note technical note techn

Evaluation of Track Monitoring and Conflict Probe Vertical Bounds in the En Route Automation Modernization

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

Brian S. Schnitzer; General Dynamics Information Technology

Loretta T. DiDonato; CSSI, Inc.

Chu Yao; FAA ANG-C41

April 2014

DOT/FAA/TC-TN13/54

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

U.S. Department of Transportation

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

NOTICE

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Technical Report Documentation Page

1. Report No. DOT/FAA/TC-TN13/54	**2. Government Accession No.**	**3. Recipient's Catalog No.**
Evaluation of Track Monitoring and Conflict Probe Vertical Bounds in the En Route Automation Modernization		**5. Report Date** April 2014 **6. Performing Organization Code** ANG-C41
7. Author(s) Christina Young, Brian Schnitzer, Loretta DiDonato, and Chu Yao		**8. Performing Organization Report No.** DOT/FAA/TC-TN13/54
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: Christina Young and Chu Yao with FAA ANG-C41; Loretta DiDonato with CSSI, Inc.; and Brian Schnitzer with General Dynamics Information Technology.

16. Abstract

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

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

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 46	**22. Price**

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

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Executive Summary

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

This paper describes the sixth in a series of integrated experiments to study these enhancements, and focuses on vertical adherence bound parameter settings. The experiment consists of simulated runs using the ERAM system with different parameter settings. The TM and CP performance of these treatment runs are compared to that of the baseline run, which has settings that match the currently deployed ERAM system, with the addition of trajectory modeling enhancements (FA32), described in [McKay, 2011] and [McKay, 2012]. The FA32 prototype is implemented in all of the treatment runs as well. The treatment runs use the recommended parameter settings determined by previous studies [Crowell *et al*., 2013]. The preferred settings are 1.0 nm lateral conformance bounds, 1.25 nm longitudinal conformance bounds, and 4|8|20 likelihood setting. All of the runs are based on the same scenario, which is generated by time-shifting real traffic data recordings to induce conflicts. The two sets of traffic data are from 2010 recordings of Washington Center and Chicago Center during peak hours.

For this experiment, the increments added to vertical conformance bounds for vertically transitioning aircraft were split into independent parts for track monitoring and conflict detection. This was done in order to determine if any additional performance gains could be observed by modifying the two parameters separately. The outcome of the analysis is a recommendation to adopt parameter values of 650 ft for the track monitoring increment and 1000 ft for the conflict detection increment.

Table of Contents

List of Figures

List of Tables

1 Introduction

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

1.1 Background to Study

The FAA's Concept Analysis Branch (ANG-C41) has published several reports of integrated experiments that were performed on recorded, time-shifted air traffic data from Air Route Traffic Control Centers (ARTCCs) [Crowell *et al*., 2011a] [Crowell *et al*., 2011b]. These reports evaluated prototypes including Growth Adherence Bound (GAB), Conflict Geometry Separation (CGS), and Forced Trajectory Rebuild (FTR) and parameter settings such as lateral and longitudinal adherence bounds and likelihood.

The study described in this paper evaluates the incremental vertical adherence bound applied to aircraft in altitude transitions. The study uses two traffic samples from the Washington (ZDC) and Chicago (ZAU) Centers. The ZDC traffic was recorded on April 30, 2010 and contains 2734 flights. The data was time-shifted to induce 239 simulated conflicts. The ZAU traffic was recorded on February 11, 2010 and contains 2234 flights and 198 simulated conflicts. Each of these time-shifted traffic samples is a six-hour recording that has been used in the previous studies [Crowell *et al*., 2012] [Crowell *et al*., 2013] [Crowell and Schnitzer, 2013].

1.1.1 Conflict Probe Parameters

The two settings manipulated in the treatment runs of this experiment are parameters of the ERAM Conflict Probe. Both are increments to be added to conformance bounds for vertically transitioning aircraft (as opposed to an aircraft in level flight). However, these parameters affect the probe in different ways and are varied independently in the experiment. One is related to track monitoring, and the second to conflict detection bounds.

Track monitoring is used to determine when a predicted trajectory is too far from the actual reported position. Figure 1 is a notional representation of the vertical distance between the actual altitude of a flight and the predicted position. If this value is greater than the stated threshold then a new trajectory will be built. Different thresholds are applied for level and vertically transitioning aircraft. This is explained in [McKay, 2012].

"Vertical adherence bounds dictate the allowable difference in track altitude and trajectory-predicted altitude before a new trajectory rebuild will be triggered. ERAM currently triggers a rebuild on the first altitude report that is out-of-vertical adherence; for a vertically transitioning flight, the present adherence bound is 1300 feet, while the adherence bound is 300 feet for aircraft in level flight. Reducing the adherence bound in the former case has the effect of increasing (on average) the number of readherences when the flight is in vertical transition and reducing the allowed track-to-trajectory vertical difference."

The difference in the applied track monitoring adherence bound between level and transitioning flights (e.g., 1300-300= 1000 feet in the excerpt above) is a factor of interest in this experiment and is referred to as the vertical TM increment.

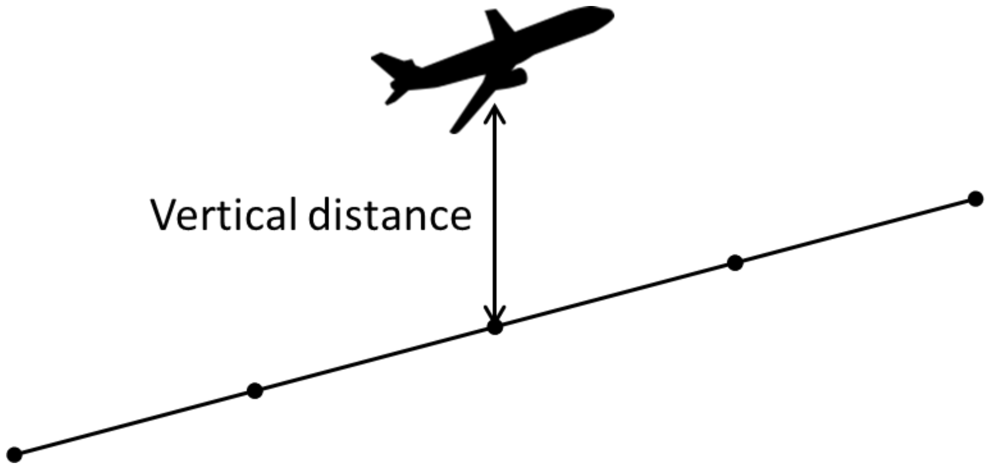

Figure 1. Track Monitoring Bounds - Track Position vs. Predicted Trajectory

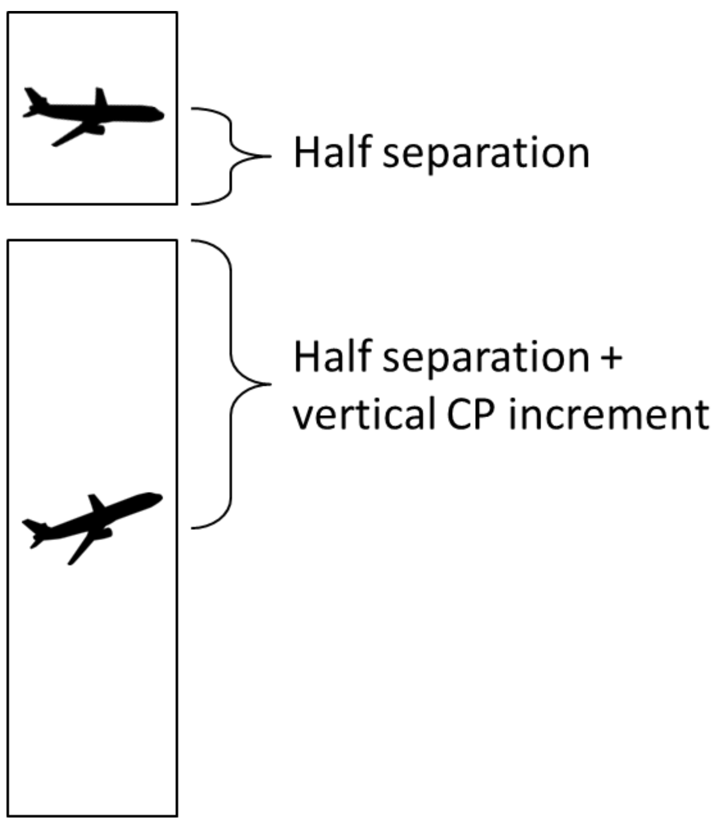

Figure 2. Conflict Detection Bounds - Alert when Boxes Overlap

Bounds are also applied in conflict detection, to determine which aircraft encounters will be alerted as potential conflicts based on aircraft trajectories. For level flight the standard conflict detection threshold is 1000 feet vertically, so flights separated by less than that threshold are considered in conflict. This is represented as a 1000 ft box centered on every aircraft, where a conflict is detected when the boxes around two aircraft touch (assuming the two aircraft are simultaneously in horizontal proximity). When an aircraft is vertically transitioning, a specified increment is added in each direction (above and below the aircraft), thereby making it more likely that a climbing or descending aircraft will be considered in conflict with another aircraft. Figure 2 depicts the conflict detection boxes, with the lower one expanded because the flight is vertically transitioning. The increment added for vertically transitioning aircraft is the second factor of interest in this experiment and is referred to as the vertical CP increment.

1.1.2 Previous Work

This study is the sixth experiment in a series. The preceding experiments are described as follows:

- Experiment 1 was performed to determine if there was a set of parameter adjustments that could be made in ERAM to improve performance of the conflict probe, using a 2005 traffic sample from Washington Center [Crowell *et al*., 2011a]
- Experiment 2 analyzed three prototype enhancements to determine if any of them provide a significant improvement to performance [Crowell *et al*., 2011b]
- Experiment 3 applied a combination of the factors (parameters and prototypes) from the first two experiments and used a 2010 traffic sample from Chicago Center [Crowell *et al*., 2012]
- Experiment 4 narrowed the focus to certain factors from previous experiments and used an updated, 2010 traffic sample from Washington Center [Crowell *et al*., 2013]
- Experiment 5 evaluated the effects of changing parameters in the likelihood threshold function [Crowell and Schnitzer, 2013]

1.2 Scope of Study

This document reports on the results of an experiment limited to two six-hour traffic samples collected from the Washington (ZDC) and Chicago (ZAU) Air Route Traffic Control Centers. To induce conflicts between aircraft and for evaluation purposes only, the data in each sample was time-shifted using the methodology documented in [Paglione *et al.*, 2003].

All of the analyses in this document were performed on a time-shifted scenario. Currently, the metrics available for analyzing conflict prediction performance require a time-shifted scenario to be used in order to generate instances of 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 values 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. Furthermore, the False Alert, Late Alert, and Missed Alert rates, as well as the warning times presented in this document do not reflect the actual values of the live ERAM system. Most of the values presented in this document are in the form of percentage change from the baseline results.

1.3 Document Organization

This technical note is organized in the following sections: Section 1 provides a high-level description of the proposed 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 Modeler (TM) and Conflict Probe (CP) performance. Finally, Section 4 presents 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 can be removed from the experiment through the application of experimental design techniques such as blocking and randomization. The output response variables are the dependent variables of the experiment, and in this case are determined by the application of various metrics.

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.	3.2.1
5 – Model Results	Examine the results of the model and discuss factor effects	3.2.1 & 3.2.2
6 – Synthesize Impact	Synthesize overall results from the model and publish conclusions.	4

For this study, the objective of designing and executing an experiment is to establish (1) which pre-determined factors and interactions of these factors have a statistically significant effect on the ERAM system's performance, and (2) the relative sizes of those 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 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 document the actual execution and analysis of the experiment and present the results.

2.1 Definition of the Problem Statement

An experiment is designed to evaluate the effect of varying TM and CP increments, for vertically transitioning flights, on trajectory accuracy and conflict probe alert performance. Trajectory accuracy is measured with vertical error and other trajectory metrics discussed in the next section. Vertical 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 studies are intended to improve False Alert performance. A significant improvement to CP performance will be recognized if it significantly improves False Alert performance and does not significantly degrade Late Alert performance. It is also desirable to avoid degrading warning time performance; but, this is not a requirement in order for a CP performance improvement to be recognized. For this study, the problem statement is expressed as follows:

Through a set of purposeful runs of ERAM, input with ZDC and ZAU time-shifted test traffic scenarios, the experiment shall evaluate the impact that vertical CP or TM adherence bounds – specifically, the increments added for vertically transitioning flights - have in terms of trajectory accuracy and conflict prediction performance.

2.2 Design of Experiment

An identical set of runs is performed for both the ZAU and ZDC data sets to broaden the scope of the experiment and see what differences may arise when comparing multiple centers. Each center has a "baseline" run in which parameter values are set to mirror the current operational settings of ERAM. However, the previous studies proposed "preferred" parameter settings which are recommended for implementation. Since these settings and prototypes were found to have distinct effects on TM and CP performance, the authors adopt the updated performance as an "experimental baseline" to compare the results of this study against. Thus, two experimental baseline scenarios are run with the preferred parameter settings and the effects being studied here are stated relative to that performance. The settings for the experimental baseline runs match those of the currently deployed ERAM with three exceptions. Lateral adherence bounds are set to 1.0nm, longitudinal bounds are set to 1.25nm, and the parameters for likelihood threshold are set to 4/8/20 minutes (corresponding to 0/0.9/1 probability values), as described in [Crowell *et al*., 2012]. The vertical TM and CP increments remain at 1000 ft for the experiment baseline, as they are in the currently deployed system. Table 2 compares various parameter settings for the currently deployed baseline version and the experimental baseline.

Table 2. Parameter Settings for Baseline Runs

	Lateral (nm)	Longitudinal (nm)	Likelihood (min)	Vertical increment (ft)
Operational Baseline (currently deployed)	2.5	1.5	10/20	1000
Experimental Baseline (preferred settings)	1.0	1.25	4/8/20	1000

2.2.1 Factors

The parameter settings for the treatment runs match the experimental baseline, with the exception of the TM and CP increments, which are the focus of this study. To evaluate their effect on trajectory and conflict probe performance, the vertical CP and TM increments are varied independently in eight treatment runs as detailed in Table 3.

Table 3. Parameter Settings for Treatment Runs

Run	Vertical TM Increment (ft)	Vertical CP Increment (ft)
ExBL	1000	1000
1	300	1000
2	300	1500
3	300	2000
4	650	1000
5	650	1500
6	650	2000
7	1000	1500
8	1000	2000

2.2.2 Model

A model is designed to integrate the results of the experiment and interpret the effects of the TM and CP increment. The constant or overall mean effect is represented in the model as μ, and $\varepsilon_{n(fg)}$ represents the assumption of independently normally distributed random error with a zero mean. All factors are assumed to be additive. The model is defined as in Eq. 1.

Response:

$$R^0_{fgn} = \mu + VertTM_f^2 + VertTM_f + VertCP_g^2 + VertCP_g + VertTM_f VertCP_g + \varepsilon_{n(fg)} \qquad \textbf{Eq. 1}$$

Where $VertTM_f$ = vertical TM increment in feet, f = 300, 650, 1000
$VertCP_g$ = vertical CP increment in feet, g = 1000, 1500, 2000
$\varepsilon_{n(fg)}$ = random error, n = 1, 2, … for all f, g

The operational baseline (OpBL) run uses different input parameter values from all other runs, as defined in Table 2, and is not part of the experiment which the model is representing. Therefore the performance of this run is not included within the model of the experimental results. However, it is important to include the OpBL in the runs and analysis as a benchmark of the current system. Performance metrics from the experiment runs are stated as percent differences from the OpBL performance.

3 Performance Evaluation

The performance evaluation analyses used in this study are similar to those used in Experiment 2 [Crowell *et al*., 2011b]. The metrics used are those described in the documentation of Experiment 1 [Crowell *et al*., 2011a]. The analysis on the trajectory modeling and conflict probe (CP) performance is described in detail.

3.1 Trajectory Modeling Analysis

The following sub-section describes the analysis of impacts to trajectory modeling from changes to the TM and CP increments.

3.1.1 Trajectory Count Comparison

An analysis of trajectory counts was carried out to observe how varying the TM and CP increment affects the frequency of generating new trajectories. For each flight in a scenario, the number of unique predicted trajectories was recorded. The mean and standard deviation of this count were calculated across 2234 flights for each of the ZAU runs. Mean and standard deviations of trajectory count per flight are presented in Table 4.

An increase in average trajectory count from 8.8 to 10.1 in the experimental baseline from the operational baseline is due to the reduced conformance bounds described in Table 2. Reducing the TM increment from 1000 to 650 ft causes an additional increase of 0.7 trajectory rebuilds, equal to a 7% increase. Reducing the TM increment from 1000 to 300 ft causes an increase of 2.3 trajectory rebuilds (23%) over the experimental baseline.

Table 4. ZAU trajectory count results

Run	Vertical TM Increment (ft)	Mean Count	Standard Deviation
OpBL	1000	8.8	4.8
ExBL	1000	10.1	5.4
1	300	12.4	6.9
2	300	12.3	6.9
3	300	12.4	6.9
4	650	10.8	5.9
5	650	10.8	5.9
6	650	10.8	5.9
7	1000	10.1	5.4
8	1000	10.1	5.4

The CP increment setting was not found to have a noticeable effect on the frequency of trajectory rebuilds, which is anticipated since it is related to the alerting function. As an example, the highlighted rows in Table 4 have different CP increment settings but the same value for TM increment. Both the mean and standard deviation are consistent across these three rows. The CP increment is applied in detecting conflicts and is not expected to affect trajectory count or accuracy.

3.1.2 Trajectory Accuracy Results

Changes to the TM increment are anticipated to affect the trajectory accuracy of a scenario while changes to the CP increment are not, as stated earlier. Trajectory accuracy is measured using trajectory metrics defined in [Paglione and Oaks, 2007] and used extensively in research studies evaluating trajectory modeling.

Vertical error is examined as a metric of interest in this study, although horizontal error metrics are also collected and presented in Appendix A. Vertical error is the time coincident altitude difference between an aircraft's position and its predicted position on a given trajectory. Figure 3 illustrates the horizontal trajectory error, which is the time coincident difference in the horizontal plane between the position of an aircraft and its predicted position on the trajectory. Note that cross and along track error are perpendicular components which comprise horizontal error.

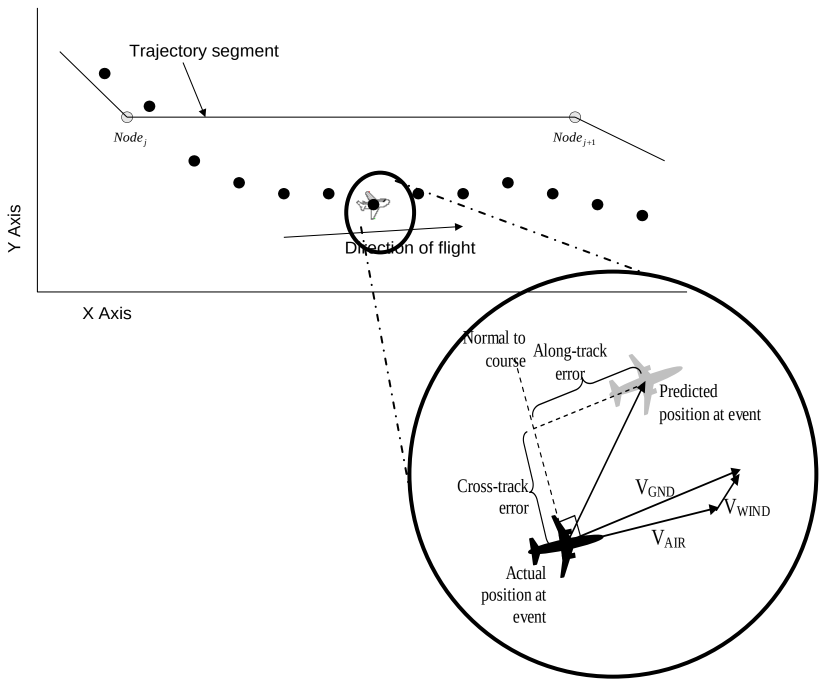

Figure 3. Diagram of horizontal trajectory error

Following the technique presented in [Paglione and Oaks, 2007], at every sampling time predicted positions along the active trajectory are compared to corresponding track position reports. Trajectory errors are calculated at the current sampling time (look ahead time of 0) and at future times along the trajectory, e.g., at 5 minutes into the future (look ahead time of 5 minutes). To get a general idea of overall trajectory accuracy for scenario comparison, the two metrics are sampled at 0, 5, 10, 15, and 20 minute look ahead times, and average values of each metric at the various look ahead times are calculated for each flight in a scenario. The average value of absolute vertical error (AAVE) is used as the main metric for comparison because the distance from centerline (zero) is of primary interest, and sign can be disregarded. Horizontal error is unsigned by definition.

A paired *t*-test compares the average error for a given flight, grouped by look ahead time, between two scenarios to determine whether the difference in trajectory performance is statistically significant. The paired *t*-test is used to benchmark the experimental baseline against the operational baseline performance, and also compare any treatment run to the experimental baseline. Run 1, which uses a TM increment of 300 ft, and Run 4, which uses a TM increment of 650 ft, are compared to the experimental baseline to reveal the effect of varying TM increment on overall vertical trajectory error. The mean difference and *p*-value results of these tests are presented in Table 5. Positive mean difference values indicate a decrease in error from the operational baseline to the experimental, and from the experimental baseline to the treatment runs. Each *p*-value is the probability of observing a discrepancy in means at least as large as that observed, even if there is no underlying difference in the means. A *p*-value less than 0.05 is typically considered to indicate statistical significance.

Table 5. Results from overall trajectory accuracy paired *t*-test analysis

Center	Comparison	*p*-value	Mean Difference (ft)
ZAU	Exp BL v Op BL	0.0049	13
	Run 1 v Exp BL	< 0.0001	40
	Run 4 v Exp BL	< 0.0001	21
ZDC	Exp BL v Op BL	0.0092	-7.7
	Run 1 v Exp BL	< 0.0001	22
	Run 4 v Exp BL	< 0.0001	10

The results of the overall trajectory accuracy analysis for ZAU and ZDC in Table 5 indicate a statistically significant difference in average vertical trajectory error between the treatment runs and the experimental baselines. The treatment runs with lower values for TM increment had lower overall trajectory errors. However, the overall effect on trajectory error was too small to be practically significant. The next section focuses on specific segments in the scenarios, and the NAS, where the impact from these parameter changes would be concentrated to get a better indication of the effect on trajectory accuracy within affected flight segments.

3.1.2.1 Analysis on Vertically Transitioning Trajectory Segments

The TM and CP increments are applied to flights that are in vertical transition, as described in Section 1.1.1. Therefore it is anticipated that the effects of the parameter changes would be concentrated in trajectory segments that are associated with vertical transitions. This section provides a subset analysis limited to points where the predicted altitude is changing by at least 2 feet between consecutive trajectory points. As in the overall trajectory accuracy analysis, metrics are sampled at 0, 5, 10, 15, and 20 minute look ahead times and the average values of vertical and horizontal error at each look ahead time are calculated for each flight in a scenario. The results for horizontal error are presented in Appendix A.

The paired *t*-test is again used to evaluate the differences in average absolute vertical error between scenarios, where the sampled error metrics are filtered to include only a subset of points, as described. The mean difference and *p*-value results of these tests are presented in Table 6.

Table 6. Results from subset trajectory accuracy paired *t*-test analysis

Center	Comparison	*p*-value	Mean Difference (ft)
ZAU	Exp BL v Op BL	< 0.0001	28
	Run 1 v Exp BL	< 0.0001	171
	Run 4 v Exp BL	< 0.0001	82
ZDC	Exp BL v Op BL	< 0.0001	15
	Run 1 v Exp BL	< 0.0001	173
	Run 4 v Exp BL	< 0.0001	78

The results of the vertically transitioning subset trajectory accuracy analysis in Table 6 indicate a statistically significant difference in average vertical trajectory error between the treatment runs and the experimental baselines in both centers. Similar to the overall results in Table 5, lower TM increments are associated with lower average error. The effect sizes are much greater in this subset analysis, which is expected because it focuses on trajectory segments where the effect would be concentrated.

The information in the paired *t*-tests is presented in graphical form by plotting the difference from each pair of error values against the average of the same pair. Each data point represents a unique combination of flight and look ahead time. Figure 4 illustrates the matched pairs in graphical form for Run 1 and Run 4 against the experimental baseline of ZAU, while Figure 5 illustrates the same for ZDC. Positive mean differences along the vertical axis indicate the experimental baseline has higher average error than the respective treatment run.

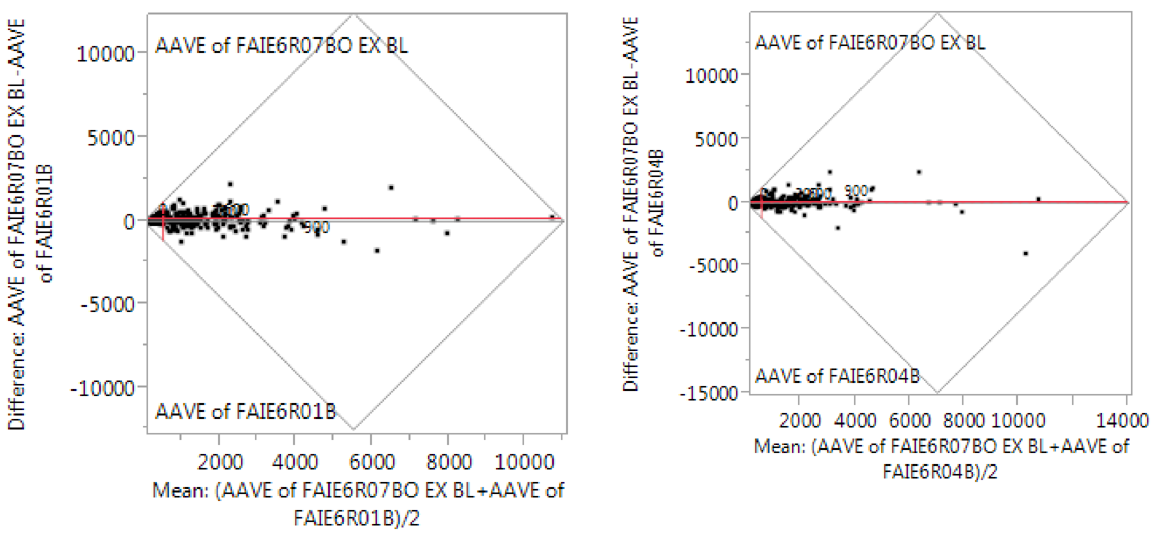

Figure 4. Matched pairs graphs for ZAU Run 1 and Run 4 vs ExBL

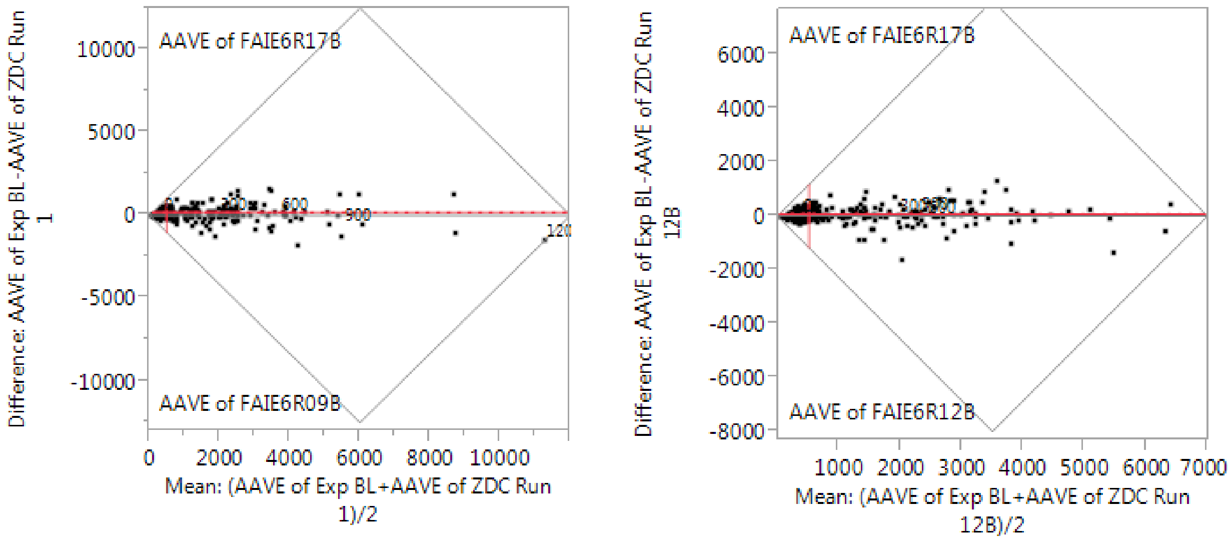

Figure 5. Matched pairs graphs for ZDC Run 1 and Run 4 vs ExBL

The series of horizontal red lines across the center of the graphs in Figures Figure 4 and Figure 5 indicate the mean calculated difference and the confidence interval around this mean difference. Although difficult to see, the red lines fall above zero in each of the four graphs, confirming that the mean difference is statistically significant and that an improvement in trajectory accuracy (indicated by a decrease in average vertical error) is observed in all cases.

3.2 Conflict Probe Analysis

Analysis was performed on the experimental data to determine the effects that independently manipulating the TM and CP increments have on the performance of the Conflict Probe (CP). The null hypothesis is as follows:

> *A statistically significant Conflict Probe performance improvement is not observed through systematic manipulation of the Track Monitoring and Conflict Probe Increments.*

Since this experiment attempts to evaluate any improvement to the CP with respect to gains already made through the introduction of prior enhancements such as a reduction in the Lateral and Longitudinal bounds [Crowell *et al*., 2011a], performance improvement is defined as no significant increases in the number of False Alerts (FA) and no significant increases in the combined number of Late Alerts and Missed Alerts (LA and MA) while maintaining a 25th percentile of Warning Time (WT) that is above the three minute threshold. All of these requirements must be met in order for it to be considered a significant improvement and to reject this null hypothesis.

The CP analysis documented here will attempt to reject this null hypothesis, therefore showing that these parameter changes do indeed provide a significant improvement to the ERAM system above and beyond those introduced in prior integrated experiments. At the very least, no significant degradation must be indicated in any of the aforementioned metrics.

3.2.1 Overall Statistics

Table 7 and Table 8 show the TM and CP settings along with alert counts, 25th percentile of warning time for Valid Alerts (excluding popups), and percent change with respect to the operational baseline (OpBL) for all runs. A few observations can be made. First, all False Alert counts are significantly improved from the OpBL (29% to 45%), though any change from the baseline TM and CP increment settings (ExBL, 1000/1000) increases the number of FAs with respect to the ExBL runs. Second, manipulating the TM and CP settings either reduces or maintains the LA/MA counts; however, the N is small, suggesting that any changes will not be found to be statistically significant. VA counts improve slightly. Finally, WT remains relatively consistent regardless of TM and CP setting which suggests robustness to the experimental parameters, and remains well above the 180 second minimum requirement. These results suggest that the only significant change to the CP resulting from experimental manipulation will be a tradeoff between FAs (slight degradation) and slight improvements to LA/MA, VA, and WT.

Table 7. Conflict probe alert statistics for ZAU runs

ZAU	OpBL	ExBL	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8
TM Inc.	1000	1000	300	300	300	650	650	650	1000	1000
CP Inc.	1000	1000	1000	1500	2000	1000	1500	2000	1500	2000
LA	6	9	5	6	4	6	6	4	6	5
MA	1	3	2	2	1	1	1	1	0	1
LA/MA	7	12	7	8	5	7	7	5	6	6
FA	1789	1001	1090	1192	1238	1046	1124	1175	1079	1152
WT	355	276.5	271.5	279	286	281	284	286	287	291.5
VA	193	188	193	192	194	192	192	195	193	193
LA %	0	71	0	14	-29	0	0	-29	-14	-14
FA %	0	-44	-39	-33	-31	-42	-37	-34	-40	-36
WT %	0	-22	-24	-21	-19	-21	-20	-19	-19	-18
VA %	0	-3	0	-1	1	-1	-1	1	0	0

Table 8. Conflict probe alert statistics for ZDC runs

ZDC	OpBL	ExBL	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8
TM Inc.	1000	1000	300	300	300	650	650	650	1000	1000
CP Inc.	1000	1000	1000	1500	2000	1000	1500	2000	1500	2000
LA	4	10	9	8	7	10	8	7	9	7
MA	1	1	2	2	2	1	1	1	1	1
LA/MA	5	11	11	10	9	11	9	8	10	8
FA	2089	1149	1232	1364	1481	1175	1295	1409	1267	1384
WT	427	332	338	348	349	345	346	349	338	346.5
VA	204	198	198	199	200	198	200	201	199	201
LA %	0	120	120	100	80	120	80	60	100	60
FA %	0	-45	-41	-35	-29	-44	-38	-33	-39	-34
WT %	0	-22	-21	-19	-18	-19	-19	-18	-21	-19
VA %	0	-3	-3	-2	-2	-3	-2	-1	-2	-1

3.2.2 Comparison of Conflict Alert Types

Next, analysis is applied to determine if there is a statistically significant difference between runs in the count of false alerts and correct no calls, as well as the count of valid and missed alerts. A chi-squared test is performed to statistically evaluate the difference in counts from the ExBL and treatment runs and determine how likely it is for this difference to be attributable to random chance. The test statistic is defined generically as follows:

$$\chi^2 = \frac{(n_{VE} - n_{EV})^2}{(n_{VE} + n_{EV})}$$

where

χ^2 is the chi-squared test statistic

n_{VE} is the number of events of evaluation code type A in Baseline and changing to type B in Treatment

n_{EV} is the number of events of evaluation code inverse of those used in n_{VE} (type A in Treatment and type B in Baseline)

The resulting test statistic can be expressed as a probability or p-value by assuming a chi-squared distribution [Agresti, 2002]. Results of the chi-squared test are shown in Table 9. FA_NC represents false alerts in the ExBL that are not called in the respective treatment scenario (reduction in FAs) whereas NC_FA represents alerts that are not called in the ExBL and are called in the treatment scenario (creating additional FAs). MA_VA represents Missed Alerts in the ExBL that are called properly (become Valid Alerts) in the treatment scenario, and VA_MA represents Valid Alerts in the ExBL that are missed in the treatment scenario. Green cells indicate that a comparison of the respective conflict counts between the treatment scenario and the ExBL found them to be statistically different, and orange cells indicate a marginal difference. It is evident that almost every treatment run results in a significantly higher FA count and that, in the ZAU scenario, VA counts increase significantly as well. A further breakdown of the number of alerts of various types in each run can be found in Appendix B.

Table 9. Test for differences in CP

		False+	False-	False Alert Test				Valid+	Valid-	Valid Alert Test		
	Ex BL vs.	NC_FA	FA_NC	Difference	Chi-Sq	*p*-value		MA_VA	VA_MA	Difference	Chi-Sq	*p*-value
ZAU	Run1	169	97	72	19.5	0.000		6	1	5	3.6	0.059
	Run2	256	109	147	59.2	0.000		6	2	4	2.0	0.157
	Run3	319	131	188	78.5	0.000		7	0	7	7.0	0.008
	Run4	112	76	36	6.9	0.009		5	0	5	5.0	0.025
	Run5	199	99	100	33.6	0.000		5	0	5	5.0	0.025
	Run6	262	115	147	57.3	0.000		7	0	7	7.0	0.008
	Run7	142	75	67	20.7	0.000		6	0	6	6.0	0.014
	Run8	226	98	128	50.6	0.000		6	0	6	6.0	0.014
ZDC	Run1	164	86	78	24.3	0.000		2	2	0	0.0	1.000
	Run2	281	87	194	102.3	0.000		3	2	1	0.2	0.655
	Run3	385	95	290	175.2	0.000		3	1	2	1.0	0.317
	Run4	91	75	16	1.5	0.214		1	1	0	0.0	1.000
	Run5	212	93	119	46.4	0.000		2	0	2	2.0	0.157
	Run6	326	100	226	119.9	0.000		3	0	3	3.0	0.083
	Run7	161	64	97	41.8	0.000		2	1	1	0.3	0.564
	Run8	264	75	189	105.4	0.000		3	0	3	3.0	0.083

3.2.3 Experimental Model Results

The design of this experiment is such that it lends itself to predictive modeling of the resultant data. Percentage of VAs, FAs, MAs and LAs and change in WT are all inputs to a second order model using the JMP® commercial statistical software package [SAS, 2010], and the output allows for the effects of any TM and/or CP setting within the experimentally tested range (300-1000 for TM, 1000-2000 for CP) to be examined. Figure 6, Figure 7, Figure 8, and Figure 9 depict the results of the model when ExBL settings (TM – 1000, CP – 1000) and recommended settings (TM – 650, CP – 1000) are selected.

For both scenarios, the LA% metric exhibits the greatest responsiveness with respect to TM and CP settings but it is also by far the metric subject to the most noise, in large part due to the low count and high variance between runs as seen in Table 7 and Table 8. In the ZAU scenarios. the modeled change in LA with respect to the OpBL ranges from a 34% decrease at TM – 650 and CP – 2000 to a 52% increase at TM – 1000 and CP – 2000. In the ZDC scenario, the modeled change in LA with respect to the OpBL ranges from a 58% increase at 800/2000 to a 125% increase at 1000/1000. In general, a middling TM setting (about 650) and a high CP setting (2000) provides the most benefit to LA, primarily due to the “u-shaped” curve in the TM Increment column of ZDC, Figure 8 and Figure 9. Since LA/MAs and WT are somewhat dependent on one another, as WT increases one expects LA% to decrease. This is demonstrated in Figure 6Figure 8 through Figure 9. However, the effects of TM and CP on WT are small. In the ZAU scenario WT reduction ranges from -24% at 300/1000 to -18% at 1000/2000 and in the ZDC scenario WT reduction ranges from -22% at 1000/1000 to -17% at 650/2000. Neither TM nor CP have a statistically significant effect on LA for the ZAU scenario, but CP increment does affect LA in the ZDC scenario ($p<.01$). CP has a minimal (significant effect on WT in both scenarios ($p<.01$).

FA% shows the best performance at the ExBL settings of 1000/1000 in both scenarios, which are the maximum and minimum TM and CP settings, respectively, as shown in Figure 6 and Figure 8. In the ZAU scenario, FA% ranges from -44% at 1000/1000 to -31% at 300/2000. In the ZDC scenario, FA% ranges from -45% at 1000/1000 to -29% at 300/2000. Both TM and CP increments significantly affect FA in both scenarios ($p<.01$).

VA% has the least amount of variability of all the metrics. However, changing the TM and CP increments do seem to affect VA% differently in the ZAU and ZDC scenarios. VA% improves as CP increases in both scenarios as expected, but an increase in the TM setting reduces VA% in the ZAU scenario while it increases VA% in the ZDC scenario. The range of change is a -2% reduction at 1000/1000 to a 1% increase at 300/2000 in ZAU and not statistically significant, and -3% at 300/1000 to -1% at 1000/2000 in ZDC. TM has no statistically significant effect on VA% in either scenario, but CP increment does have an effect on VA% in the ZDC scenario.

Overall, a moderate TM increment setting of 650 provides a balance between FAs and the other 3 metrics. Applying the prior reduction in Latitudinal and Longitudinal conformance bounds along with the new 4/8/20 Likelihood functions provided a reduction in FA % of 44% and 45% (788 and 940 alerts) to the ZAU and ZDC scenarios, respectively, and LAs/MAs increased by 71% and 120% (5 and 6 alerts). WT 25th percentile decreased by 22% in both scenarios, 79 seconds to 276.5 sec in ZAU and 95 seconds to 332 seconds in ZDC. VA% decreased by 3% in both scenarios, 5 alerts in ZAU and 6 alerts in ZDC. By reducing the TM increment to 650 and leaving the CP at 1000, some gains in LAs/MAs, VAs, and WT can be made by giving back a small amount of FAs.

- FAs changes from 44% to 42% (increase of 45 False Alerts) in the ZAU scenario and from 45% to 44% (increase of 26 False Alerts) in the ZDC scenario
- LA% changes from an increase in 52%to 14% (decrease of 2.5 Late and Missed Alerts) in the ZAU scenario and from 125% to 113% (decrease of 1.3 Late and Missed Alerts) in the ZDC scenario
- WT% changes from -22% to -21% (increase of 5 seconds) in the ZAU scenario and from -22% to -20% (increase of 10 seconds)
- VA% changes from -2% to -1% (increase of 3 valid alerts) in the ZAU scenario and remains virtually unchanged the ZDC scenario

However, if the primary goal is to minimize False Alerts, then it is recommended that the TM and CP increment settings remain unchanged. The Experimental Baseline (ExBL) settings of 1000/1000 provide a 44-45% reduction in FAs. While LAs/MAs do increase and VAs are reduced, either the number or percentage is relatively small:

- 5-6 more LA/MA
- 5-6 less VAs (about 3%)
- 22% decrease to WT (still 277 sec and 332 sec, well above the 180 second threshold

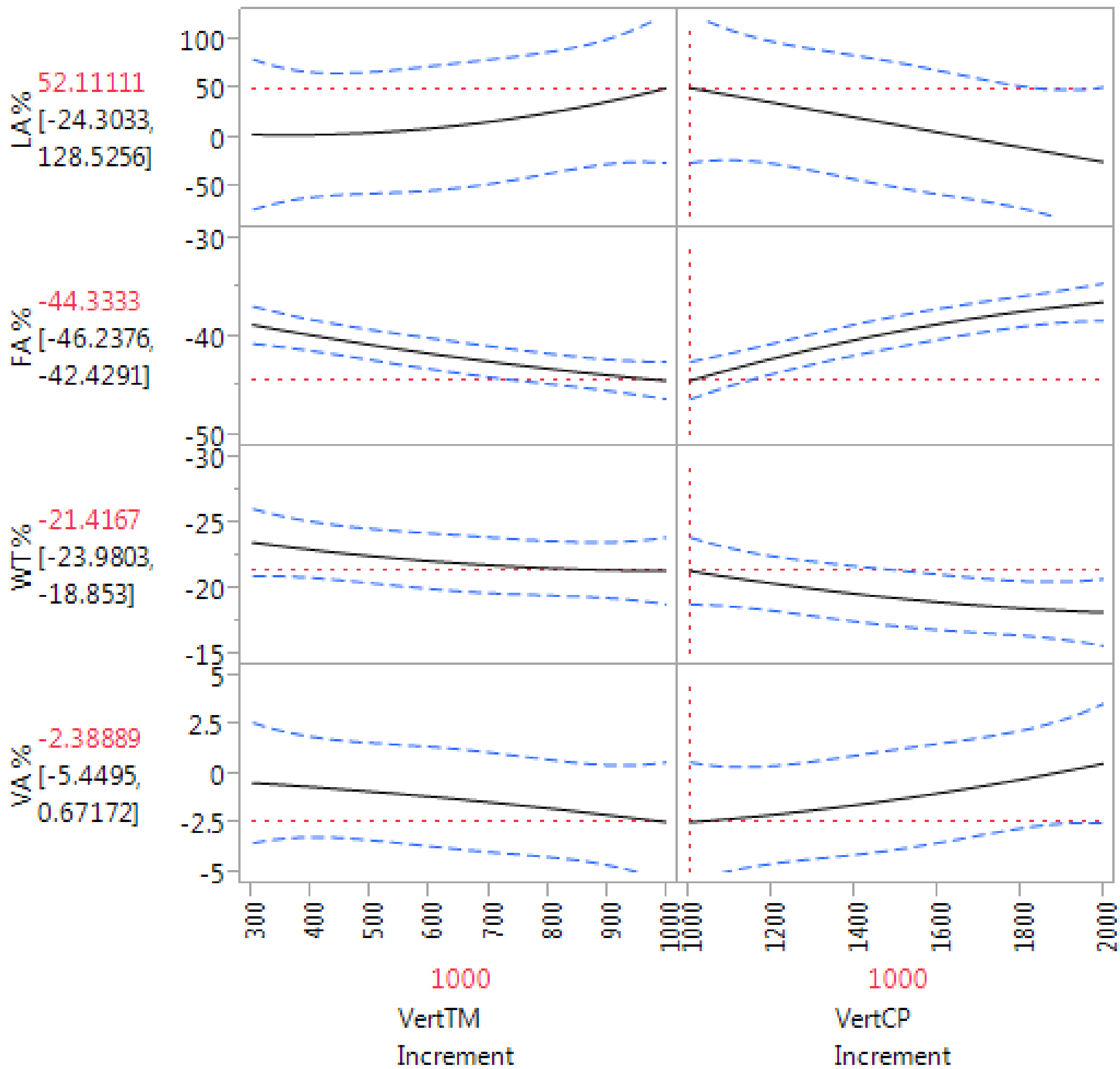

Figure 6. Prediction Profiler: ZAU Experimental Baseline settings

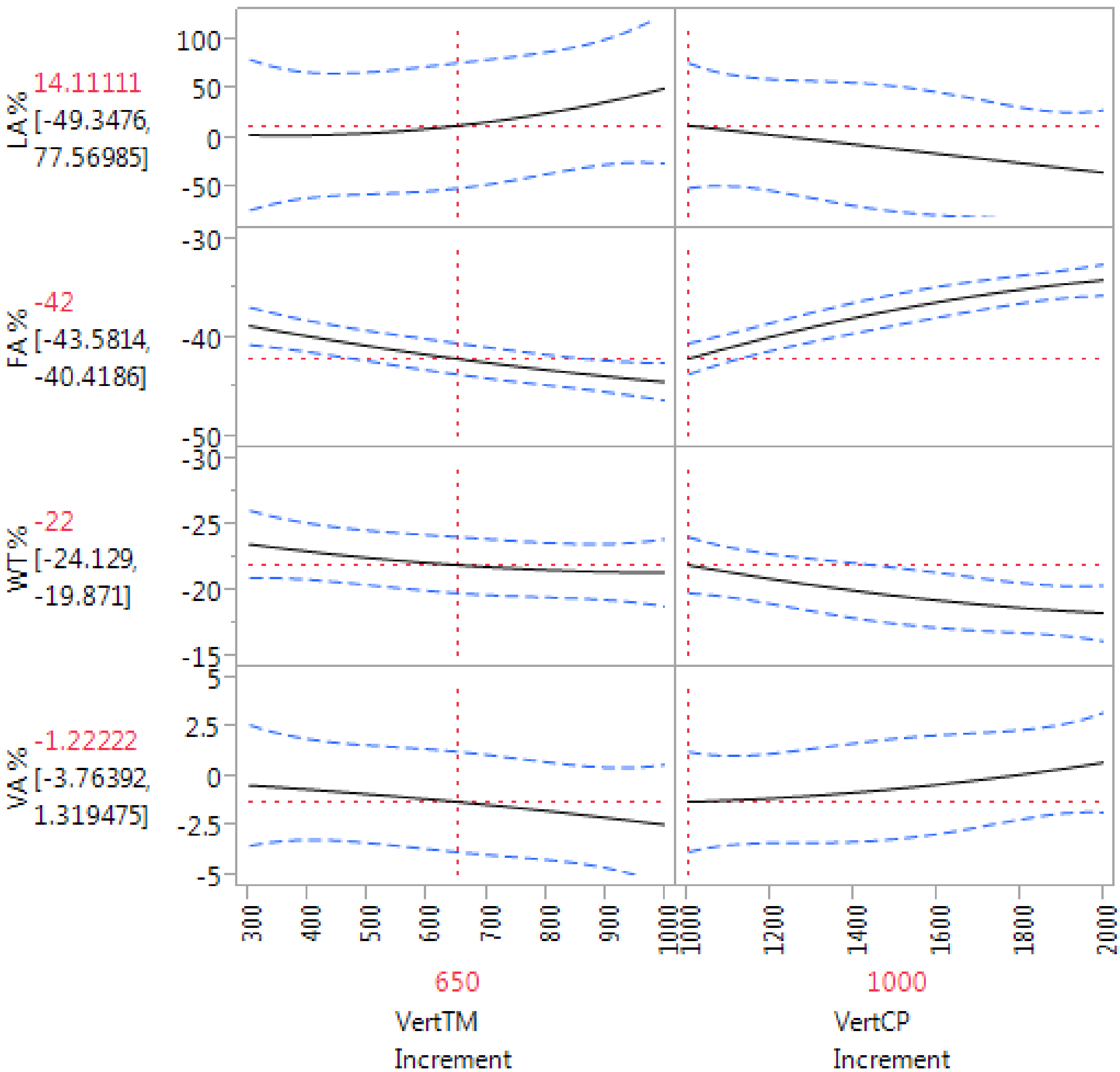

Figure 7. Prediction Profiler: ZAU recommended settings

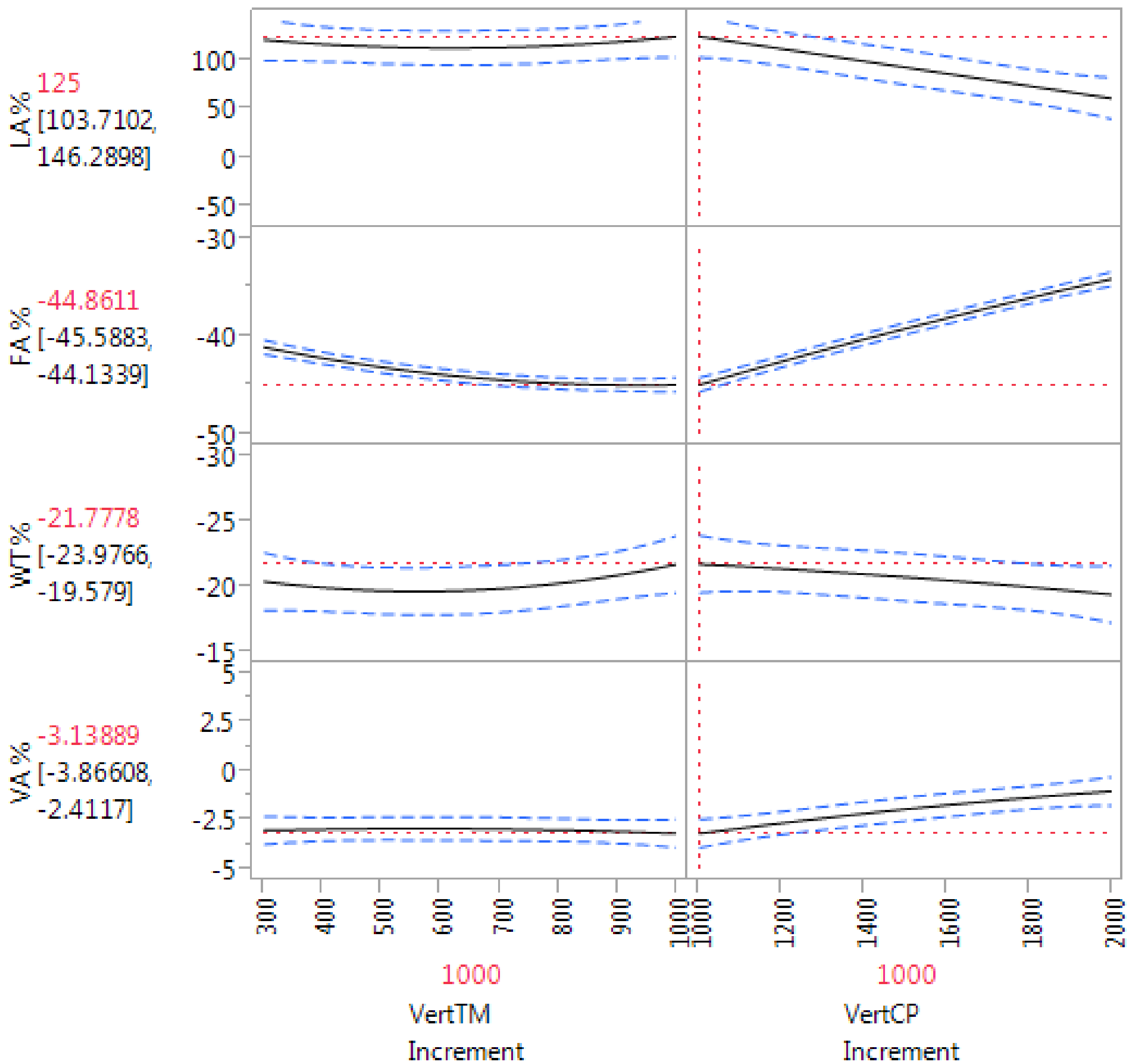

Figure 8. Prediction Profiler: ZDC Experimental Baseline settings

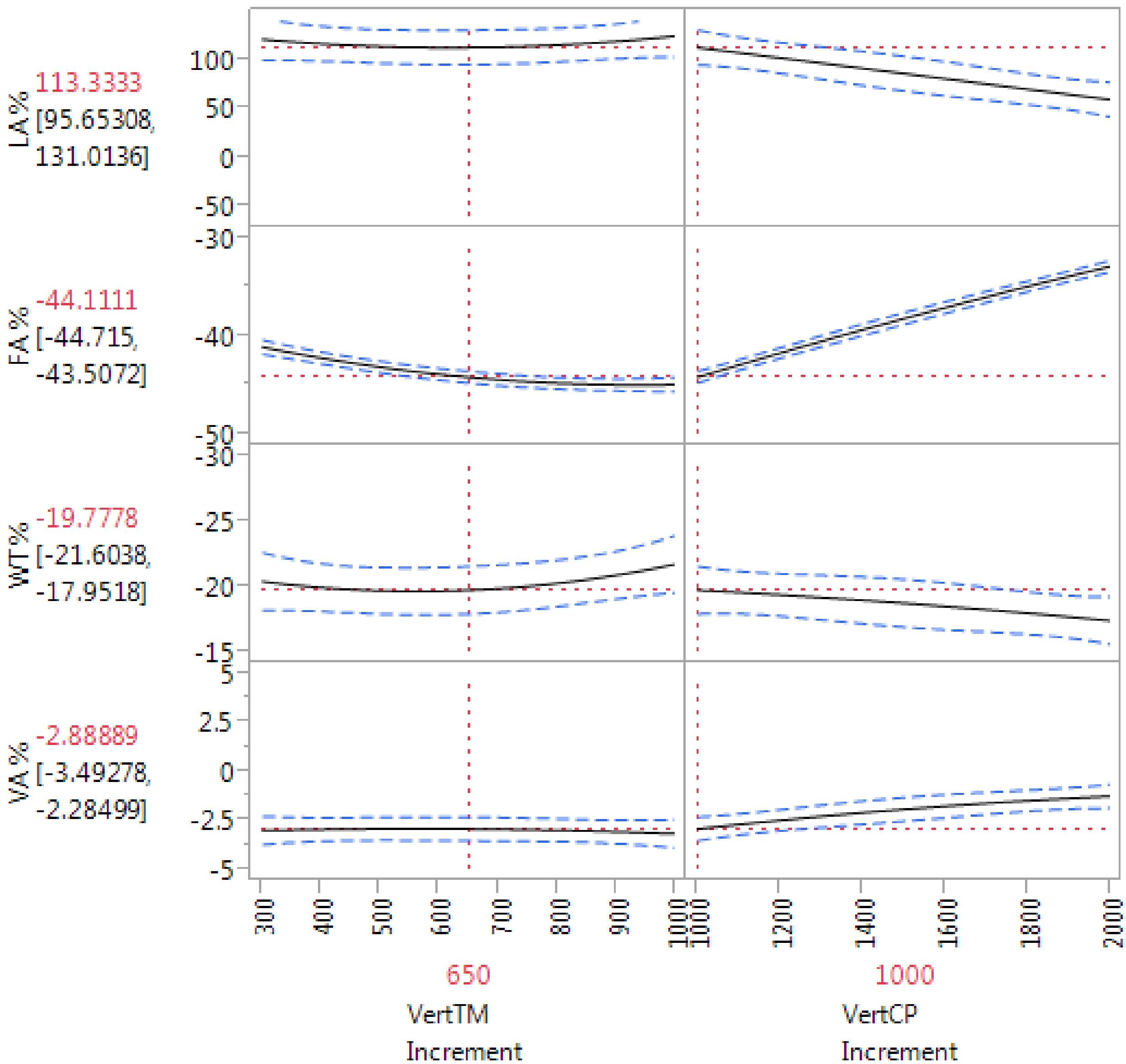

Figure 9. Prediction Profiler: ZDC recommended settings

3.3 *Flight Examples*

The flight examples depicted below demonstrate some of the effects that manipulating the TM increment have on trajectories and conflict prediction. All of the examples are from ZAU Run 1, with TM increment set to 300 ft and CP increment set at the baseline value of 1000 ft. This run was chosen so that the effects of reducing the TM, which primarily include forcing trajectories to be built more frequently during vertical transitions, could be explored. Increasing the CP increment has no effect on trajectories themselves and only serves to expand the zone around a given trajectory where alerts may be detected. CP expansion results in more alerts being generated and with higher warning time. Examples of this effect are not of interest.

In the following CP examples, two flights are shown. The reported track positions of each are represented by dotted lines, the trajectories by wireframe boxes, the minimum separation boundary as solid cylinders, and the conflict probe bounds as rectangular regions with rounded corners. Black coloration is used to represent level flights, which show no effect of the reduced TM increment. Blue colors are used to represent the baseline (TM increment of 1000 ft) trajectories and alert bounds. Red colors are used to represent the treatment (TM increment of 300 ft) trajectories and alert bounds. The first two sections illustrate encounters where reduction of the TM increment result in Missed Alerts (MA) being called with sufficient warning time in the treatment run to be reclassified as a Valid Alert (VA). The third section illustrates an encounter in which a False Alert (FA) is generated in the baseline scenario but not in the treatment scenario, which is termed a correct No Call (NC). The fourth section points out an example of trajectory modeling improvement resulting from a smaller TM increment resulting in an earlier trajectory build.

3.3.1 CP Example 1 – Late Missed Alert to Valid Alert

CP example 1 (illustrated in Figure 10 and Figure 11) depicts an encounter between CP01_F1 and CP01_F2 at 69337 sec. CP01_F2 is an Embraer 170 (E170) recently leveled at FL 300 out of Milwaukee (KMKE) and en route to Kansas City (KMCI). CP01_F2 is indicated by black coloring in the figures. CP01_F1 is a Boeing 737-700 out of Kansas City (KMCI) beginning its descent into Milwaukee (KMKE). Blue coloring represents flight CP01_F1 when the TM vertical increment is set at 1000 ft (baseline), and red represents trajectories built when a 300 ft TM increment is applied. Dots represent the track of each flight, wireframes represent the trajectories, and rectangular boxes represent the point in the future at which loss of separation is predicted. As depicted, flight CP01_F1 is currently descending from FL 350 to FL 240. Horizontal overlap at a look ahead time of 212 seconds demonstrates a loss of horizontal separation for both scenarios in Figure 10, as both the black-blue and black-red rectangles are overlapping. However, the synchronous side-view shown in Figure 11 depicts a loss of vertical separation only in the treatment (black-red) scenario due to a slightly better vertical trajectory. The 300 ft TM increment lowers the threshold of deviation that prompts trajectory rebuilds, resulting in more frequent and potentially more accurate trajectories. In this case, the result is a correct conflict probe alert (Valid Alert). The baseline (blue) trajectory rebuilds later and does generate an alert, but the alert is presented with only 114 seconds of warning time and therefore classified as a Missed Alert (LATE_MA).

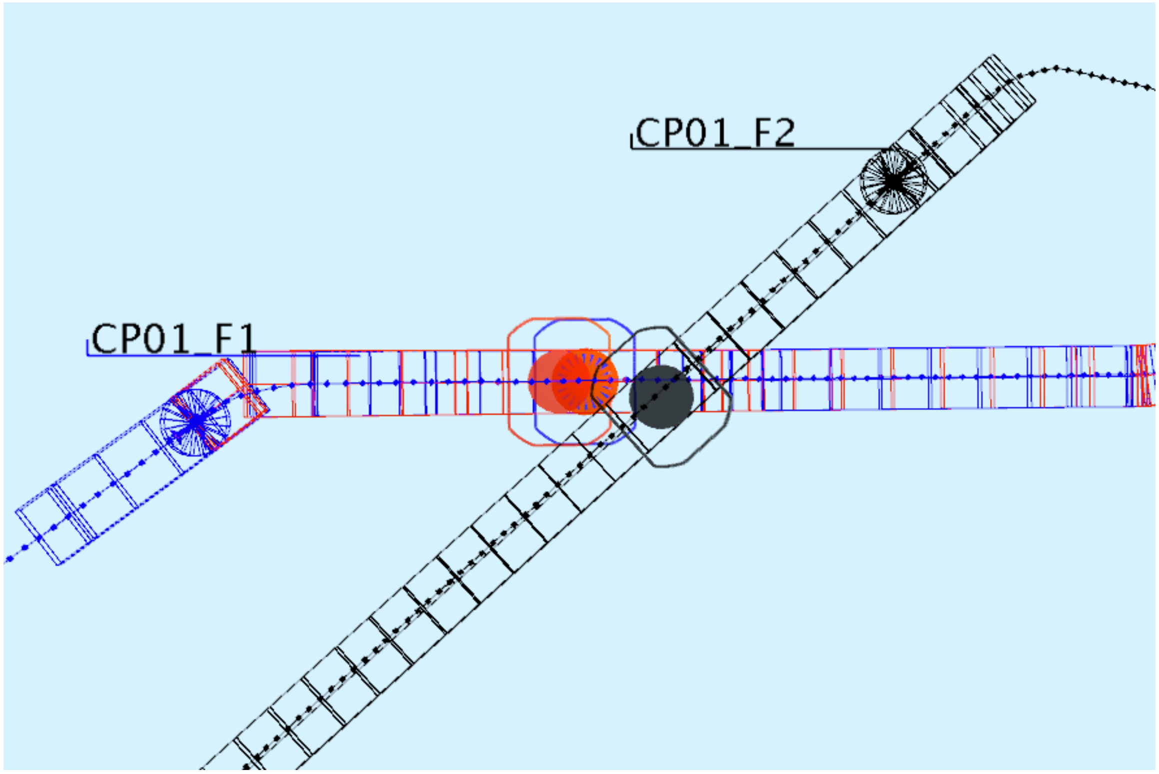

Figure 10. CP example 1, MA to VA, top down view

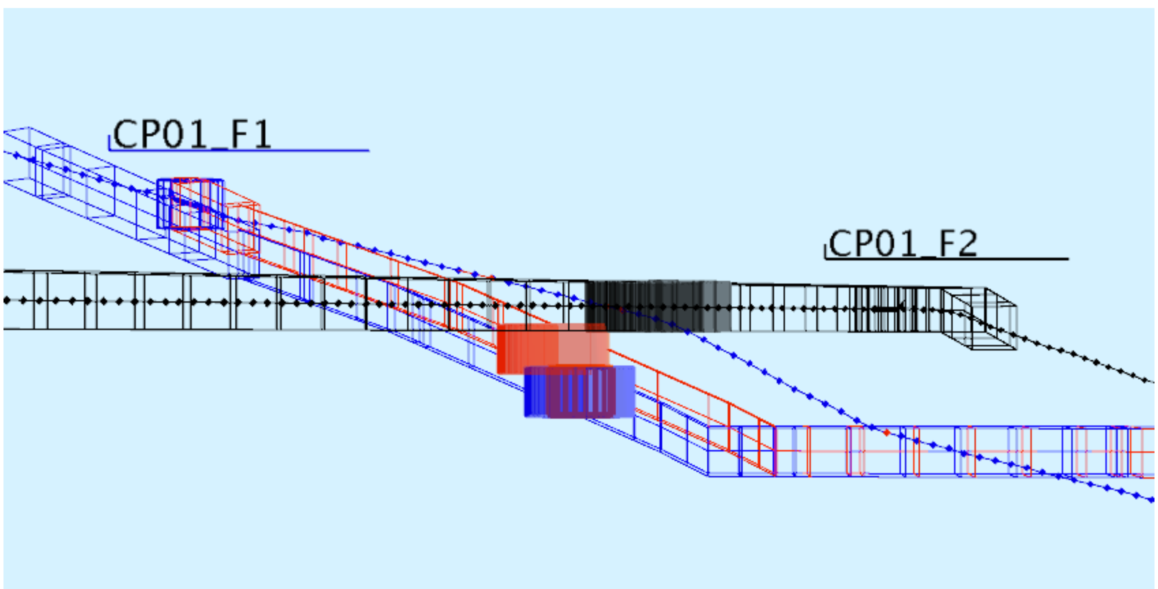

Figure 11. CP example 1, MA to VA, side view

3.3.2 CP Example 2 – No Call Missed Alert to Valid Alert

CP example 2 (illustrated in Figure 12 and Figure 13) depicts an encounter between CP02_F1 and CP02_F2 at 73609 sec. CP02_F1 is a Boeing 737-800 (B738) flying level at FL 370 en route from Salt Lake City (KSLC) to New York City (KJFK) and is represented by black coloring. CP02_F2 is a Dessault-Breguet Falcon 900 (F900) climbing out of Racine (KRAC) en route to Aspen (KASE) and is represented by blue (baseline) and red (treatment) coloring. As depicted in the side view (Figure 13), CP02_F2 is approaching the top of its climb. An alert is posted in the treatment scenario, warning that a loss of separation will occur in 205 seconds. In the baseline scenario an alert is not yet posted. This is due to the fact that while in the top view horizontal loss of separation is indicated for both scenarios, in Figure 13 vertical separation is at 1000 ft and CP02_F2 is predicted to have moved beyond 5 NM of the top of climb in the baseline trajectory shown. While within 5 NM of either the beginning or the end of a vertically transitioning trajectory segment, the Conflict Probe is expanded by the CP increment (1000 ft in the run shown), and since the treatment (red) trajectory predicts a top of climb further along the route than does the baseline trajectory due to the more stringent 300 ft TM resulting in more frequent trajectory rebuilds, the CP is still expanded at the look ahead shown. The CP is not expanded in the baseline case, and while an alert is eventually posted in the baseline scenario it is not posted until 172 seconds before the conflict start time, which is less than the minimum required warning time of 180 seconds.

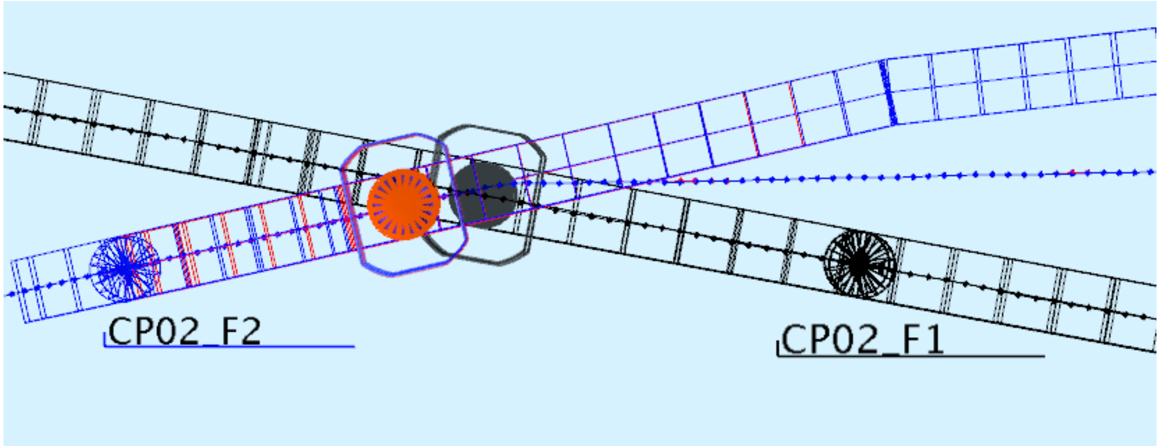

Figure 12. CP example 2, MA to VA, top down view

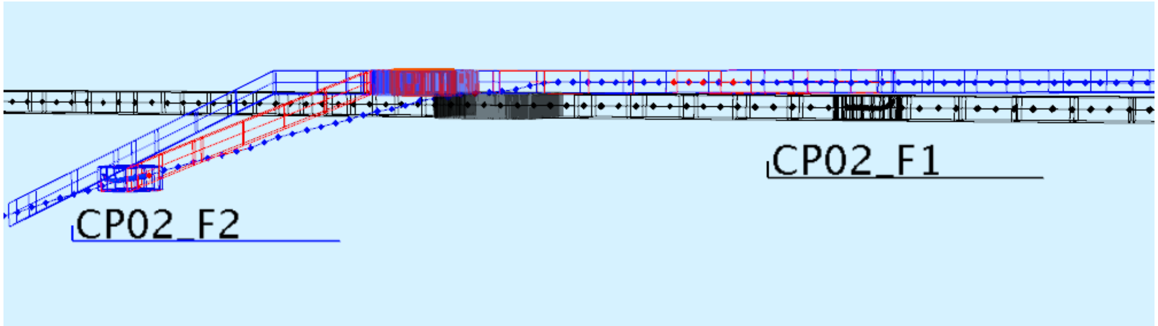

Figure 13. CP example 2, MA to VA, side view

3.3.3 CP Example 3 – False Alert Removed

CP example 3 (illustrated in Figure 14) depicts an encounter between CP03_F1 and CP03_F2 at 58592 sec. CP03_F2 is an Airbus 320 (A320) flying level at FL 380 en route from Tampa (KTPA) to Minneapolis (KMSP) and is represented by black coloring. CP03_F1 is a Boeing 737-700 (B737) climbing out of Milwaukee (KMKE) en route to Orlando (KMCO) and is represented by blue (baseline) and red (treatment) coloring. As depicted in the side view, CP03_F1 is at about 31,000 ft and is currently cleared to FL 340. A muted alert exists at the current time in the baseline scenario, which predicts a loss of separation where CP03_F1 baseline is predicted at 36,007 ft (above the currently cleared altitude). In the treatment scenario, the more recently built and slightly more accurate (red) trajectory predicts that CP03_F1 will be slightly lower at 35,601 ft which is beyond the CP window, and no alert exists. One second later, CP03_F1 receives an interim altitude deletion clearance and the cleared altitude reverts to the previous flight plan altitude of FL 370 (note that sometime after this point CP03_F1 receives another clearance for FL 390, where it finally levels). In the baseline scenario, the muted alert is then unmuted and an alert is revealed, which is categorized as a false alert. In the treatment scenario, on the other hand, no alert is ever predicted.

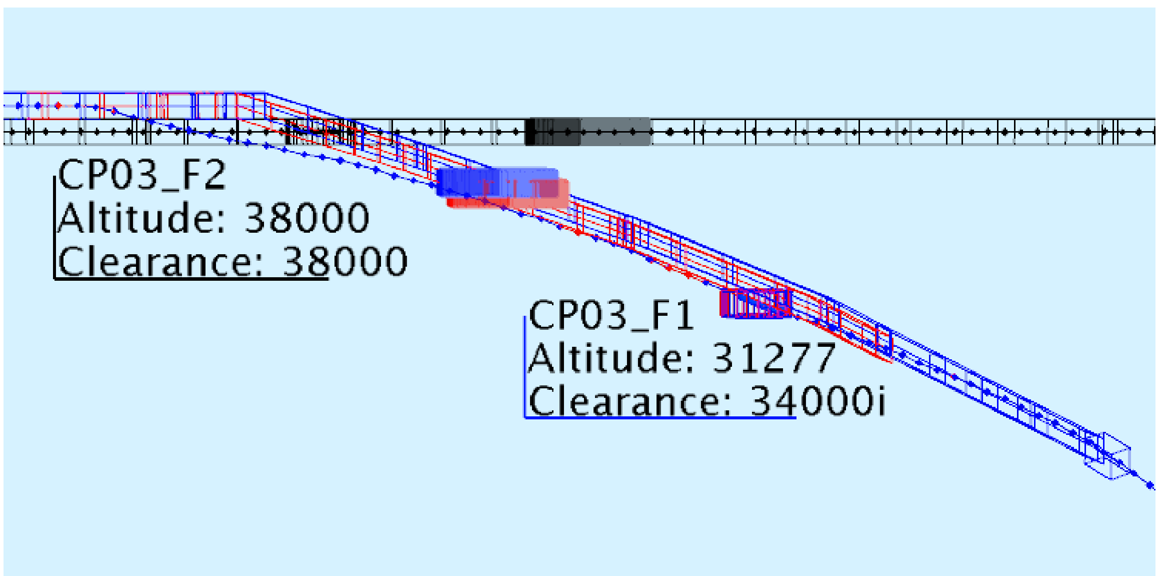

Figure 14. CP example 3, FA to NC, side view

3.3.4 Example 4 – Improvement in Trajectory Accuracy

Trajectory example 4 (illustrated in Figure 15) depicts the difference in vertical trajectory accuracy caused by varying the TM increment for TRAJ01_F1, an Embraer 170 (E170) climbing out of Chicago (KORD) en route to Pittsburgh (KPIT). In this example, the flight's track is represented by blue dots. The two predicted trajectories are depicted as wireframes, with the baseline in blue and treatment scenario in red. The baseline trajectory was built at 60534 sec and the treatment trajectory was built at 60546 sec. At the time shown in Figure 15 (60555 sec) it is evident that the more recently built treatment trajectory is more accurate. With a more restrictive TM increment such as in this treatment scenario, trajectories are built more often during vertical transitions which results in improved accuracy. The original baseline trajectory persists until

60710 (176 seconds), during which conflict probe predictions are based on this less accurate trajectory that does not represent the path that TRAJ01_F1 actually flew.

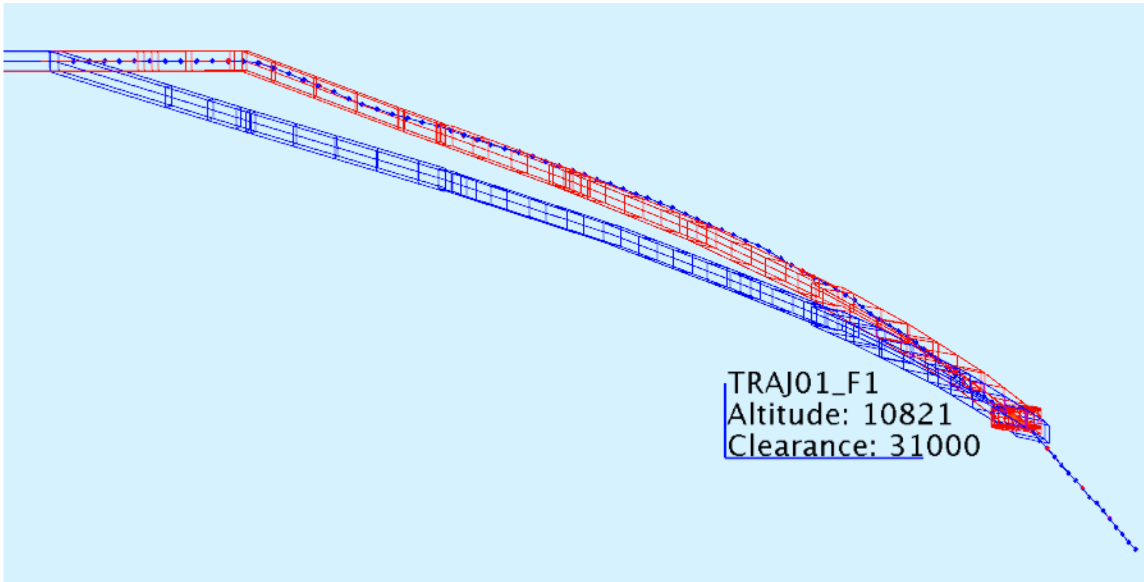

Figure 15. Trajectory example 4, side view

4 Recommendations and Future Work

An experimental study was designed to test the impact of independently varying the TM and CP increments to adherence bounds for vertically transitioning flights. Radar data collected from the ZAU and ZDC centers was time-shifted and simulated as incoming track position reports to test the performance of the Trajectory Modeler and Conflict Probe sub-systems of ERAM and compare the performance across various scenario runs. Two baseline scenarios were used in this comparison: an operational baseline with the same settings as the currently deployed version and an experimental baseline with the preferred parameter and prototype settings from previous Separation Management reports, as described in Section 2.2.

The trajectory accuracy analysis described in Section 3.1.2 demonstrates an overall reduction in vertical error that is statistically significant but small in magnitude. However, when narrowing the focus to trajectory segments that are vertically transitioning, significant improvement in trajectory accuracy is observed when the TM increment is decreased. It is noted that setting the TM increment to 650 ft reduces average vertical error in this subset of data by about 80 feet, averaged across centers.

The conflict probe alert analysis is described in Section 3.2. Overall, the results indicate that it is possible to increase valid alerts and decrease late alerts by decreasing the TM increment, although this improvement in valid and late alerts is accompanied by some increase in false alerts. It is also possible to improve valid and late alert performance by increasing the CP increment, but there is a larger increase in the false alert rate. The recommendations from previous studies, reflected in the experimental baseline, are shown to significantly reduce false alerts with a marginal loss of performance in valid and late alerts. Changing the TM increment to 650 ft would regain some of the performance in valid and late alerts while making the tradeoff of a slightly increased rate of false alerts. Thus, the recommended settings are 650 ft for TM increment and 1000 ft for CP increment.

The conflict probe alert performance was observed to vary across the two centers, sometimes greatly. Further studies with more centers are recommended to quantify the variation and determine the optimal settings. In addition, while this study focused on the effect of lowering the TM increment and increasing the CP increment, it may be of interest to observe the effect of a lower CP increment as well in future studies.

5 List of Acronyms and Abbreviations

AJE-15	FAA Domain Engineering Group
ANG-C41	FAA Concept Analysis Branch
ARTCC	Air Route Traffic Control Center
ATC	Air Traffic Control
ATO-E	Air Traffic Organization En Route Program Office
CD	Conflict Detection
CP	Conflict Probe
ERAM	En Route Automation Modernization
ExBL	Experimental Baseline
FA	False Alert
FA18	Function Area 18
FA32	Function Area 32
FAA	Federal Aviation Administration
Horz	Horizontal
IQR	Inter-quartile Range
JPDO	Joint Planning and Development Office
LA	Late Alert
Lat	Lateral
Llh	Likelihood
LM	Lockheed Martin Corporation
Long	Longitudinal
MA	Missed Alert
MITRE	The MITRE Corporation
NAS	National Airspace System
NC	Correct no-call
NextGen	Next Generation Air Transportation System
nm	Nautical miles
OpBL	Operational Baseline
SME	Subject Matter Expert
TBO	Trajectory Based Operations
TM	Trajectory Modeler
VA	Valid Alert
Vert	Vertical
WT	Warning Time

6 References

Agresti, A. 2002. *Categorical Data Analysis*, 2nd ed., John Wiley and Sons.

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

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

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

Crowell, Andrew, Brian Schnitzer. April 2013. “Analysis of Likelihood Parameter Changes in the En Route Automation Modernization,” FAA Technical Note, DOT/FAA/TC-TN13/13.

Crowell, Andrew, Brian Schnitzer, Mike Paglione. February 2013. “Evaluation of Prototype Enhancements to the En Route Automation Modernization’s Conflict Probe with an Updated Traffic Sample from Washington, DC,” FAA Technical Note, DOT/FAA/TC-TN13/18.

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

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

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

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

McKay, Edward. January 2012. “ERAM Aircraft Trajectory Horizontal Modeling: Candidate Enhancements and Prototype Results,” Lockheed Martin, Document Number FAA-ERAM-2012-0092.

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

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

Paglione, Mike M., Robert D. Oaks. August 2007. “Implementation and Metrics for a Trajectory Prediction Validation Methodology,” American Institute of Aeronautics and Astronautics (AIAA) Guidance, Navigation, and Control Conference, Hilton Head, South Carolina.

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.

Ryan, 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.

SAS Institute Inc. 2010. *JMP® 9 Design of Experiments Guide*, SAS Institute Inc., Cary, NC.

Appendix A

Analysis of Average Horizontal Error (AHE)

This appendix presents an analysis of the horizontal error metric, which was collected from the same experimental runs as the main analysis. Horizontal error is defined in [Paglione and Oaks, 2007] as the time coincident difference between the position of an aircraft and its predicted position on the trajectory. As in the analysis for vertical error, trajectory errors are calculated at the current sampling time (look ahead time of 0) and at future times along the trajectory. To get an idea of overall trajectory accuracy for scenario comparison, horizontal error is sampled at 0, 5, 10, 15, and 20 minute look ahead times, and from these samples a value for average horizontal error (AHE) at each look ahead time is calculated for each flight in a scenario. A paired *t*-test is applied, which compares the average error for a given flight and look ahead time between two scenarios to determine whether a statistically significant difference exists. The results of the matched pairs analysis are presented graphically in Figure 16 for ZAU and Figure 17 for ZDC, where the underlying metric is average horizontal error per flight and look ahead time.

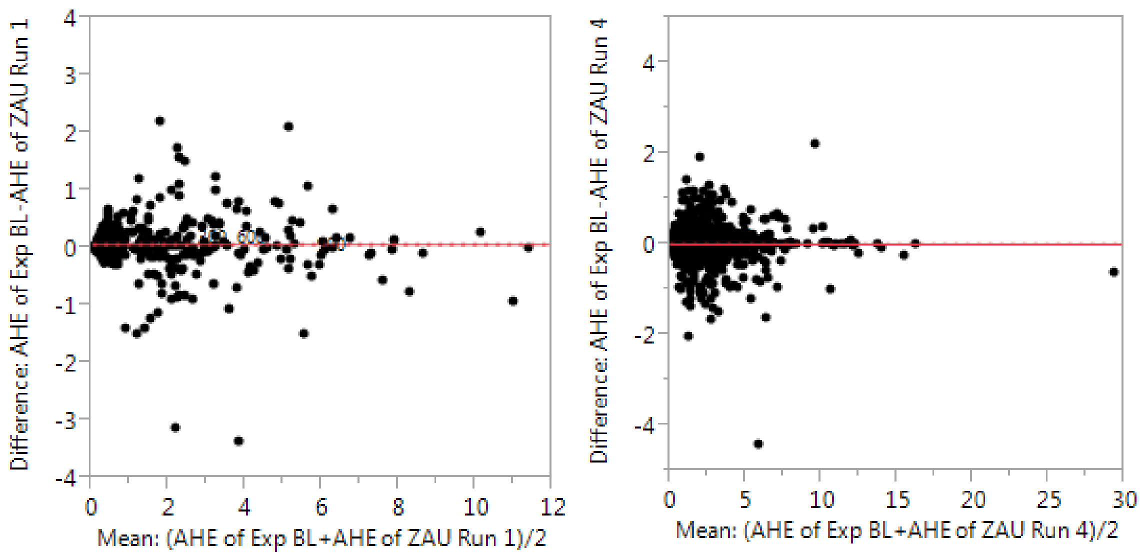

Figure 16. Matched pair data for AHE in ZAU

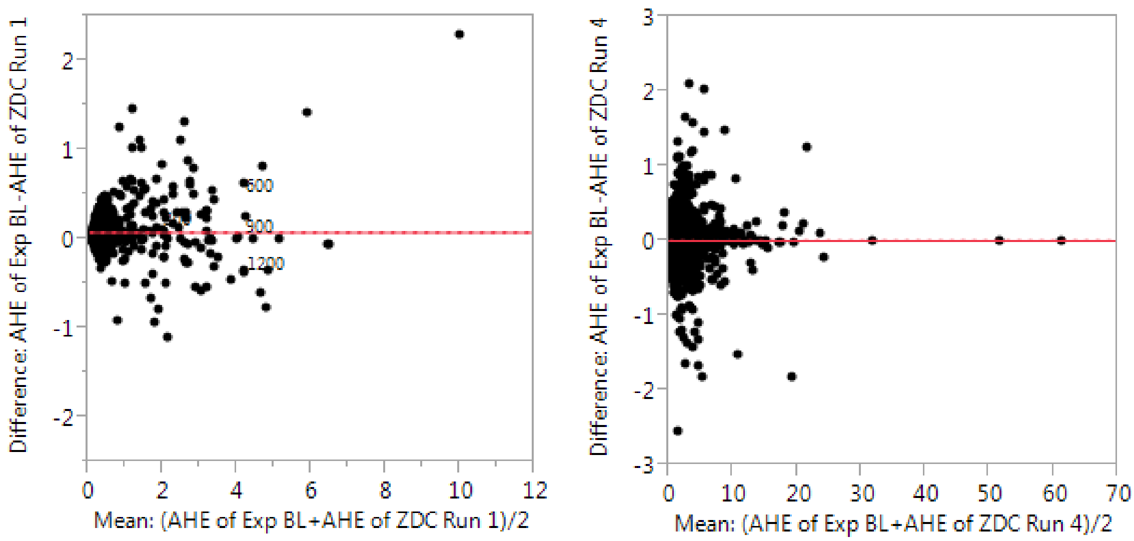

Figure 17. Matched pair data for AHE in ZDC

Table 10 shows the *p*-value and mean differences across the scenarios being compared. Positive mean difference values indicate a decrease in error from the operational baseline to the experimental, and from the experimental baseline to the treatment runs. Each *p*-value is the probability of observing a discrepancy in means at least as large as that observed, even if there is no underlying difference in the means. A *p*-value less than 0.05 is typically considered to indicate statistical significance.

Table 10. Analysis of average horizontal error

Center	Comparison	*p*-value	Mean Difference (nm)
ZAU	Exp BL v Op BL	< 0.0001	-0.1800
	Run 1 v Exp BL	< 0.0001	0.0253
	Run 4 v Exp BL	0.0109	0.0067
ZDC	Exp BL v Op BL	< 0.0001	0.1100
	Run 1 v Exp BL	< 0.0001	0.0230
	Run 4 v Exp BL	0.0234	0.0037

Based on the matched pair analysis, it was determined that the differences between ExBL and the treatment runs are statistically significant; however, the mean differences in error are too small to be practically significant.

Appendix B

Detailed Conflict Probe Alert Comparisons

The following tables contain category definitions (Table 11) and detailed counts of the alerts present in the various runs for the ZAU (Table 12) and ZDC (Table 13) scenarios. OpBL is the operational baseline run and ExBL is the experimental baseline run. The specific TM and CP parameters for each run can be found in Table 3.

Table 11. Alert category definitions

CFL_FA_DISCARD	False Alert notified beyond last conflict actual start time so discard
CLR_FA_DISCARD	Retracted False Alert assigned by an ATC clearance so discard
IN_APDIA_FA	False alert generated but predicted conflict start time determined to be inside an APDIA
LATE_MA	Late alert – alert presented with less than the minimum required warning time
LATE_VA	Late Valid Alert, Valid since conflict was determined a pop-up
NO_CALL_MA	Missed Alert due to no call (no alert at all before the actual conflict start time)
NO_TRK_FA_DISCARD	No post processed track at predicted conflict start time so discard
RETRACT_FA	Retracted False Alert, notification end time earlier than predicted conflict start time
STD_FA	Standard False Alert
STD_VA	Standard Valid Alert

Table 12. Detailed alerts for ZAU runs

ZAU	OpBL	ExBL	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8
CFL_FA_DISCARD	15	11	10	11	9	9	9	11	12	13
CLR_FA_DISCARD_A	477	268	243	258	249	264	274	260	271	266
IN_APDIA_FA	186	165	183	194	208	182	184	196	170	184
LATE_MA	8	11	7	8	5	7	7	5	6	5
LATE_VA	28	23	23	21	18	24	21	17	21	18
NO_CALL_MA	2	1	0	0	0	0	0	0	0	1
NO_TRK_FA_DISCARD_A	0	0	0	0	0	0	0	0	0	0
NO_TRK_FA_DISCARD_B	109	69	64	77	81	65	78	82	78	82
NO_TRK_FA_DISCARD_C	456	374	387	398	409	384	389	417	389	402
RETRACT_FA	559	380	447	467	446	412	415	392	386	377
STD_FA	709	456	460	531	584	452	525	587	523	591
STD_VA	162	165	170	171	176	168	171	178	172	175

Table 13. Detailed alerts for ZDC runs

ZDC	OpBL	ExBL	Run1	Run2	Run3	Run4	Run5	Run6	Run7
CFL_FA_DISCARD	15	11	15	15	14	14	9	9	14
CLR_FA_DISCARD_A	467	344	344	359	354	347	346	359	361
IN_APDIA_FA	78	86	91	95	96	89	94	100	88
LATE_MA	5	11	10	9	8	11	9	8	10
LATE_VA	16	13	13	13	12	14	14	13	13
NO_CALL_MA	1	0	1	1	1	0	0	0	0
NO_TRK_FA_DISCARD_A	80	91	83	90	95	83	88	92	89
NO_TRK_FA_DISCARD_B	272	174	176	190	210	176	189	207	187
NO_TRK_FA_DISCARD_C	496	370	409	431	434	382	390	418	386
RETRACT_FA	484	475	565	584	609	503	518	547	476
STD_FA	860	588	576	685	776	583	683	762	703
STD_VA	187	185	185	186	188	184	186	188	186