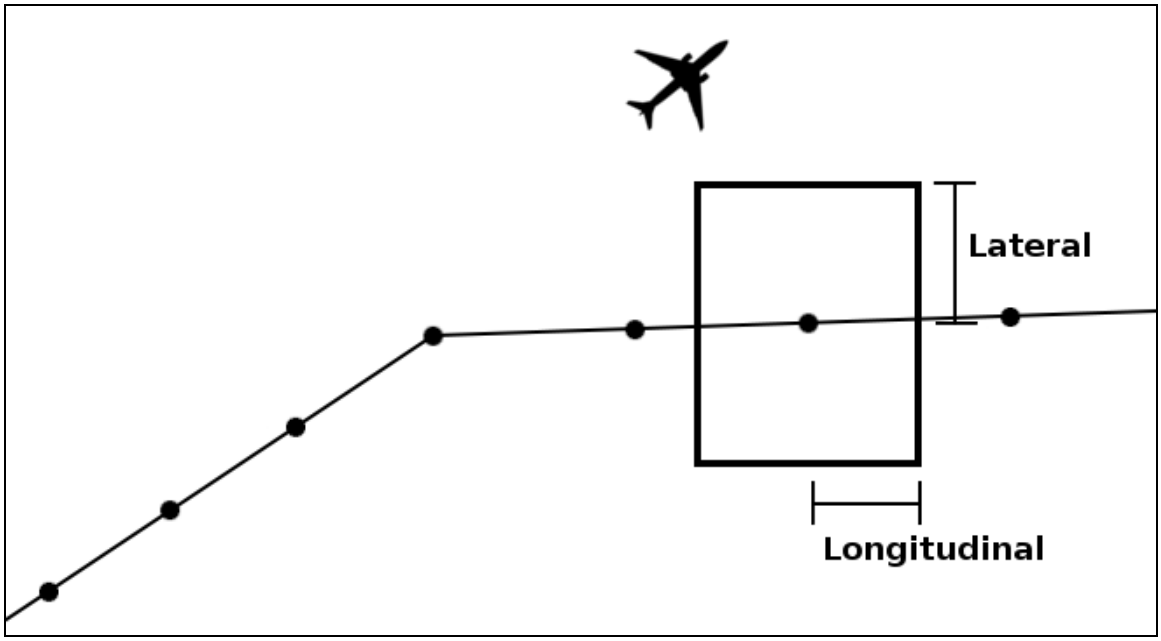

Figure 2. Conformance bounds used for re-adherence of the trajectory to the flight path.

The use of conformance bounds for the Conflict Probe is much more complex. Several levels of filters are used first. 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 as shown in Figure 3. One of the boxes is given rounded corners while the other box is given square corners. If these two boxes intersect each other and it passes all filters, then a conflict prediction is made.

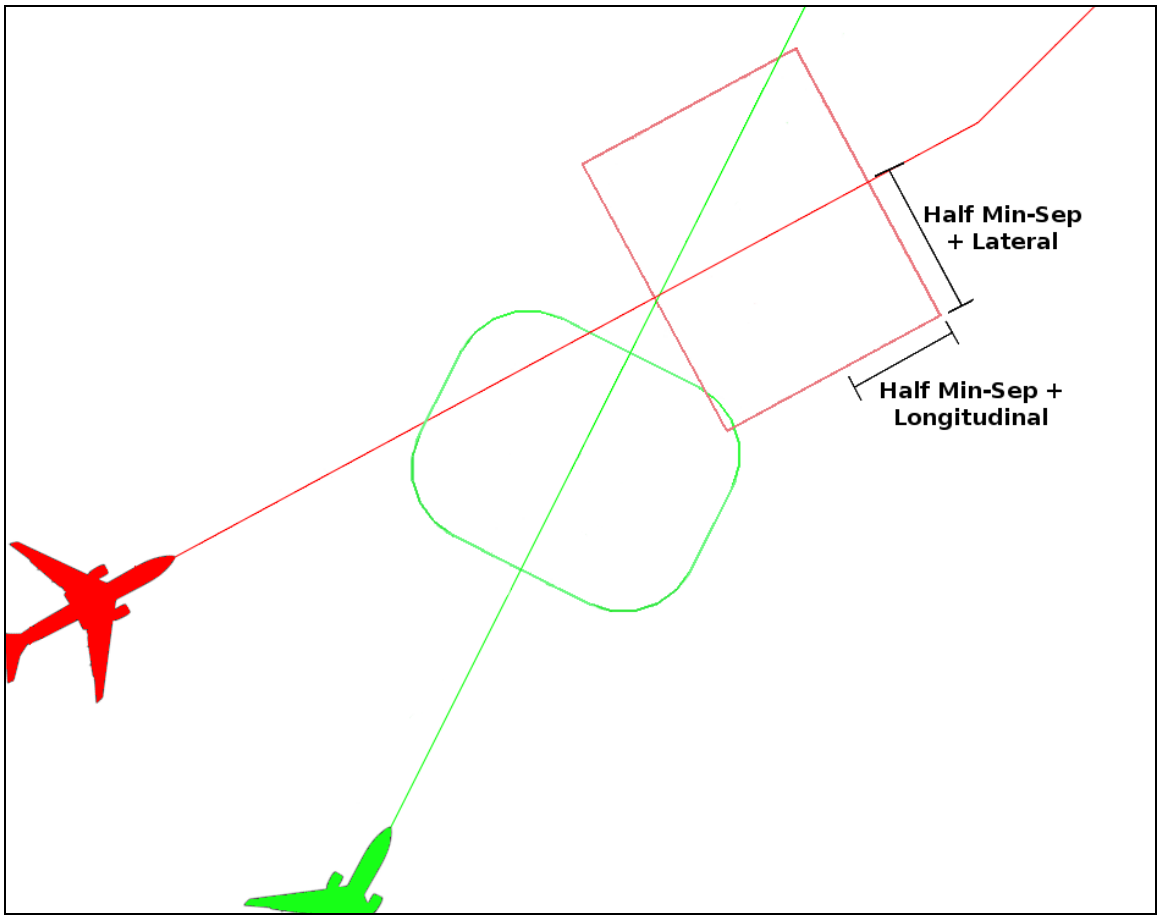

Figure 3. A simplified depiction of the conformance bounds being used for the conflict probe.

1.1.2.2 Likelihood

Likelihood is a value determined by the conflict probe of how likely the conflict prediction 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 at what likelihood the conflict prediction will be notified to the air traffic controller.

The likelihood setting is a piecewise linear function of likelihood over warning time as shown in Figure 4. If the likelihood value determined 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*. *a* is the minimum time in minutes at which likelihood is considered. At and below this time, a likelihood value of 0.0 will still cause a notification. *b*, in this experiment, is the time in minutes at which a 0.8 likelihood setting 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. At any time greater than *c* a value of 1.0 likelihood is required for a notification. If no *b* is included, then the function is linear from *a* to *c*, and the value of *b* can be determined by linearly interpolating.

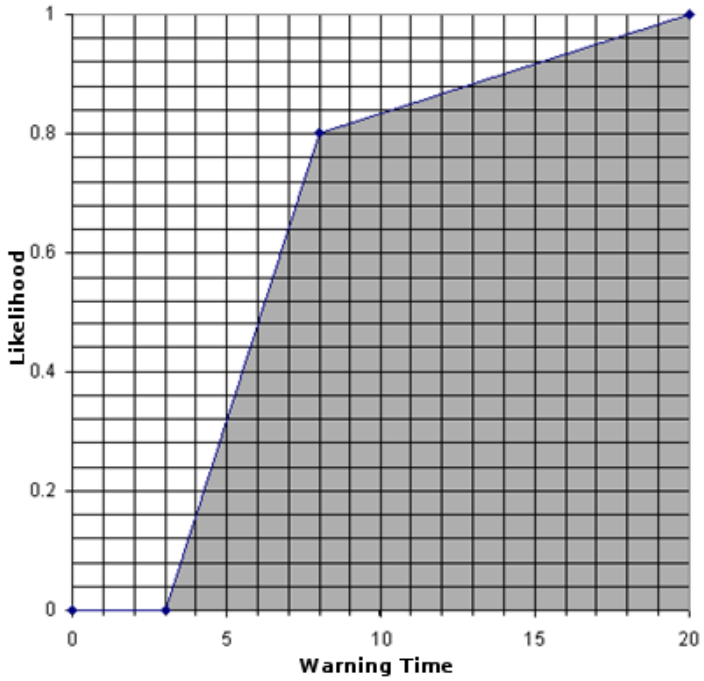

Figure 4. Piecewise linear likelihood function.

1.1.3 Previous Work

This study, designated Experiment 2, is the main study for FA18 prototyping efforts. It was preceded by Experiment 1, which studied the effects of parameter settings on the performance of the ERAM CP. Experiment 1 was an integrated experiment that used a d-optimal design [NIST/SEMATECH, 2011] with twelve treatment runs and a single baseline run. The parameters varied in the treatment runs were longitudinal conformance bound, lateral conformance bound, and likelihood setting. The purpose of Experiment 1 was to determine if a significant performance gain could be achieved by only altering parameters that can be changed in the current system, with no enhancements required.

The lateral conformance bound had a large impact on the False Alerts. Setting the lateral conformance bound significantly lower than the 2.5 nm default reduced the False Alerts by as much as 26%. The effect of the longitudinal conformance bound was found to be much different, having a large effect on the Late Alerts. It was found that setting the lateral conformance bound slightly lower, to a setting of around 1.25 nm, provided an improvement to the Late Alert performance as well as the False Alert performance. The likelihood was slightly more difficult to draw a conclusion on. Though a major effect on performance was seen when changing the likelihood, it was determined that a future study is required that will help to understand the likelihood further and provide information for recommending a setting.

The final result of the experiment was that a significant improvement can indeed be achieved by altering these parameters.

1.2 Scope of Study

This document reports on the results of an experiment limited to one large six hour traffic sample collected in May 2005 from the Washington Air Route Traffic Control Center (ARTCC). To

induce conflicts between aircraft and for evaluation purposes only, the data sample was time shifted using a methodology documented in [Paglione, 2003].

There are several important details of this experiment that should be noted before any results are presented. First and foremost is that this experiment is an initial experiment and will be followed by others. This particular experiment contains a single scenario, so conclusions can only be made on this scenario. The experiments to follow will contain multiple scenarios that will expand on the conclusions of this document. This initial experiment provides insight to how the prototypes and parameter settings affect the probe, and will help to design the follow-up experiments.

The distribution of actual conflicts in this scenario does not provide a significant number of late alerts or missed alerts. For this reason, the confidence intervals of the experiment for the Late Alerts values are very large in comparison to those for False Alerts. The follow-up experiments will help to improve this confidence, and the statistical significance of the results. This initial experiment helped to determine an area of parameter settings that should be focused on in the follow-up experiments.

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.

Lastly, there were two discoveries made while analyzing Experiment 1 and 2. When designing the likelihood (Llh) factor, two additional settings were chosen that were seen as an extreme setting and a mid-range setting, with the original setting being considered the most conservative. However, something that was overlooked when choosing these two additional settings was the effect of the minimum value on the late alerts. The minimum value for both additional settings was chosen at three minutes which is the same as the warning time requirement. This resulted in some cases that delayed the alerts until the three minute minimum was reached. However, the likelihood uses the prediction, and the warning time requirement is on the actual. Often the actual conflict starts slightly before the prediction, which causes these to be considered late alerts because the likelihood function delayed them too long. It is expected that much better Late Alert performance can be gained by increasing the minimum value slightly, but this will need to be determined in a future experiment. As a result of this, the main conclusion that can be made on likelihood at this time is that it can have a large impact on performance, but a specific setting cannot yet be recommended. However, it is clear from both Experiment 1 and 2, likelihood is an extremely important function that must be integrated in the experiments with the prototype functions because of its potential for interaction with them. At this time, there are also plans to perform additional experiments just on likelihood as well.

Similar to the issue with likelihood, the Conflict Geometric Separation (CGS) prototype was given a delay value for in-trail conflicts of three minutes, also causing it to delay many alerts to beyond the minimum three minute warning time. As a result, no conclusion can be made on this particular prototype, and the analysis will have to be redone in order to make a recommendation.

In total, there are a number of noteworthy considerations listed above affecting the scope of judgments that can be drawn from this experiment. However, this study is just the second in a series of experiments each building on the previous. This study is a leap forward in the application of experimental design techniques and use of advanced metrics on ERAM development. They were first introduced in Experiment 1 [Crowell, 2011, B] and now are being applied to complex algorithmic prototypes. Thus, the considerations listed do not necessarily limit the scope of the results being presented but provide context and document the issues to be resolved in follow-up experiments already being planned.

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). This analysis section also includes Section 3.3 which presents detailed flight examples providing illustrations of the prototype enhancements performance on ERAM's predictions. 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.	3.2.3.1
5 – Model Results	Examine the results of the model and discuss factor effects	3.2.3.2 & 3.2.3.3
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 determine (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 expands upon that used in Experiment 1 by adding three new factors that represent the three prototype enhancements: Forced Trajectory Rebuild (FTR), Growth Adherence Bounds (GAB), and Conflict Geometric Separation (CGS). The main purpose of this experiment is to determine the effects of these enhancements on the trajectory modeling and conflict probe performance. In order to evaluate this, it is also necessary to determine how each of the factors interacts with one another.

The factors for the prototype enhancements are binary and indicate whether that particular prototype enhancement is on or off. Given the three new binary factors, combined with the one ternary factor and two continuous factors from Experiment 1, the total number of runs required for a full factorial (assuming three samplings of the continuous functions) would be 216. 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 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 desirable to not degrade warning time performance, also, 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:

> *Through a set of purposeful runs of ERAM, input with the legacy ZDC 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 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. 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, 2011, A] [Lapihuska, 2011, A] that decoupled the TM bounds from the CP bounds. Instead all changes to the conformance bounds affected both the TM and CP bounds. Both of these factors are continuous factors, modeled using a quadratic equation. The ranges of 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	0.5 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 5. 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 0.8. This results in a function like the one shown in Figure 6. The settings used for likelihood are shown in Table 3.

Table 3. Nominal factors of the Integrated Experiment.

Factor	Settings		
Likelihood Function	10/20	3/8/20	3/8/10

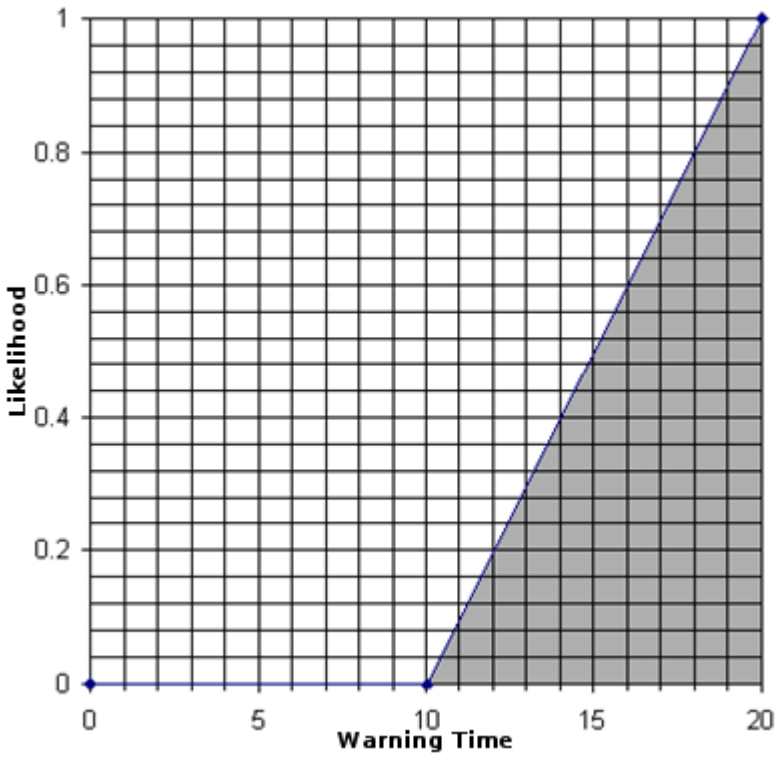

Figure 5. Linear likelihood function for 10/20 setting.

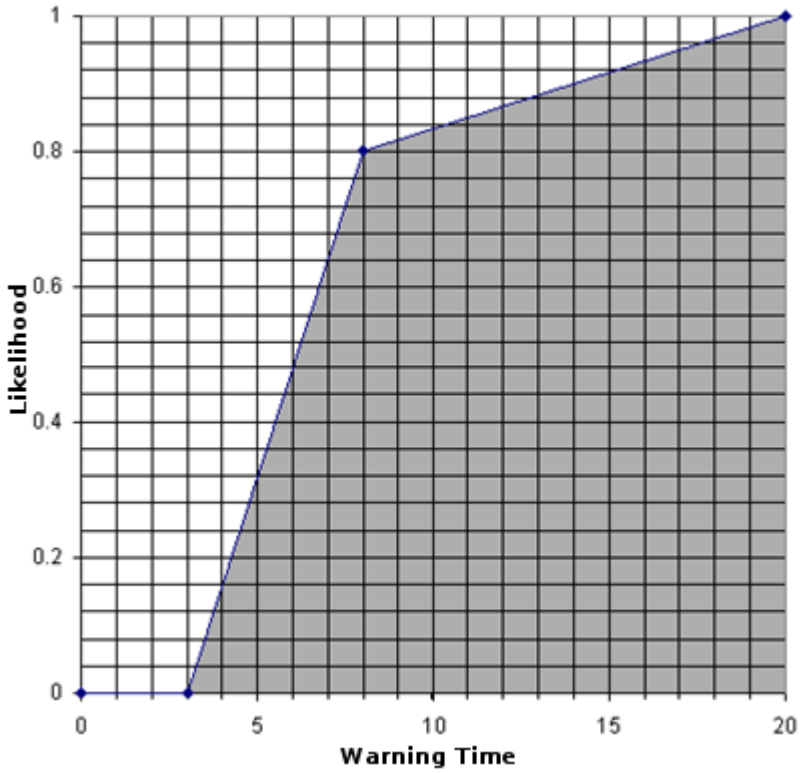

Figure 6. Piecewise linear likelihood function for 3/8/20 setting.

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

These settings resulted in the 30 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 or FA32 Baseline (32BL).

Table 4. Thirty runs for the integrated experiment and the baseline run.

Run	FTR	GAB	CGS	Lat	Long	Likelihood
1	Off	Off	Off	0.5	1.0	3/8/20
2	On	Off	Off	2.5	1.0	3/8/20
3	On	Off	Off	0.5	1.5	10/20
4	Off	On	Off	1.5	1.5	3/8/20
5	On	On	Off	2.5	0.5	10/20
6	On	Off	Off	0.5	0.5	3/8/10
7	Off	On	On	0.5	1.0	3/8/20
8	On	On	On	0.5	1.5	10/20
9	On	On	On	0.5	0.5	3/8/10
10	Off	On	Off	2.5	0.5	3/8/20
11	Off	Off	On	0.5	0.5	3/8/10
12	Off	On	On	2.5	1.5	3/8/10
13	On	Off	On	1.5	1.0	3/8/10
14	Off	Off	On	2.5	1.0	3/8/20
15	On	On	On	2.5	1.0	3/8/20
16	Off	On	On	2.5	0.5	10/20
17	On	On	Off	0.5	1.0	3/8/20
18	On	Off	On	2.5	0.5	3/8/10
19	On	On	Off	2.5	1.5	3/8/10
20	Off	On	Off	0.5	0.5	10/20
21	On	Off	On	2.5	1.5	10/20
22	On	Off	On	0.5	0.5	10/20
23	Off	On	Off	0.5	1.5	3/8/10
24	Off	On	Off	2.5	0.5	3/8/10
25	Off	Off	On	0.5	1.5	10/20
26	Off	Off	Off	2.5	1.5	3/8/10
27	Off	On	Off	2.5	1.5	10/20
28	On	Off	On	0.5	1.5	3/8/20
29	Off	Off	Off	2.5	0.5	10/20
30	On	Off	On	1.5	0.5	3/8/20
BL	Off	Off	Off	2.5	1.5	10/20

2.2.1.2 Model

The initial model allowed both continuous factors to have at most a quadratic effect. It was assumed all factors could interact with each other 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 additive. The model is defined as in Eq. 1.

Response:

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

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, 1.5, 2.5
$Long_j$ = longitudinal conformance bounds in nautical miles, j = 0.5, 1.0, 1.5
$Like_k$ = likelihood, k = "10/20", "3/8/10", "3/8/20"
$\varepsilon_{n(fghijk)}$ = random error, n = 1, 2, … for all f, g, h, i, j, k

Upon testing this model, it was found that the prototype choice interacted with $Lat_i Like_k$, but not enough data was collected to model all of these effects. Fortunately, it was found that CGS_h could account for the majority of this interaction after inserting $FTR_f Lat_i Like_k$, $GAB_g Lat_i Like_k$, and $CGS_h Lat_i Like_k$ into the model one at a time and comparing the effects. This leads to the first model being defined as in Eq. 2.

Response:

$$
R_{fghijk} = R^0_{fghijk} + CGS_h Lat_i Like_k \qquad \textbf{Eq. 2}
$$

Where the factors are as described in Eq. 1.

There are five response variables, all of which are addressed by this model. R_{fghijk} will represent the percentage improvement in traditional and fuzzy false alert counts, traditional and fuzzy late alert counts, or the 25th percentile of warning time from the baseline depending on context.

3 Performance Evaluation

The performance evaluation analyses used in this study are similar to those used in Experiment 1. The metrics used in Experiment 1 will also be used in this experiment so a description of them can be found in [Crowell, 2011, B]. An integrated experiment was designed, similar to Experiment 1, but containing three new factors, each of which represents one of the prototype enhancements described in the previous section. This integrated experiment allowed the analysts to determine the effects of the prototype enhancements on the trajectory modeling and conflict probe performance.

3.1 Trajectory Modeling Analysis

Trajectory errors from 43 runs were input to the model. This was done by taking the average absolute error at a specific look ahead time for each flight in the scenario. Then, the total average of these per-flight averages was calculated for every look ahead time value. The look ahead times employed in this analysis were 0, 300, 600, 900, and 1200 (every 5 minutes). The results of the prediction profiler are depicted in Figure 7.

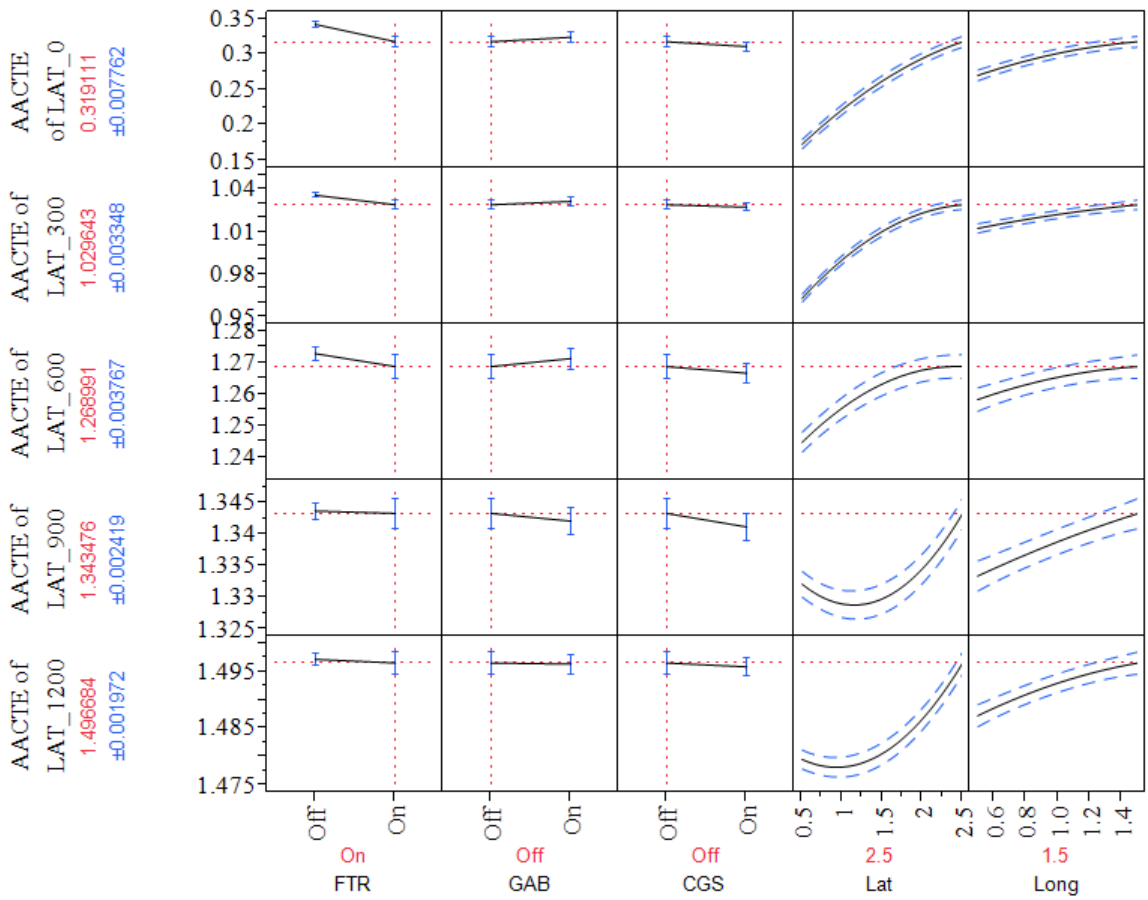

Figure 7. Prediction Profiler for Average Absolute Cross Track Error

The prediction profiler results in Figure 7 demonstrate the values of predicted average absolute cross track error at various look ahead times, starting at 0 seconds in the row at the top of the graph and increasing to 1200 seconds at the bottom. The effect from FTR decreasing error is very small, and is observed to be less strong than the effect from adjusting the lateral or longitudinal

conformance bound. The strength of the FTR effect decreases with increasing look ahead time; increasing look ahead time is itself tied to increased cross track error. Also, when either the lateral or longitudinal bound is decreased, the effect from FTR appears to be reduced.

To get a more focused measure of 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 one with FTR off, with identical settings for other parameters. This process was repeated for scenario pairs that effectively isolate GAB in the same way.

Table 5. Matched Pair Test Results.

Isolated Binary Factor	Lat	Long	Mean Diff (Cross)	*p*-value (Cross)	Mean Diff (Along)	*p*-value (Along)
FTR	0.5	1.5	0.00295	0.0034	0.05817	<0.0001
FTR	0.5	0.5	-0.00003	0.9741	-0.01800	0.4488
FTR	2.5	1.0	0.00963	<0.0001	0.02593	<0.0001
GAB	1.5	1.5	0.00021	0.7932	-0.0205	0.3870
GAB	2.5	0.5	0.00120	0.1859	-0.0229	0.3351

Positive mean difference values in Table 5 indicate an improvement in trajectory accuracy (decrease in average absolute error by flight) when the algorithm enhancement is enabled. The preliminary results indicate that the effect of the FTR enhancement in this sample data is statistically significant in some cases with large lateral and longitudinal. For example, in average absolute along track error in the first data row the mean difference (decrease) between scenarios of ~0.06nm has an associated *p*-value of less than 0.0001, which means the effect is statistically significant. However, at magnitudes below 0.1nm, these effects are not practically significant because they are on the scale of radar noise. Future analysis will investigate the mechanisms of how flights are affected by the enhancements, and possibly add terms to account for variation between flights in the model. Further work is needed to provide proof of increased trajectory accuracy.

3.2 Conflict Probe Analysis

Many different analyses were performed on the data in order to determine the effects that each factor has on the performance of the conflict probe. These analyses are similar to those performed in [Crowell, 2011, B], but expanded to include the three additional factors. The null hypothesis is similarly:

> *A significant Conflict Probe performance improvement cannot be gained through the addition of the forced trajectory rebuild, growth adherence bounds, or conflict geometric separation prototype enhancements to the conflict probe with some combination of lateral adherence bound, longitudinal adherence bound, and likelihood function.*

Note that the effects of adherence bounds and the likelihood function are worded differently in this null hypothesis. The main focus of Experiment 2 is to determine if the prototype enhancements can increase performance, whereas the adherence bounds and likelihood factors are included because the best performance gain of a particular prototype may be noticed at different settings than the baseline has. There may indeed be some interaction between the altered settings and the prototype enhancements.

The study performed in this document will attempt to prove this null hypothesis false. This will be proven false by a significant decrease in False Alert Rate (FAR) as well as either a decrease or no increase in Late Alert Rate (LAR), and an increase or no decrease in Warning Time (WT). If these three requirements cannot be met, then the null hypothesis cannot be rejected.

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 of the current probe being run in the live ERAM system. This baseline run does not include any prototype performance enhancements. The second baseline run is referred to as the FA32 Baseline (32BL, BL). This baseline uses the settings of the current live ERAM system, but includes performance enhancements under Function Area 32 (FA32) [Crowell, 2011, B] [McKay, 2011]. This was included as a baseline because all of the thirty treatment runs also included the FA32 enhancements. Since this document is concerned with the FA18 enhancements described in Section 1.1.1, the main comparisons will be between the treatment runs and the 32BL, which will show only the performance effects of the FA18 enhancements.

The alert type counts of each of the thirty runs and the two baselines are shown in Table 6. With the 32 rows to study, this table is harder to gather information from than its counterpart in Experiment 1. It is easy, however, to recognize that all treatment runs reduce the False Alert (FA) count from the 1394 in the 32BL. Runs 1, 6, 7, 9, 11, 13, 15, 17, 23, and 30 reduce the FA count to 720 or below. Something that immediately stands out about these runs is that none of them have a likelihood setting of 10/20. From the Experiment 1 results this is expected since likelihood was determined to be a major factor in the reduction of False Alerts. As far as the prototype factors, there is no pattern in these ten runs that is immediately recognizable.

Another observation that can be gathered from this table is that all treatment runs increase the Late Alert (LA) count over the 32BL, though Runs 3, 23, and 27 do not increase it significantly. Run 23 was also included in the list of those that have the greatest reduction to False Alerts.

Table 6. Alert type counts of the thirty treatment runs and the two baselines.

Run	FTR	GAB	CGS	Lat	Lon	Llh	VA	Std VA	LA	MA	FA
1	Off	Off	Off	0.5	1	3/8/20	158	129	17	3	718
2	On	Off	Off	2.5	1	3/8/20	158	130	14	6	981
3	On	Off	Off	0.5	1.5	10/20	159	132	13	6	987
4	Off	On	Off	1.5	1.5	3/8/20	157	128	16	5	878
5	On	On	Off	2.5	0.5	10/20	154	123	20	4	1166
6	On	Off	Off	0.5	0.5	3/8/10	152	121	21	5	614
7	Off	On	On	0.5	1	3/8/20	154	124	21	3	666
8	On	On	On	0.5	1.5	10/20	150	121	23	5	892
9	On	On	On	0.5	0.5	3/8/10	146	116	27	4	596
10	Off	On	Off	2.5	0.5	3/8/20	154	123	19	5	889
11	Off	Off	On	0.5	0.5	3/8/10	148	117	27	3	606
12	Off	On	On	2.5	1.5	3/8/10	151	121	23	4	788
13	On	Off	On	1.5	1	3/8/10	148	121	25	5	701
14	Off	Off	On	2.5	1	3/8/20	153	123	21	5	835
15	On	On	On	2.5	1	3/8/20	150	120	24	4	668
16	Off	On	On	2.5	0.5	10/20	147	116	26	5	949
17	On	On	Off	0.5	1	3/8/20	156	126	18	4	662
18	On	Off	On	2.5	0.5	3/8/10	141	111	32	5	736
19	On	On	Off	2.5	1.5	3/8/10	155	126	15	8	783
20	Off	On	Off	0.5	0.5	10/20	154	123	20	4	870
21	On	Off	On	2.5	1.5	10/20	153	127	18	6	1094
22	On	Off	On	0.5	0.5	10/20	146	115	29	3	857
23	Off	On	Off	0.5	1.5	3/8/10	161	132	13	4	699
24	Off	On	Off	2.5	0.5	3/8/10	154	123	19	5	865
25	Off	Off	On	0.5	1.5	10/20	150	120	24	4	953
26	Off	Off	Off	2.5	1.5	3/8/10	159	131	15	4	1150
27	Off	On	Off	2.5	1.5	10/20	162	134	12	4	1253
28	On	Off	On	0.5	1.5	3/8/20	149	119	25	4	734
29	Off	Off	Off	2.5	0.5	10/20	155	125	18	5	1294
30	On	Off	On	1.5	0.5	3/8/20	145	116	28	4	720
32BL	Off	Off	Off	2.5	1.5	10/20	163	135	11	4	1394
IBL	Off	Off	Off	2.5	1.5	10/20	158	134	12	4	1470

As always, it is important to also 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 7. 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. Looking at this metric we can see that Run 23 has a value of 11.2, which is still only a slight increase over the 32BL's value of 8.95.

Table 7. Adjusted Late Alert value of the treatment runs compared to the LA+MA counts.

Run	FTR	GAB	CGS	Lat	Long	Llh	LA+MA	Adj LA
1	Off	Off	Off	0.5	1.0	3/8/20	20	13.74
2	On	Off	Off	2.5	1.0	3/8/20	20	12.27
3	On	Off	Off	0.5	1.5	10/20	19	13.57
4	Off	On	Off	1.5	1.5	3/8/20	21	12.85
5	On	On	Off	2.5	0.5	10/20	24	12.67
6	On	Off	Off	0.5	0.5	3/8/10	26	16.96
7	Off	On	On	0.5	1.0	3/8/20	24	15.71
8	On	On	On	0.5	1.5	10/20	28	16.18
9	On	On	On	0.5	0.5	3/8/10	31	18.61
10	Off	On	Off	2.5	0.5	3/8/20	24	12.73
11	Off	Off	On	0.5	0.5	3/8/10	30	16.38
12	Off	On	On	2.5	1.5	3/8/10	27	14.11
13	On	Off	On	1.5	1.0	3/8/10	30	17.53
14	Off	Off	On	2.5	1.0	3/8/20	26	15.22
15	On	On	On	2.5	1.0	3/8/20	28	15.98
16	Off	On	On	2.5	0.5	10/20	31	16.27
17	On	On	Off	0.5	1.0	3/8/20	22	14.61
18	On	Off	On	2.5	0.5	3/8/10	37	19.70
19	On	On	Off	2.5	1.5	3/8/10	23	14.89
20	Off	On	Off	0.5	0.5	10/20	24	15.26
21	On	Off	On	2.5	1.5	10/20	24	14.13
22	On	Off	On	0.5	0.5	10/20	32	19.14
23	Off	On	Off	0.5	1.5	3/8/10	17	11.20
24	Off	On	Off	2.5	0.5	3/8/10	24	13.31
25	Off	Off	On	0.5	1.5	10/20	28	14.20
26	Off	Off	Off	2.5	1.5	3/8/10	19	10.43
27	Off	On	Off	2.5	1.5	10/20	16	9.78
28	On	Off	On	0.5	1.5	3/8/20	29	16.51
29	Off	Off	Off	2.5	0.5	10/20	23	13.23
30	On	Off	On	1.5	0.5	3/8/20	32	20.30
32BL	Off	Off	Off	2.5	1.5	10/20	15	8.95
IBL	Off	Off	Off	2.5	1.5	10/20	16	8.79

The LA and FA Rates are shown in Table 8. LA Rate is calculated by summing the LA and MA count then dividing by the total number of conflict events (LA + MA + VA). FA Rates are calculated by dividing the number of FAs in a particular separation category by the number of correct no call events and FAs in the same category. Correct no call events are pairs of flights that pass a particular filter of separation but do not ever become a conflict and are never predicted as such. The filter used in this study is 5000 ft vertically and 30 nm horizontally. Since less than 5 nm horizontal and less than 1000 ft vertical is an actual conflict, this column is expected to be very close to or exactly 1.0 for all runs. This is because there is very little chance of getting a correct no-call, so the denominator becomes only the number of FAs. However, because the

minimum duration of loss of separation to be considered a conflict is 10 seconds, there is a possibility of a correct no-call. If the loss of separation is less than 10 seconds (only one sample point), then it is not considered a conflict, even though the minimum separation is less than 5 nm and 1000 ft. This is rare and only occurs once each in two of the 32 runs.

Table 8. Overall Alert Rates and False Alert Rates by separation category.

			FA Rate									
			Horz	< 5	< 5	5 ≤ h < 8	5 ≤ h < 8	8 ≤ h ≤ 13	8 ≤ h ≤ 13	> 13	> 13
Run	LA Rate	FA Rate	Vert	<1000	≥1000	<1000	≥1000	<1000	≥1000	<1000	≥1000
1	0.112	0.042		1.000	0.117	0.872	0.307	0.172	0.036	0.043	0.010
2	0.112	0.058		1.000	0.097	0.932	0.328	0.441	0.138	0.043	0.012
3	0.107	0.058		1.000	0.102	0.907	0.324	0.272	0.056	0.061	0.025
4	0.118	0.052		1.000	0.122	0.969	0.349	0.336	0.088	0.038	0.011
5	0.135	0.068		1.000	0.129	0.942	0.349	0.369	0.090	0.068	0.030
6	0.146	0.036		1.000	0.122	0.714	0.226	0.147	0.029	0.035	0.010
7	0.135	0.039		1.000	0.116	0.852	0.286	0.145	0.028	0.039	0.010
8	0.157	0.052		1.000	0.114	0.869	0.269	0.192	0.041	0.051	0.025
9	0.175	0.035		1.000	0.120	0.704	0.214	0.117	0.026	0.032	0.011
10	0.135	0.052		1.000	0.119	0.916	0.336	0.316	0.084	0.051	0.012
11	0.169	0.035		1.000	0.117	0.721	0.239	0.093	0.031	0.035	0.010
12	0.152	0.046		1.000	0.110	0.883	0.306	0.303	0.075	0.033	0.011
13	0.169	0.041		1.000	0.106	0.871	0.306	0.191	0.049	0.030	0.009
14	0.145	0.049		1.000	0.110	0.904	0.319	0.310	0.075	0.038	0.011
15	0.157	0.039		1.000	0.101	0.857	0.263	0.191	0.042	0.037	0.010
16	0.174	0.056		1.000	0.128	0.872	0.316	0.207	0.052	0.056	0.025
17	0.124	0.039		1.000	0.112	0.836	0.254	0.143	0.035	0.040	0.011
18	0.208	0.043		1.000	0.103	0.838	0.280	0.204	0.062	0.038	0.010
19	0.129	0.046		1.000	0.105	0.903	0.297	0.295	0.072	0.038	0.012
20	0.135	0.051		1.000	0.117	0.808	0.243	0.228	0.039	0.069	0.023
21	0.136	0.064		1.000	0.104	0.917	0.341	0.374	0.099	0.051	0.026
22	0.180	0.050		0.957	0.135	0.754	0.237	0.172	0.032	0.062	0.024
23	0.096	0.041		1.000	0.104	0.893	0.298	0.231	0.045	0.034	0.009
24	0.135	0.051		1.000	0.121	0.919	0.335	0.304	0.082	0.037	0.012
25	0.157	0.056		1.000	0.107	0.888	0.328	0.262	0.054	0.050	0.023
26	0.107	0.068		1.000	0.107	0.994	0.364	0.609	0.185	0.037	0.014
27	0.090	0.074		1.000	0.114	0.970	0.355	0.532	0.135	0.060	0.031
28	0.163	0.043		0.944	0.101	0.862	0.314	0.174	0.045	0.033	0.011
29	0.129	0.076		1.000	0.122	0.964	0.360	0.492	0.134	0.069	0.030
30	0.181	0.042		1.000	0.127	0.827	0.278	0.163	0.043	0.045	0.011
32BL	0.084	0.082		1.000	0.113	0.994	0.365	0.648	0.192	0.058	0.031
IBL	0.092	0.087		1.000	0.114	1.000	0.381	0.650	0.206	0.053	0.033

Table 9 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, but the metric of interest is the 25th percentile. This metric tells how the lower end of the alerts performed for warning time. This metric is used because it represents how close the lower end of warning time 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) tells the range between the 75th percentile and the 25th

percentile. This value can help explain some differences between runs. A larger IQR indicates that a conflict probe increases the warning time of those that already have a lot of warning time, while decreasing the warning time of those that have little. The IQR can be observed along with the 25th percentile value to get an idea of the shape of the curve of warning times. From here on, the 25th percentile of warning time will be used as the main warning time metric.

Table 9. Median, inter-quartile range, and 25th percentile of conflict warning time.

Run	Median	IQR	25th %
1	371.0	240.0	262.0
2	406.0	267.5	305.0
3	443.0	396.0	300.0
4	402.0	278.0	293.5
5	445.5	441.0	311.0
6	364.5	255.0	253.0
7	312.0	209.0	197.0
8	336.0	393.0	198.0
9	288.5	219.0	193.0
10	389.5	294.0	278.0
11	305.5	220.0	192.0
12	337.0	262.0	196.0
13	313.5	214.0	204.0
14	328.0	244.0	201.0
15	330.0	256.0	203.0
16	328.5	385.0	195.0
17	374.0	245.0	263.0
18	320.0	235.0	182.0
19	408.0	262.0	309.0
20	406.5	391.0	258.0
21	355.0	377.0	228.0
22	300.0	314.0	193.0
23	383.0	238.0	274.5
24	391.0	261.0	282.0
25	354.0	314.0	201.0
26	410.0	258.5	311.5
27	486.0	450.0	321.5
28	310.0	231.0	195.0
29	426.0	423.0	306.0
30	315.5	231.0	192.0
32BL	492.0	440.0	331.5
IBL	460.0	378.5	320.5

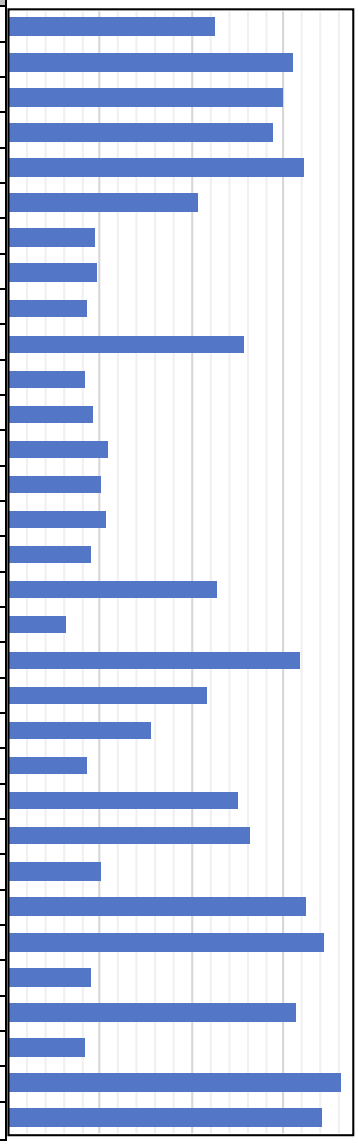

Figure 8 shows the Hit Rate versus the False Alert Rate for each of the 32 runs. The Hit Rate is *1 – LA Rate*. This is the rate at which the CP correctly predicts conflicts. Each point on the plot represents one of the 32 runs and is labeled by the run number or baseline name. The color of each point represents the percentage difference of warning time from the 32BL scenario. The warning time metric used is the 25th percentile. The legend on 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. Given these goals, Run 23

once again comes to the front as the best performing run. Run 23 has a lateral bound of 0.5 nm, longitudinal bound of 1.5 nm, likelihood of 3/8/10 and only has Growth Adherence Bounds turned on. These settings, minus the GAB setting, are the same as Run 6 from Experiment 1 which was also found to be the best performing of that experiment. Run 23 in this experiment has a Hit Rate of 0.904, a FA Rate of 0.041 and a warning time of 274.5, which is slightly better in both Hit Rate and FA Rate than Run 6 from Experiment 1 that had a Hit Rate of 0.886, an FA Rate of 0.045, and a warning time of 275. This perhaps indicates that GAB does have a positive effect on performance.

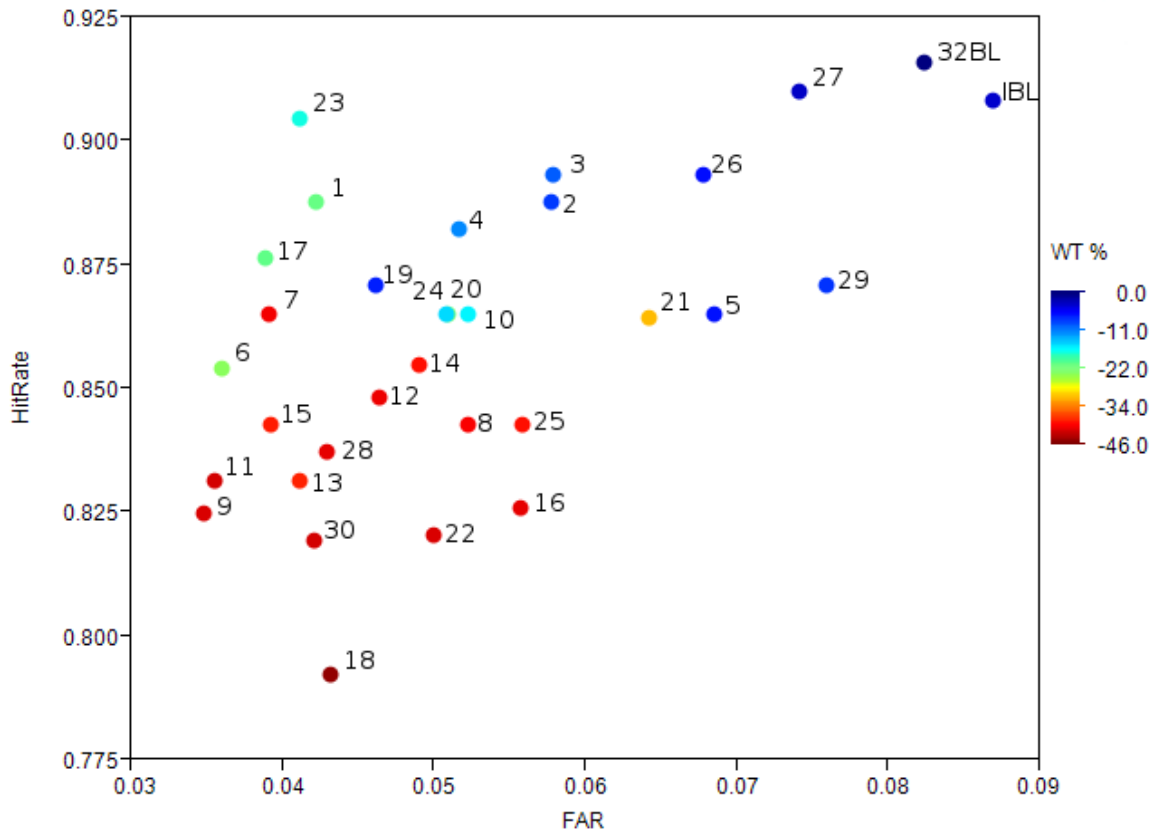

Figure 8. Hit Rate vs. False Alert Rate colored by Warning Time. The top-left corner of the graph and a blue color are the most desirable.

3.2.1 Fuzzy Analysis Results

The fuzzy metrics used in this study were defined in [Crowell, 2011, B]. They are continuous measurements of the results of a conflict probe, as opposed to the binary values used in the traditional metrics. Each event is assigned some value between 0.0 and 1.0 that represents how severe an event is and how strongly an alert should be generated for that event. Each alert is then assigned a value that represents how strongly the conflict probe predicted a conflict. Using a continuous value allows the metrics to be less susceptible to the small amount of noise that is generated each time ERAM is run. It allows traditional Late Alerts that are just barely late to count only slightly against the LA value, and allows traditional False Alerts that were very nearly actual conflicts to count only slightly, or not at all, against the False Alert value. The fuzzy metrics allow these cases to be "gray" whereas the traditional metrics require them to be "black or white."

Table 10 shows the fuzzy metrics for each of the treatment and baseline runs. Looking through this table, there are two runs that are of interest. Runs 4 and 26 each have LA values that are right around the values of the baselines, but they also have significantly reduced FA values. Run 23, the run of most interest from the traditional analyses, has an even lower FA value but increases the LA value slightly to 22.95. This shift in the LA performance is most likely due to better warning time performance in Runs 4 and 26 than in Run 23, which the fuzzy Late Alerts are weighted by. Looking back at Table 9, we can see that the 25th percentile of warning time in Run 23 is 274.5, whereas the values for Runs 4 and 26 are 293.5 and 311.5 respectively.

Table 10. Fuzzy Alert values for each of the treatment runs and the baselines.

Run	FTR	GAB	CGS	Lat	Long	Llh	VA	LA	FA
1	Off	Off	Off	0.5	1.0	3/8/20	266.58	25.30	565.11
2	On	Off	Off	2.5	1.0	3/8/20	257.16	27.14	676.90
3	On	Off	Off	0.5	1.5	10/20	264.51	26.31	792.96
4	Off	On	Off	1.5	1.5	3/8/20	265.22	19.03	656.85
5	On	On	Off	2.5	0.5	10/20	264.20	24.87	882.68
6	On	Off	Off	0.5	0.5	3/8/10	247.68	39.12	489.61
7	Off	On	On	0.5	1.0	3/8/20	259.81	27.90	519.58
8	On	On	On	0.5	1.5	10/20	255.54	30.39	720.32
9	On	On	On	0.5	0.5	3/8/10	250.02	40.65	467.64
10	Off	On	Off	2.5	0.5	3/8/20	261.16	25.35	654.16
11	Off	Off	On	0.5	0.5	3/8/10	252.37	35.61	476.71
12	Off	On	On	2.5	1.5	3/8/10	250.67	29.12	573.69
13	On	Off	On	1.5	1.0	3/8/10	254.65	32.06	533.10
14	Off	Off	On	2.5	1.0	3/8/20	258.42	26.52	618.98
15	On	On	On	2.5	1.0	3/8/20	251.51	31.34	502.58
16	Off	On	On	2.5	0.5	10/20	252.70	30.80	766.15
17	On	On	Off	0.5	1.0	3/8/20	258.65	29.69	520.64
18	On	Off	On	2.5	0.5	3/8/10	252.76	38.55	557.54
19	On	On	Off	2.5	1.5	3/8/10	253.07	28.08	555.18
20	Off	On	Off	0.5	0.5	10/20	257.57	29.94	734.05
21	On	Off	On	2.5	1.5	10/20	251.35	32.69	807.37
22	On	Off	On	0.5	0.5	10/20	254.54	39.85	720.35
23	Off	On	Off	0.5	1.5	3/8/10	260.06	22.95	543.12
24	Off	On	Off	2.5	0.5	3/8/10	264.87	25.07	629.35
25	Off	Off	On	0.5	1.5	10/20	262.27	27.49	760.40
26	Off	Off	Off	2.5	1.5	3/8/10	257.16	20.83	766.01
27	Off	On	Off	2.5	1.5	10/20	258.89	23.18	887.91
28	On	Off	On	0.5	1.5	3/8/20	249.38	32.99	566.87
29	Off	Off	Off	2.5	0.5	10/20	268.94	23.43	950.38
30	On	Off	On	1.5	0.5	3/8/20	258.33	35.76	557.11
32BL	Off	Off	Off	2.5	1.5	10/20	260.31	20.55	949.11
IBL	Off	Off	Off	2.5	1.5	10/20	257.33	19.08	994.56

Given the fuzzy metric results, two of the thirty treatment runs meet the requirements of reducing the False Alerts without degrading in LA performance. However, the purpose of this study is to determine if a performance gain can be achieved from any of the three prototype enhancements, and Run 26 has only parameter adjustments, but all prototypes disabled. Run 4 has the Growth Adherence Bounds prototype enabled, which is also the only prototype that is enabled in Run 23.

Table 11 shows the alert rates for the fuzzy metrics for each of the treatment runs and baselines. LA Rate is calculated by dividing the LA value from Table 10 by the total signal of the run. FA Rate is calculated by dividing the FA value by the total event count minus the signal. The event count minus signal gives the value of non-conflict events. This rate is calculated in total and also for the eight separation categories shown in Table 11.

Table 11. Overall Fuzzy Alert Rates and Fuzzy False Alert Rates by separation category.

				FA Rate							
			Horz	< 5	< 5	5 ≤ h < 8	5 ≤ h < 8	8 ≤ h ≤ 13	8 ≤ h ≤ 13	> 13	> 13
Run	LA Rate	FA Rate	Vert	<1000	≥1000	<1000	≥1000	<1000	≥1000	<1000	≥1000
1	0.087	0.033		0.852	0.198	0.749	0.265	0.165	0.035	0.043	0.010
2	0.095	0.040		0.831	0.186	0.763	0.271	0.264	0.092	0.038	0.010
3	0.090	0.047		0.802	0.336	0.766	0.273	0.236	0.051	0.055	0.023
4	0.067	0.039		0.810	0.205	0.809	0.291	0.271	0.071	0.033	0.010
5	0.086	0.052		0.771	0.360	0.797	0.291	0.281	0.067	0.057	0.026
6	0.136	0.029		0.848	0.207	0.583	0.194	0.145	0.029	0.035	0.010
7	0.097	0.031		0.844	0.195	0.716	0.247	0.140	0.027	0.038	0.009
8	0.106	0.043		0.758	0.341	0.713	0.231	0.169	0.039	0.047	0.024
9	0.140	0.028		0.822	0.207	0.566	0.183	0.115	0.026	0.032	0.011
10	0.088	0.039		0.831	0.209	0.747	0.280	0.247	0.062	0.046	0.011
11	0.124	0.028		0.832	0.201	0.565	0.207	0.092	0.030	0.035	0.010
12	0.104	0.034		0.779	0.195	0.714	0.262	0.212	0.057	0.028	0.010
13	0.112	0.032		0.816	0.180	0.736	0.260	0.161	0.043	0.028	0.009
14	0.093	0.037		0.814	0.200	0.765	0.272	0.227	0.059	0.034	0.010
15	0.111	0.030		0.790	0.181	0.683	0.224	0.167	0.038	0.034	0.010
16	0.109	0.045		0.734	0.346	0.739	0.273	0.176	0.047	0.049	0.024
17	0.103	0.031		0.880	0.202	0.709	0.220	0.140	0.034	0.039	0.011
18	0.132	0.033		0.767	0.185	0.707	0.235	0.167	0.054	0.035	0.010
19	0.100	0.033		0.761	0.191	0.736	0.245	0.208	0.054	0.032	0.010
20	0.104	0.043		0.820	0.340	0.688	0.209	0.226	0.039	0.068	0.023
21	0.115	0.048		0.778	0.322	0.748	0.279	0.252	0.074	0.037	0.023
22	0.135	0.042		0.797	0.361	0.627	0.205	0.169	0.032	0.061	0.024
23	0.081	0.032		0.845	0.175	0.761	0.258	0.208	0.043	0.032	0.009
24	0.086	0.037		0.830	0.202	0.744	0.280	0.233	0.061	0.032	0.010
25	0.095	0.045		0.791	0.327	0.735	0.280	0.223	0.050	0.046	0.022
26	0.075	0.045		0.800	0.199	0.802	0.297	0.347	0.120	0.030	0.011
27	0.082	0.053		0.728	0.337	0.792	0.296	0.333	0.093	0.047	0.025
28	0.117	0.033		0.824	0.189	0.722	0.268	0.151	0.041	0.031	0.010
29	0.080	0.056		0.777	0.358	0.807	0.298	0.339	0.095	0.056	0.026
30	0.122	0.033		0.828	0.215	0.662	0.238	0.147	0.040	0.042	0.011
32BL	0.073	0.057		0.761	0.330	0.798	0.298	0.389	0.124	0.045	0.024
IBL	0.069	0.059		0.771	0.334	0.817	0.312	0.367	0.130	0.039	0.026

Figure 9 shows the fuzzy hit rate versus fuzzy FA rate for each of the thirty treatment runs and two baselines. The hit rate is *1 – LA Rate*. Each point on the graph represents one of the treatment runs or baselines, and is colored based on the percent difference from 32BL of the 25th percentile of warning time. Warning time values are unchanged from the traditional metrics because the fuzzy metrics do not affect warning time. A high hit rate, low FA rate, and high warning time is

most desirable, so the top left corner of the plot is the most desirable location and blue is the most desirable color.

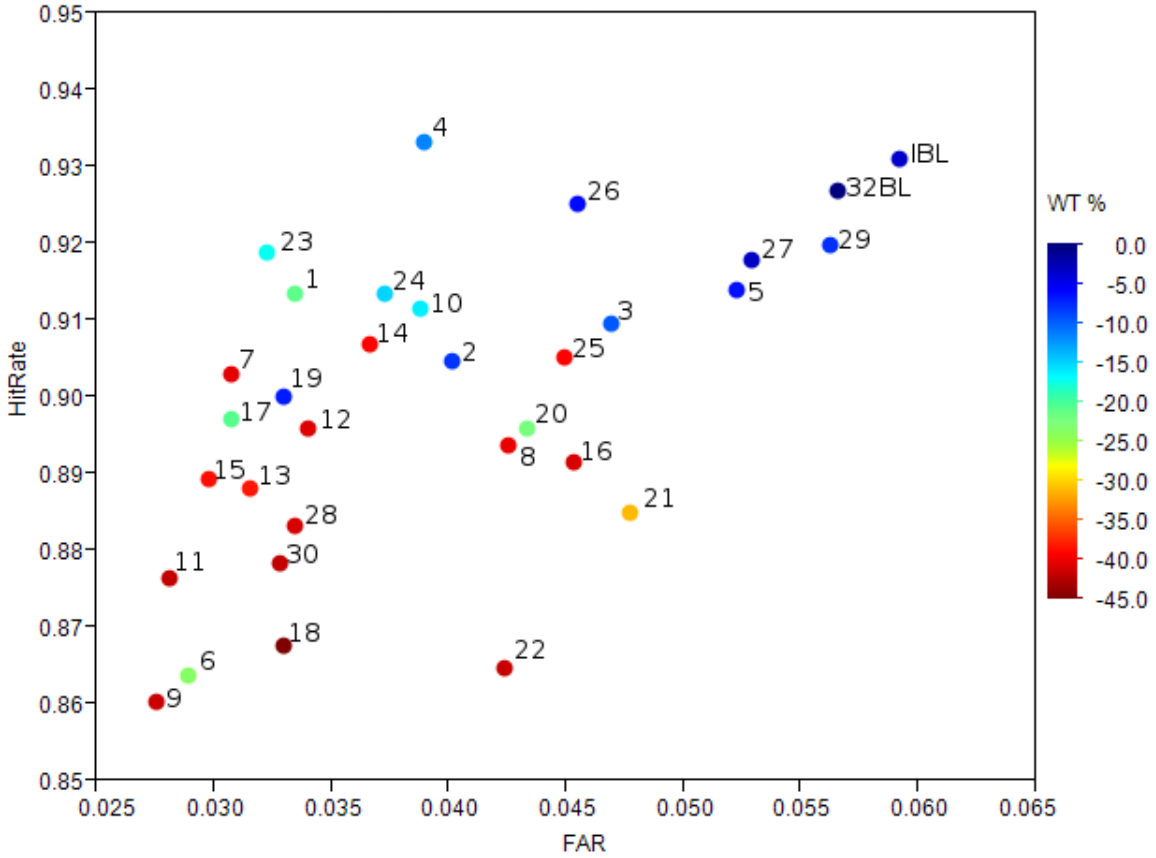

Figure 9. Fuzzy Hit Rate vs. Fuzzy False Alert Rate

Similar to the previous analyses, Runs 23, 4, and 26 stand out as the best performers. Runs 4 and 26 perform far better in warning time than Run 23, and in the case of the fuzzy metrics, also perform better in hit rate. Run 23, on the other hand, performs far better than the other two in FA rate. Keeping in mind the requirements of increasing FA performance without reducing MA performance, Run 4 is the only run that meets those requirements completely. Though the other two perform well in comparison to the other 27 treatment runs, they still degrade, if only slightly, in MA performance.

3.2.2 Direct Comparisons

Experiment 2 was designed as an extension to Experiment 1. The baselines of both experiments are the same. The only difference between the two experiments is that Experiment 2 includes treatment runs that have the prototype enhancements turned on. All the treatment runs in Experiment 1 are directly comparable to those in Experiment 2 which simply have the prototypes turned off. This allows us to combine the results of both experiments and provide direct comparisons in which the only difference between two runs is whether a prototype is turned off or not. Doing these direct comparisons can help to gather some information about how the prototypes affect performance and can help validate the results of the model analysis in the next section.

Table 12. FA, LA, and WT of Experiments 1 and 2 treatment runs and FA32 Baseline

Exp	Run	FTR	GAB	CGS	Lat	Long	Llh	FA %	LA %	WT %
2	1	Off	Off	Off	0.5	1.0	3/8/20	-48.49	33.33	-20.97
2	2	On	Off	Off	2.5	1.0	3/8/20	-29.63	33.33	-7.99
2	3	On	Off	Off	0.5	1.5	10/20	-29.20	26.67	-9.50
2	4	Off	On	Off	1.5	1.5	3/8/20	-37.02	40.00	-11.46
2	5	On	On	Off	2.5	0.5	10/20	-16.36	60.00	-6.18
2	6	On	Off	Off	0.5	0.5	3/8/10	-55.95	73.33	-23.68
2	7	Off	On	On	0.5	1.0	3/8/20	-52.22	60.00	-40.57
2	8	On	On	On	0.5	1.5	10/20	-36.01	86.67	-40.27
2	9	On	On	On	0.5	0.5	3/8/10	-57.25	106.67	-41.78
2	10	Off	On	Off	2.5	0.5	3/8/20	-36.23	60.00	-16.14
2	11	Off	Off	On	0.5	0.5	3/8/10	-56.53	100.00	-42.08
2	12	Off	On	On	2.5	1.5	3/8/10	-43.47	80.00	-40.87
2	13	On	Off	On	1.5	1.0	3/8/10	-49.71	100.00	-38.46
2	14	Off	Off	On	2.5	1.0	3/8/20	-40.10	73.33	-39.37
2	15	On	On	On	2.5	1.0	3/8/20	-52.08	86.67	-38.76
2	16	Off	On	On	2.5	0.5	10/20	-31.92	106.67	-41.18
2	17	On	On	Off	0.5	1.0	3/8/20	-52.51	46.67	-20.66
2	18	On	Off	On	2.5	0.5	3/8/10	-47.20	146.67	-45.10
2	19	On	On	Off	2.5	1.5	3/8/10	-43.83	53.33	-6.79
2	20	Off	On	Off	0.5	0.5	10/20	-37.59	60.00	-22.17
2	21	On	Off	On	2.5	1.5	10/20	-21.52	60.00	-31.22
2	22	On	Off	On	0.5	0.5	10/20	-38.52	113.33	-41.78
2	23	Off	On	Off	0.5	1.5	3/8/10	-49.86	13.33	-17.19
2	24	Off	On	Off	2.5	0.5	3/8/10	-37.95	60.00	-14.93
2	25	Off	Off	On	0.5	1.5	10/20	-31.64	86.67	-39.37
2	26	Off	Off	Off	2.5	1.5	3/8/10	-17.50	26.67	-6.03
2	27	Off	On	Off	2.5	1.5	10/20	-10.11	6.67	-3.02
2	28	On	Off	On	0.5	1.5	3/8/20	-47.35	93.33	-41.18
2	29	Off	Off	Off	2.5	0.5	10/20	-7.17	53.33	-7.69
2	30	On	Off	On	1.5	0.5	3/8/20	-48.35	113.33	-42.08
1	1	Off	Off	Off	0.5	0.5	10/20	-37.02	53.33	-20.36
1	2	Off	Off	Off	2.5	1.5	3/8/10	-17.65	26.67	-7.24
1	3	Off	Off	Off	0.5	1.5	10/20	-27.33	20.00	-7.99
1	4	Off	Off	Off	1.5	0.5	3/8/20	-41.61	80.00	-19.76
1	5	Off	Off	Off	1.5	1.0	10/20	-23.53	20.00	-6.18
1	6	Off	Off	Off	0.5	1.5	3/8/10	-44.48	13.33	-17.04
1	7	Off	Off	Off	1.5	1.5	3/8/20	-28.69	46.67	-15.84
1	8	Off	Off	Off	2.5	0.5	10/20	-7.60	46.67	-5.88
1	9	Off	Off	Off	2.5	1.0	3/8/20	-21.66	33.33	-8.60
1	10	Off	Off	Off	2.5	0.5	3/8/10	-29.63	60.00	-16.14
1	11	Off	Off	Off	0.5	0.5	3/8/10	-54.59	86.67	-22.78
1	12	Off	Off	Off	0.5	1.0	3/8/20	-48.49	33.33	-20.06
-	32BL	Off	Off	Off	2.5	1.5	10/20	0.00	0.00	0.00

Table 12 shows all the treatment runs from both experiments along with the FA32 Baseline (32BL). The metrics shown are percentage change from the FA32 Baseline. A negative value

indicates the metric was lowered in the treatment run. The following subsections will draw comparisons between runs from both experiments to analyze the performance differences achieved by each of the three prototypes. The treatment runs will be referred to in the format E#-##, where the first # will be 1 or 2 for Experiment 1 or Experiment 2, respectively, and the second ## will be 1 through 30 representing the respective run number in the specified experiment. So, E2-12 would represent Experiment 2, Run 12, whereas E1-3 would represent Experiment 1, Run 3. Since the baseline run is not unique to either experiment it will still be referred to as BL.

3.2.2.1 Forced Trajectory Rebuild

There are three sets of runs that only differ in their Forced Trajectory Rebuild (FTR) setting. Runs E2-6 and E1-11 have lateral and longitudinal settings of 0.5 nm and a likelihood of 3/8/10. The FA performance is improved slightly by FTR, reducing it from -54.59% to -55.95%. The LA performance is also improved reducing it from 86.67% to 73.33%. Finally the warning performance is degraded slightly, decreasing the value from -22.78% to -23.68%.

Runs E2-3 and E1-3 also match with a lateral setting of 0.5 nm, longitudinal bound of 1.5 nm, and a likelihood setting of 10/20. Once again FTR slightly improves the FA performance, reducing it from -27.33% to -29.20%. LA performance and warning time performance are both degraded, however. The LA percentage is increased from 20.00% to 26.67%, and the warning time is reduced from -7.99% to -9.50%.

Finally, Runs E2-2 and E1-9 match with a lateral setting of 2.5 nm, a longitudinal of 1.0 nm, and a likelihood setting of 3/8/20. False Alert performance is once again improved by FTR which reduces it from -21.66% to -29.63%. Late Alert performance is unchanged from 33.33%. Warning time performance is slightly improved from -8.60% to -7.99%.

Table 13 shows the results of all three comparisons. The metrics are calculated by subtracting the value of the run without FTR from the value of the run with FTR. So a negative value is desired for FA and LA, whereas a positive value is desired for warning time (WT).

Table 13. Performance comparisons for runs with all common settings except for FTR.

FTR On	FTR Off	Lat	Lon	Llh	FA	LA	WT
E2-6	E1-11	0.5	0.5	3/8/10	-1.36	-13.34	-0.90
E2-3	E1-3	0.5	1.5	10/20	-1.87	6.67	-1.51
E2-2	E1-9	2.5	1.0	3/8/20	-7.97	0.00	0.61

These results indicate that FTR tends to improve the FA performance slightly. It also indicates that neither Late Alerts nor warning time is significantly negatively affected by the FTR prototype. A value of 6.67% represents an increase of one Late Alert, and is not statistically significant.

3.2.2.2 Growth Adherence Bounds

For the Growth Adherence Bounds prototype, there are four sets of common runs that differ only in their GAB setting. Runs E2-20 and E1-1 have lateral and longitudinal conformance bounds of 0.5 nm and a likelihood of 10/20. The False Alert performance is improved slightly from -37.02% to -37.59%. LA performance is degraded from 53.33% to 60.00% and warning time is degraded from -20.36% to -22.17%.

Runs E2-23 and E1-6 have common settings of 0.5 nm lateral, 1.5 nm longitudinal, and 3/8/10 likelihood setting. The FA performance is improved again from -44.48% to -49.86%. LA

performance is unchanged from 13.33%, and warning time performance is insignificantly reduced from -17.04% to -17.19%.

Runs E2-4 and E1-7 have lateral and longitudinal settings of 1.5 nm and a likelihood setting of 3/8/20. False Alert performance is improved from -28.69% to -37.02%. LA performance is also improved slightly from 46.67% to 40.00%. Warning time is also improved from -15.84% to -11.46%.

Finally, Runs E2-24 and E1-10 have common settings of 2.5 nm lateral, 0.5 nm longitudinal, and 3/8/10 likelihood. Again, FA performance is improved, reducing it from -29.63% to -37.95%. LA performance is unchanged from 60.00% and warning time performance is improved slightly from -16.14% to -14.93%.

Table 14 shows the results of all four comparisons. The metrics are calculated by subtracting the value of the run without GAB from the value of the run with GAB. So a negative value is desired for FA and LA, whereas a positive value is desired for warning time (WT).

Table 14. Performance comparisons for runs with all common settings except for GAB.

GAB On	GAB Off	Lat	Lon	Llh	FA	LA	WT
E2-20	E1-1	0.5	0.5	10/20	-0.57	6.67	-1.81
E2-23	E1-6	0.5	1.5	3/8/10	-5.38	0.00	-0.15
E2-4	E1-7	1.5	1.5	3/8/20	-8.33	-6.67	4.38
E2-24	E1-10	2.5	0.5	3/8/10	-8.32	0.00	1.21

These results indicate that GAB has a positive effect on FA performance. With the smaller sample size of Late Alerts, a change of 6.67% indicates only a difference of one alert, which can usually be considered insignificant. Likewise the effects on warning time performance are insignificant.

3.2.2.3 Conflict Geometric Separation

For Conflict Geometric Separation (CGS) there are three sets of runs that have all but their CGS setting in common. Runs E2-11 and E1-11 have lateral and longitudinal conformance bounds of 0.5 nm and a likelihood setting of 3/8/10. The FA performance is improved slightly from -54.59% to -56.53%. LA performance is degraded, increasing from 86.67% to 100.00%. Warning time performance is also significantly degraded, reducing from -22.78% to -42.08%.

Runs E2-25 and E1-3 have common lateral settings of 0.5 nm, longitudinal settings of 1.5 nm, and likelihood settings of 10/20. False Alert performance is improved from -27.33% to -31.64%. LA performance is significantly degraded, increasing from 20.00% to 86.67%. Warning time performance is also significantly reduced from -7.99% to -39.37%.

Finally, Runs E2-14 and E1-9 have common settings of 2.5 lateral, 1.0 longitudinal, and 3/8/20 likelihood. FA performance is once again significantly improved from -21.66% to -40.10%. The LA performance is again significantly degraded, increasing from 33.33% to 73.33%. Warning time is also significantly reduced from -8.60% to -39.37%.

Table 15 shows the results of all three comparisons. The metrics are calculated by subtracting the value of the run without CGS from the value of the run with CGS. So a negative value is desired for FA and LA, whereas a positive value is desired for warning time (WT).

Table 15. Performance comparisons for runs with all common settings except for CGS.

CGS On	CGS Off	Lat	Lon	Llh	FA	LA	WT
E2-11	E1-11	0.5	0.5	3/8/10	-1.94	13.33	-19.30
E2-25	E1-3	0.5	1.5	10/20	-4.31	66.67	-31.38
E2-14	E1-9	2.5	1.0	3/8/20	-18.40	40.00	-30.77

These results indicate that the CGS prototype will create a performance increase for False Alerts, but only at the sacrifice of a relatively large degradation for LA and warning time performance. However, while observing the effects of this prototype, it was determined that the one of the settings internal to the prototype was set too low. The prototype delays notification of alerts for conflicts that are determined to be in-trail. The prototype in these treatment runs was set to delay notification until three minutes prior to the start of the conflict. It was determined in Experiment 1 when observing the effects of the likelihood setting that delaying notification until the required warning time will result in many Late Alerts, because of the amount of error that can exist in the longitudinal direction.

In order to better understand this effect, Run E2-25 was re-run with the CGS prototype set to delay only to four minutes (240 seconds). This new run will be referred to as E2-25B. The results of the new run can be seen in Table 16. With this new run, the FA performance is improved, though not quite as much, decreasing them from -27.33% to -30.92% as opposed to the -31.64% in E2-25. However, the LA degradation is far less than in E2-25, increasing from 20.00% in E1-3 to 46.67% in E2-25B instead of the 86.67% in E2-25. Likewise, warning time is not degraded quite as much, reducing it from -7.99% to -25.19% instead of the -39.37% in E2-25.

Table 16. The results of re-running Run E2-25 with notification delay set to four minutes.

FA %	LA %	WT %
-30.92	46.67	-25.19

Table 17 shows the comparison results of each of the runs with CGS on to the run with CGS off. In this table, it can be seen that, though CGS still has a significant negative effect on LA performance, it is much less extreme when the notification delay time used is greater than the required warning time. The effects on warning time are exactly what are expected because of the way the CGS prototype works by delaying the notification and therefore intentionally reducing the warning time of the alert.

Table 17. Performance comparisons for the CGS run that was re-run with four minutes notification delay and the original run with three minute delay.

Run	FA	LA	WT
E2-25	-4.31	66.67	-31.38
E2-25B	-3.59	26.67	-17.20

3.2.3 Model Analysis

The model analysis of the integrated experiment is the main focus of this study. While the other analyses performed up to this point are important, their main purpose is to support the findings of the model. The model uses linear and quadratic interpolation of the effects to allow any combination of values in the six-dimensional hypercube to be modeled accurately. This allows the analyst to determine what the performance of the probe would be at many different combinations of settings that were not run through ERAM. This is absolutely necessary to determine the effects of each factor on the performance of the probe without running an infeasible

number of simulations. With the thirty treatment runs, combined with the twelve treatment runs from Experiment 1 and the FA32 Baseline, we are able to model any combination of the prototype enhancements with lateral setting between 0.5 and 2.5 nm, longitudinal setting between 0.5 and 1.5 nm, and at any of the three likelihood settings.

3.2.3.1 Implementation of Statistical Model

The traditional metrics will be modeled separately from the fuzzy model (using the same warning time percentage in both models since it does not change). Using the data collected from the 42 runs and the baseline, the following leverage plots are obtained (Figure 10). In these plots the measured values (*y*-axis) are plotted against their modeled values (*x*-axis). Therefore, all points that fall on the diagonal are exactly modeled. The horizontal blue line represents the mean value of the samples and the red curves indicate the 95% confidence interval. The model captures between 98% and 100% of the variation in the study, which can be seen from the RSq value under each plot (RSq is the coefficient of determination R^2), and all of the responses are statistically significant. The significance is quickly determined in a leverage plot by determining if the confidence interval intersects the mean, where no intersection indicates insignificance.

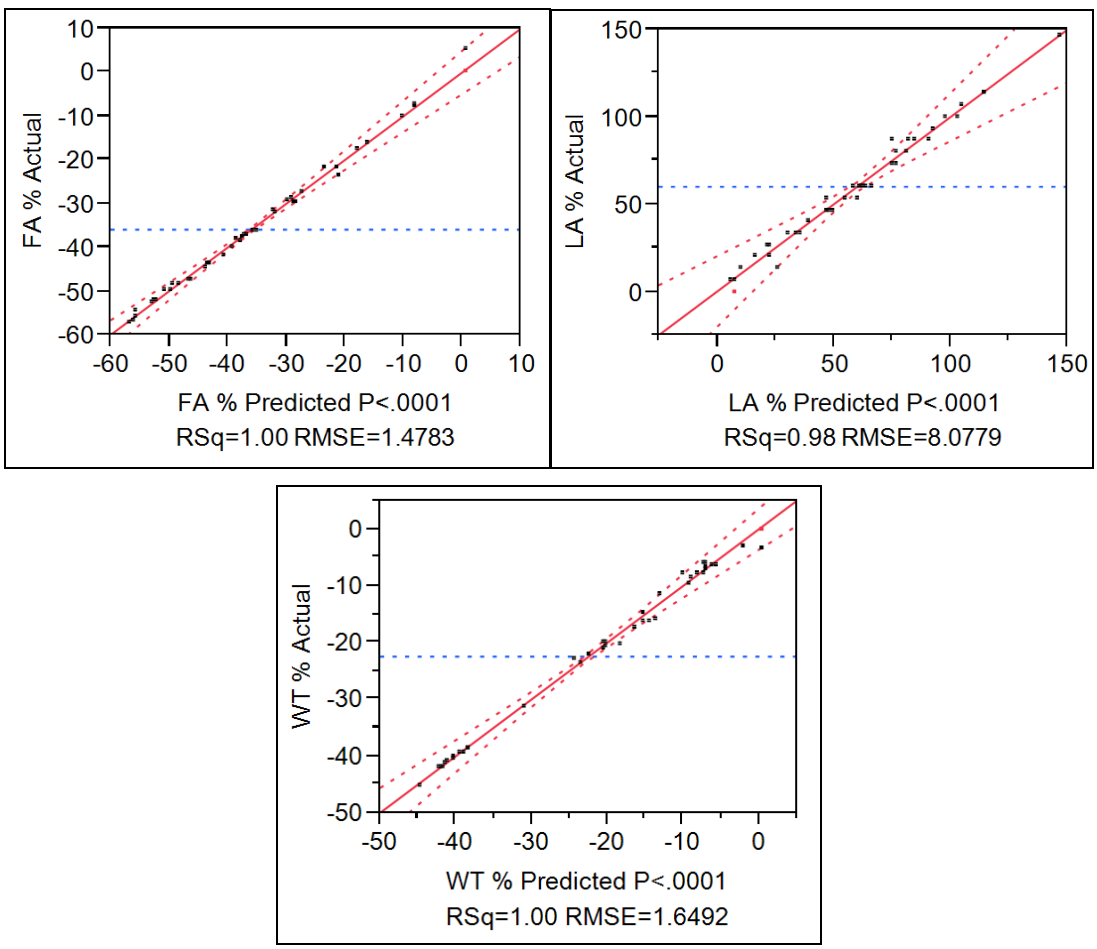

Figure 10. Traditional FA % and LA % metrics and WT % metric leverage plots.

By checking residuals, it is decided that run 1.11 is a considerable outlier, and since it is extra to the original requirements of the model, it is excluded, giving the following revised leverage plots (Figure 11).

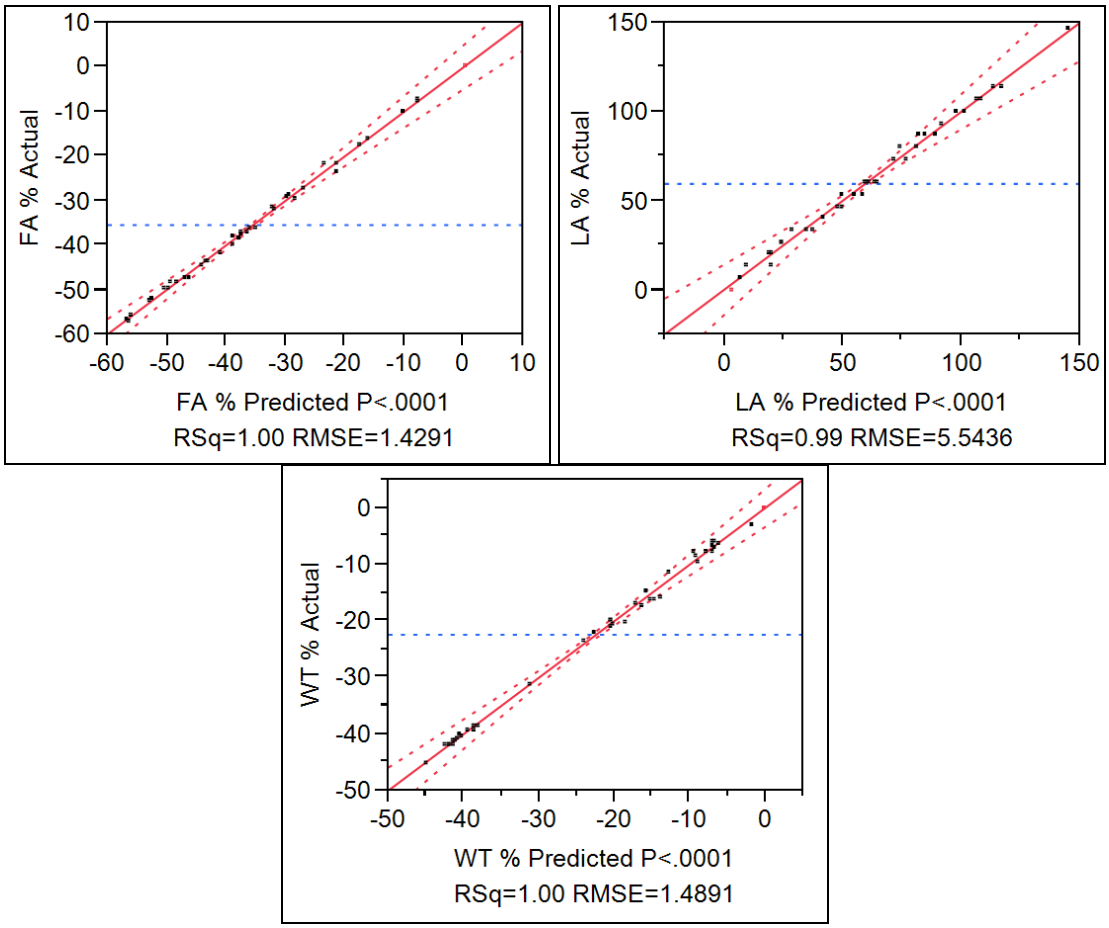

Figure 11. Traditional FA % and LA % metrics and WT % metric leverage plots with outlier removed.

The following tables (Table 18, Table 19, Table 20) list the effect tests for each interaction of the various factors for each of the three responses currently being investigated. Highlighted are those effects that have a *p*-value of at least 0.7000, where $p > 0.0500$ indicates statistical insignificance. CGS and GAB interacting with longitudinal was decided to be worth attempting to remove to gain some confidence in the weakest model (LA %). This is shown to be beneficial enough to be justified by examining the new leverage plots in Figure 12.

Table 18. Initial Model Effect Test for the FA % Response Variable.

Source	DF	Sum of Squares	F Ratio	Prob > F
FTR	1	67.2828	32.9458	0.0002
GAB	1	268.4345	131.4421	<0.0001
CGS	1	440.1402	215.5197	<0.0001
Lat(0.5,2.5)	1	1373.0169	672.3136	<0.0001
Long(0.5,1.5)	1	163.9831	80.2962	<0.0001
Likelihood	2	1793.6834	439.1489	<0.0001
FTR*GAB	1	2.3426	1.1471	0.3093
FTR*CGS	1	2.8848	1.4126	0.2621
FTR*Lat	1	26.9948	13.2183	0.0046
FTR*Long	1	35.1000	17.1871	0.0020
FTR*Likelihood	2	18.0396	4.4166	0.0422
GAB*CGS	1	5.2087	2.5505	0.1413
GAB*Lat	1	101.0828	49.4964	<0.0001
GAB*Long	1	5.5265	2.7061	0.1310
GAB*Likelihood	2	11.8488	2.9010	0.1015
CGS*Lat	1	261.7271	128.1577	<0.0001
CGS*Long	1	3.8507	1.8855	0.1997
CGS*Likelihood	2	26.1078	6.3920	0.0163
Lat*Long	1	5.1288	2.5114	0.1441
Lat*Likelihood	2	23.4130	5.7322	0.0219
Long*Likelihood	2	10.5825	2.5909	0.1240
Lat*Lat	1	6.0657	2.9702	0.1155
Long*Long	1	14.9960	7.3430	0.0219
Lat*Likelihood*CGS	2	4.3267	1.0593	0.3826

Table 19. Initial Model Effect Test for the LA % Response Variable.

Source	DF	Sum of Squares	F Ratio	Prob > F
FTR	1	1263.0900	41.1015	<0.0001
GAB	1	4.7540	0.1547	0.7023
CGS	1	17290.0210	562.6244	<0.0001
Lat(0.5,2.5)	1	36.2860	1.1808	0.3027
Long(0.5,1.5)	1	3552.6920	115.6060	<0.0001
Likelihood	2	732.3380	11.9153	0.0023
FTR*GAB	1	68.2780	2.2218	0.1669
FTR*CGS	1	40.9100	1.3312	0.2754
FTR*Lat	1	2.5050	0.0815	0.7811
FTR*Long	1	36.4630	1.1865	0.3016
FTR*Likelihood	2	551.4400	8.9720	0.0059
GAB*CGS	1	39.2900	1.2785	0.2846
GAB*Lat	1	26.9130	0.8757	0.3714
GAB*Long	1	1.5700	0.0511	0.8257
GAB*Likelihood	2	145.3130	2.3643	0.1443
CGS*Lat	1	104.6110	3.4041	0.0948
CGS*Long	1	0.7740	0.0252	0.8771
CGS*Likelihood	2	475.8300	7.7419	0.0093
Lat*Long	1	50.2220	1.6342	0.2300
Lat*Likelihood	2	1217.2000	19.8041	0.0003
Long*Likelihood	2	228.0540	3.7105	0.0623
Lat*Lat	1	208.2520	6.7766	0.0263
Long*Long	1	1581.0350	51.4475	<0.0001
Lat*Likelihood*CGS	2	64.7560	1.0536	0.3844

Table 20. Initial Model Effect Test for the WT % Response Variable.

Source	DF	Sum of Squares	F Ratio	Prob > F
FTR	1	13.1995	5.9523	0.0349
GAB	1	9.4655	4.2685	0.0657
CGS	1	5214.4206	2351.4530	<0.0001
Lat(0.5,2.5)	1	255.9998	115.4436	<0.0001
Long(0.5,1.5)	1	135.4824	61.0960	<0.0001
Likelihood	2	137.3479	30.9686	<0.0001
FTR*GAB	1	0.2260	0.1019	0.7561
FTR*CGS	1	0.0687	0.0310	0.8638
FTR*Lat	1	0.3012	0.1358	0.7202
FTR*Long	1	9.3573	4.2197	0.0670
FTR*Likelihood	2	6.0758	1.3700	0.2980
GAB*CGS	1	0.0119	0.0054	0.9431
GAB*Lat	1	0.0475	0.0214	0.8866
GAB*Long	1	3.9137	1.7649	0.2135
GAB*Likelihood	2	16.2844	3.6717	0.0637
CGS*Lat	1	162.5594	73.3065	<0.0001
CGS*Long	1	1.5790	0.7120	0.4185
CGS*Likelihood	2	36.5934	8.2509	0.0076
Lat*Long	1	4.8039	2.1663	0.1718
Lat*Likelihood	2	47.5188	10.7143	0.0033
Long*Likelihood	2	4.5494	1.0258	0.3934
Lat*Lat	1	0.0421	0.0190	0.8931
Long*Long	1	22.3339	10.0715	0.0099
Lat*Likelihood*CGS	2	20.6971	4.6667	0.0370

From Figure 12, it can be seen that R^2 has not degraded and the confidence intervals have been improved enough in the LA % to be justifiable. The new model is therefore Eq. 3.

Response:

$$\begin{aligned} R_{fghijk} = {} & \mu + Lat_i^2 + Lat_i + Long_j^2 + Long_j + Lat_i Long_j \\ & + Like_k + Lat_i Like_k + Long_i Like_k + FTR_f GAB_g + FTR_f CGS_h \\ & + FTR_f Lat_i + FTR_f Long_j + FTR_f Like_k + FTR_f + GAB_g CGS_h \\ & + GAB_g Lat_i + GAB_g Like_k + GAB_g + CGS_h Lat_i + CGS_h Like_k \\ & + CGS_h + CGS_h Lat_i Like_k + \varepsilon_{n(fghijk)} \end{aligned} \qquad \textbf{Eq. 3}$$

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, 1.5, 2.5
$Long_j$ = longitudinal conformance bounds in nautical miles, j = 0.5, 1.0, 1.5
$Like_k$ = likelihood, k = "10/20", "3/8/10", "3/8/20"
$\varepsilon_{n(fghijk)}$ = random error, n = 1, 2, … for all f, g, h, i, j, k

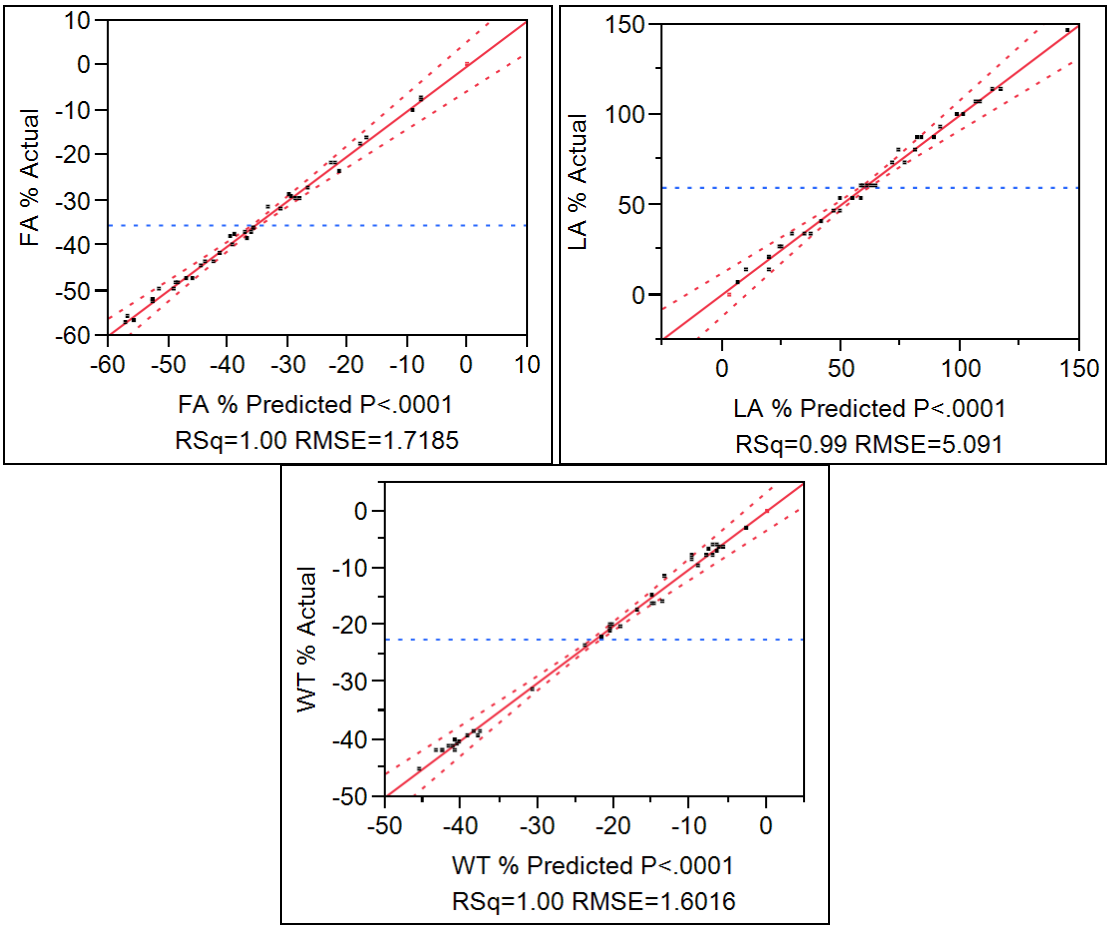

Figure 12. Traditional FA % and LA % metrics and WT % metric in the refined model.

The model also relies on the assumption that the random error $\varepsilon_{n(fghijk)}$ is normally distributed. The residual errors should therefore be tested for normalcy. Figure 13 shows normal probability plots, box plots, and histograms fitted to a normal density curve. The histograms and box plots do trend toward a symmetric distribution.

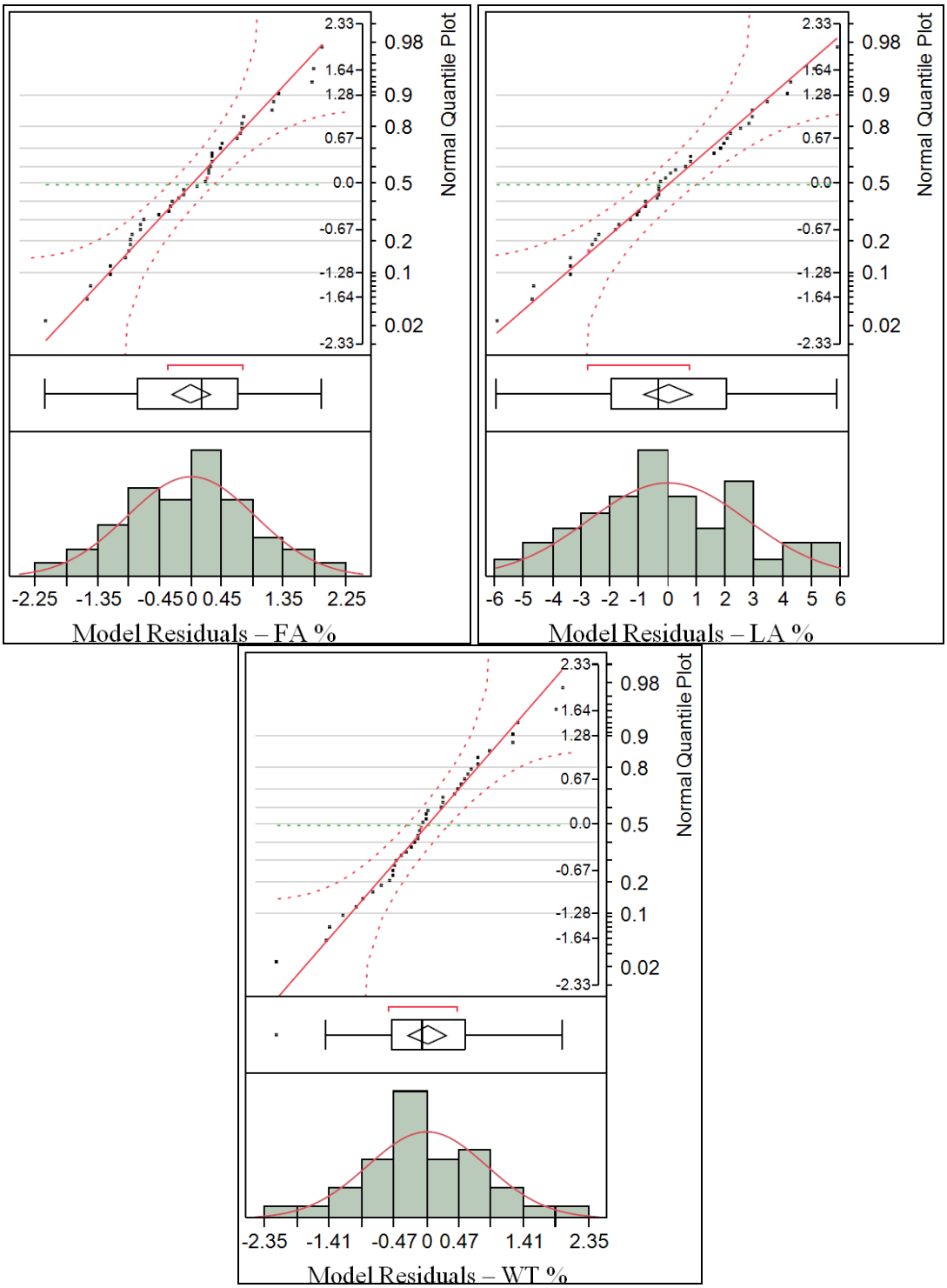

Figure 13. Residual Error Distributions for the Response Variables from the Refined Model.

Also, 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 model is indeed appropriate.

Table 21 shows the *p*-values for the interactions, with those less than 0.05 highlighted. Of the interactions that are not highlighted, there is not much good evidence for removing them, so the model will not be further refined by removing interactions.

Table 21. Summary of Refined Model Effect Tests for Response Variables.

Source	FA % P-Value	LA % P-Value	WT % P-Value
FTR	0.0004	<0.0001	0.0426
GAB	<0.0001	0.6911	0.0568
CGS	<0.0001	<0.0001	<0.0001
Lat(0.5,2.5)	<0.0001	0.2444	<0.0001
Long(0.5,1.5)	<0.0001	<0.0001	<0.0001
Likelihood	<0.0001	0.0005	<0.0001
FTR*GAB	0.7205	0.0984	0.4737
FTR*CGS	0.2710	0.2403	0.7637
FTR*Lat	0.0052	0.7151	0.5390
FTR*Long	0.0003	0.1130	0.0223
FTR*Likelihood	0.0506	0.0020	0.4000
GAB*CGS	0.3388	0.2167	0.7319
GAB*Lat	0.0001	0.3009	0.9152
GAB*Likelihood	0.1769	0.0946	0.0670
CGS*Lat	<0.0001	0.0611	<0.0001
CGS*Likelihood	0.0513	0.0036	0.0067
Lat*Long	0.0939	0.1183	0.2279
Lat*Likelihood	0.0227	<0.0001	0.0014
Long*Likelihood	0.2646	0.0374	0.3623
Lat*Lat	0.0850	0.0125	0.6420
Long*Long	0.0808	<0.0001	0.0200
Lat*Likelihood*CGS	0.1191	0.0840	0.0050

Since it is unknown at this point how the fuzzy counts will interact with the model, the initial model will be used for the starting point again. Since WT % is the same as the first model, it is already known that Run 1.11 should be excluded. The same three-way interaction is found to be most significant in this case as well. Figure 14 shows the leverage plots for the fuzzy response variables. The fits are much better for these values compared to the traditional metrics without further need to reduce the model. The model is defined as in Eq. 4.

Response:

$$\begin{aligned} R_{fghijk} = {} & \mu + Lat_i^2 + Lat_i + Long_j^2 + Long_j + Lat_i Long_j \\ & + Like_k + Lat_i Like_k + Long_i Like_k + FTR_f Lat_i + FTR_f Long_j \\ & + FTR_f Like_k + GAB_g Lat_i + GAB_g Long_j + GAB_g Like_k \\ & + CGS_h Lat_i + CGS_h Long_j + CGS_h Like_k + FTR_f GAB_g \\ & + FTR_f CGS_h + GAB_g CGS_h + FTR_f + GAB_g + CGS_h \\ & + CGS_h Lat_i Like_k + \varepsilon_{n(fghijk)} \end{aligned} \qquad \textbf{Eq. 4}$$

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, 1.5, 2.5
$Long_j$ = longitudinal conformance bounds in nautical miles, j = 0.5, 1.0, 1.5
$Like_k$ = likelihood, k = "10/20", "3/8/10", "3/8/20"
$\varepsilon_{n(fghijk)}$ = random error, n = 1, 2, … for all f, g, h, i, j, k

In the fuzzy model, it is noticed that the interaction of the lateral and longitudinal conformance bounds must now be included in the model.

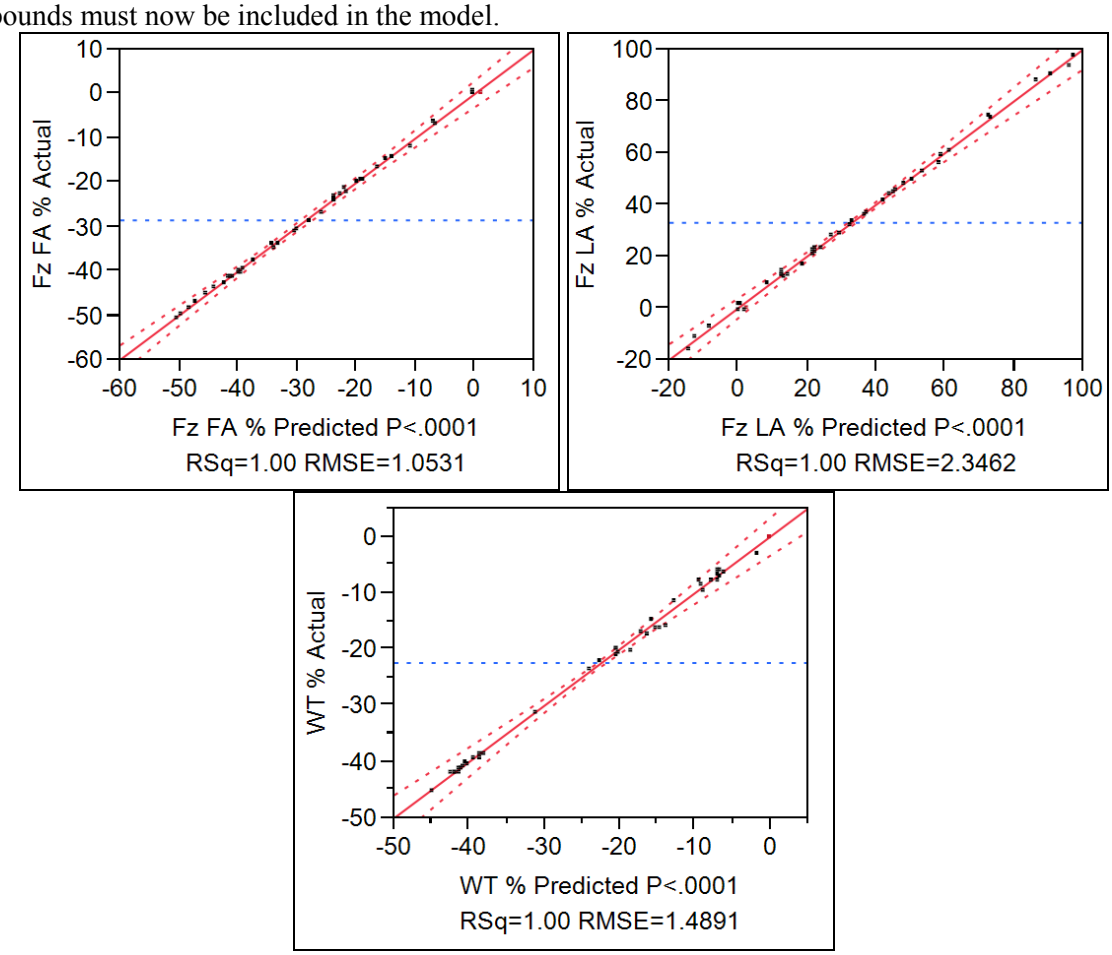

Figure 14. Fuzzy FA % and LA % metrics and WT % metric leverage plots.

Table 22 and Table 23 give the effects tests for the fuzzy responses, and Table 20 can be used again for the WT % response. These tables suggest that this model is already as refined as possible.

Table 22. Initial Model Effect Test for the Fuzzy FA % Response Variable.

Source	DF	Sum of Squares	F Ratio	Prob > F
FTR	1	68.8956	62.1274	<0.0001
GAB	1	203.3366	183.3611	<0.0001
CGS	1	320.2065	288.7498	<0.0001
Lat(0.5,2.5)	1	655.1947	590.8291	<0.0001
Long(0.5,1.5)	1	68.2935	61.5844	<0.0001
Likelihood	2	2948.4760	1329.4100	<0.0001
FTR*GAB	1	1.0251	0.9244	0.3590
FTR*CGS	1	1.6064	1.4486	0.2565
FTR*Lat	1	23.7657	21.4310	0.0009
FTR*Long	1	32.3985	29.2157	0.0003
FTR*Likelihood	2	21.8357	9.8453	0.0043
GAB*CGS	1	2.2613	2.0391	0.1838
GAB*Lat	1	52.8315	47.6413	<0.0001
GAB*Long	1	2.1096	1.9024	0.1979
GAB*Likelihood	2	6.8846	3.1041	0.0894
CGS*Lat	1	114.4165	103.1764	<0.0001
CGS*Long	1	5.5145	4.9727	0.0498
CGS*Likelihood	2	20.0013	9.0182	0.0058
Lat*Long	1	32.3584	29.1795	0.0003
Lat*Likelihood	2	5.5237	2.4905	0.1325
Long*Likelihood	2	35.0411	15.7993	0.0008
Lat*Lat	1	0.1223	0.1103	0.7466
Long*Long	1	16.5598	14.9330	0.0031
Lat*Likelihood*CGS	2	6.4397	2.9035	0.1013

Table 23. Initial Model Effect Test for the Fuzzy LA % Response Variable.

Source	DF	Sum of Squares	F Ratio	Prob > F
FTR	1	4419.3602	802.8399	<0.0001
GAB	1	53.5828	9.7341	0.0109
CGS	1	5094.4590	925.4813	<0.0001
Lat(0.5,2.5)	1	1247.3770	226.6039	<0.0001
Long(0.5,1.5)	1	2402.7561	436.4950	<0.0001
Likelihood	2	465.5771	42.2894	<0.0001
FTR*GAB	1	110.5316	20.0796	0.0012
FTR*CGS	1	4.0359	0.7332	0.4119
FTR*Lat	1	9.7673	1.7744	0.2124
FTR*Long	1	22.8521	4.1514	0.0689
FTR*Likelihood	2	410.2031	37.2596	<0.0001
GAB*CGS	1	12.7704	2.3199	0.1587
GAB*Lat	1	40.5717	7.3704	0.0218
GAB*Long	1	8.2858	1.5052	0.2480
GAB*Likelihood	2	6.3833	0.5798	0.5778
CGS*Lat	1	174.0261	31.6143	0.0002
CGS*Long	1	1.9060	0.3462	0.5693
CGS*Likelihood	2	141.1467	12.8207	0.0017
Lat*Long	1	826.1702	150.0856	<0.0001
Lat*Likelihood	2	0.6606	0.0600	0.9421
Long*Likelihood	2	51.0126	4.6336	0.0377
Lat*Lat	1	1116.2831	202.7888	<0.0001
Long*Long	1	297.4725	54.0401	<0.0001
Lat*Likelihood*CGS	2	23.2103	2.1082	0.1722

Again, the random error is tested for normalcy, where it is already known from Figure 13 that the residual error for the WT % response has evidence for normalcy.

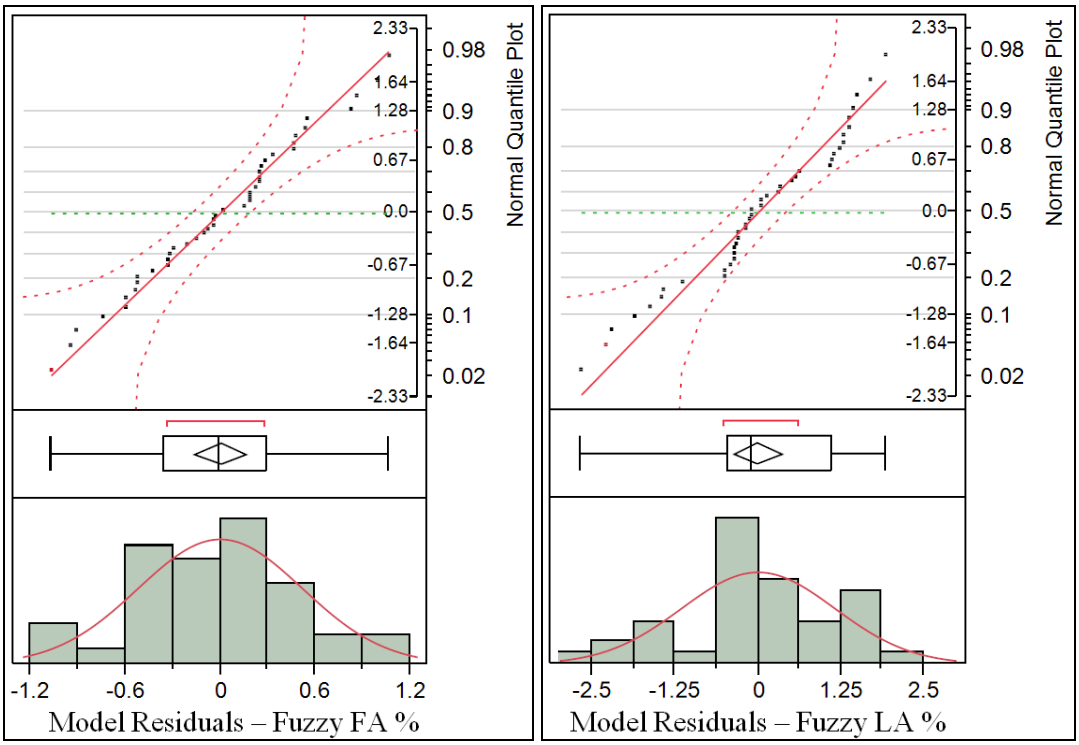

Figure 15. Residual Error Distributions for the Fuzzy Response Variables from the Model.

Figure 15 shows a nice distribution for the Fuzzy FA %, though the Fuzzy LA % is not as nicely distributed; in fact it barely passes a Shapiro-Wilk Test for Normalcy with a *p*-value of 0.08. This suggests that although the model fits the data very well for the Fuzzy LA %, not all of the error is being explained by the model.

Table 24 summarizes the *p*-values for the significance of interaction effect tests on the factors of the fuzzy model. All effects that have at least one *p*-value less than 0.05 are highlighted. The CGS interactions with FTR and GAB do not have a large effect on the system, so in hopes of improving the normalcy test, they are removed.

Table 24. Summary of Fuzzy Model Effect Tests for Fuzzy Response Variables.

Source	Fz FA % P-Value	Fz LA % P-Value	WT % P-Value
FTR	<0.0001	<0.0001	0.0349
GAB	<0.0001	0.0109	0.0657
CGS	<0.0001	<0.0001	<0.0001
Lat(0.5,2.5)	<0.0001	<0.0001	<0.0001
Long(0.5,1.5)	<0.0001	<0.0001	<0.0001
Likelihood	<0.0001	<0.0001	<0.0001
FTR*GAB	0.3590	0.0012	0.7561
FTR*CGS	0.2565	0.4119	0.8638
FTR*Lat	0.0009	0.2124	0.7202
FTR*Long	0.0003	0.0689	0.0670
FTR*Likelihood	0.0043	<0.0001	0.2980
GAB*CGS	0.1838	0.1587	0.9431
GAB*Lat	<0.0001	0.0218	0.8866
GAB*Long	0.1979	0.2480	0.2135
GAB*Likelihood	0.0894	0.5778	0.0637
CGS*Lat	<0.0001	0.0002	<0.0001
CGS*Long	0.0498	0.5693	0.4185
CGS*Likelihood	0.0058	0.0017	0.0076
Lat*Long	0.0003	<0.0001	0.1718
Lat*Likelihood	0.1325	0.9421	0.0033
Long*Likelihood	0.0008	0.0377	0.3934
Lat*Lat	0.7466	<0.0001	0.8931
Long*Long	0.0031	<0.0001	0.0099
Lat*Likelihood*CGS	0.1013	0.1722	0.0370

The refined fuzzy model is defined as in Eq. 5.

Response:

$$\begin{aligned} R_{fghijk} &= \mu + Lat_i^2 + Lat_i + Long_j^2 + Long_j + Lat_i Long_j \\ &+ Like_k + Lat_i Like_k + Long_i Like_k + FTR_f Lat_i + FTR_f Long_j \\ &+ FTR_f Like_k + GAB_g Lat_i + GAB_g Long_j + GAB_g Like_k \\ &+ CGS_h Lat_i + CGS_h Long_j + CGS_h Like_k + FTR_f GAB_g \\ &+ FTR_f + GAB_g + CGS_h + CGS_h Lat_i Like_k + \varepsilon_{n(fghijk)} \end{aligned} \qquad \textbf{Eq. 5}$$

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, 1.5, 2.5
$Long_j$ = longitudinal conformance bounds in nautical miles, j = 0.5, 1.0, 1.5
$Like_k$ = likelihood, k = "10/20", "3/8/10", "3/8/20"
$\varepsilon_{n(fghijk)}$ = random error, n = 1, 2, … for all f, g, h, i, j, k

The leverage plots are now given in Figure 16. There is little change to the fit of the model.

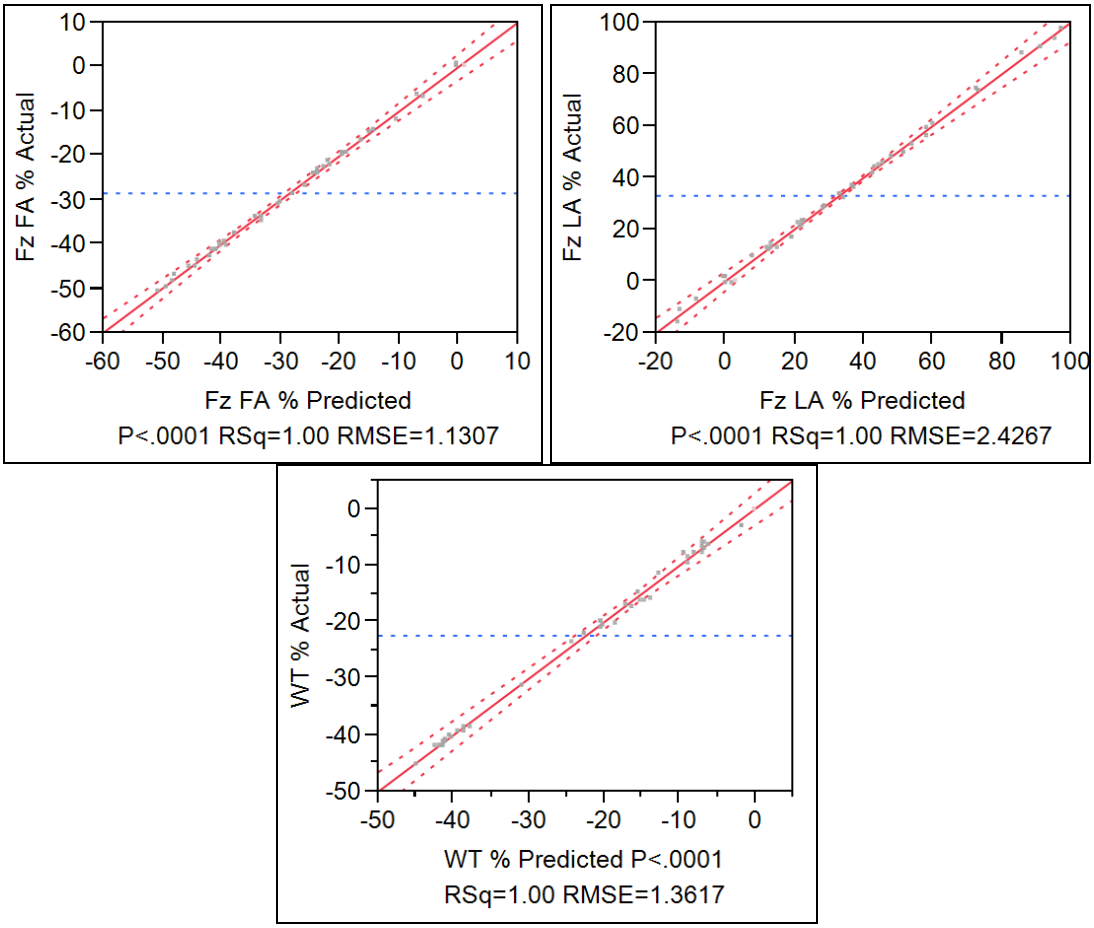

Figure 16. Fuzzy FA % and LA % metrics and WT % metric refined leverage plots.

Table 25 summarizes the *p*-values for the significance of interaction effect tests on the factors of the refined fuzzy model. All effects that have at least one *p*-value less than 0.05 are highlighted. The three interactions that are not highlighted cannot be removed without severely impacting the quality of the Fuzzy LA % response, so they are left alone. Figure 17 gives the normalcy test for this refined model, where the performance is much better. Now the Fuzzy LA % response passes the Shapiro-Wilk Normalcy Test with a *p*-value of 0.39, which is a big improvement from the previous version of the model.

Table 25. Summary of Refined Fuzzy Model Effect Tests for Fuzzy Response Variables.

Source	Fz FA % P-Value	Fz LA % P-Value	WT % P-Value
FTR	<0.0001	<0.0001	0.0185
GAB	<0.0001	<0.0001	0.0437
CGS	<0.0001	<0.0001	<0.0001
Lat(0.5,2.5)	<0.0001	<0.0001	<0.0001
Long(0.5,1.5)	<0.0001	<0.0001	<0.0001
Likelihood	<0.0001	<0.0001	<0.0001
FTR*GAB	0.7923	0.7923	0.7287
FTR*Lat	0.0002	0.0002	0.6982
FTR*Long	0.0003	0.0003	0.0421
FTR*Likelihood	0.0041	0.0041	0.2300
GAB*Lat	<0.0001	<0.0001	0.8289
GAB*Long	0.2186	0.2186	0.1610
GAB*Likelihood	0.1259	0.1259	0.0373
CGS*Lat	<0.0001	<0.0001	<0.0001
CGS*Long	0.0777	0.0777	0.3738
CGS*Likelihood	0.0081	0.0081	0.0029
Lat*Long	0.0003	0.0003	0.1333
Lat*Likelihood	0.1346	0.1346	0.0010
Long*Likelihood	0.0010	0.0010	0.3265
Lat*Lat	0.5776	0.5776	0.9163
Long*Long	0.0031	0.0031	0.0185
Lat*Likelihood*CGS	0.1309	0.1309	0.0437

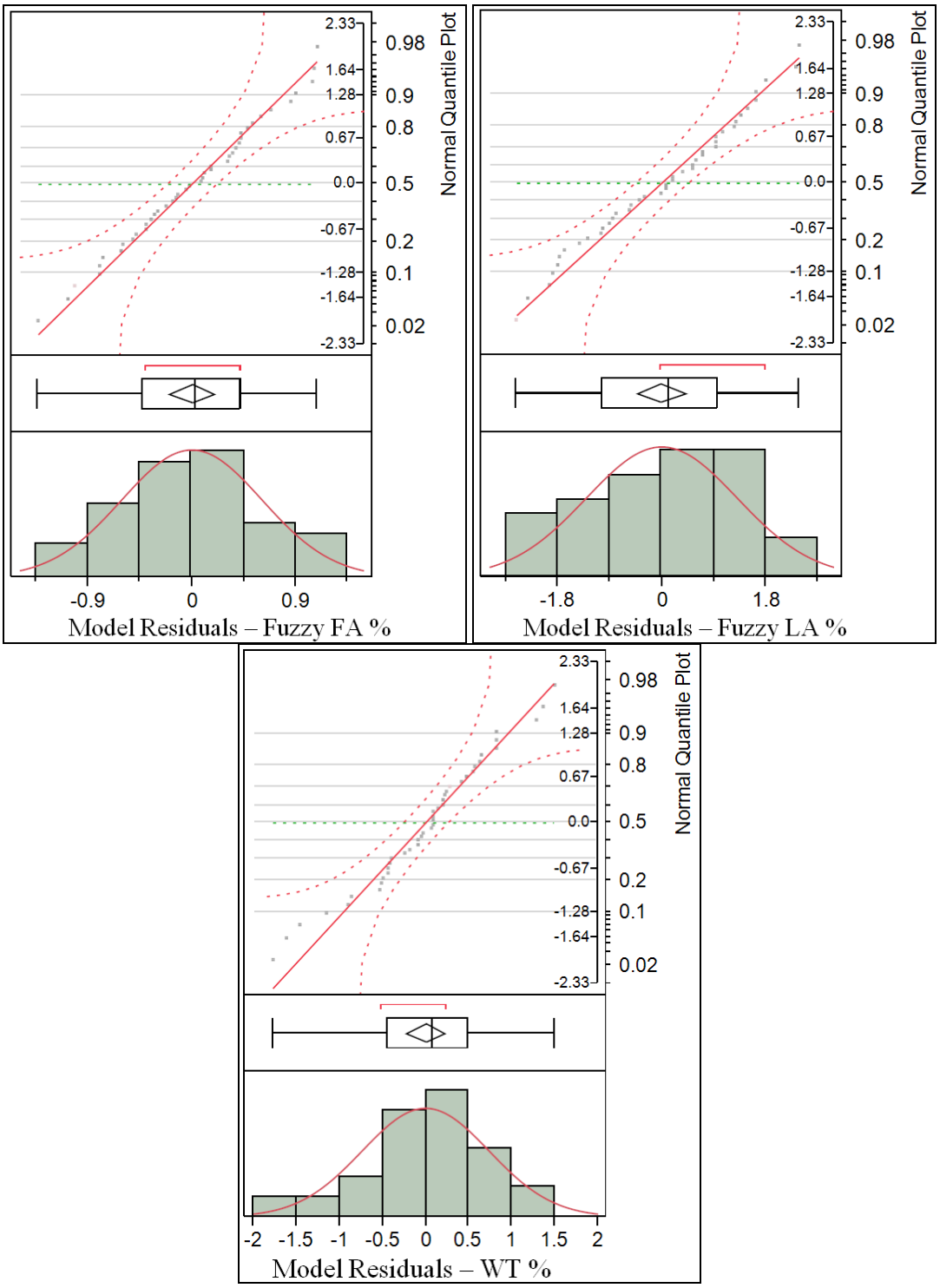

Figure 17. Residual Error Distributions for the Fuzzy Response Variables from the Refined Model.

3.2.3.2 Examine Model Results

Figure 18 shows the prediction profiler[‡] of the model set to the baseline settings. This shows that the model does a good job at predicting the results of the baseline. It is off by -0.23% on the FA performance, 2.75% on LA performance, and 0.08% on warning time performance. The slightly larger error in the LA is expected because of the smaller sample sizes.

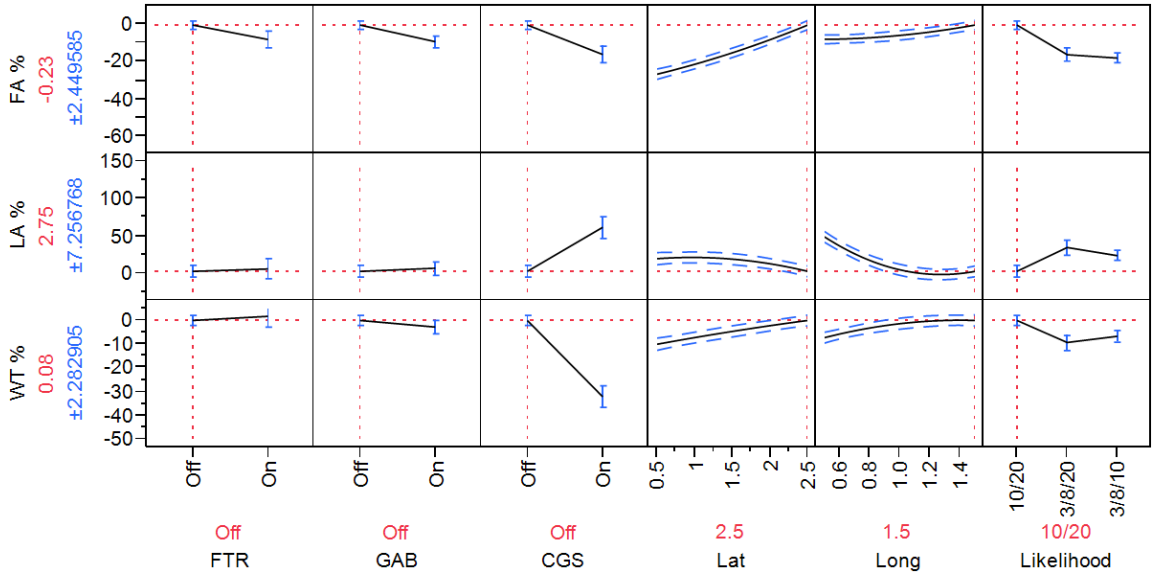

Figure 18. Prediction profiler of the model set to the settings of the baseline run.

With the baseline settings, it seems that FTR and GAB have the same effects on FA and LA performance when turned on, but FTR slightly increases warning time, whereas GAB slightly decreases it. In order to explore this further, we will set the lateral, longitudinal, and likelihood to the settings used in Experiment 1.

Figure 19 shows the performance if we were to turn on all prototype enhancements without modifying any parameters. False Alert performance is significantly improved, reducing it to -30.49%, but LA performance is significantly degraded, increasing to 69.7%. Consistent with the LA performance, warning time is significantly reduced to -34.48%. Though a significant improvement in FA performance is gained, it is clear that the prototype enhancements cannot simply be turned on without adversely affecting the performance of the probe on Late Alerts.

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

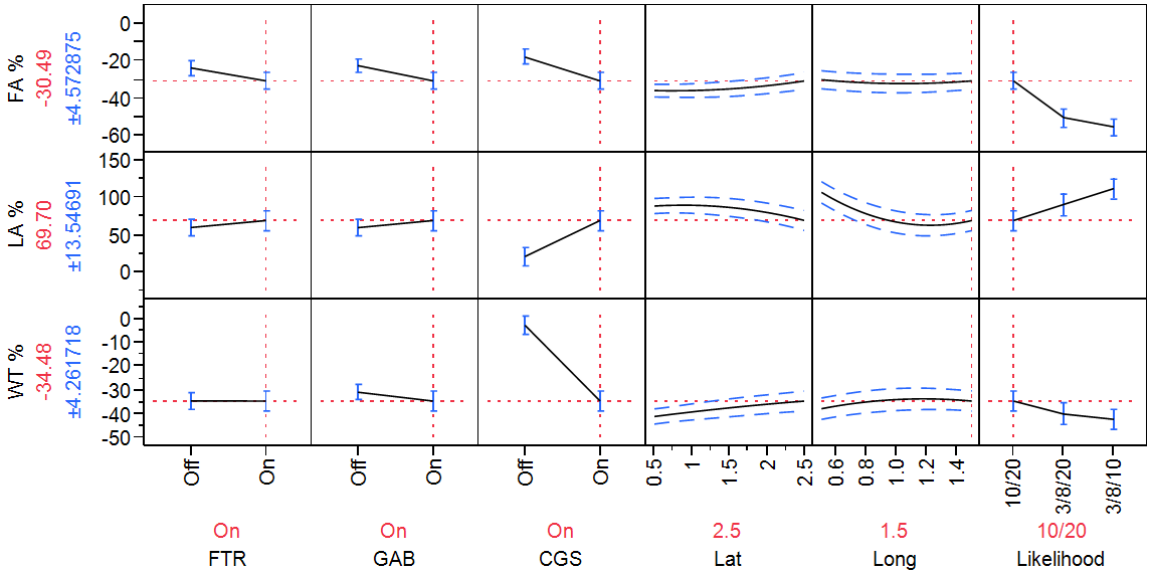

Figure 19. Prediction profiler set to the baseline settings for parameters but with all three prototype enhancements turned on.

Figure 20 shows the prediction profiler with the settings for the optimal LA performance from Experiment 1. There is not much difference in these settings from the baseline. Only the longitudinal is changed reducing it to a setting of 1.28 nm. FA and LA performance are improved slightly, decreasing them from -0.23% and 2.75% respectively to -3.11% and -1.51%. Warning time is reduced slightly from 0.08% to 0.03%. The LA performance optimal settings for Experiment 2 were determined to be the same as Experiment 1.

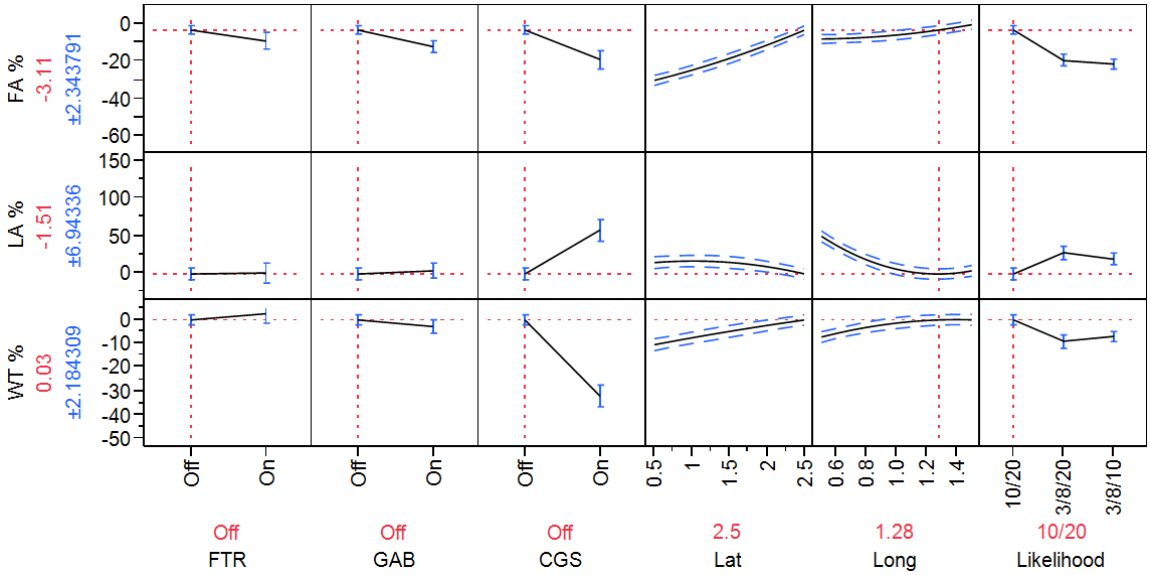

Figure 20. Prediction profiler set to the optimal settings to reduce Late Alerts according to Experiments 1 and 2.

Figure 21 shows what were determined to be the optimal settings for reducing False Alerts in Experiment 1. The resulting numbers are similar to those found in Experiment 1. Interestingly, at these parameter settings, the effects of the prototype have changed significantly. The effects of all three prototypes on FA % seem almost negligible in contrast to the previous figures. The GAB

prototype no longer causes an increase in LA % when turned on, and actually reduces it slightly. Likewise, the warning time is nearly unaffected by either the FTR or GAB prototypes. The CGS prototype still has an adverse effect on LA and warning time performance, though the LA performance degradation may be due to the issue with the minimum warning time discussed in Section 3.2.2.3.

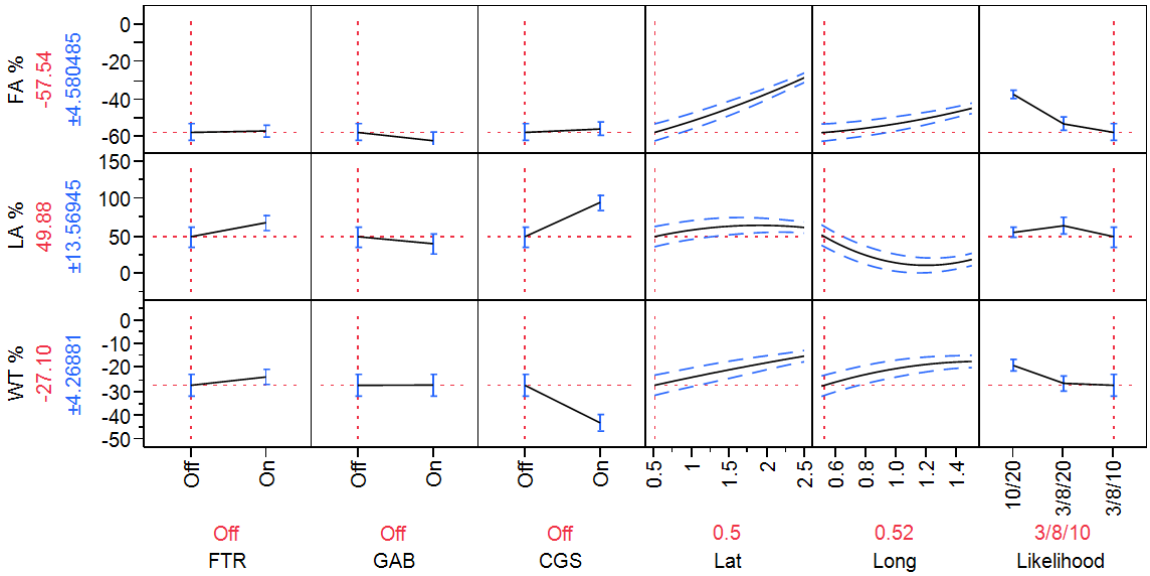

Figure 21. Prediction profiler set to the optimal settings to reduce False Alerts according to Experiment 1.

In contrast, Figure 22 shows the optimal settings for reducing False Alerts according to the Experiment 2 model. False Alert performance is improved more to -62.1% as opposed to the -57.54% in Experiment 1. The most interesting change is that, though LA performance is still adversely affected, it is better than in the Experiment 1 results. LA performance is 42.77% instead of the 49.88% previously. Warning time is nearly unchanged from -27.1% to -27.39%.

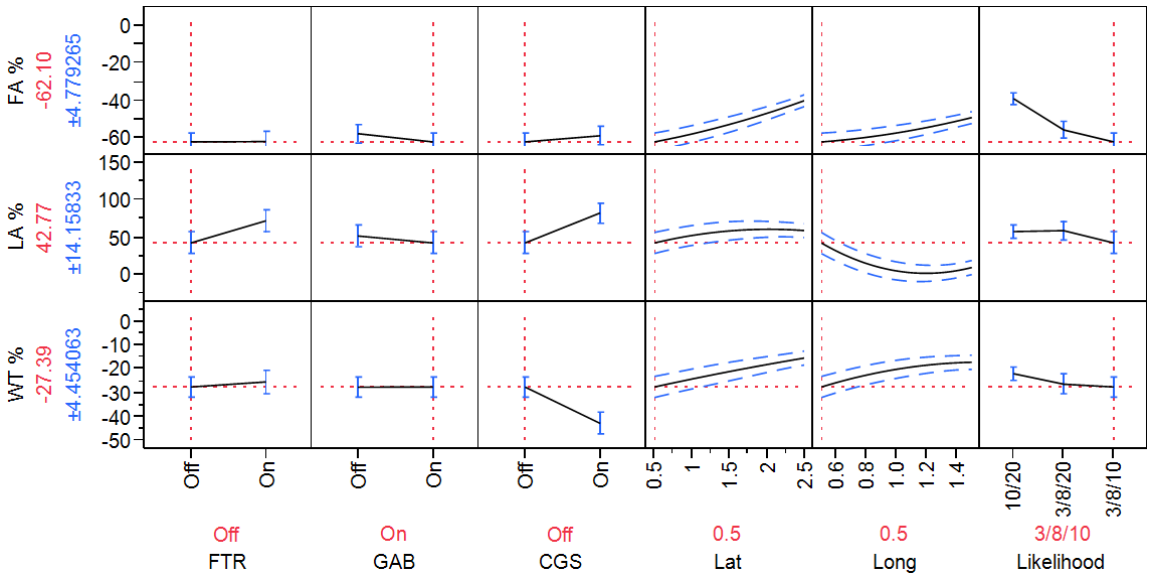

Figure 22. Prediction profiler set to the optimal settings to reduce False Alerts according to Experiment 2.

We know from Experiment 1 that longitudinal should not be lowered much further than 1.25 nm, and decreasing it does not have a large positive effect on FA performance but does have a large adverse effect on LA performance.

Figure 23 shows the results when the longitudinal is changed to a value of 1.2 nm, which provides the best LA performance when no other factors are changed from the previous settings. The FA performance is only degraded slightly from -62.1% to -54.38%, but the LA performance is greatly improved. LA % is reduced from 42.77% to 2.06% and warning time is increased from -27.39% to -18.19%.

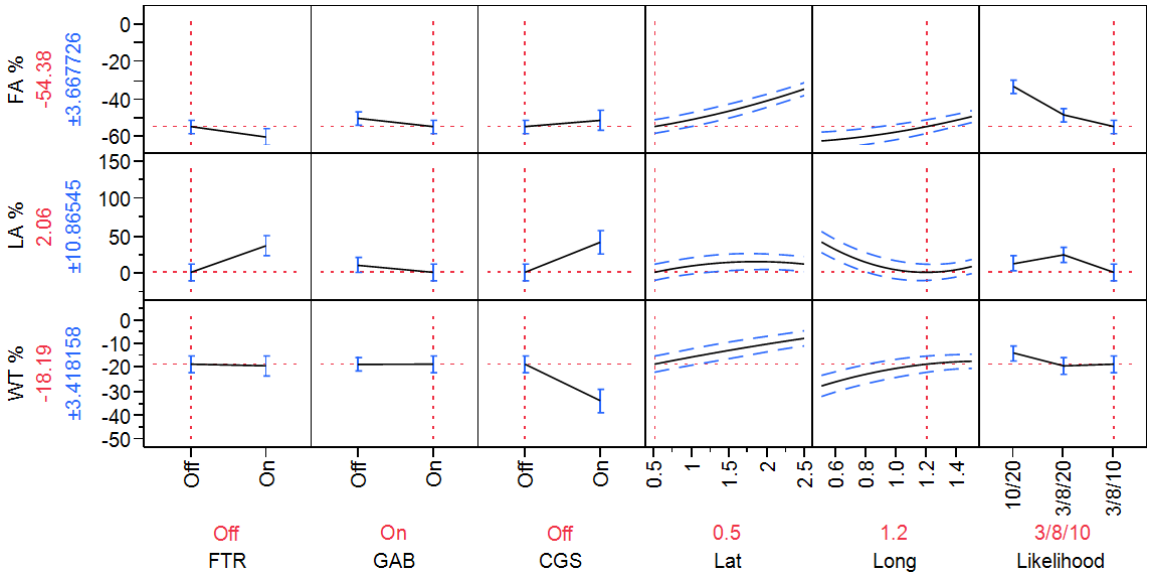

Figure 23. Prediction profiler set to one possible combination of the optimal settings.

The 2.06% increase to Late Alerts represents less than one Late Alert count and is statistically insignificant. The prediction profiler indicates that there is indeed a setting at which the Growth Adherence Bound prototype can provide some improvement to performance. The slope of the line indicates that it is a relatively small improvement.

Figure 24 shows the results of turning off GAB, with no other changes in the settings. The FA improvement drops from -54.38% to -49.95%, a difference of 4.43%, which is a significant change. More importantly, the Late Alert performance is degraded to 11.47%, which is a significant increase from the baseline. Warning time performance is nearly unchanged with a value of -18.28%. This figure indicates that the GAB prototype not only improves the FA performance, but can also improve the LA performance at certain parameter settings.

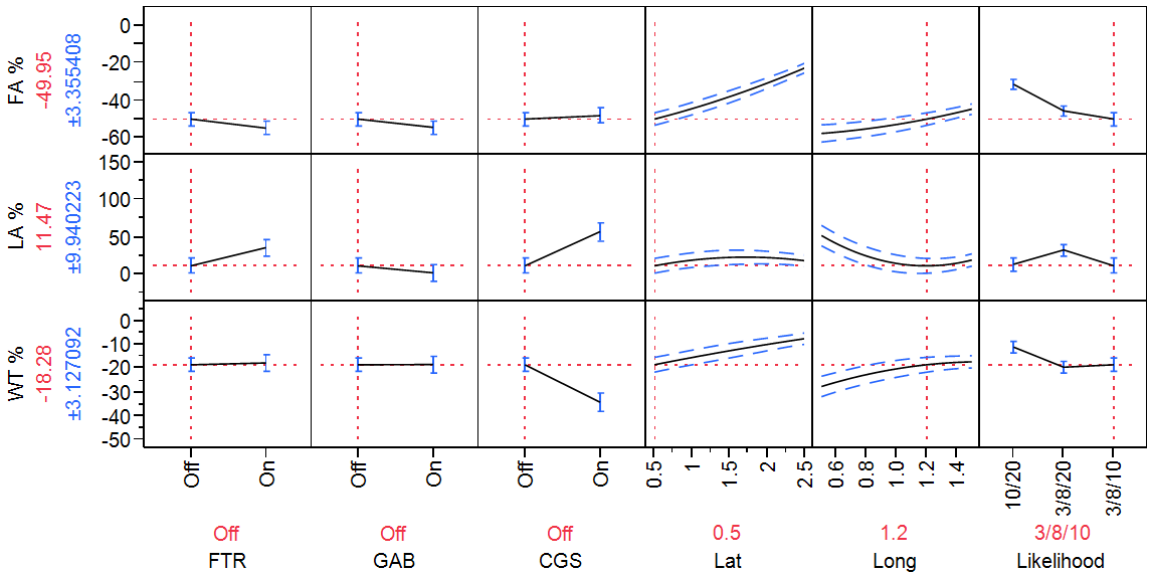

Figure 24. Prediction profiler with settings to greatly improve FA performance and only slightly adversely affect LA performance.

It is clear from the previous figures of the prediction profiler that there are some interactions between the factors. In order to better understand these, we can observe the interaction profiles for each of the response variables. In Figure 25, several significant interactions can be seen. The most significant interactions are the interaction between lateral and CGS, and the interaction between lateral and GAB. In both cases, reducing the lateral conformance bound greatly reduces the effect of the prototype enhancement on reducing the False Alerts. This is an expected interaction because the lateral conformance bound has the largest effect of all the factors on FA performance. By only setting the lateral conformance bound to 0.5 nm, we can reduce the False Alerts by nearly 28%, leaving far less potential for performance increases from the rest of the factors. Since the effect of the lateral factor flattens out as the prototypes are turned on, this indicates that reducing the lateral conformance bound is removing many of the same FAs that are removed when the prototype enhancements are turned on.

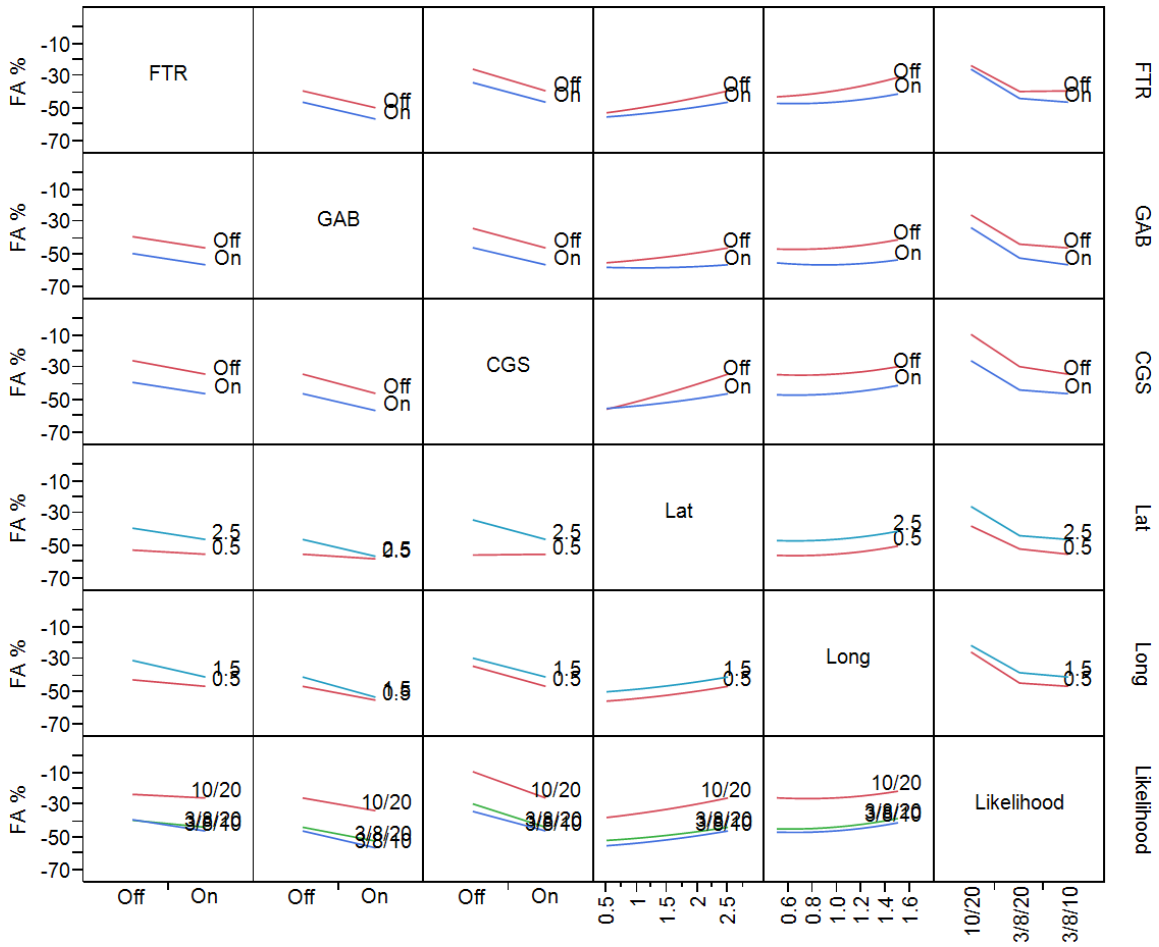

Figure 25. Interaction profile for FA % metric.

In Figure 26, the interactions of the factors in regards to the LA % metric are shown. Several interactions can be seen, all of which are related to likelihood. All of the other metrics except for longitudinal have some interaction with likelihood. The most extreme interaction is between likelihood and CGS. It is difficult to see in the figure, but 3/8/20 likelihood (the green line in the bottom row of plots) is almost a flat line whereas 10/20 and 3/8/10 have similar slopes that significantly increase when CGS is turned on. The 3/8/20 likelihood setting provides less Late Alerts than the 3/8/10 likelihood when CGS is turned on, but 3/8/10 provides less when likelihood is turned off. Unfortunately, there is no simple explanation for this interaction, partly due to the lack of understanding of the effects of likelihood. There is a plan in future work to perform an experiment that will gather more understanding of the effects of the likelihood setting.

Another significant interaction is seen between likelihood and the FTR prototype. Both 3/8/10 and 3/8/20 likelihood settings cause an increase in Late Alerts to occur when FTR is turned on. The 10/20 likelihood causes almost no change to occur when FTR is turned on.

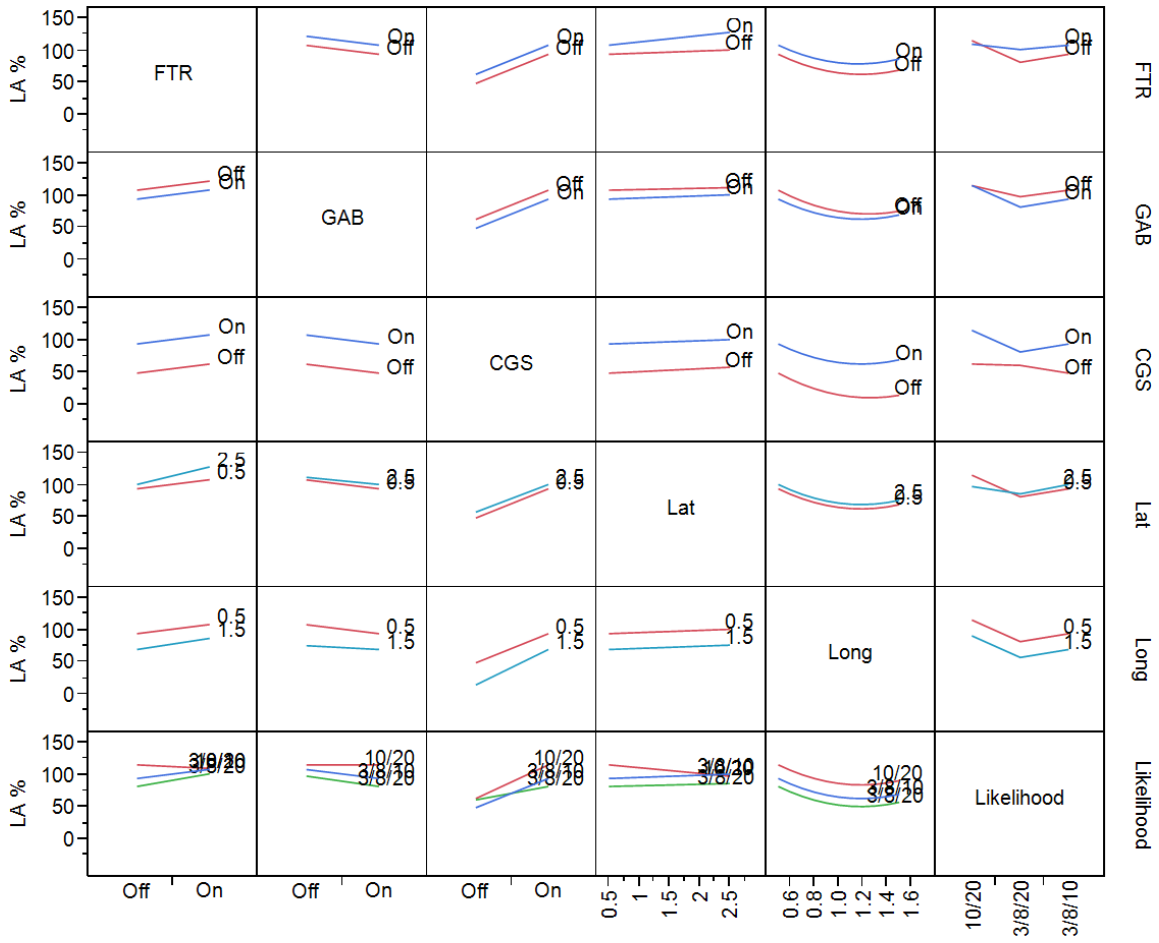

Figure 26. Interaction profile for LA % metric.

Figure 27 shows the interactions between factors for the WT % metric. There are no significant interactions for this metric. Something that is interesting in this plot is that, in comparison to CGS, none of the other factors have a significant impact on warning time performance. This indicates that most of the increase in warning time, apart from that due to CGS, is due to a combination of several settings rather than a single setting.

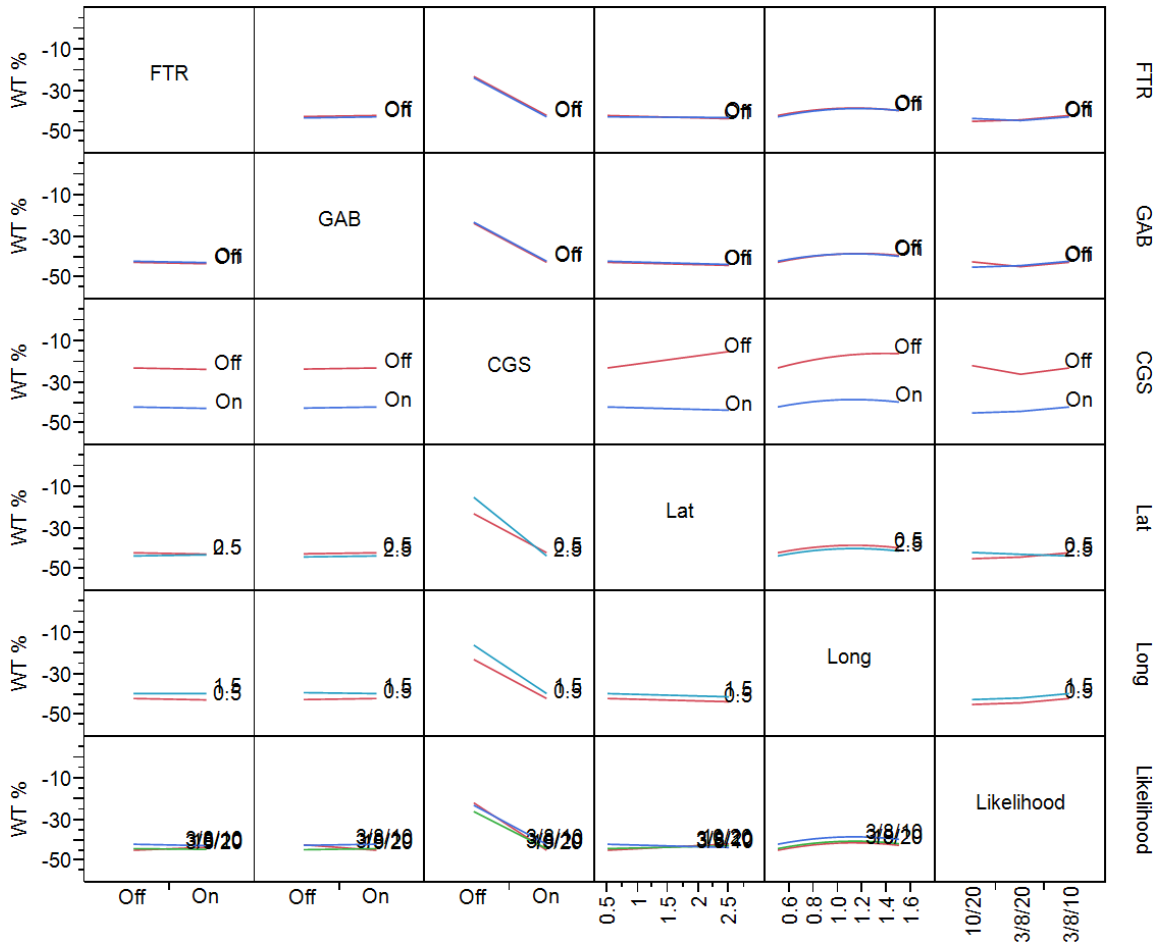

Figure 27. Interaction profile for WT % metric.

With the information gathered so far, the Growth Adherence Bounds prototype enhancement seems like it could provide the best performance gains. It can reduce the False Alerts without increasing Late Alerts or significantly reducing warning time performance. We can explore this further to determine how the lateral, longitudinal, and likelihood settings affect the performance when GAB is turned on.

From Experiment 1, as well as some previous analyses in Experiment 2, we know that reducing longitudinal below a setting of 1.25 nm increases Late Alerts, but does not significantly reduce False Alerts. So, in this analysis, we will keep the longitudinal setting around 1.25 nm. We also know that decreasing the lateral conformance bound greatly reduces the False Alerts, so we will alter that and the likelihood to find settings that could provide performance gains to the probe.

Figure 28 shows how the GAB prototype would affect the performance of the probe if it was turned on today, without any other changes to settings. A significant False Alert improvement is gained, reducing them by -9.17%. Late Alert performance is degraded by 6.94% which represents an increase of about one Late Alert and is statistically insignificant. Warning time is reduced by an insignificant -2.71%.

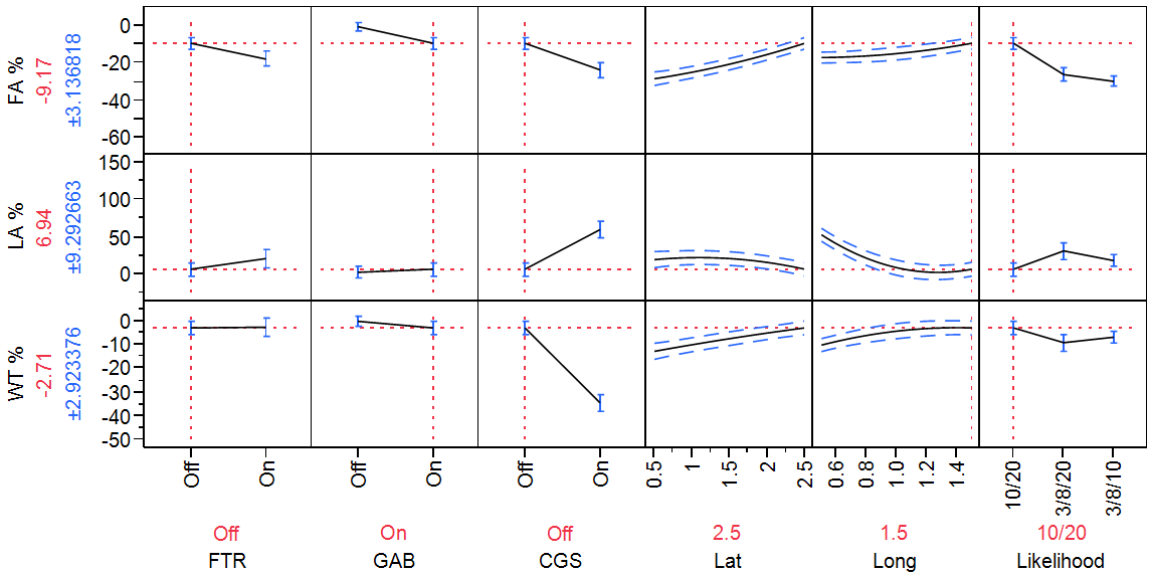

Figure 28. Baseline settings with only GAB prototype turned on.

Figure 23 already showed the best Late Alert performance when GAB is turned on and it can be contrasted with Figure 20, which shows the overall optimal settings for Late Alerts. Longitudinal is changed slightly from 1.28 nm to 1.2 nm. Lateral is set to 0.5 nm, likelihood to 3/8/10, and GAB is turned on. Late Alert performance is degraded slightly, increasing from -1.51% to 2.06%, which is a statistically insignificant increase. False Alert performance is greatly improved from -3.11 to -54.38%. Warning time performance is reduced significantly from 0.03% to -18.19%.

Figure 22 showed the overall optimal settings for reducing False Alerts, which had only the GAB prototype turned on. It was already determined that the longitudinal bound should not be set much lower than 1.25 nm. Figure 29 is a repeat of Figure 23 since these were the settings that were found to provide the best FA performance without a significant degradation to LA performance. There is still a very significant FA improvement of -54.38. LA performance is degraded slightly from the baseline by a value of 2.06%, but this number is not statistically significant, and represents less than a single Late Alert. Warning time is reduced by a value of -18.19%, which is a 25th percentile warning time value of 271.2 seconds. This is a significant decrease in warning time from the baseline's 331.5 seconds, but is still significantly greater than the 180 second warning time requirement used in this study.

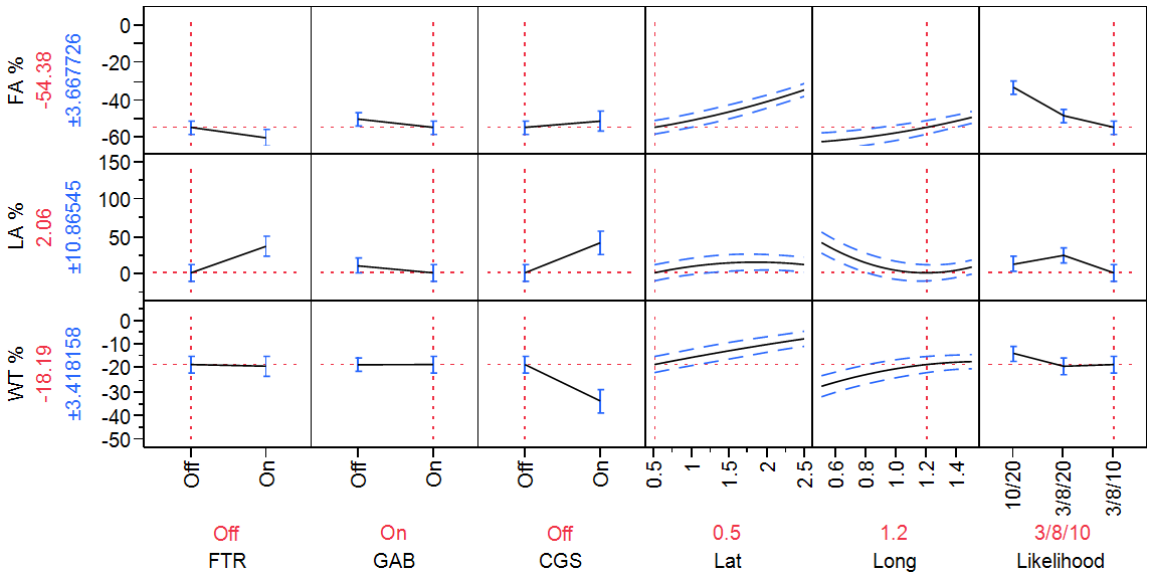

Figure 29. Settings that provide a good performance to FA with GAB turned on.

3.2.3.3 Fuzzy Model Analysis

The fuzzy metrics can also be analyzed using the model. The metrics to be used in this section are the Fuzzy FA % and Fuzzy LA %. These values are calculated by subtracting the Fuzzy FA and Fuzzy LA value of the FA32 Baseline from the respective metric value for the treatment run, then dividing by the value for the FA32 Baseline and multiplying by 100 to get a percent value. As seen in Table 10, the Fuzzy FA value for the 32BL is 949.11, and the Fuzzy LA value is 20.55. In the figures of the prediction profiler shown in this section, the WT % metric will also be included. This is the same warning time metric from the previous section, and is included to provide perspective.

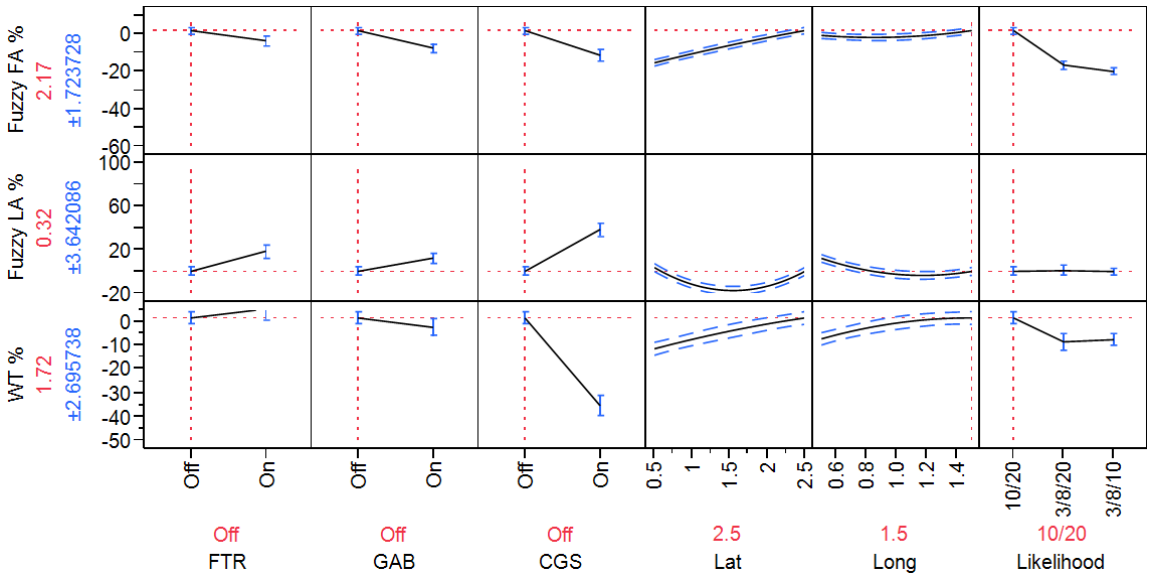

Figure 30. Fuzzy metric values of the model at baseline settings.

Figure 30 shows the model results for the fuzzy metrics at the baseline settings. The warning time is slightly different in this model than in the traditional model because of the different effects

used in this model, as discussed in Section 2.2.1.2. In this prediction profiler we can see that the models are far more precise than the models for the traditional metrics. Fuzzy FA % has a confidence interval of 1.7% and Fuzzy LA % has a confidence interval of 3.6%. Once again, because of the smaller sample size, the LA metric is expected to be less confident. However, the fuzzy metrics make use of a larger sample size than the traditional metrics, allowing the confidence interval to be much smaller than in the traditional metrics.

In Figure 31, all three prototypes are turned on at the baseline parameter settings. This result is consistent with the traditional metrics. A significant performance increase is gained for the False Alerts, but far greater performance degradation is seen in the Late Alerts, indicating this is not a viable option for the live system.

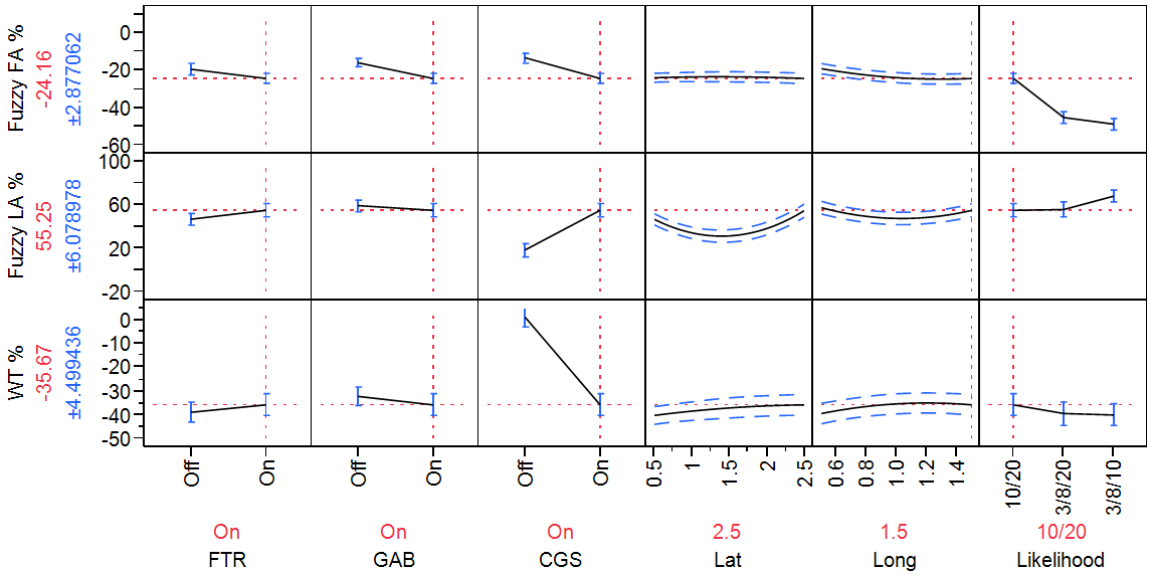

Figure 31. Fuzzy values with all prototypes turned on at the baseline settings.

Next, we can optimize the performance of the Late Alerts independently. Figure 32 shows the results of that analysis. Interestingly, the results are slightly different than the traditional metrics. The longitudinal conformance bound is set to a value of 1.36 nm, which is similar to the value of 1.28 nm determined using the traditional metrics, and is consistent with the conclusion that the longitudinal bound should be lowered slightly from its value today in order to optimize Late Alert performance.

The interesting setting is the lateral conformance bound. All analysis of the traditional metrics indicated that lowering the lateral conformance bound at all reduces the Late Alert performance slightly. The model for the Fuzzy Late Alerts indicates that the performance is much more like a parabola, increasing performance in the middle and degrading at the ends. This model is a far better model than the traditional metrics, so it is possible that this is the correct shape of the model, and the noisiness of the traditional metrics prevent it from capturing the effects properly. This shape may also be an effect of weighting the late alerts by their warning time, which would indicate that, although the Late Alert count increases slightly when lateral is reduced, the warning time of all the Late Alerts is actually increased, which in turn reduces the overall value for the Fuzzy LA. A lateral conformance bound of 1.6 nm was determined to perform best for fuzzy Late Alerts.

These settings improve the FA performance to -6.65%, which, though small in comparison with other settings, is statistically significant. The Late Alerts are reduced by -17.96%, which once again is statistically significant and represents a reduction of almost three Late Alerts.

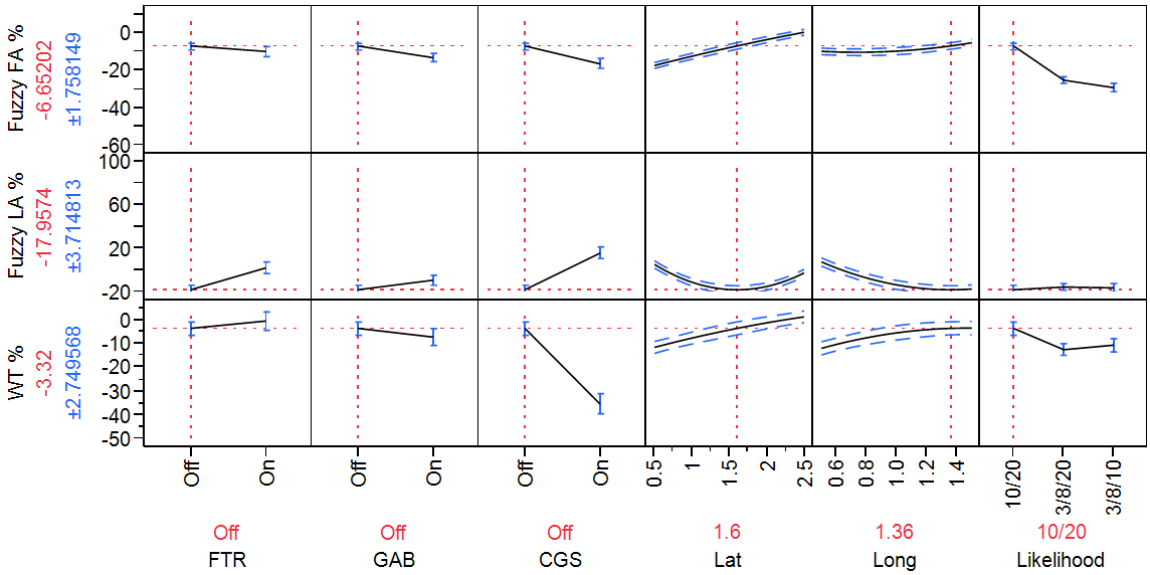

Figure 32. Fuzzy values with optimized performance for only Late Alerts.

Next, False Alert performance is optimized using the model. Figure 33 shows the results of this analysis. There are several settings that differ from the optimal for the traditional metrics. The traditional metrics had the FTR prototype disabled, since it had an insignificant effect on FA performance, but it does have a slightly significant effect in the fuzzy metrics. Also, the longitudinal is set higher than in the traditional. As can be seen in the figure, there is no significant improvement to FA performance by reducing longitudinal below a value of 0.88 nm.

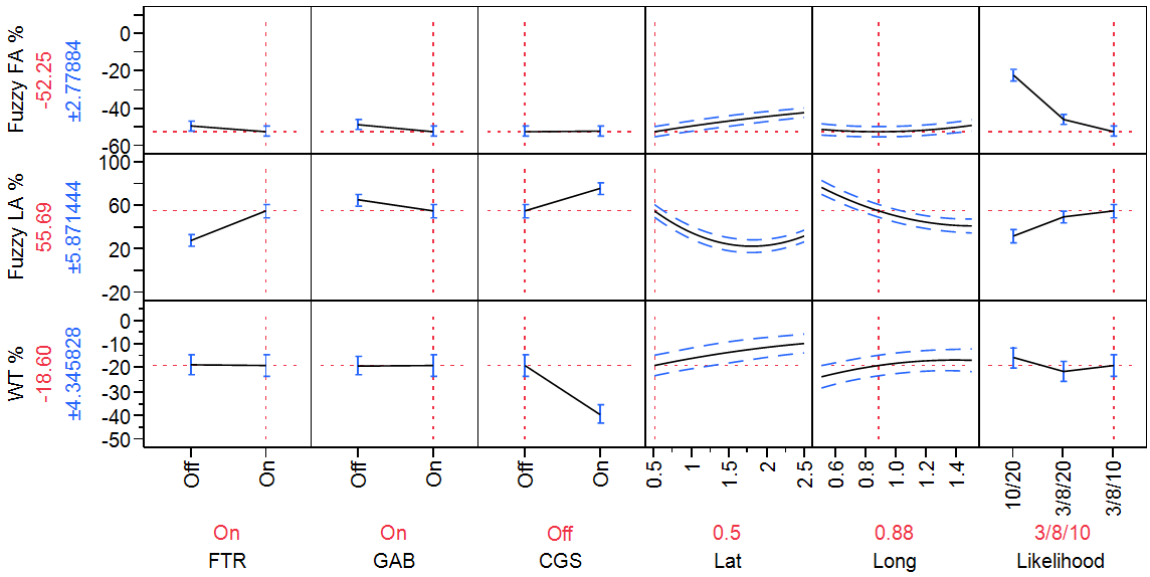

Figure 33. Fuzzy values with optimized performance for False Alerts only.

The FA performance is improved to a value of -52.25%, but LA performance is degraded to 55.69%. However, from the prediction profiler, it seems that the FTR prototype has a large

negative impact on the LA performance. It does improve FA performance, but the improvement is minimal in comparison to the LA performance degradation.

Figure 34 shows the results of turning off FTR. FA performance is only lowered slightly from -52.25% to -49.14%. This number is statistically significant, but represents only 29.5 False Alert value, which is a small price to pay for a much better LA performance. LA performance is improved from the previous settings to 28.29%, but this value still represents an increase over the baseline of 5.7 Late Alerts, which is unacceptable.

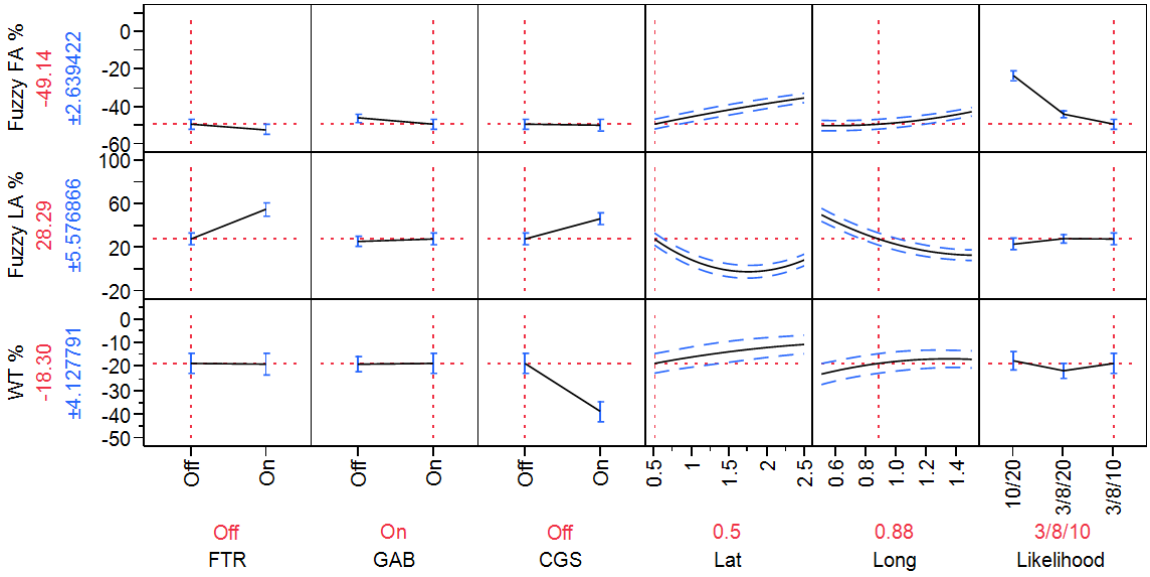

Figure 34. Fuzzy values with mostly optimal settings for FA performance, but FTR turned off to improve LA performance.

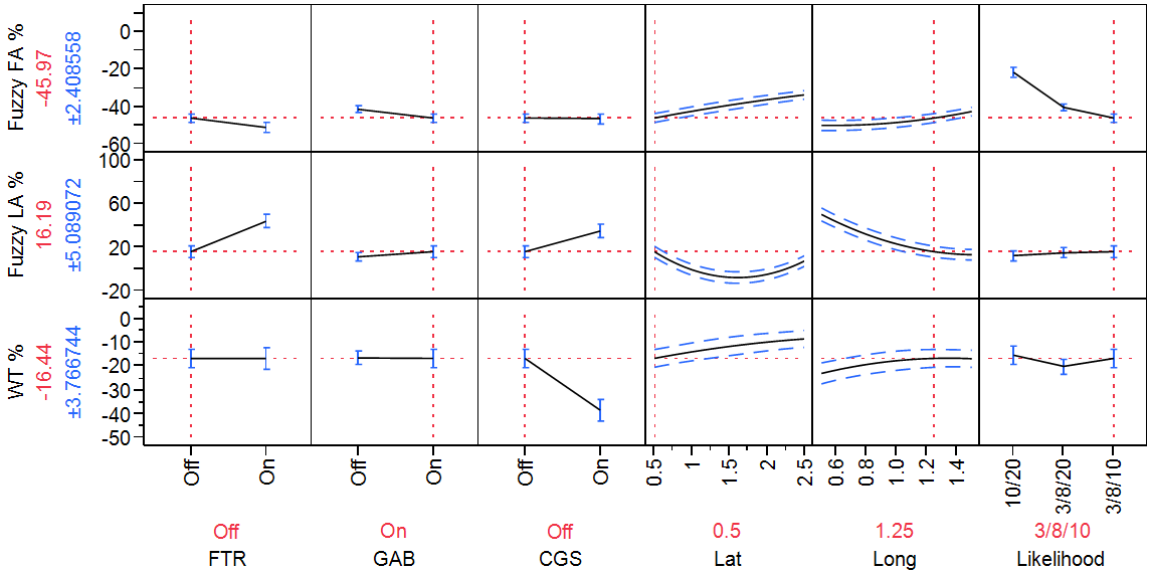

Figure 35. Fuzzy values with good FA performance and FTR off and longitudinal set to 1.25 nm to improve LA performance.

It has been determined previously that the best setting for longitudinal is around 1.25 nm. So, in Figure 35 the longitudinal conformance bound is adjusted to a value of 1.25 nm. This once again lowers the FA performance slightly from the previous settings to a value -45.97%, but it also reduces the LA performance degradation to 16.19%. This still is a significant increase in Late Alerts and is not acceptable. Better LA performance could be achieved by increasing the longitudinal bound further, but looking at the prediction profiler, it would seem that more performance can be achieved by raising the lateral setting.

Figure 36 adjusts the lateral setting to what will provide the best Late Alert performance. With a lateral setting of 1.59 nm, the Late Alert performance is improved by -7.65% over the baseline. This is a statistically significant improvement and represents a reduction of almost two Late Alerts. These settings still provide a good FA performance with a reduction of -38.49% over the baseline.

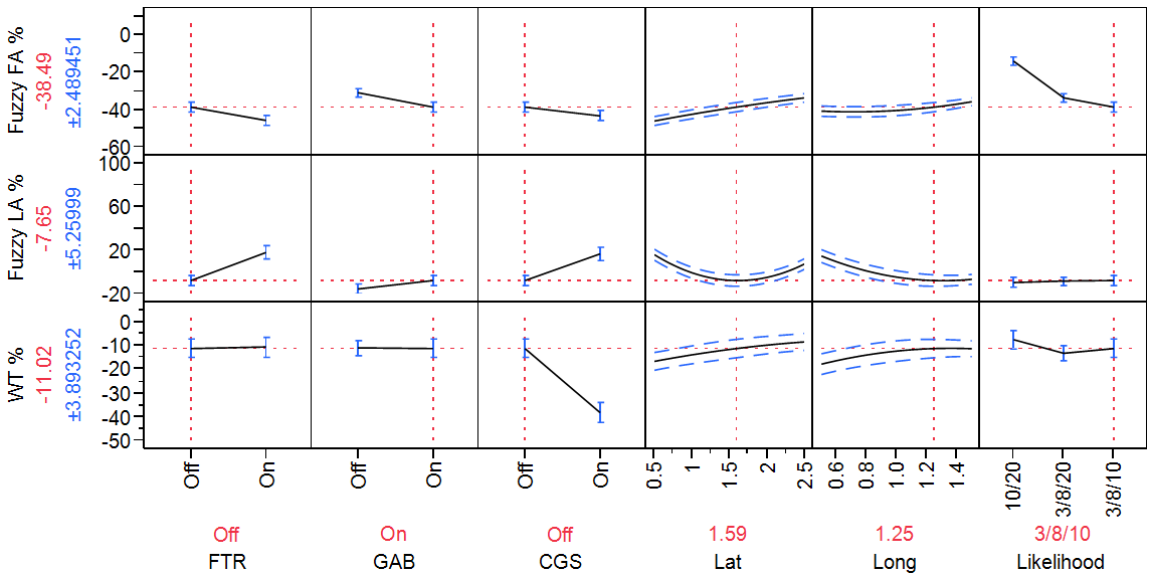

Figure 36. Fuzzy values with only the lateral adherence bound adjusted from the previous settings to achieve the best Late Alert performance.

According to the fuzzy metrics, these settings would provide a good performance of the conflict probe. However, the purpose of the prototypes being analyzed in this study is to improve FA performance while not degrading LA performance. The LA performance value of -7.65% provides room for improvement in the FA performance while sacrificing LA performance.

Figure 37 shows the best settings, according to the fuzzy metrics, for improving False Alert performance without hurting Late Alert performance. A lateral setting increased to 0.99 improves the FA performance to -42.4% while having no significant impact on LA performance. With these settings the warning time is reduced by 13.73%, which is significant. It represents a 25th percentile of warning time of 286 as opposed to the 331.5 of the baseline.

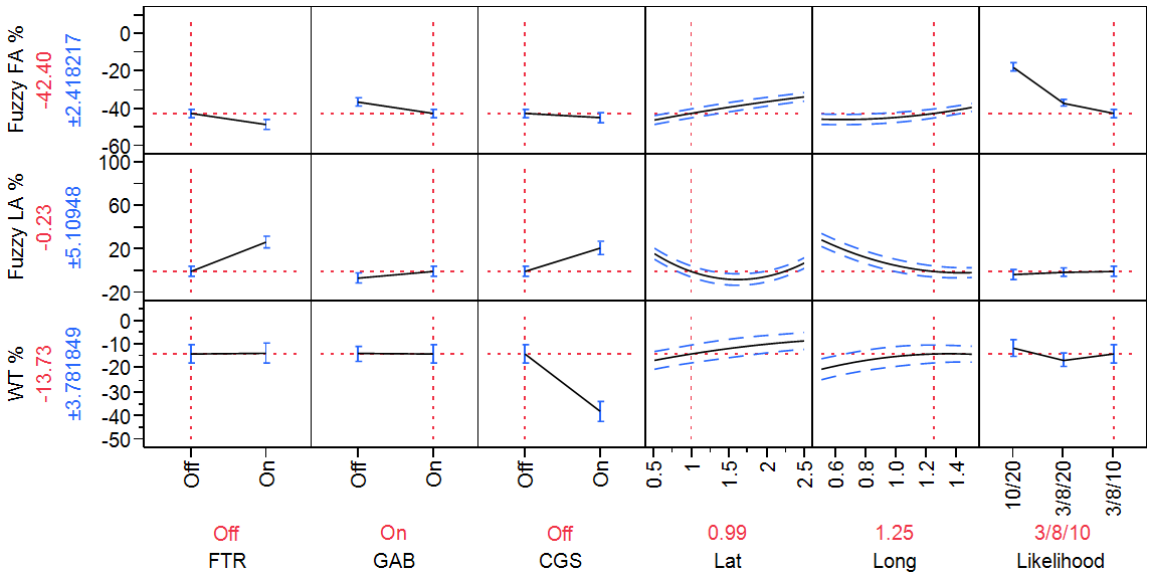

Figure 37. Settings of the probe that most reduce the False Alerts without degrading the Late Alerts, according to Fuzzy metrics.

Figure 38 shows the results of the traditional metrics for the same settings. The False Alert and warning time performances are similar to the fuzzy metrics, but the LA performance is much different. The traditional metrics have an increase of 10.74% in Late Alerts. However, there is also a confidence interval of ±10.55%. The value is only 0.19% greater than the confidence interval, so it is just barely statistically significant. The fuzzy metrics have not been proven enough to consider them instead of the traditional metrics, but the amount of noise in the model for the traditional LA metric is a powerful argument for using the results of the fuzzy metrics.

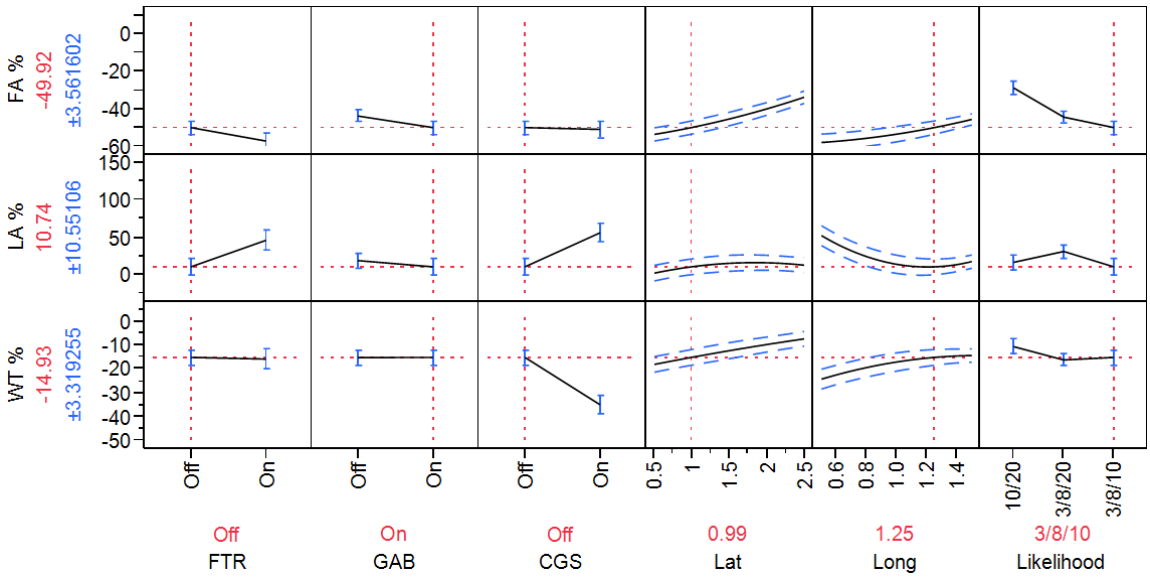

Figure 38. Traditional metrics with settings of the probe that most reduce False Alerts without degrading Late Alerts according to Fuzzy Metrics.

3.2.4 Factor Effects

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

The Forced Trajectory Rebuild (FTR) prototype mostly improves the FA performance, but the effects decrease to an insignificant improvement as the conformance bounds are reduced. At the most it can improve it by up to 12%. By only turning on the FTR and leaving all other factors at the baseline settings, the False Alerts are decreased by about 8%.

Though no negative effects of FTR on LA performance were found in the direct comparisons, the model clearly shows that FTR adversely affects the performance when 3/8/10 likelihood is used. When 10/20 likelihood is used, LA performance is statistically unchanged when FTR is turned on. FTR can increase the Late Alerts by as much as 33%.

For the most part, FTR has very little effect on the warning time performance. Though some settings do show a minor improvement in the warning time performance as a result of turning FTR on, none of these improvements are larger values than the confidence interval of the model and therefore are not statistically significant improvements.

The Growth Adherence Bounds prototype shows some improvement in performance, and may be the most promising of the prototypes. Two of the most interesting runs from the data exploration analyses had only the GAB prototype turned on. It consistently improves the FA performance across all other settings. It can improve the FA performance by as much as 12% and at the baseline settings improves the FA performance by 9%.

Late Alerts are negatively affected much less by the GAB prototype compared to the other prototypes. It can increase the Late Alerts by up to 15%, but at non-standard settings it can reduce the Late Alerts by as much as 10%. At the baseline settings, increases the Late Alerts by a statistically insignificant 4%.

Warning time is mostly unaffected by the GAB prototype, but it can be reduced by as much as 4%. At the baseline settings it is reduced by 3%, which is about the same as the confidence intervals of the model, indicating it is not significant, statistically. The 3% translates to a value of 10 seconds, which is hardly practically significant.

The CGS prototype is much harder to draw conclusions based on the analyses performed in this study. As discussed in Section 3.2.2.3, the treatment runs were run through ERAM using a setting for the CGS prototype that was later determined to be inappropriate. This setting resulted in many alerts being delayed too long, which caused them to be Late Alerts. So, although we do see a dramatic improvement of up to 15% in FA performance, LA performance is degraded by as much as 68%. Warning time is also decreased significantly, but this is expected since the prototype works by delaying warning of the alert, which decreases the warning time. It is quite obvious that the CGS prototype should not be used at the settings used in this study, but from the analysis done in Section 3.2.2.3, it appears that it does have some promise if the delay time is set greater than the minimum warning time.

3.3 Example Flights

A detailed statistical analysis has been provided in the previous sub-sections of Section 3 on the entire sample scenario of air traffic. This section is dedicated to an in depth description of a set of

selected example flights, their flight plans in the NAS, their ERAM trajectory and conflict predictions, and their accuracy results. It illustrates how the performance evaluation is being applied on these individual flights and how the prediction errors manifest themselves in the metric calculations.

3.3.1 Forced Trajectory Rebuild Example

The first prototype to test is FTR. Here it is expected that the trajectories in the FTR prototype scenarios will be more accurate than the scenarios without FTR, thus causing fewer false alerts. Runs in which only turning on FTR differs are runs E1-3 and E2-3 where conformance bounds are 0.5 by 1.5 nautical miles and likelihood is 10/20. The following example is a flight pair that was given a false alert without FTR, but a matching false alert was not generated with FTR.

Flight A (red) is an Airbus A321 with aircraft equipage /Q flying from Charlotte-Douglas International Airport (KCLT) to Bradley International Airport (KBDL). In each figure, it is moving from bottom-left to top-right in-trail to Flight B.

Flight B (green) is a Boeing 737-300 also with aircraft equipage /Q flying from KCLT to General Edward Lawrence Logan International Airport (KBOS). In each figure, it is moving from bottom-left to top-right leading Flight A.

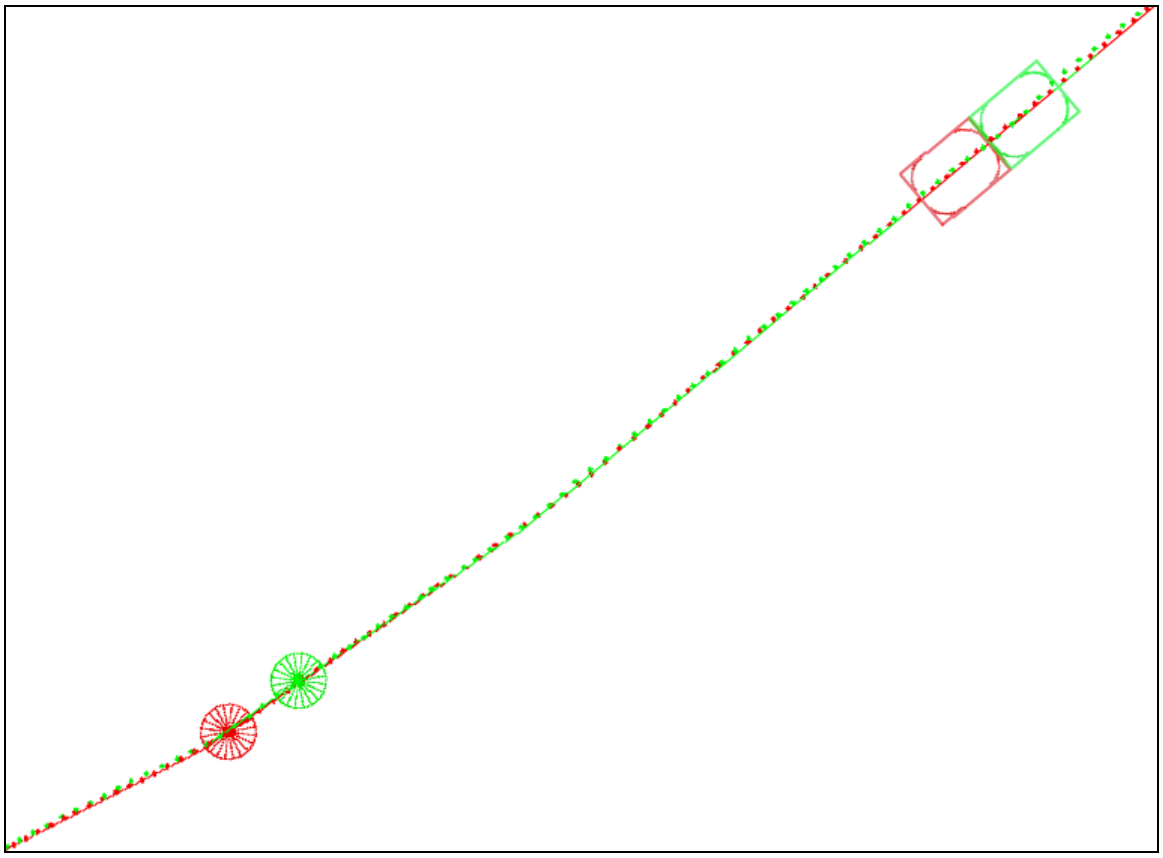

Figure 39. FTR not used in conflict prediction.

Figure 39 shows a conflict prediction being generated, later realized to be a false alert.

The accuracy, as was expected, is increased in Figure 40 so that a false alert will clearly not be generated any longer.

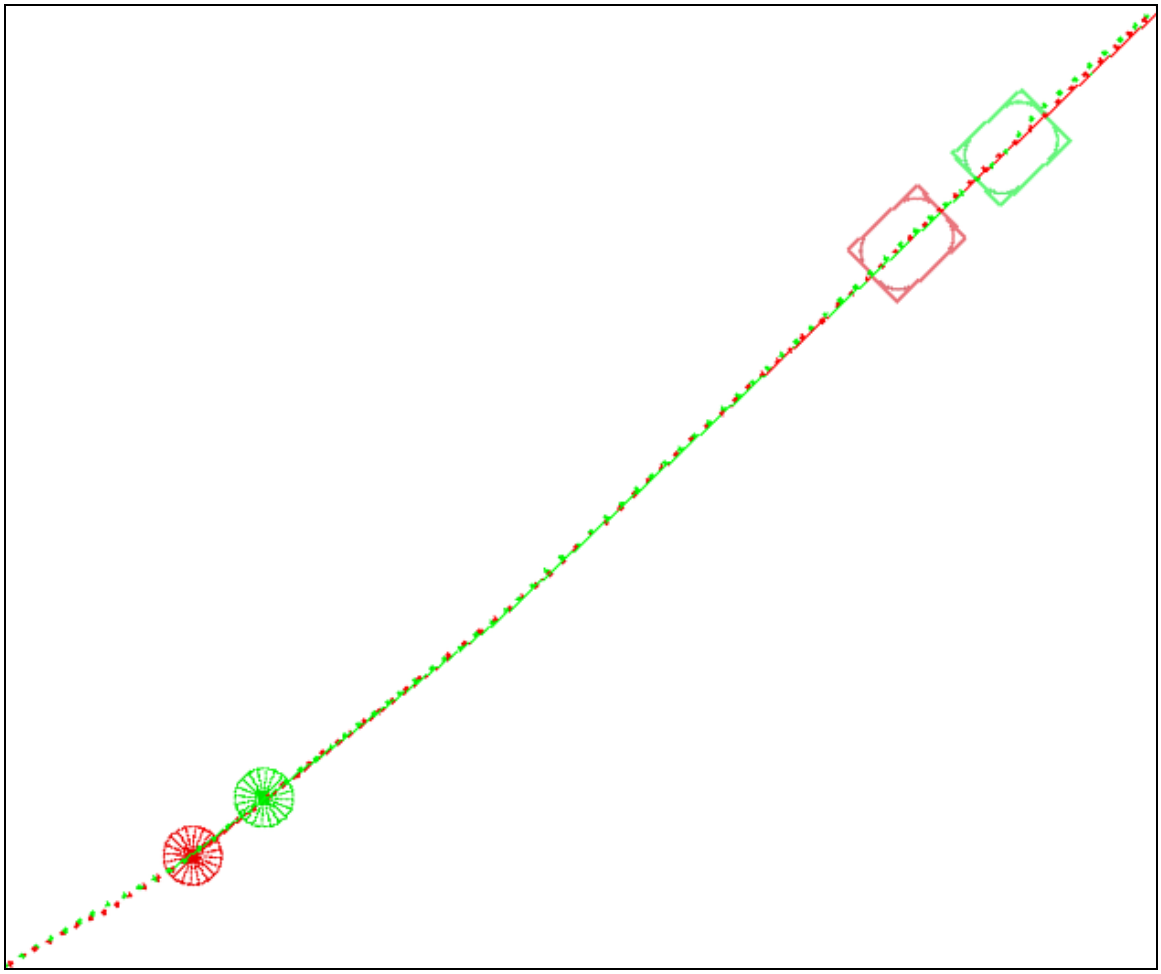

Figure 40. FTR use increases prediction accuracy.

3.3.2 Growth Adherence Bounds Example

The next prototype is GAB, which grows the adherence bounds from small values to the full adherence bounds as the trajectory ages. A run matched the baseline for all parameters except having GAB used as well (Run E2-27).

Flight A (red) is a McDonnell Douglas MD-82 equipped with RVSM (Code /W) flying from Dallas/Fort Worth International Airport (KDFW) to Philadelphia International Airport (KPHL). In each figure, it is moving towards the top-right leading Flight B.

Flight B (green) is an Airbus A320 with aircraft equipage /Q flying from Phoenix Sky Harbor International Airport (KPHX) to KPHL. In each figure, it is trailing Flight A.

The baseline detects a conflict immediately upon a trajectory rebuild, as can be seen in Figure 41. Since it is the beginning of a trajectory, GAB will have much smaller boxes than these, and the detection will not be made in the run.

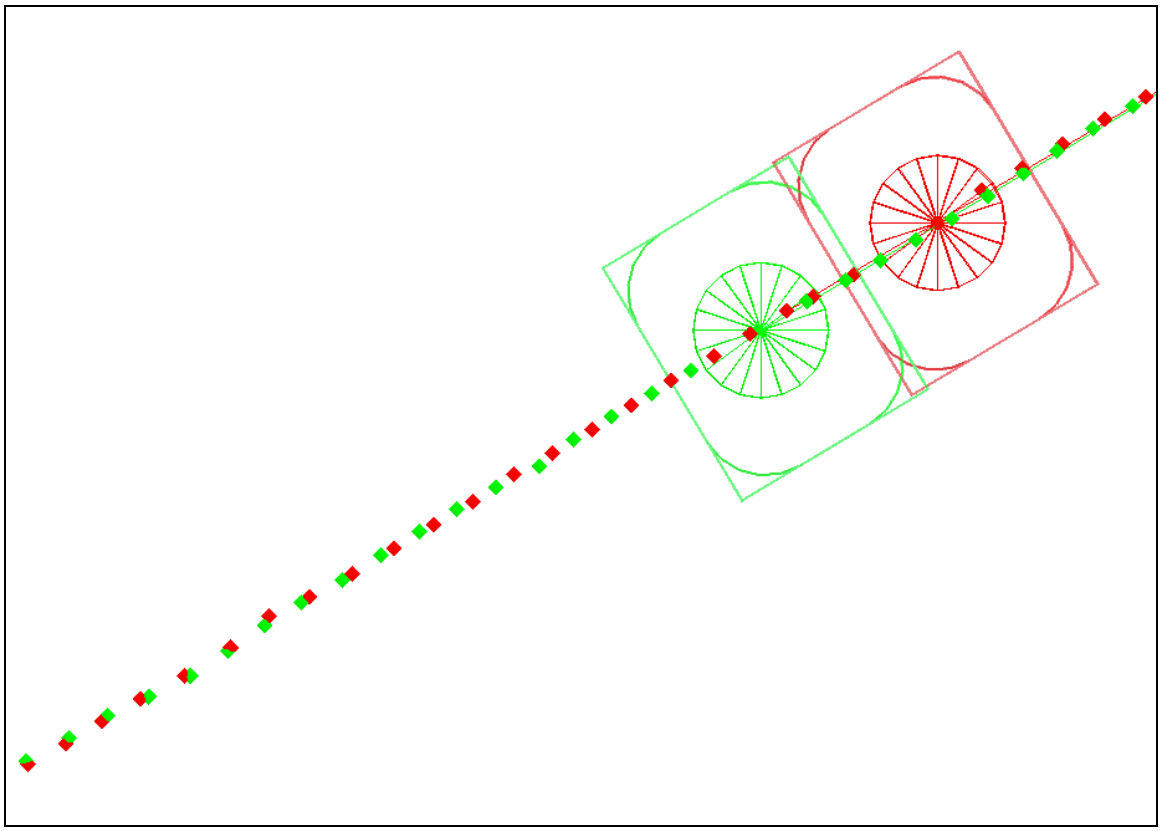

Figure 41. Baseline conflict detection (2.5 by 1.5 conformance bounds) of second aircraft pair.

3.3.3 Conflict Geometric Separation Example

The final prototype is CGS. One effect of CGS is to delay in-trail alerts for as long as possible, so a pair of in-trail aircraft was chosen in which a run without CGS generated a false alert, and another run did not. Run E1-3 can be used for this example as well, this time pairing to Run E2-25. They only differ in the use of CGS.

Both flights are Boeing 717-200s with aircraft equipage /Q flying to Hartsfield-Jackson Atlanta International Airport (KATL). Flight A (red) originated from KBOS and is leading Flight B from the top-right to the bottom-left. Flight B (green) is trailing Flight A and originated from Newark Liberty International Airport (KEWR).

Figure 42 shows the initial conflict alert generation, later determined to be a false alert. Figure 43 shows the same conflict alert being generated, but the alert is not being posted since CGS is waiting as long as possible. In this case, waiting revealed it to be a false alert and no warning was ever posted.

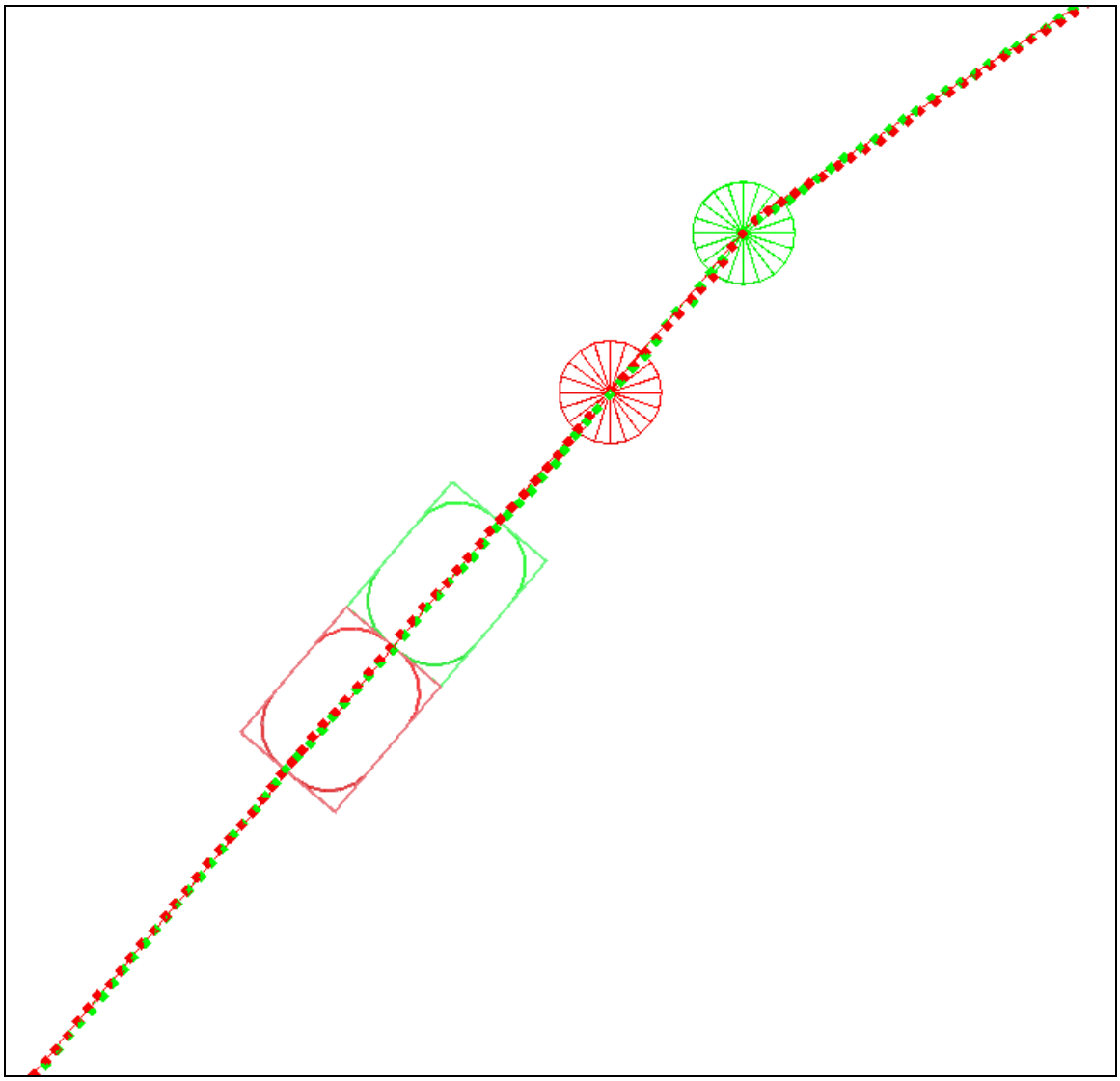

Figure 42. Conflict detection without CGS.

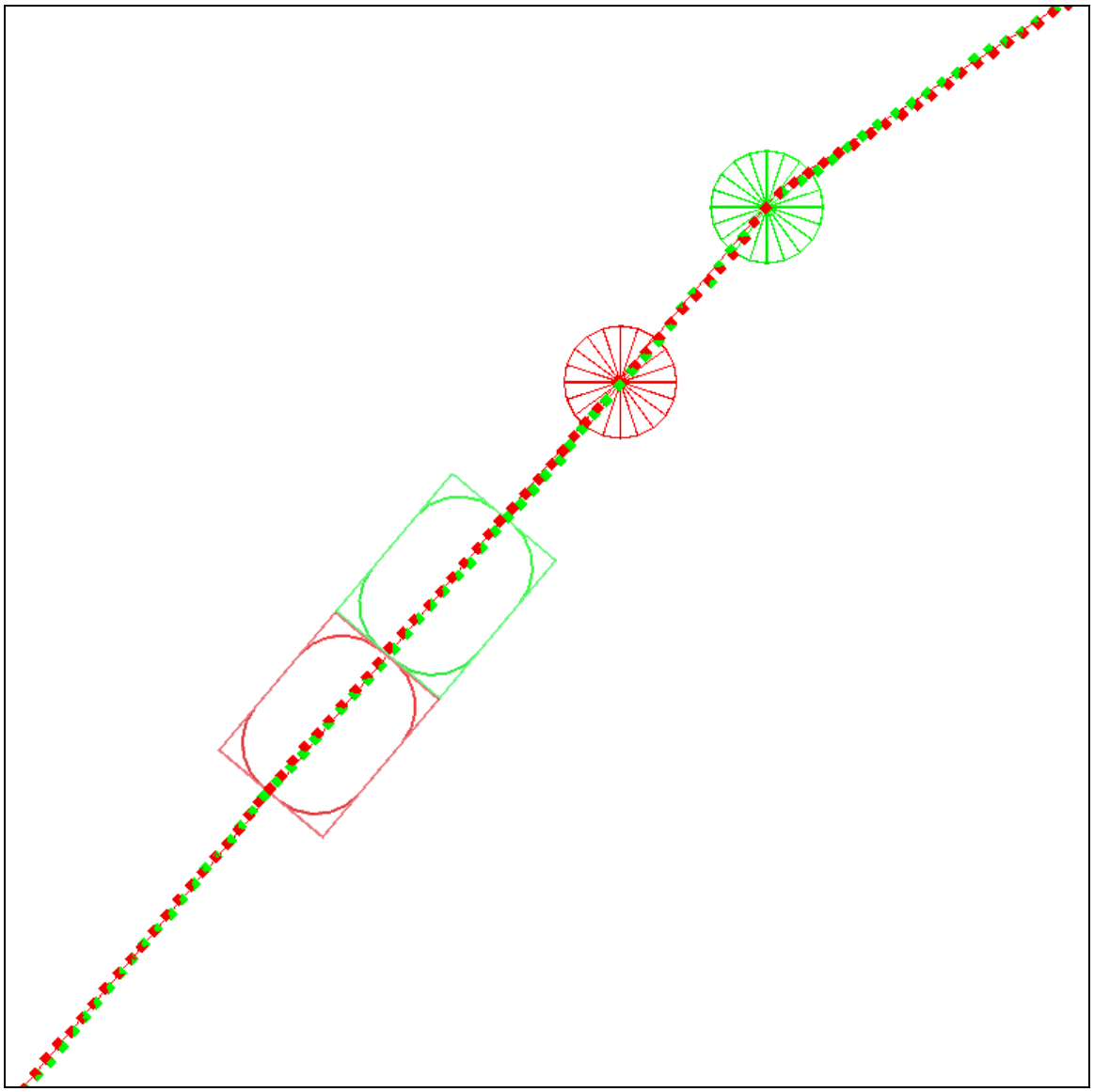

Figure 43. Same scenario with CGS, where even though the boxes overlap, an alert does not (and will not) generate due to finding this to be a false alert while it is still waiting to present the prediction.

4 Recommendations and Future Work

Based on the scenario used in this study, the Growth Adherence Bound (GAB) prototype provides the largest positive impact on the CP for the prototypes evaluated, since it can decrease the False Alerts (FA) without significantly increasing Late Alerts (LA). Turning on this prototype with all other settings unchanged can provide a 9% FA reduction, with no significant performance degradation to Late Alerts or warning time. However, a much larger FA reduction can be gained by changing the likelihood setting, slightly reducing the longitudinal conformance bound to around 1.25 nm and greatly reducing the lateral conformance bound to 1.0 nm or lower. These parameters can be set in such a way that the LA performance is not significantly degraded and FA performance is greatly improved. Though the warning time is reduced by these parameter changes, it is still significantly greater than the 180 second warning time requirement that was used in this study, which is reflected by the lack of an increase in Late Alerts.

It was discovered that the Conflict Geometric Separation (CGS) prototype was run at inappropriate settings that would greatly increase the Late Alerts. CGS was found to have the capability of reducing False Alerts by as much as 18%, but with the settings used, only at the expense of an increase to Late Alerts. As a result of a few additional runs with corrected settings, CGS was shown to reduce False Alerts without the negative impact on Late Alerts. To fully verify this, another study must be performed. There were 15 treatment runs with the CGS prototype turned on. Each of these should be re-run with the appropriate settings so the experiment can be repeated. Based solely on the experiment performed for this study, a recommendation cannot be made for CGS.

The Forced Trajectory Rebuild (FTR) prototype was found to provide a minor improvement to FA performance, but at some settings sacrifices LA performance for the FA reduction. There is some interaction of the FTR prototype with the likelihood setting. At this time it is not recommended to use the FTR prototype, but it cannot yet be ruled out as a viable addition to the probe. Once the experiments have been run on multiple scenarios, a final recommendation can be made.

Overall, Experiment 2 provides an in depth analysis of three significant prototype enhancements to ERAM. Even though some of the conclusions require additional follow-up experiments to completely evaluate the algorithms' efficacy, the advancements documented provide a new benchmark on how these prototypes need to be examined. It includes integrated experiments where the longitudinal and lateral adherence bounds are altered. It also includes various settings of the likelihood function. All must be examined simultaneously with these prototypes since their interactions provide the full benefit required. Additionally, advanced metrics were developed and employed to investigate many aspects of the system's performance. In conclusion, this study presents interim results, provides some recommendations on the prototypes examined, and details a number of objectives for the next series of experiments currently being planned.

List of Acronyms and Abbreviations

32BL	FA**32 B**ase**l**ine
AJE-15	FAA Domain Engineering Group
ANG-C41	FAA Concept Analysis Branch
ANSP	**A**ir **N**avigation **S**ervice **P**rovider
ARTCC	**A**ir **R**oute **T**raffic **C**ontrol **C**enter
ATC	**A**ir **T**raffic **C**ontrol
ATO-E	Air Traffic Organization En Route Program Office
BL	FA32 **B**ase**l**ine
CGS	**C**onflict **G**eometric **S**eparation
CP	**C**onflict **P**robe
DST	**D**ecision **S**upport **T**ool
ERAM	**E**n **R**oute **A**utomation **M**odernization
FA	**F**alse **A**lert
FA18	**F**unction **A**rea **18**
FA32	**F**unction **A**rea **32**
FAA	**F**ederal **A**viation **A**dministration
FAR	**F**alse **A**lert **R**ate
FTR	**F**orced **T**rajectory **R**ebuild
GAB	**G**rowth **A**dherence **B**ounds
Horz	**Hor**i**z**ontal
IBL	**I**nitial **B**aseline
IQR	**I**nter-**q**uartile **R**ange
JPDO	**J**oint **P**lanning and **D**evelopment **O**ffice
LA	**L**ate **A**lert
LAR	**L**ate **A**lert **R**ate
Lat	**Lat**eral
LIh	**L**ike**lih**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 13, 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 2011. "Evaluation of Parameter Adjustments to the En Route Automation Modernization's Conflict Probe," FAA Technical Note, DOT/FAA/TC-TN12/2.

Federal Aviation Administration. December 12, 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 13, 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 18, 2011. "FA18 Reduced Strategic Alerts Prototype Summary Interim 2," Lockheed Martin Report, Separation Management Modern Procedures.

Lapihuska, Charlie. November 15, 2011. "ERAM Conflict Probe Enhancement: Prototype of Reduced Nuisance Alert Options," Lockheed Martin Report, Separation Management Modern Procedures.

McKay, Edward, Ph. D. February 7, 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 16, 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.

Appendix A

Table 26. Comparison results of each of the 30 treatment runs to the FA32 Baseline run.

	Lat	0.5	2.5	0.5	1.5	2.5	0.5	0.5	0.5	0.5	2.5	0.5	2.5	1.5	2.5	2.5	2.5	0.5	2.5	2.5	0.5	2.5	0.5	0.5	2.5	0.5	2.5	2.5	0.5	2.5	1.5
	Lon	1.0	1.0	1.5	1.5	0.5	0.5	1.0	1.5	0.5	0.5	0.5	1.5	1.0	1.0	1.0	0.5	1.0	0.5	1.5	0.5	1.5	0.5	1.5	0.5	1.5	1.5	1.5	1.5	0.5	0.5
	LIh	3/8/20	3/8/20	10/20	3/8/20	10/20	3/8/10	3/8/20	10/20	3/8/10	3/8/20	3/8/10	3/8/10	3/8/10	3/8/20	3/8/20	10/20	3/8/20	3/8/10	3/8/10	10/20	10/20	10/20	3/8/10	3/8/10	10/20	3/8/10	10/20	3/8/20	10/20	3/8/20
	Enh		F	F	G	FG	F	GC	FGC	FGC	G	C	GC	F C	C	FGC	GC	FG	F C	FG	G	F C	F C	G	G	C		G	F C		F C
BL	**Prot**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
FA	**FA**	551	804	699	747	897	436	504	614	418	717	446	703	549	703	530	710	495	562	643	594	824	534	566	696	717	1116	1158	540	1012	525
FA	**N/A**	774	521	619	570	407	898	829	706	925	591	894	633	784	644	791	605	840	773	651	717	493	800	749	618	608	267	178	797	311	819
N/A	**FA**	102	104	184	72	184	107	99	185	108	113	100	44	83	75	71	155	96	101	59	187	186	214	85	108	160	21	62	122	203	120
LA	**LA**	12	15	12	14	15	13	12	13	13	14	14	15	13	14	15	15	13	14	15	13	14	13	12	14	13	15	15	13	14	13
LA	**VA**	3	0	3	1	0	2	3	2	2	1	1	0	2	1	0	0	2	1	0	2	1	2	3	1	2	0	0	2	1	2
VA	**LA**	8	5	7	7	9	13	12	15	18	10	16	12	17	12	13	16	9	23	8	11	10	19	5	10	15	4	1	16	9	19
VA	**VA**	155	158	156	156	154	150	151	148	144	153	147	151	146	150	150	147	154	140	155	152	152	144	158	153	148	159	162	147	154	143
VA	**N/A**	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
N/A	**VA**	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DSC	**DSC**	435	517	710	533	867	326	385	635	299	490	329	479	369	461	384	700	386	344	439	627	785	526	442	445	703	701	1134	384	904	354
DSC	**FA**	65	73	104	59	85	71	63	93	70	59	60	41	69	57	67	84	71	73	81	89	84	109	48	61	76	13	33	72	79	75
FA	**DSC**	69	69	76	77	90	60	61	74	51	86	54	58	61	47	73	79	59	59	100	83	77	60	79	80	69	11	58	57	71	50
DSC	**N/A**	772	682	458	680	320	875	824	544	903	723	883	752	834	754	821	488	815	855	752	556	403	637	782	766	493	558	105	816	289	843
N/A	**DSC**	63	53	123	48	113	57	50	96	53	56	57	28	48	43	43	92	61	49	50	128	90	103	50	56	83	8	39	56	127	52
% FA Imp		48.2	29.9	31.2	35.7	16.0	56.7	52.4	37.4	58.6	34.3	57.0	42.3	50.3	40.8	51.6	32.3	53.4	48.2	42.5	38.0	22.0	42.0	47.6	36.6	32.1	17.6	8.3	48.4	7.7	50.1
% LA Imp		-33.3	-33.3	-26.7	-40.0	-60.0	-73.3	-60.0	-86.7	-106.7	-60.0	-100.0	-80.0	-100.0	-73.3	-86.7	-106.7	-46.7	-146.7	-53.3	-60.0	-60.0	-113.3	-13.3	-60.0	-86.7	-26.7	-6.7	-93.3	-53.3	-113.3